Analysis of a Multiple Access Scheme for Broadband Indoor Wireless ATM LAN*

Ashraf S. Mahmoud, David D. Falconer and Samy A. Mahmoud

Department of Systems and Computer Engineering, Carleton University
1125 Colonel By Dr., Ottawa, Canada, K1S 5B6

Abstract

A dynamic Time Division Multiple Access (TDMA) polling-based scheme is proposed for an indoor broadband wireless ATM LAN. The protocol attempts to bridge the existing and rapidly-growing ATM technology to that of personal communications systems and should be capable of handling variable bit rate as well as constant bit rate connections. The system assumes a micro-cellular architecture with a low frequency reuse such as 3 or 4. To combat the severe co-channel interference, which is expected to be the dominant transmission impairment, switched sector-beam antennas are used on both portables and base stations.

The chosen polling scheme is Gated-Limited where service limits are imposed by the central server or base station. This service discipline, with fixed TDMA frame length, is not optimal in terms of overall traffic delay, but it provides to the wireless terminals controlled access to the available capacity and is appropriate for the transport of delay sensitive traffic.

Analytic evaluation of the embedded queueing system is performed under some assumptions, and expressions relating cell loss probability, frame length and structure, and the input traffic statistics to the mean and variance of the buffer size of the served wireless terminals are obtained. The analysis considered typical discrete-time renewal models such as simple Bernoulli, batched Bernoulli, and periodic models. The analysis shows how increases in traffic cell loss probability, which is assumed to be the polling token loss probability, lead to longer survivor (complementary cumulative distribution) function that in turn impacts adversely the buffer size distribution. The analysis supplements the simulation results published previously in [1].

To consider Markov-renewal traffic models such as the Batched Markovian Arrival Process (BMAP), the matrix analytic technique is used to obtain the buffer size distribution. The BMAP model possesses correlation in the number of arriving traffic cells from one polling cycle to the next and thus is more bursty than typical renewal models.

1. Introduction

The last decade witnessed a rapid growth in the area of wireless communications systems, and new applications are being spawned at a great rate. One emerging technology is the specialized low-peak power wireless Local Area Networks (LANs) with various data rates ranging from few kilobits per second to several mega bits per second. Although these wireless LANs are considered as an interpretation [2] of the many that belong to the second generation technology, they share distinguishing characteristics. These wireless LANs are designed with much higher capacities when compared to the voice-oriented systems, and are an attempt to provide service to multimedia capable terminals, so that they can be easily integrated or be an extension to a Broadband Integrated Services Digital Network (B-ISDN).

* This work was supported by a grant from the Canadian Institute for Telecommunications research (CITR) under the NCE program of the Government of Canada.

New standards have been developed for wireless LANs under the IEEE 802.11 in the United States and under ETSI/RES10 or High Performance Radio LAN (HIPER-LAN) in Europe [3]. Japan is also considering standardizing two types of wireless LANs: one for rates less than 2 Mbps and the other for rates greater than 10 Mbps. The IEEE 802.11 and the HIPER-LAN are going to support different network architectures including point-to-point, point-to-multipoint, and broadcast services. Examples of existing wireless LANs include NCR WaveLAN, Motorola's ALTAIR, and FiRLAN.

In this paper we present a concise recounting of the wireless system proposed in [1] together with the analysis of the suggested multiple access scheme. The system provides broadband service at a rate of 155.52 Mb/s in an indoor environment for wireless terminals of various traffic rates. The system attempts to bridge the gap between the ATM technology and the wireless LANs by providing transparent medium of communication between multimedia capable terminals and a high speed ATM backbone network [4]. The second section of this paper describes briefly the overall system and the procedures used in combating errors on the wireless link. The multiple access scheme is discussed in the third section. This section also prepares the ground for the analysis that is conducted in the fourth section. Finally conclusions are presented.

2. Mm-waveband Cellular Architecture for WLAN[1]

Multiple base stations or servers are deployed in a cellular fashion to provide service in an indoor environment such as factories, hospitals, computer laboratories, etc. with each base station covering a microcells or a domain of up to 50 meters radius. On the ground wireless workstations or terminals communicate with each other or with the outside world via the central switch which is connected to the base stations. Due to the high bandwidth of the system and the wide frequency allocation it is expected to operate in the 20-60 GHz band. With a large bandwidth allocated to each cell, a low frequency reuse factor such as 3 or 4 is expected and thus co-channel interference is going to be the main signal impairment. To cope with this high interference environment, both base stations and terminals are equipped with sectorized antennas to reduce outage probability. The system also supports handoff and power control to further interference levels and facilitate the operation of error control procedures. A fixed-length Time-Division Multiple-Access (TDMA) frame which is divided into sector segments is used for communication between the server and the terminals on the ground. The order of the sector segments is randomized from frame to frame. The system uses Time-Division Duplex (TDD) to multiplex up link and down link traffic on the same wireless channel. The system relies on error control procedures to provide the necessary transparency required for augmenting the wireless link with the wireline backbone network. In [1] it was shown through simulation that a selective repeat request protocol is suitable in providing error free link while not incurring excessive retransmissions Other enhancements like hybrid ARQ [5], which employs a combination of Forward Error Control (FEC) and retransmission, may help increase utilization. A detailed description of the system and the multiple access accompanied with simulation study are found in [1], [6], and [7]. Reference [7] is a more comprehensive reference that contains also the material in this paper. Reference [8] describes the overall context of the broadband indoor wireless system.

3. Multiple Access Method

One of the most obvious requirements for this system is to provide service for both non real-time and real-time traffic and to be able to utilize the available capacity optimally. Non real-time traffic like computer data, file transfers, etc. generally are very tolerant to delays (relative to real-time traffic) where the expected response time is on the order of one second or so. However, they are not tolerant to bit errors and/or packet loss. On the other hand, real-time traffic, such as voice, real-time video, etc., has stringent delay requirements; usually a maximum delay figure is associated with the traffic stream and is compared against a time stamp (generation time) of each packet.

In general there are different types of services each with a different set of service requirements. This suggests that the multiple access method should be able to prioritize service and effectively isolate traffic streams so that adverse effects from one stream to another are minimized or neutralized. This leads to the preference of centralized multiple access. The other requirement is the bounded delay or access time. This suggests that a preemptive type of multiple access is required. This manifests itself in the choice of a fixed-length frame. Based on these arguments, contention based methods like ALOHA and CSMA/CA in their original forms are ruled out because they do not meet the centralized and bounded delay criteria. CDMA and FDMA are also ruled out by complexity considerations.

A dynamic TDMA polling based method is chosen. The basic model in polling schemes assumes that a central server polls every user before service commences. Then according to the service policy used, traffic from each user is served. When the service time is completed, the server moves to the next user. Polling schemes have been suggested for other broadband applications [9] and [10].

Polling has been around for many years and many variations have been analyzed and studied. A comprehensive study of polling schemes and their performance can be found in [11][12][13][14]. In general there are three main types of service policies: exhaustive, gated, and limited. The first two are non preemptive service policies whereas the limited is a preemptive policy. Using the exhaustive service, users are allowed to empty all their buffer contents, while for the gated service only traffic existing prior to the polling instant is served. Although these two policies are optimal in terms of the over all average traffic delay, they do not provide a fair service; When a heterogeneous traffic mixture is served, the available capacity tends to be captured by high data rate users, increasing the traffic delays for the low bit users. In the limited service, only a specified number of traffic cells are served every polling cycle. We propose the use of gated-limited service discipline where only traffic present in terminal's buffer at the polling instant is served up to a maximum service limit of cells per visit.

The following two subsections prepare the ground for the section in which the queueing analysis is conducted. The first describes the TDMA frame and characterizes the intervisit time, an important parameter that reflects the structure of the frame. The second subsection introduces the basic queueing model embedded in the chosen multiple access method.

Throughout the coming sections it is assumed that the time axis is divided into equal length slots. For analysis purposes we assume also that all *wireless envelopes*, the units of transmission on the wireless link, have the same size whether they are traffic or control (beacons, polling tokens, etc.). Furthermore, the server and the wireless terminals are synchronized at the slot level.

3.1 TDMA Frame Structure and Intervisit Time

The fixed-length TDMA frame of FL time slots is divided into S sector segments or subframes not necessarily of equal lengths. During the i^{th} subframe the base station uses only the i^{th} antenna sector. This serves to reduce the impact of interference. Furthermore, the order of these subframes is randomized from one frame to the next to average co-channel interference experienced by users and facilitate the operation of error control procedures.

Each sector segment starts with a beacon envelope which carries information specific to that segment such as synchronization word for the antenna sector, power level, traffic condition, etc. Then a number of registration slots are provided for new users to contend on using a slotted Aloha protocol with backoff procedures to minimize collision probability. The remaining part of the sector segment is used for down link and up link traffic to or from users registered to that particular subframe. The frame structure is depicted in Figure 1.

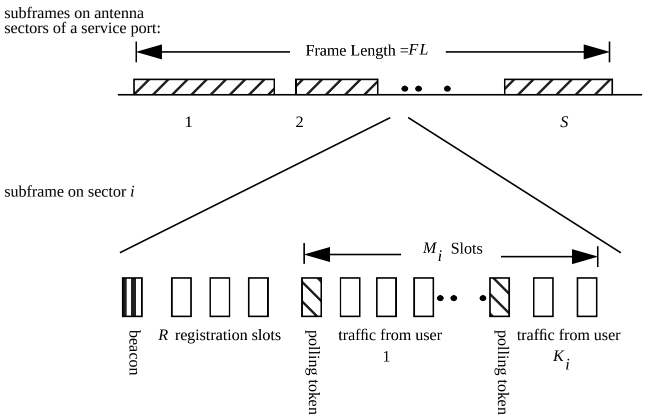


Figure 1: Frame structure and polling cycle.

Let us assume that wireless envelopes (traffic or otherwise) are incorrectly received with probability α and this loss probability is independent from one time slot to the next. This independency claim is justified by assuming that co-channel interference is different from frame to frame, if the order of sectors and slot occupancy is randomized from frame to frame, and also the bit/packet-level interleaving that may be applied as a requirement for FEC. A user registered in a particular segment is sent a polling token every time its turn is reached when that particular segment is served. The user may or may not correctly receive the polling token. The i^{th} intervisit time, v_i, is defined as the duration in time slots between the i^{th} and the $(i-1)^{\text{st}}$ correctly received polling tokens. To characterize intervisit time two distinct situations are recognized. The first is when the sector segments are equal in length. This is the case when traffic load is spatially symmetric or uniformly distributed in the service area, i.e all sectors of space are equally utilized. The second case, assumes that only one sector segment is occupied while the remaining ones have very little or no traffic. This may represent a case where a high bit-rate user is occupying most of the frame. Let the probability generating functions of the intervisit random variable for the two cases be $V_1(z)$ and $V_2(z)$, respectively. These are given by

$$V_1(z) = M(E_1(z)) \tag{1}$$

and

$$V_2(z) = M(E_2(z)) \tag{2}$$

where $M(z) = \frac{(1-\alpha)z}{1-\alpha z}$, for $|\alpha z| < 1$ and $E_m(z) = \sum_{\forall i} e_{m,i} z^i$ for $m = 1, 2$. $e_{m,i}$ is defined as

$$e_{1,i} = \frac{S-|k-S|}{S^2}, i = k(FL/S), k = 1, 2, \ldots, 2S-1 \quad \text{or}$$

$$e_{2,i} = \frac{S-|k-(S-1)|}{S^2}, i = k\varnothing_1 + \varnothing_2, k = 0, 1, \ldots, 2S-2.$$

$\varnothing_m$ is the lengths of the unbusy ($m = 1$) or busy ($m = 2$) sector segment.

3.2 Queueing Model

The server cycles through the wireless terminals, sending a polling token to each, signalling permission to receive down link and/or transmit uplink traffic. In the remainder of this paper we will concentrate on the uplink assuming that the down link can be dealt with in a similar manner. That is, the server queues traffic destined for a particular terminal in a separate buffer and transmits it in groups of up to a maximum number of cells each time the wireless terminal is polled.

As mentioned in the previous subsection, from the point of view of the terminal, the token is received at intervisit of v_i time slots. Assuming that the terminal is allocated a service limit of Y cells per polling cycle and following the gated-limited discipline, the number of traffic cells to be served or transmitted on the up link in the i^{th} polling cycle, b_i, is determined by

$$b_i = \min(x_i, Y) \tag{3}$$

where x_i is the number of traffic cells present in the terminal buffer upon the receipt of the token. The terminal transmits those b_i cells but does not discard then from the buffer; a copy is kept for ARQ purposes. When the terminal is polled in the next cycle, it is informed of which cells were corrupted by RF interference, allowing the terminal to discard cells that were correctly received. Let the number of cells that are corrupted by interference be l_i. Then the new buffer size, x_{i+1} follows from

$$x_{i+1} = \begin{cases} n_0 & i = 0 \\ x_i - b_i + l_i + n_i & i = 1, 2, \ldots, \end{cases} \tag{4}$$

where n_i is the number of cells that arrive during the $(i+1)^{\text{st}}$ intervisit time. The above describes the buffer size evolution equation and it is the mathematical representation of the queueing model shown in Figure 2.

The number of retransmissions, l (dropping the polling cycle index) is a binomial random variable of the following parameters: the total transmissions, b, and the cell loss probability, α. This is a consequence of the independent cell loss probability assumption stated earlier. The number of traffic cells arriving during an intervisit time, n, is not only a function of the traffic source characteristics but also the randomization of the subframes. That arrival stream is modulated by the polling token arrival. In terms of PGFs, one can write

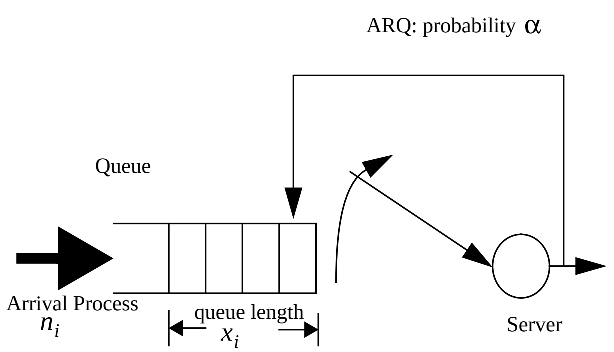


Figure 2: Queueing model.

$$A(z) = V(\tilde{A}(z)) \tag{5}$$

where $A(z)$ is the PGF of the variable n, $V(z)$ is the PGF of the intervisit time and $\tilde{A}(z)$ is the PGF of the number of traffic cells arriving in a time slot.

4. Queueing Analysis

Before we embark on analyzing the queue evolution equation, we consider two types of the $\{n_i; i = 0, 1, \ldots\}$ series. The first corresponds to a *renewal* model [15] where the arrivals n_i's are independent from one intervisit duration to the next; i.e. there is no temporal correlation in the arrivals. The auto-correlation function $R_n(k) = E[n_{i+k} n_i]$ is only non zero at the origin. $E[\cdot]$ is the expectation operator. For the second type of model however, $R_n(k)$ is non zero for non zero lags. These models are known as *non renewal* models. Examples of these models include Markov and Markov renewal models. Because of the temporal dependencies amongst arrivals usually these models are mathematically less tractable than renewal models and thus the lack of closed form solutions for many queueing cases involving the non renewal models. It should be pointed out that it has been recognized that the main contributors to the *burstiness* of a particular model are the shape of the probability mass function of the arrivals and the temporal dependencies [16]. By definition, the non renewal models possess both components (contrary to the renewal models) and thus lead to greater queue lengths and delays than that caused by a renewal model for a given traffic load.

4.1 Analysis For Renewal Model

Renewal models have been the basis for most of the results in classical queueing theory [17] for their mathematical tractability. In discrete time queueing theory the Bernoulli model [18] and its derivatives are typical models to consider. We are going to consider a generalized Bernoulli model, where at any time slot a batch of (one or more) cells may arrive. Let the PGF of the batch size be given by $G(z)$ and β be the mean batch size. For a terminal of bit rate equal to R b/s, then the following holds

$$\beta p = \frac{R}{R_{\text{total}}} \tag{6}$$

where p is the probability of batch arrival and R_{total} is the total capacity of the link. From the above, the PGF of the total arrivals in a time slots, $\tilde{A}(z)$ is given by

$$\tilde{A}(z) = 1 - p + pG(z) \tag{7}$$

For a Bernoulli model where batches arrive in fixed size, β cells per batch, $G(z)$ is equal to z^{β}, while if the batch size is geometrically distributed then $G(z) = \frac{(1-\zeta)z}{1-\zeta z}$ and $\zeta = 1 - 1/\beta$. Another model of interest in this category is the constant bit rate or the periodic model. According to this model, only a fraction r $(r < 1)$ of a traffic cells is generated during any time slot. Hence $\tilde{A}(z)$ is given by

$$\tilde{A}(z) = z^{p} \tag{8}$$

Examining the evolution relation of the queue ((4)), one can recognize that the sequence of buffer sizes $\{x_i; i = 1, 2, \ldots\}$ constitute a discrete-time Markov chain. The state transition probabilities are given by

$$p_{j,k} = \text{Probability}[x_{i+1} = k / x_i = j] = \begin{cases} \sum_{m=0}^{k} a_{k-m} L_{jm} & k \ge 0,\ j \le Y \\ \sum_{m=0}^{k-j+Y} a_{k-j+Y-m} L_{Ym} & k \ge j - Y > 0,\ m \le Y \end{cases} \tag{9}$$

where $a_k = \lim_{i \to \infty} \text{Probability}[n_i = k]$, L_{jm} is the probability of having to retransmit m cells when j are served. Let the steady state probability distribution of the buffer size, at the polling instant, be

$$q_k = \lim_{i \to \infty} \text{Probability}[x_i = k] \qquad , k = 0, 1, \ldots, \infty \tag{10}$$

and the corresponding PGF be $Q(z)$. q_k can be rewritten in terms of the transition probabilities as

$$q_k = \sum_{j=0}^{\infty} q_j p_{j,k} \qquad k = 0, 1, \ldots, \infty \tag{11}$$

Substituting (9) into (11), rearranging terms and applying the z-transform yields

$$Q(z) = \frac{A(z)\left[z^{Y}\sum_{j=0}^{Y-1} q_j L_j(z) - Q_Y(z)L_Y(z)\right]}{z^{Y} - A(z)L(z)} \qquad (12)$$

The new terms in (12) $L_j(z)$ and $Q_Y(z)$ are given by $\sum_{m=0}^{j} L_{jm}z^{m} = (1-\alpha+\alpha z)^{j}$, the generating function describing the ARQ process and $\sum_{j=0}^{Y-1} q_j z^{j}$, respectively. $\{q_j, j = 0, 1, \ldots, Y-1\}$ are the first Y points or boundary probabilities of the required distribution. They are solved for exploiting the analytic property of $Q(z)$ in the region enclosed by $|z| = 1$. The Y roots $\{z_m; m = 0, 1, \ldots, Y-1\}$ of $z^{Y} - A(z)L_Y(z)$ can be found using Lagrange's theorem or more conveniently using packages like matlab. The numerator of (12) also vanishes at these roots resulting in a set of Y linear equations which can be solved for the boundary probabilities.

On the stability issue, one can show by applying Rouche's theorem on the expression of $Q(z)$, the condition

$$\frac{\bar{a}}{(1-\alpha)} < Y \qquad (13)$$

is necessary for the existence of $Q(z)$. $\bar{a} = \lim_{z\to 1} A^{(1)}(z)$, the mean number of arrivals during an intervisit time. The superscript (n) over a PGF function $X(z)$ indicates the n^{th} derivative with respect to z. This condition translated to the fact that the mean arrivals plus the retransmitted traffic each polling cycle should be less than the allocated bandwidth for the buffer to be stable. Let us define the traffic load to be the ratio of the mean arrivals rate to the capacity allocated to the terminal. Using (13), it can be seen that the load is always less than or equal to $(1-\alpha)$.

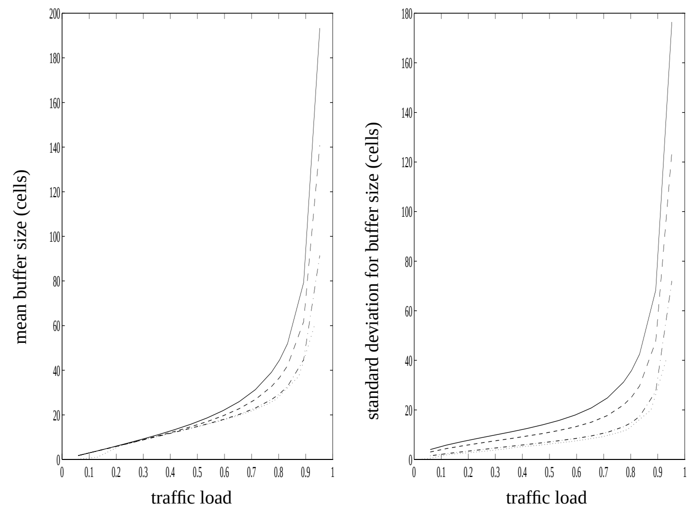
Expressions for the mean queue length $\bar{q}$ and standard deviation σ_q can be easily, albeit tediously, obtained using

$$\bar{q} = \lim_{z\to 1} Q^{(1)}(z) \qquad (14)$$

and

$$\sigma_q = \sqrt{\lim_{z\to 1} Q^{(2)}(z) + \bar{q} - \bar{q}^{2}} \qquad (15)$$

Figure 3 shows the mean and standard deviation for different traffic models. These results assume 10 Mb/s terminals which are generating traffic that is packetized as 500 bits per wireless envelope, 48 byte of which is actual payload. The bit rate of the channel is 155.52 Mb/s and the TDMA frame is divided into $S = 6$ sector segments each 34 time slots in length. The terminal is allocated $Y = 29$ time slots. It can be noticed that the curves increase exponentially as the traffic load approaches $(1-\alpha)$. Furthermore, for both the mean and standard deviation figures, the batched Bernoulli model leads to the highest values while the constant bit rate is the lowest.



a) mean **b) standard deviation**

Figure 3: Mean and standard deviation of buffer size at polling instant for cell loss probability of 1% (solid line: batched-geometric Bernoulli with mean batch size equal to 5, dashed line: batched-constant Bernoulli with mean batch size equal to 5, dashed-dotted line: simple Bernoulli, dotted line: constant bit rate).

Another figure of comparison between the traffic models that is indicative of burstiness is the complementary cumulative distribution function for the buffer occupancy. This is shown in Figure 4 for two different cell loss probabilities. It can be noticed that the batched Bernoulli source has the longest tail, especially, when the batch sizes are geometrically distributed as opposed to constant size bursts. The slope of the complementary cumulative function decreases as the batch size increases. The shortest distribution corresponds for the case of a constant bit rate source.

Regarding the second case of traffic spatial distribution, the service limit for the terminal occupying the busy sector is very large compared to the previous case. Hence, most of the time, if not all, the terminal does not use all the allocated capacity. Writing the buffer evolution equation for this case and the state transition probabilities of the Markov chain describing the buffer occupancy, one can show that the buffer size PGF satisfies the functional form

$$Q(z) = Q(1-\alpha+\alpha z)A(z) \qquad (16)$$

Through direct substitution the solution for $Q(z)$ is given by

$$Q(z) = \prod_{k=0}^{\infty} \widehat{A}(\alpha^k(1-z)) \quad (17)$$

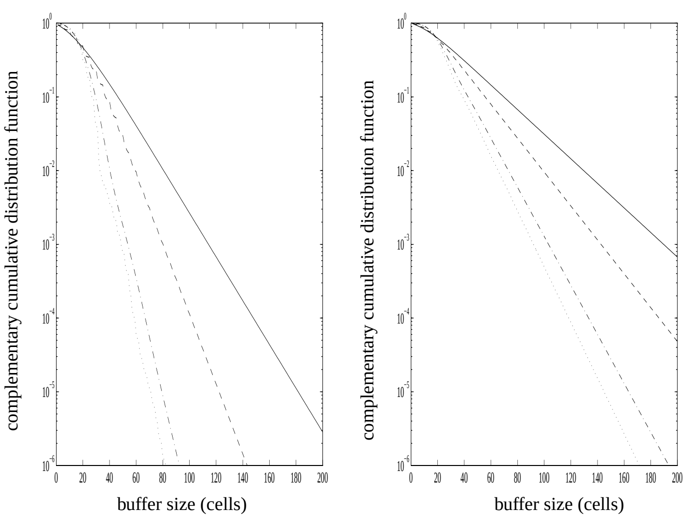
where $\widehat{A}(z) = A(1-z)$. Invoking the geometric distribution approximation of the intervisit time, the mean queue length and the variance are given by

$$\bar{q} = \frac{\bar{a}}{1-\alpha} \quad (18)$$

and

$$\sigma_q^2 = \frac{\overline{a^2}}{(1-\alpha^2)} + \frac{\bar{a}\alpha(2\bar{a}+1-\alpha)}{(1-\alpha)(1-\alpha^2)} - \frac{\bar{a}^2}{(1-\alpha)^2} \quad (19)$$

respectively. $\overline{a^2}$ is the second moment of the modulated arrival process.



a) $\alpha = 0.01$ **b)** $\alpha = 0.1$

Figure 4: Complementary cummulative distribution function for the buffer size (solid line: batched-geometric Bernoulli with mean batch size equal to 5, dashed line: batched-constant Bernoulli with mean batch size equal to 5, dashed-dotted line: simple Bernoulli, dotted line: constant bit rate).

4.2 Analysis for BMAP Model

Departing from the classical renewal models model where only the marginal distribution is of significance, many studies considered the effect of correlation on the queue statistics. Assuming the same model in Figure 2 applies and the input process is modeled by a Batched Markov Arrival Process (BMAP), a solution for the distribution buffer size is described. The BMAP model assumes that arrivals during an intervisit times follow a Markov process with a finite number of states. The states in this chain correspond to the number of arrivals and the transition probabilities from one state to another characterize the second order statistics, the covariance, of the source in addition to the first order statistics, the marginal traffic arrival distribution. In other words, the system considered here is the BMAP/G/1 queue. To solve this queue a very powerful technique called “Matrix Geometric” is used. This technique was introduced by M. Neuts in [19] and then it was applied to BMAP/G/1 by Neuts in [20] and Lucantoni in [21].

Ahmadi and Guérin [22] present the solution procedure to a problem very similar to ours except that their queue evolution equation does not contain the ARQ term (refer to (4)). To solve the resulting queue they apply the theory of the matrix geometric technique and obtain the stationary buffer size distribution. Here their model is extended to include the retransmissions incurred because of the ARQ strategy and a new queue transition matrix is built. The queue transition matrix is the key matrix required for solving complex queueing problems using the matrix geometric method.

Consider our queue evolution equation, namely,

$$x_{i+1} = \max(0, x_i - Y) + l_i + n_i \quad (20)$$

which is rewritten here in a different form than the one in (4). Let n_i be generated through a Markov chain with a finite number of states $\{0, 1, \ldots, N-1\}$ where the k^{th} state corresponds to the event of k traffic cells arriving*. Furthermore, let p_{ij} be the transition probability between state i and j, $0 \le i, j \le N-1$. Using these definitions (x_i, n_i) forms a discrete-time Markov chain. The transition probability matrix of this chain has a block-partitioned structure similar to that of an M/G/1 queue, which is of the form

$$Q = \begin{bmatrix} B_0 & B_1 & B_2 & . & B_{M-1} & 0 & 0 & 0 & . \\ A_0 & A_1 & A_2 & . & . & A_M & 0 & 0 & 0 & . \\ 0 & A_0 & A_1 & . & . & A_{M-1} & A_M & 0 & 0 & . \\ 0 & 0 & A_0 & A_1 & . & . & . & A_{M-1} & 0 & . \\ . & . & . & . & . & . & . & . & . & . \end{bmatrix} \quad (21)$$

where A_i for $0 \le i \le M$ and B_i for $0 \le i \le M-1$ are square nonnegative submatrices. This is true when the following mapping is used: the transition probability from state (i_1, j_1) to state (i_2, j_2) is the entry in the $(Ni_1 + j_1)^{\text{th}}$ row and the

* The study in [22] considers a more general definition of the Markov chain in the sense that state k corresponds to $S_k \ge 0$ traffic cells arriving. However, since our arrival process will vary from 0 to $N-1$ traffic cells, we chose the definition that state k corresponds to the arrival of k traffic cells.

$(Ni_2 + j_2)^{th}$ column of the matrix Q. i.e the states $(i, j); 0 \le i, 0 \le j \le N-1$ are ordered lexicographically as $\{(0, 0), (0, 1), \ldots, (0, N-1), (1, 0), (1, 1), \ldots, (1, N-1), \ldots\}$.

From the evolution equation one can deduce that the transition probability is given by

$$q_{(i_1, j_1), (i_2, j_2)} = \begin{cases} p_{j_1, j_2} L_{i_1, k} & i_2 = \max(0, i_1 - Y) + j_2 + k \\ 0 & \text{else} \end{cases} \quad (22)$$

$k = 0, 1, \ldots, \min(i_1, Y)$, and Y is the service limit every polling cycle. As before, the variable $L_{i,k}$ is the probability that k cells are in error when i are transmitted. Substituting in (22) and building the matrix Q, and after recognizing the block structure, one finds out that the submatrices A_i and B_i are of size $YN \times YN$ and M is equal to $\left\lceil \frac{N}{Y} \right\rceil + 1$, where $\lceil x \rceil$ is the smallest integer greater or equal to x.

It should be noted that the matrix Q is stochastic and its invariant probability vector $\underline{q}$ can be obtained through solving the system described by $Q\underline{q} = \underline{q}$ and $\underline{e}\underline{q} = 1$, where $\underline{e}$ is a row of ones. However, since the dimensions of Q are infinite, the matrix geometric technique relies on the submatrices to obtain the vector $\underline{q}$.

Applying the matrix geometric technique to our queueing problem and considering the same bit rate and service limit namely, 10 Mb/s and 29 cells per polling cycle, will result in square submatrices that have 9×10^6 elements. Defining and manipulating tens of these matrices is beyond the capability of the available computational resources. Thus we resort to scaling down the problem. Below we show the results for up to 2 Mb/s terminal with a service limit of 5 cells per polling cycle. The mean and standard deviation figures of buffer occupancy for some traffic loads are listed in Table 1. The table serves also as a comparison between simulation and analysis results. Simulation results are averaged over 10^6 iterations. It can be noticed that numerical errors are more evident as the traffic load ρ increases. This is because of the large size of the submatrices that have to be considered. For example, for $\rho = 0.86$, the traffic source is represented by a 50 states Markov chain; and since Y is equal to 5, the components of the queue transition matrix are 23 distinct submatrices each of size 250×250. Also shown in the table (last two colunms) are the results obtained using the classical Bernoulli model with zero correlation between arrivals. Clearly, the Markovian model leads to greater buffer occupancy for the same traffic load.

Table 1: Mean and standard deviation results for buffer size using a Markov traffic model (service limit = 5 cells per polling cycle, cell loss prob. = 10%)

Terminal Rate R Mb/s	Traffic Intensity ρ	Analytical		Simulation		Using (14) & (15)	
		q	σ_q	q	σ_q	q	σ_q
0.5	0.29	1.1	1.2	1.1	1.2	1.1	1.1
1.0	0.48	2.3	2.0	2.3	2.0	1.1	1.8
1.5	0.67	4.7	4.4	4.7	4.6	3.9	2.9
2.0	0.86	14.4	16.1	16.5	17.8	8.4	6.8

The traffic model used to produce the results in the above table has the marginal distribution of a simple Bernoulli source. The correlations were generated using the Transform-Expand-Sample (TES) method [23].

5. Conclusions

This paper presented a study of a multiple access method for broadband wireless ATM LAN. The proposed TDMA polling-based scheme accommodates relatively high levels of interference through the use of switched sectored antennas and the random scheduling of subframes within the fixed-length frame. Other procedures such as handoffs, power control, and ARQ collaborate to reduce the impact of the radio interference on the performance of the wireless link. Given the various polling schemes, we opt for the preemptive gated-limited method due to its suitability and flexibility in providing a controlled access to the available capacity. The queueing model embedded in this system is analyzed and results for the buffer size at polling instant are obtained. The analysis considered classical models such as Bernoulli and generalized Bernoulli models in addition to batched Markovian arrival process model. Results give an insight into the bandwidth (service limit) required by different traffic sources and relates cell loss probability, frame structure, and ARQ to the performance of the multiple access scheme. These results can be used in call admission procedures in wireless or even wireline systems.

Acknowledgement

The authors would like to thank Prof. I. Lambadaris for the valuable discussions and comments, and Prof. H. Rubin from Purdue university for his over-the-internet mathmetical hint.

References

[1] A.S. Mahmoud, D.D. Falconer, and S.A. Mahmoud, “A Multiple Access scheme for Wireless Access to Broadband ATM LAN Based on Polling and Sectored Antennas,” IEEE Journal on Selected Areas in Communications, Vol. 14, No. 4, May 1996, pp. 596-

608.

[2] D.C. Cox, “Wireless Network Access for Personal Communications,” *IEEE Communications Magazine*, Vol. 30, No. 12, December, 1992, pp. 96-115.

[3] R.O. LaMaire, A. Krishna, P. Bhagwat, and J. Panian, “Wireless LANs and Mobile Networking: Standards and Future Directions,” IEEE Communications Magazine, Vol. 34, No. 8, August 1996, pp. 86-94.

[4] D. D. Falconer, G. Stamatelos, “Wireless Access to Broadband Services Through Microcellular Indoor Architectures”, Can. J. Elect. & Comp. Eng., Vol. 19, No. 1, January 1994.

[5] Y.M. Wang and S. Lin, “A Modified Selective-Repeat Type-II Hybrid ARQ System and its performance analysis,” IEEE Trans. on Communications, COM-31, No. 5, May 1983, pp. 593-608.

[6] A.S. Mahmoud, D.D. Falconer, and S.A. Mahmoud, “A Multiple Access scheme for Wireless Access to Broadband ATM LAN Based on Polling and Sectored Antennas,” PIMRC‘95, Toronto, Canada, Vol. 3, pp. 1047-1051.

[7] A. S. Mahmoud, “Cellular Architecture for Wireless Indoor Broadband Communication Systems in the mm-waveband,” PhD Thesis, Department of Systems and Computer Engineering, Carleton University, Ottawa, Ontario, Canada, January 1997.

[8] D.D. Falconer, “A system Architecture for Broadband Millimeter Wave Access to an ATM LAN,” IEEE Personal Communications, Vol. 3, No. 4, August 1996, pp. 36-41.

[9] Z. Zhang and A.S. Accampora, “Performance of a Modified Polling Strategy for Broadband Wireless LANs in a Harsh Fading Environment,” Proc. IEEE Globcom‘91, Phoenix, December 1991, pp. 1141-1146.

[10] J.M. Karlsson, H.G. Perros, and I. Viniotis, “Adaptive Polling Schemes for an ATM Bus with Bursty Arrivals,” Computer Networks and ISDN Systems 24, North Holland, 1992, pp. 223-229.

[11] Hideaki Takagi, “Queueing analysis of Polling models: an update,” in Stochastic Analysis of Computer and Communication Systems, Edited by H. Takagi, Elsevier Science Publishing Company Inc. 1990.

[12] Hideaki Takagi, “Queueing Analysis of Polling Models,” *ACM COMP. Survey*, Vol. 20, No. 1, March 1988, pp. 5-28.

[13] Hideaki Takagi, Analysis of Polling Systems, The MIT Press, London, England, 1986.

[14] Hideaki Takagi, Queueing Analysis, a Foundation of Performance Evaluation, volumes 1, 2, and 3, 1993 Elsevier Science Publishers B.V.

[15] V.S. Frost, B. Melamed, “Traffic Modeling For Telecommunications Networks,” IEEE Communications Magazine, Vol. 32, No. 3, March 1994, pp. 70-81.

[16] M. Livny, B. Melamed, and A.K. Tsiolis, “The Impact of autocorrelation on Queueing Systems,” Management Science, Vol. 39, No. 3, 1993, pp. 322-339.

[17] Queueing Systems - Volume 1: Theory, L. Kleinrock, John Wiley & Sons, 1975.

[18] T. Meisling, “Discrete-Time Queuing Theory,” Operation Res. 6 (1958), pp. 96-105.

[19] Marcel F. Neuts, “A versatile Markovian point process,” *Journal of Applied Probability*, (16) 1979, pp. 764-779.

[20] Marcel F. Neuts, Matrix Geometric Solutions in Stochastic Models. Johns Hopkins University Press, Baltimore, Maryland, 19981.

[21] D.M. Lucantoni, “New Results on the Single Server Queue with a Batch Markovian Arrival Process,” *Commun. Statist.-Stochastic Models*, 7(1), 1991, pp. 1-46.

[22] Hamid Ahmadi and Roch Guérin, “Analysis of a Class of Buffer Storage Systems with Markov-Correlated Input and Bulk Service,” Data Communication Systems nd Their Performance, G. Pujoll and R. Puigjaner (editors), IFIP, 1991, pp. 109-125.

[23] B. Melamed, “TES: A class of methods for generating autocorrelated uniform variates,” *ORSA J. Comp*. Vol. 3, No. 4, 1991, pp. 317-329.