Performance Analysis of Prioritized MAC in UWB WPAN With Bursty Multimedia Traffic

Kuang-Hao Liu, Xinhua Ling, Xuemin (Sherman) Shen, *Senior Member, IEEE*, and Jon W. Mark, *Life Fellow, IEEE*

***Abstract*—Ultra-wideband (UWB) is expected to be the transmission technology of future wireless personal area networks (WPANs), carrying various multimedia streams. Recently, the WiMedia Alliance has launched its standard for UWB WPANs, where the prioritized channel access (PCA) protocol is specified to provide differentiated medium access control (MAC) in a distributed manner. For time-sensitive multimedia traffic, the total delay, including the frame service time and the frame waiting time, is an important metric for quality-of-service (QoS) provisioning. This paper presents a performance analysis for the PCA protocol, considering the bursty nature of multimedia traffic. The mean frame service time and the mean waiting time of frames belonging to different traffic classes are obtained. Simulation results are given to verify the analytical results and demonstrate that the effect of the traffic differentiation mechanism in PCA is magnified when the interarrivals are highly bursty and correlated. In addition, the characteristics of multimedia traffic have a significant impact on the mean frame waiting time. Finally, our analytical model is applied to delay-sensitive traffic for QoS provisioning.**

***Index Terms*—Markov Modulated Poisson Process (MMPP), multimedia traffic, prioritized channel access (PCA), queueing analysis, ultra-wideband (UWB), wireless personal area networks (WPANs).**

I. Introduction

RECENT advances in semiconductor technology have made ultra-wideband (UWB) technology ready for commercial applications [1], [2]. Consumer UWB products and prototypes that deliver high-data-rate ($>$ 100 Mb/s) multimedia traffic over a short distance ($\leq$ 10 m) with very low power consumption have been emerging. In future wireless personal area networks (WPANs) or broadband home networks, multiple UWB devices can exchange high-volume multimedia traffic or deliver high-volume data to/from the Internet [3].

To support high-data-rate multimedia applications in a personal/home network, the WiMedia Alliance recently launched its physical (PHY) layer and medium access control (MAC) protocol specifications [4] based on the UWB orthogonal frequency-division multiplexing (OFDM) technology for emerging high-rate WPANs. In WiMedia MAC, a fixed-length superframe structure, consisting of a beacon period (BP) and a data transfer period (DTP), is defined to maintain coordination among communication devices and allow an efficient power-saving mode. Each device will first listen to at least one beacon frame, if available, which contains information for synchronization, device discovery, sleep-mode operation, and reservation announcement. Without a centralized controller, each device needs to broadcast its beacon frame in BP and observes its own BP length. Detailed discussions of the operation of BP are given in [4] and [5]. Transmissions in the DTP can use a contention-based channel access mechanism called prioritized channel access (PCA) or a contention-free channel access called the distributed reservation protocol (DRP). The basic principle of DRP is similar to time-division multiple access, which is suitable for real-time traffic with a constant bit rate. On the other hand, the PCA is based on carrier-sense multiple access with collision avoidance (CSMA/CA) and employs different contention parameters to support both non-realtime and real-time data transfer. Most of the previous work on the CSMA/CA protocol and its variants assumes saturation stations and independent interarrivals. Multimedia applications, however, generally exhibit strong burstiness/correlations between interarrivals that violate the above assumptions.

In UWB WPAN, various multimedia applications may be carried in the network. The traffic arrivals of multimedia application are generally bursty and correlated. Therefore, the resulting arrival process significantly deviates from the renewal process, where the arrivals in consecutive slots are independently and identically distributed (i.i.d.), e.g., the Bernoulli and Poisson processes. The nonrenewal arrival process resulting from multimedia traffic has a profound impact on the queueing statistics, as has been confirmed by many studies (see [6] and the references therein). While modeling the multimedia traffic as nonrenewal processes is preferable to capturing the real characteristic of multimedia applications than the renewal counterpart, the exact queueing analysis is quite difficult and generally incurs a high computational burden. An alternative is to seek some acceptable approximations with close enough performance characteristics to those of the original system.

In this paper, we study the performance of the PCA protocol in which the arrival process is bursty/correlated. The user traffic is classified into two classes, whereby multimedia traffic, such as voice and video streaming, has higher priority to access the channel than the data traffic such as file transfer has lower priority. We model the backoff and channel access behavior of a tagged station in each class and obtain the probability generating function (PGF) of the MAC service time distribution. The arrival process is described by a Markov Modulated

Manuscript received July 18, 2007; revised August 2, 2007. This work was supported by a research grant from the Natural Science and Engineering Research Council (NSERC) of Canada. The review of this paper was coordinated by Dr. P. Lin.

The authors are with the Centre for Wireless Communications, Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: k8liu@bbcr.uwaterloo.ca; x2ling@bbcr.uwaterloo.ca; xshen@bbcr.uwaterloo.ca; jwmark@bbcr.uwaterloo.ca).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVT.2007.912139

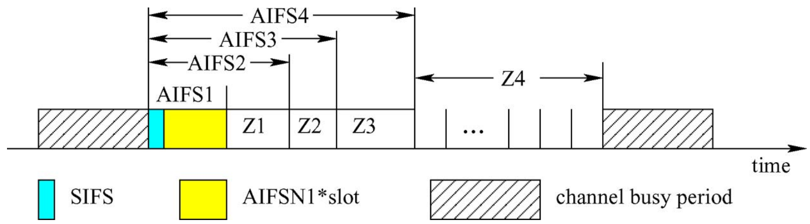


Fig. 1. PCA for different ACs.

Poisson Process (MMPP) for its versatility of modeling various traffic sources and the capability of capturing the burstiness and correlation in the arrival stream. The mean waiting time is obtained by three approximation approaches, and their accuracy is comprehensively studied.

The remainder of this paper is organized as follows. The PCA protocol and related work are briefly reviewed in Section II. The traffic and network models are described in Section III. The analysis of service time is presented in Section IV, followed by that of waiting time in Section V. Numerical results are given in Section VI. Section VII gives the concluding remarks.

II. PCA Protocol and Related Work

A. PCA Protocol

In PCA, the user traffic is differentiated into different access categories (ACs), such as voice, video, best effort, and background [4]. Each station regulates its frame transmission using the contention parameters associated with each AC. When a station has a frame at the MAC sublayer buffer, it will first sense the channel. If the channel is busy, it performs the backoff procedure by first setting the backoff counter to an integer sampled from the minimum contention window (CW) size. Therefore, the first differentiation mechanism is the assignment of higher priority ACs with a smaller value of minimum CW size such that the higher priority ACs statistically spend less time on backoff. After the channel becomes idle for an arbitrary interframe space (AIFS),[1] the station can count down the backoff counter at the beginning of each idle slot and the first slot of a channel busy period. Since the higher priority ACs are assigned with shorter AIFS, they obtain higher chances to access the channel than low-priority ACs. Fig. 1 shows an example of four ACs, where AC_1 has the highest priority. To illustrate the effect of different AIFS lengths, the time between two busy periods, except AIFS_1, is divided into four contention zones, Z_i, $i = 1, 2, 3, 4$. In Z_1, only the AC_1 stations are allowed to contend for channel access, whereas in Z_2, the competitions are between AC_1 and AC_2, i.e., contentions in Z_i involve AC_j, $j \leq i$. Consequently, each AC encounters different contentions in its allowable contention zones. After one station succeeds in contending for channel access, it can transmit for a duration up to the transmission opportunity (TXOP). Different TXOP durations can be assigned to different ACs to further differentiate the service.

[1]In [4], the length of AIFS is determined by $\mathrm{AIFS} = \mathrm{SIFS} + \mathrm{AIFSN} \times \sigma$, where $\mathrm{SIFS} = 10\ \mu\mathrm{s}$ is the short interframe spacing, AIFSN is an integer between [1, 7], and $\sigma = 9\ \mu\mathrm{s}$ is the slot time duration.

B. Related Work

The PCA defined in the WiMedia specification is a CSMA/CA-based MAC protocol with traffic prioritization. There has been a tremendous amount of research studying the performance of CSMA/CA protocols and its variants, such as the distributed coordination function (DCF) in IEEE 802.11 and the enhanced distributed channel access (EDCA) in IEEE 802.11e. Two major approaches have been employed in deriving the average MAC service time, namely the discrete Markov modeling [7]–[12] and the mean value analysis [13], [14]. Most of the work is concerned with the asymptotic performance, where each station in the network is saturated with traffic arrivals; thereby, the mean service time can be found equal to the reciprocal of the throughput. In practice, however, the station queues may not always be full; thus, the inverse relation between the average service time and throughput does not exist. Another approach has been proposed in [14], where the mean service time for both saturated and unsaturated stations can be successfully captured based on renewal theory.

Recently, the emergence of multimedia applications in the wireless domain has drawn much attention on studying the quality-of-service (QoS) provisioning for delay-sensitive traffic. In addition to the service time, the waiting time (i.e., queueing time) of a MAC frame has a significant impact on the delay performance, which is not only dependent on its service time that the network provides but is also affected by the incoming traffic characteristics. Several works on queueing analysis for DCF and EDCA have appeared, where the arrival process is always assumed uncorrelated [12], [15]. For multimedia traffic, however, the packet interarrivals are typically correlated and bursty in nature. In [16], a nonrenewal MMPP arrival process is considered, resulting in the use of MMPP$/G/1/K$ modeling. These studies obtain the collision probability as a function of the station idle probability (i.e., when the MAC buffer at the tagged station is empty), which is dependent on both the service time distribution and the characteristics of arrival process. Thus, the studies rely on certain recursive algorithms to find the collision probability, and the resultant computation is normally high. In addition, the impacts of burstiness and correlation in interarrival streams have not been explored; thus, their results may not be so useful for assessing the delay performance of multimedia traffic with bursty/correlated arrivals.

III. Traffic and Network Models

A. Incoming Traffic Model

Multimedia streams usually possess correlated and bursty characteristics that can significantly affect system performance (e.g., delay outage probability and throughput) [17], [18]. By burstiness, it is meant that one can observe the clustering phenomenon of arrivals on the time line [19]. A highly bursty arrival process tends to have a higher variance-to-mean ratio of the interarrival time. Letting X denote the interarrival time process, burstiness can be characterized by the squared coefficient of variation of the interarrival time [20], i.e.,

$$c^2 = \frac{\mathrm{Var}(X)}{\mathbb{E}^2(X)} \tag{1}$$

where $\mathbb{E}(X)$ and $\mathrm{Var}(X)$ are the mean and variance of the random variable X. The other important feature of multimedia traffic, particularly the variable bit rate streams, is the high correlation between interarrival times that produces long-range dependence into the arrival process and, hence, cumulative effect on the queueing system. The degree of correlation between interarrival times is typically measured by the correlation coefficient of X.

In this paper, the arrival process of multimedia traffic is represented by an MMPP. The reason for using MMPP is twofold. First, many studies have shown that MMPP has enough flexibility to describe a wide variety of traffic with correlated and bursty arrival processes, such as voice, video, and data [21]. Second, the queueing-related results of MMPP have been well studied [22]–[24]. Therefore, the use of MMPP offers versatility in the modeling environment and allows the achievement of analytical tractability while preserving the actual traffic characteristics [6].

The MMPP model is a nonrenewal doubly stochastic process where the rate process is determined by the state of a continuous-time Markov chain. An m-state MMPP is characterized by the following two elements: the infinitesimal generator $\mathbf{Q}$ given by

$$\mathbf{Q} = \begin{bmatrix} -\sigma_1 & \sigma_{12} & \cdots & \sigma_{1m} \\ \sigma_{21} & -\sigma_2 & \cdots & \sigma_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{m1} & \sigma_{m2} & \cdots & -\sigma_m \end{bmatrix} \tag{2}$$

where $\sigma_i = \sum_{j=1,j\neq i}^{m} \sigma_{ij}$, and σ_{ij} governs the transition rate from state i to state j; and the Poisson arrival rate matrix $\mathbf{\Lambda}$ given by

$$\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_m) \tag{3}$$

where λ_i is the rate of a Poisson arrival process at state i of the Markov chain. The steady-state probability vector $\mathbf{\Pi}$ of the Markov chain can be determined using the relations

$$\mathbf{\Pi Q} = \mathbf{\Pi} \tag{4}$$

$$\mathbf{\Pi e} = 1 \tag{5}$$

where $\mathbf{e} = (1, 1, \ldots, 1)^T$.

B. Network Model

We consider a network with two classes of stations (N_i stations in AC_i, $i = 1, 2$). Without loss of generality, let the AC_1 stations have high priority and the AC_2 stations have low priority in accessing the channel. The operation of the beacon group specified in PCA ensures that there are no hidden terminals in the network [4]. Time is discretized into *generic slots* denoted as ϕ, which may have different lengths Δ, T_s, and T_c, which correspond to the different channel status of idle, successful transmission, and collision, respectively. In addition, all the stations are synchronized, and they can correctly sense the channel status at the beginning of the slots. An ideal wireless channel without transmission error is assumed so that all the transmitted frames may be lost only due to collisions caused by simultaneous transmissions from multiple stations. The effect of imperfect channels can be embedded in our analysis following the approach presented in [25]. For simplicity, all the MAC frames are assumed to have the same fixed length. The case of different frame lengths (and thus the analysis of TXOP) can be incorporated in our model following the work in [26].

IV. MAC Service Time Analysis

This section derives the probability distribution of the MAC service time, and the result will be used in Section V for waiting time analysis. The modeling of MAC performance heavily relies on two key probabilities, i.e., the station transmitting probability and the frame collision probability, conditioned on there being at least one frame in the station's buffer to be served. For an AC_i station, the former is denoted by τ_i, and the latter is denoted by p_i. Considering a lossless queueing system, the probability of a nonempty buffer is given by the server utilization factor $\rho = \lambda_a \cdot \overline{Z}$, where λ_a is the mean frame arrival rate, and $\overline{Z}$ is the mean frame service time. The probability that an unsaturated AC_i station transmits in a randomly chosen generic slot is thus $\tau_i\rho_i, \rho_i \in (0, 1]$. We follow the approach proposed in [14] to obtain p_i and τ_i. With the probabilities τ_i and p_i, we then proceed to derive the PGF (equivalently, the Z-transform) of the MAC service time for both classes. By numerical inversion of the Z-transform, the probability mass function (PMF) and the corresponding moments can then be obtained.

A. Transmission and Collision Probabilities

We assume that the probability of a station to initiate a transmission in a given backoff slot is constant in all its backoff slots [8], [11]. Since the channel access procedure of the tagged station regenerates itself for each new MAC frame, the complete service periods for MAC frames form renewal cycles in the renewal process. The average length of the renewal cycle is thus the average frame service time [14]. According to the renewal reward theorem, in a randomly chosen slot, the transmitting probability τ_i of an AC_i station can be obtained as the average reward during the renewal cycle, i.e.,

$$\tau_i = \frac{\mathbb{E}[R_i]}{\mathbb{E}[R_i] + \mathbb{E}[B_i]} \tag{6}$$

where $\mathbb{E}[R_i]$ is the expected number of transmission trials for a frame, and $\mathbb{E}[B_i]$ is the expected number of total backoff slots experienced by the frame. Assuming an average collision probability of p_i for the frames of AC_i stations, R_i follows a truncated geometric distribution, and $\mathbb{E}[R_i]$ is given by

$$\mathbb{E}[R_i] = \sum_{j=0}^{m-1} (p_i)^j. \quad (7)$$

Similarly, $\mathbb{E}[B_i]$ can be obtained as

$$\mathbb{E}[B_i] = \sum_{j=0}^{m-1} b_j (p_i)^j \quad (8)$$

where $b_j = CW_j/2$ is the average number of backoff slots in the backoff stage j, $j = 0, \dots, m$, and m is the retry limit, as stated earlier. Notice that the class-dependent CW parameters have been included in the analysis. The collision probability of AC_2 can be obtained by

$$p_2 = 1 - (1 - \rho_1\tau_1)^{N_1}(1 - \rho_2\tau_2)^{N_2-1} \quad (9)$$

considering that the AC_2 station can only transmit in zone 2 with possible collisions with one or more of the other stations from any class. The computation of the collision probability of AC_1 is more involved, as its transmissions may take place in either zone 1 or zone 2 with collision probabilities $p_{1,1}$ and $p_{1,2}$, respectively, where

$$p_{1,1} = 1 - (1 - \rho_1\tau_1)^{N_1-1} \quad (10)$$
$$p_{1,2} = 1 - (1 - \rho_1\tau_1)^{N_1-1}(1 - \rho_2\tau_2)^{N_2}. \quad (11)$$

Supposing that zone 1 contains M slots (the difference between AIFSN_1 and AIFSN_2), the frame transmission may take place in zone 2 if neither itself nor any of the other AC_1 stations transmit in zone 1 with probability denoted by θ_2, i.e.,

$$\theta_2 = \left((1 - \tau_1\rho_1)^{N_1-1}(1 - \tau_1)\right)^M. \quad (12)$$

Otherwise, the transmission occurs in zone 1 with probability $\theta_1 = 1 - \theta_2$. The average collision probability of an AC_1 station can thus be given by

$$p_1 = \theta_1 p_{1,1} + \theta_2 p_{1,2}. \quad (13)$$

By jointly solving (6), (9), and (13), we can obtain $(\tau_1, \tau_2, p_1, p_2)$.

B. PGF of Frame Service Time

Now we proceed to derive the PGF of the frame service time. Similar to [27], we work on a discrete-time system where the time interval in our analysis is approximated as multiples of a common quantity, representing the smallest granularity that can be observed by our model. Thus, the frame service time is a discrete random variable and leads to a Z-transform-based analysis. For a tagged station of AC_i, it spends an amount of time $Z_i = B_i + R_i$ to successfully transmit a frame, where $B_i(R_i)$ is the random variable representing the amount of time attributed to backoff (transmission trials). Moreover, the introduction of AIFS causes further delay to AC_2 stations, as explained in Section II. This additional amount of time is referred to as a "pre-backoff waiting" period [14], which is denoted as Z'. Therefore, the PGF of the frame service time can be written as

$$G_{Z_1}(z) = G_{B_1}(z)G_{R_1}(z)$$
$$G_{Z_2}(z) = G_{B_2}(z)G_{R_2}(z)G_{Z'}(z). \quad (14)$$

In the following, we derive each component in $G_{Z_i}(z)$, where the subscript i will be omitted for notation brevity.

1) $G_\phi(z)$: The time unit is measured in a generic slot ϕ, as defined in Section III-B. For a randomly chosen slot, the channel status may be in one of the following three mutually exclusive events: being idle (I); having a successful transmission (S); or having a collision (C). The length of a generic slot ϕ can be expressed as

$$\phi = \Omega_I \sigma + \Omega_S T_s + \Omega_C T_c \quad (15)$$

where Ω_e is a binary variable that takes the value of 1 if the event $e \in \{\mathrm{I}, \mathrm{S}, \mathrm{C}\}$ occurs, and zero otherwise. Thus, the PGF of ϕ takes the form

$$G_\phi(z) = p_I z^\sigma p_S G_{T_s}(z) p_C G_{T_c}(z) \quad (16)$$

where p_I, p_S, and p_C are class-dependent, as given by

$$p_{I,1} = (1 - \rho_1\tau_1)^{N_1} \quad (17)$$
$$p_{I,2} = (1 - \rho_1\tau_1)^{N_1}(1 - \rho_2\tau_2)^{N_2} \quad (18)$$
$$p_{S,1} = N_1\rho_1\tau_1(1 - \rho_1\tau_1)^{N_1-1} \quad (19)$$
$$p_{S,2} = N_1\rho_1\tau_1(1 - \rho_1\tau_1)^{N_1-1}(1 - \rho_2\tau_2)^{N_2} \quad (20)$$
$$+ N_2\rho_2\tau_2(1 - \rho_2\tau_2)^{N_2-1}(1 - \rho_1\tau_1)^{N_1} \quad (21)$$
$$p_{C,i} = 1 - p_{I,i} - p_{S,i}, \qquad i = 1, 2. \quad (22)$$

2) Backoff Period $G_B(z)$: Between two successful transmissions, the time contributed by backoff is

$$B = \sum_{j=1}^{N_C} \phi_j \quad (23)$$

where N_C is the overall number of generic slots between two successful transmissions, given that a frame transmission undergoes C trials ($C \in \{1, 2, \dots, m\}$) before success, and ϕ_j is the length of the jth generic slot (we assume ϕ is an i.i.d. random variable). Using the conditional expectation, the PGF of B can be written as

$$G_B(z) = \mathbb{E}\left[z^{\sum_{j=1}^{N_C}\phi_j}\right] = \sum_{c=1}^{m} \mathbb{E}\left[z^{\sum_{j=1}^{N_c}\phi_j}\right]\mathbb{P}\left[C = c\right]. \quad (24)$$

Similar to the argument of R_i in (7), C is a geometric random variable with successful probability $1 - p$, i.e., $\mathbb{P}[C = c] = p^{c-1}(1 - p)$. For the first term in (24), the sum of a random number N_C of i.i.d. random variables ϕ, $S_{N_C} = \sum_{j=1}^{N_C} \phi_j$ has the property $G_{S_{N_C}}(z) = G_{N_C}(G_\phi(z))$. Consequently, we can obtain $G_B(z)$ as

$$G_B(z) = (1-p)\sum_{c=1}^{m} p^{c-1} G_{N_C}\left(G_\phi(z), c\right) \tag{25}$$

where we use the notation $G_{N_C}(G_\phi(z), c)$ to indicate that it is a function of c, which can be derived according to [27]. Let x_j be the number of generic slots contained in backoff stage j, $j = 0, \ldots, C-1$. According to the exponential binary backoff, x_j is uniformly distributed over $[0, CW_j - 1]$, where $CW_j = \min\{2^j CW_0, 2^{m'} CW_0\}$, with m' being the maximum backoff stage. The PGF of x_j can be derived as

$$G_{x_j}(z) = \sum_{k=0}^{CW_j - 1} \frac{z^k}{CW_j} = \frac{1 - z^{CW_j}}{(1-z)CW_j}. \tag{26}$$

The random variable N_C can be expressed as

$$N_C = \sum_{j=0}^{C-1} x_j \tag{27}$$

such that the corresponding PGF is given by

$$G_{N_C}(z, c) = \frac{1}{(1-z)^{c+1}} \prod_{j=0}^{c-1} \frac{1 - z^{CW_j}}{CW_j}. \tag{28}$$

3) Retry Period $G_R(z)$: Given that there are C transmission trials encountered before a successful frame transmission, the random variable R representing the total time contributed by transmission trials can be written as

$$R = \sum_{j=1}^{C} \phi_j. \tag{29}$$

Therefore, the PGF of R is given by

$$G_R(z) = G_C\left(G_\phi(z)\right). \tag{30}$$

According to the fact that C follows a truncated geometric distribution, its PGF is derived as

$$\begin{aligned} G_C(z) &= \sum_{k=1}^{m} z^k \mathbb{P}[C = k] + p^m z^m \\ &= (1-p)\sum_{k=1}^{m} z^k p^{k-1} + (pz)^m. \end{aligned} \tag{31}$$

4) Pre-Backoff Period $G_U(z)$: For AC_2, it undergoes pre-backoffs that introduce further waiting time U. Using a similar argument, we can derive $G_U(z)$ as (see Appendix)

$$G_U(z) = (1-p)\sum_{c=1}^{m} p^{c-1} G_{N_{U(c)}}(G_\eta(z)) + (pz)^m z^{\Delta}. \tag{32}$$

Thus far, we have derived the PGF of the frame service time for each priority class. However, it is often very difficult, or even impossible, to analytically invert the Z-transform of a discrete probability distribution. Several numerical inversion algorithms have been proposed to address this difficulty. Next, we employ the approach in [28] and [29] to obtain the nth moment of a discrete random variable from its PGF.

C. Numerical Evaluation of the Frame Service Time

The PMF of the frame service time Z_i, $i = 1, 2$, which was derived in the previous subsection, can be obtained by the numerical algorithm reported in [29]

$$\begin{aligned} Z_i(k) = \frac{1}{2klr^k} \Bigg[&\beta_0(k,l,r) + (-1)^k \beta_k(k,l,r) \\ &+ 2\sum_{j=1}^{k-1} (-1)^j \mathrm{Re}\Big(\beta_j(k,l,r)\Big) \Bigg] \end{aligned} \tag{33}$$

where $\beta_j(k,l,r) = \sum_{j_1=0}^{l-1} e^{-\pi i j_1/l} Z(\mathrm{r}e^{\pi \mathrm{j}(j_1 + l j_2)/lk})$, $j = \sqrt{-1}$, $1 \le j_2 \le k$ for real r and integer l. As suggested in [29], the algorithm can achieve a low error estimate (less than 10^{-8}) by setting $l = 1$ and $r = 10^{-4/k}$, reducing to the simplified formula

$$\begin{aligned} Z_i(k) = \frac{1}{2kr^k} \Bigg[&Z_i(r) + (-1)^k Z(\mathrm{r}e^{\pi i}) \\ &+ 2\sum_{j=1}^{k-1} (-1)^j \mathrm{Re}\left(Z(\mathrm{r}e^{\pi i j/k})\right) \Bigg]. \end{aligned} \tag{34}$$

The nth moment μ_n is obtained by numerically inverting $Z(z')$, $z' = e^z$ [28], i.e.,

$$\begin{aligned} \mu_n = \frac{n!}{2nlr_n^n} \Bigg\{ &Z(r_n) + (-1)^n Z(-r_n) \\ &+ 2\sum_{j=1}^{nl-1} \mathrm{Re}\left[Z(r_n e^{\pi i j/nl}) e^{\pi i j/l}\right] \Bigg\} - \bar{e}. \end{aligned} \tag{35}$$

V. Mean Waiting Time Analysis

Our mean waiting time analysis is obtained by modeling each station as a $G/G/1$ queue. It is well known that there is no exact expression for the mean waiting time of the $G/G/1$ queue. In what follows, we consider three approximate queueing systems and summarize them in Table I.

A. MMPP/G/1

The MMPP/$G/1$ model is parameterized by the service time distribution and its Laplace–Stieltjes transform $\tilde{h}(s)$. The arrival process is parameterized by $\mathbf{Q}$ and $\mathbf{\Lambda}$ (see Section III-A). The mean waiting time $\overline{W}$ can be found as [30, Sec. 3.1.4.1]

$$\begin{aligned} \overline{W} = \frac{1}{\rho}\Bigg[\frac{1}{2(1-\rho)}\Bigg[&2\rho + \lambda_a h^{(2)} - 2h^{(1)}\Big((1-\rho)\boldsymbol{g} + h^{(1)}\mathbf{\Pi\Lambda}\Big) \\ &\times\ (\mathbf{Q} + \mathbf{e}\Pi)^{-1}\boldsymbol{\lambda}\Bigg] - \frac{1}{2}\lambda_a h^{(2)}\Bigg] \end{aligned} \tag{36}$$

TABLE I
CONSIDERED QUEUEING SYSTEMS

Notation	Queue	Service time distribution	Mean waiting time
Q^{Γ}_{MMPP}	*MMPP*/ Γ/ 1	Gamma $f_{\Gamma}(x) = \frac{\beta^{\alpha} x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}$ $\tilde{h}_{\Gamma}(s) = (\frac{\beta}{s+\beta})^{\alpha}$	Eq. (36)
Q^{M}_{MMPP}	*MMPP*/ M/ 1	Exponential $f_M(x) = 1/m_1 e^{-x/m_1}$ $\tilde{h}_{\Gamma}(s) = (\frac{\beta}{s+\beta})^{\alpha}$	Eq. (36)
Q^{Γ}_{heavy}	*MMPP*/ Γ/ 1 ($\rho \to 1$)	Gamma	Eq. (37)
Q^{M}_{heavy}	*MMPP*/ M/ 1 ($\rho \to 1$)	Exponential	Eq. (37)
Q^{M}_{PMRQ}	*PMRQ*/ M/ 1	Exponential	Eq. (40)

where $\boldsymbol{g}$ is a vector that can be obtained by the iterative algorithm provided in [30, Sec. III-B].

We consider two approximations for the service time, i.e., the exponential and gamma distributions. They are considered because the Laplace transform of either has a closed-form expression, which is required in computing the vector $\boldsymbol{g}$ in (36). In addition, these two distributions are representative in the sense that they rely on different orders of moments to model the distribution, that is, the exponential distribution can be modeled just by the first-order statistics, whereas the gamma distribution needs the first two moments to describe its distribution. Thus, we have two queueing systems, i.e., the MMPP arrival process with gamma service time, which is denoted as $Q^{\Gamma}_{\mathrm{MMPP}}$, and the MMPP arrival process with exponential service time, denoted as Q^{M}_{MMPP}. Their distributions and corresponding Laplace transforms are listed in Table I, where m_1 and m_2 represent the first two moments of the service time obtained from (35), $\alpha = m_1^2/(m_2 - m_1^2)$, and $\beta = m_1/(m_2 - m_1^2)$.

B. Heavy Traffic Approximation

For the heavy traffic case (or saturated stations), i.e., $\rho \to 1$, $\overline{W}$ can be approximated by [31]

$$\overline{W} \approx \frac{\rho}{1-\rho} \frac{h^{(1)} \left(c_X^2 + c_Y^2\right)}{2} \tag{37}$$

where $c_X = \lambda_X \sigma_X$ and $c_Y = \lambda_Y \sigma_Y$ denote the coefficients of variation of the interarrival time and service time, respectively. Combining with the exponential service time approximation, we obtain the queueing systems Q^{M}_{heavy}. Likewise, $Q^{\Gamma}_{\mathrm{heavy}}$ represents the heavy traffic approximation with gamma-distributed service time.

C. PMRQ Approximation

The exact analysis of the queueing system with autocorrelated arrival processes (thus nonrenewal) is generally hard and incurs a high computational burden. Approximating the nonrenewal arrival process by a renewal counterpart is a commonly used approach to deal with the complex queueing system. Recently, Jagerman *et al.* [6] proposed a renewal approximation to analyze delay systems with autocorrelated arrival processes. The property of the correlated interarrival time is first captured by the peakedness function, as defined in [32]. By mapping a $G/G/1$ queue to an approximating $GI/G/1$ queue called the Peakedness Matched Renewal Queue (PMRQ), which preserves the peakedness of the original arrival process and its arrival rate, it is shown that the approximate $GI/G/1$ queue achieves close enough performance measures to those of the original system. In this paper, we adopt the peakedness matching technique proposed in [6] to estimate the mean waiting time in our system, leading to the $Q^{\Gamma}_{\mathrm{PMRQ}}$ approximation with gamma service time distribution and the Q^{M}_{PMRQ} with exponential service time distribution. Note that the PMRQ approximation has also been applied to a recent work [33], studying the impact of correlated wireless channel variations to queueing systems. In the following, we give the gist of the PMRQ approximation relevant to our study.

The object of the PMRQ approximation is to approximate a general arrival process X by a renewal process X', considering the fact that X' is generally analytically simpler than X. The approximation is achieved by matching the peakedness function of X, which is denoted as $z_X(s)$, to that of X', which is denoted as $z_{X'}(s)$. It has been shown that the Laplace transform of X' takes the form of

$$\tilde{a}_{X'}(s) = \frac{\lambda_X \alpha_E + (\lambda_X + A_E \alpha_E) s}{\lambda_E \alpha_E + (\lambda_E + \alpha_E + A_E \alpha_E) s + s^2}, \qquad s \geq 0 \tag{38}$$

where λ_X is the average arrival rate of X, and A_E and α_E are estimated from $z_X(s)$. To obtain the mean waiting time, first consider the complementary stationary distribution of the waiting time W that is asymptotically approximated as

$$\mathbb{P}[W > t] \approx \Gamma_W e^{-\theta_W t}, \qquad t \geq 0 \tag{39}$$

where Γ_W is referred to as the asymptotic coefficient, and θ_W is called the critical decrement. The corresponding mean waiting time can be approximated by

$$\overline{W} \approx \frac{\Gamma_W}{\theta_W}. \tag{40}$$

Given the Laplace transform of the approximate renewal process $\tilde{a}_{X'}(s)$ and that of the service time distribution $\tilde{h}(s)$, one can compute θ_W as the smallest positive root of

$$\tilde{a}_{X'}(\theta) \tilde{b}(-\theta) = 1, \quad \theta > 0. \tag{41}$$

On the other hand, the asymptotic coefficient Γ_W can be found from

$$\Gamma_W = 2\frac{\tilde{k}_+(0) - \tilde{k}_+(\theta_W)}{\tilde{k}_+(-\theta_W) - \tilde{k}_+(\theta_W)} \tag{42}$$

where $\tilde{k}_+$ can be obtained by decomposing the kernel transform

$$\tilde{k}(s) = \tilde{a}_{X'}(-s)\tilde{h}(s) \tag{43}$$

into $\tilde{k}(s) = \tilde{k}_-(s) + \tilde{k}_+(s)$. By inserting $\tilde{a}_{X'}(s)$ and $\tilde{h}(s)$ into (43) and using the partial fraction decomposition technique, one can obtain the decomposition of $\tilde{k}(s)$ as

$$\tilde{k}_+(s) = \tilde{k}(s) - \tilde{k}_-(s) \tag{44}$$

$$\begin{aligned}\tilde{k}_-(s) = &\frac{\lambda_X\alpha_E - (\lambda_X + A_E\alpha_E)r_1}{r_1 - r_2}\cdot\frac{\tilde{h}(r_1)}{s - r_1}\\ &+ \frac{\lambda_X\alpha_E - (\lambda_X + A_E\alpha_E)r_2}{r_2 - r_1}\cdot\frac{\tilde{h}(r_2)}{s - r_2}\end{aligned} \tag{45}$$

where (r_1, r_2) are the roots of the quadratic function $\lambda_X\alpha_E - (\lambda_X + \alpha_E + A_E\alpha_E)s + s^2 = 0$. Notice that a typo in [6, eq. (9.5)] has been fixed here.

VI. Numerical Results and Discussions

In this section, we first validate the efficacy of our analytical results through simulations. We then study the effects of traffic characteristics, namely the burstiness and correlations, to the MAC layer performance. We focus on the temporal performance metrics, i.e., frame service time and waiting time, whereas other metrics, such as throughput and efficiency, can be readily obtained but are omitted here due to space limitation.

The traffic arrival process is modeled by a two-state MMPP, which has been widely used as a building block for the construction of various multimedia sources such as voice, video, and Internet traffic [34], [35, Sec. IV-A]. The use of the two-state MMPP model also enables simple and explicit forms of important parameters that facilitate our demonstration. A two-state MMPP is characterized by the infinitesimal generator $\mathbf{Q} = [q_{ij}]$ given by

$$\mathbf{Q} = \begin{bmatrix} -\sigma_1 & \sigma_1 \\ \sigma_2 & -\sigma_2 \end{bmatrix}, \tag{46}$$

a diagonal matrix $\mathbf{\Lambda}$ of the Poisson arrival rates given by

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \tag{47}$$

and the initial probability vector

$$\pi_0 = \frac{1}{\lambda_1\sigma_2 + \lambda_2\sigma_1}[\lambda_1\sigma_2 \quad \lambda_2\sigma_1]. \tag{48}$$

The steady-state probability vector $\mathbf{\Pi}$ is given by

$$\mathbf{\Pi} = (\pi_1, \pi_2) = \frac{1}{\sigma_1 + \sigma_2}(\sigma_2, \sigma_1). \tag{49}$$

TABLE II
Parameters Used in the Performance Evaluation

Channel rate	110 Mbps	Retry limit $[m]$	7
Slot time $[\sigma]$	9 μs	Max. backoff stage $[m']$	6
SIFS	10 μs	Min. contention window	32
PHY header	13.125 μs	Frame payload	500 bytes

The mean arrival rate λ_a is given as

$$\lambda_a = \frac{\sigma_1\lambda_2 + \sigma_2\lambda_1}{\sigma_1 + \sigma_2}. \tag{50}$$

In addition, two parameters are used to describe the burstiness and the correlation of the arrival process. The burstiness is characterized by the squared coefficient of variation of the interarrival time c^2, as defined in (1), and has the form [36]

$$c^2 = 1 + \frac{2\sigma_1\sigma_2(\lambda_1 - \lambda_2)^2}{(\sigma_1 + \sigma_2)^2(\lambda_1\lambda_2 + \lambda_1\sigma_2 + \lambda_2\sigma_1)}. \tag{51}$$

The one-step correlation coefficient r_1 is used to describe the correlation between interarrival times, as given by [36]

$$\begin{aligned} r_1 &= \frac{\mathbb{E}\left[(t_{n-1} - \mathbb{E}[X_{t-1}])\,(t_n - \mathbb{E}[t_n])\right]}{\mathrm{Var}[t_n]} \\ &= \frac{\lambda_1\lambda_2(\lambda_1 - \lambda_2)^2\sigma_1\sigma_2}{c^2(\sigma_1 + \sigma_2)^2(\lambda_1\lambda_2 + \lambda_1\sigma_2 + \lambda_2\sigma_2)^2} \end{aligned} \tag{52}$$

where t_n denotes the nth interarrival time. Based on the interrelation between c^2 and r_1, we can generate the arrival processes with the same mean arrival rate but different bursty/correlation characteristics, as suggested in [36]. In our experiments, we fix $\lambda_1 = 1$ and find the corresponding MMPP parameters (σ_1, σ_2) as a function of λ_2 from (50) and (51). Subsequently, the relation between r_1 and λ_2 can be obtained by (52). The value of c^2 is chosen from $\{2, 10, 20\}$, which represents different degrees of burstiness. It is reported in [37] that $c^2 = 18.1$ is very large compared to that of a Poisson process, which has a c^2 value of 1.0. The corresponding correlation r_1 is then obtained as long as the inequality $\lambda_i < \sigma_i$ is satisfied.

The PCA protocol in [4] is simulated using our event-driven simulator. All the numerical results reported here are obtained based on the PHY and MAC parameters listed in Table II. Both RTS/CTS handshake and the contention-free burst functionality [4] are disabled. Because of space limitation, we fix the minimum CW size for all ACs and only report the results relevant to the impact of AIFS. In all the experiments, we consider the following setting: The number of stations $N_1 = N_2 = 5$; each AC_1 station carries a traffic flow driven by the two-state MMPP with the same parameters r_1 and c^2; and AC_2 stations are saturated such that there are always frames in their MAC buffers. Such a setting mimics the scenario where the station carrying multimedia traffic has a higher priority, and the traffic delivered by other stations is considered as the background traffic with low priority. We are interested in the mean waiting time and mean service time of the high-priority multimedia traffic.

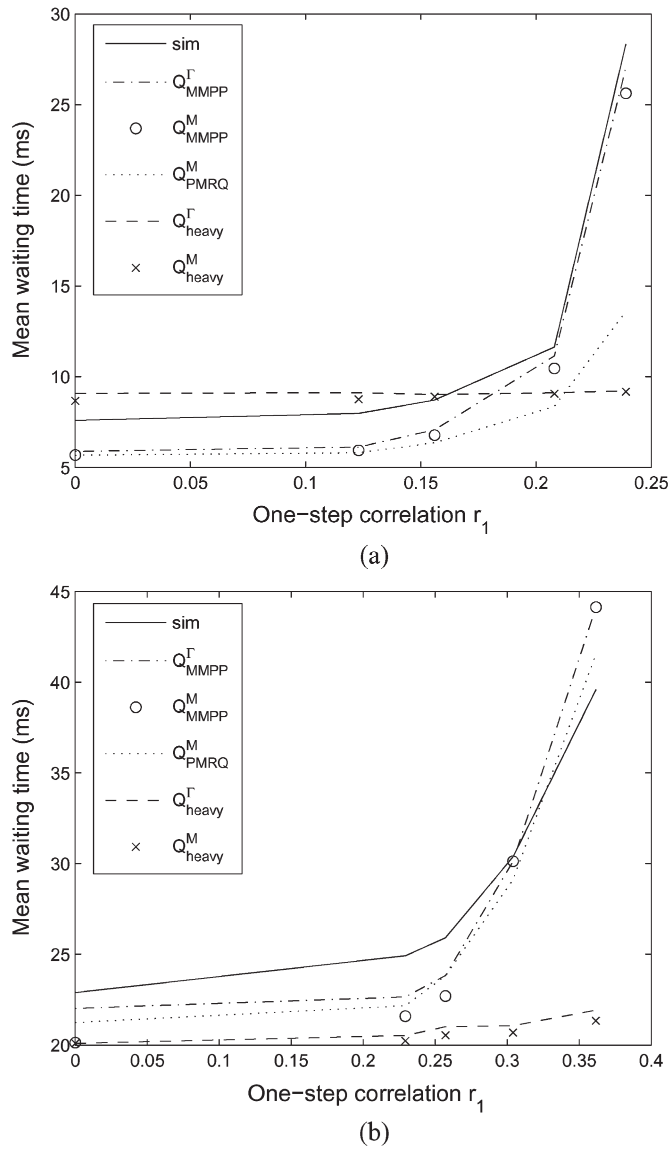


Fig. 2. Comparisons of mean waiting time obtained from simulations and from the approximated queueing systems (listed in Table I) in different degrees of burstiness, with $\lambda_a = 0.6$, and $M = 2$. (a) Low bursty traffic $c^2 = 2$. (b) High bursty traffic $c^2 = 10$.

A. Model Validation

To verify the efficacy of our analysis, we compare the mean waiting time obtained from simulations and those obtained from the aforementioned approximation methods, as summarized in Table I. We fix the mean arrival rate of $\lambda_a = 0.6$ and consider two burstiness levels, i.e., $c^2 = 2$ and $c^2 = 10$. Fig. 2(a) displays the result of a low bursty case with $c^2 = 2$. The simulation results show that the mean waiting time tends to increase as the one-step correlation increases, and a rapid increase can be found for $r_1 > 0.2$. It can be seen that $Q^{\Gamma}_{\mathrm{MMPP}}$ and Q^{M}_{MMPP} can reasonably capture this increasing trend, whereas $Q^{\Gamma}_{\mathrm{MMPP}}$ slightly outperforms Q^{M}_{MMPP}. Q^{M}_{PMRQ} performs similar to the previous two approximations for low and medium correlation r_1 and loses its accuracy for a high correlation range. The heavy traffic approximation performs close to the simulation results for low and medium r_1, but the flat curve indicates that this approximation cannot properly reflect the impact of correlation (here the server utilization factor ρ is about 0.7). Fig. 2(b) displays the results of higher bursty traffic with $c^2 = 10$. Similar to the low bursty case, $Q^{\Gamma}_{\mathrm{MMPP}}$ and Q^{M}_{MMPP} well approach the simulated mean waiting time curve for all ranges of the correlation r_1. Q^{M}_{PMRQ} performs very close to the previous two approximations in this setting. Again, the heavy approximation does not effectively reflect the impact of correlation in interarrival times. The above results suggest that $Q^{\Gamma}_{\mathrm{MMPP}}$ and Q^{M}_{MMPP} can capture the impact of traffic characteristics to the mean waiting time with reasonable accuracy. In particular, the former performs slightly better than the latter. For brevity, in what follows, we will only report the results for the gamma approximation.

1) Remark on the Q^{M}_{PMRQ}: It can be found that Q^{M}_{PMRQ} is also effective in responding to the effect of bursty/correlation in the arrival process. Although Q^{M}_{PMRQ} tends to underestimate the mean waiting time for low and medium levels of correlation, the use of *GI* arrival approximation helps to reduce the computational burden. Its inaccuracy should be due to the peakedness function obtained from the exponential service time approximation.

B. Burstiness/Correlation Versus Mean Waiting Time

To further explore this performance characteristic, we present the results of different traffic densities, i.e., $\lambda_a = 0.3$ in Fig. 3(a) and $\lambda_a = 0.6$ in Fig. 3(b), respectively. Comparing the effects of burstiness c^2 and correlation r_1, both figures show that, for low and medium correlation r_1, the traffic burstiness dominates the mean waiting time. For highly correlated traffic, the mean waiting time exponentially grows, which implies that the correlation r_1 between interarrival times has stronger effects on the mean waiting time. These results confirm the importance of taking into account the second-order statistics (e.g., burstiness/correlation) of the multimedia traffic in estimating the mean waiting time. On the other hand, according to our simulation results, the mean service time is not sensitive to the burstiness/correlation properties of interarrivals. Hence, the assumption of a Poisson arrival process, which has $r_1 = 0$ and $c^2 = 1$, is reasonably valid to obtain the mean service time estimation. However, this assumption greatly underestimates the mean waiting time of the incoming traffic with bursty/correlated arrivals and, thus, compromises its usage in evaluating the multimedia traffic performance. For instance, the video traffic generally has a strict delay bound, where a video frame may become useless if it cannot arrive at the decoding buffer in time. Proactively dropping the video frame that has a high probability of exceeding the deadline has been an effective approach to improve the video quality and bandwidth utilization in wireless transmissions [38]. In this context, an accurate estimate about the mean frame waiting time can assist in designing an effective transmission policy.

C. Impact of AIFS

The impact of AIFS on the mean waiting time is investigated, and the results are shown in Fig. 4(a) and (b) for $\lambda_a = 0.3$ and $\lambda_a = 0.6$, respectively. The label of the horizontal axis

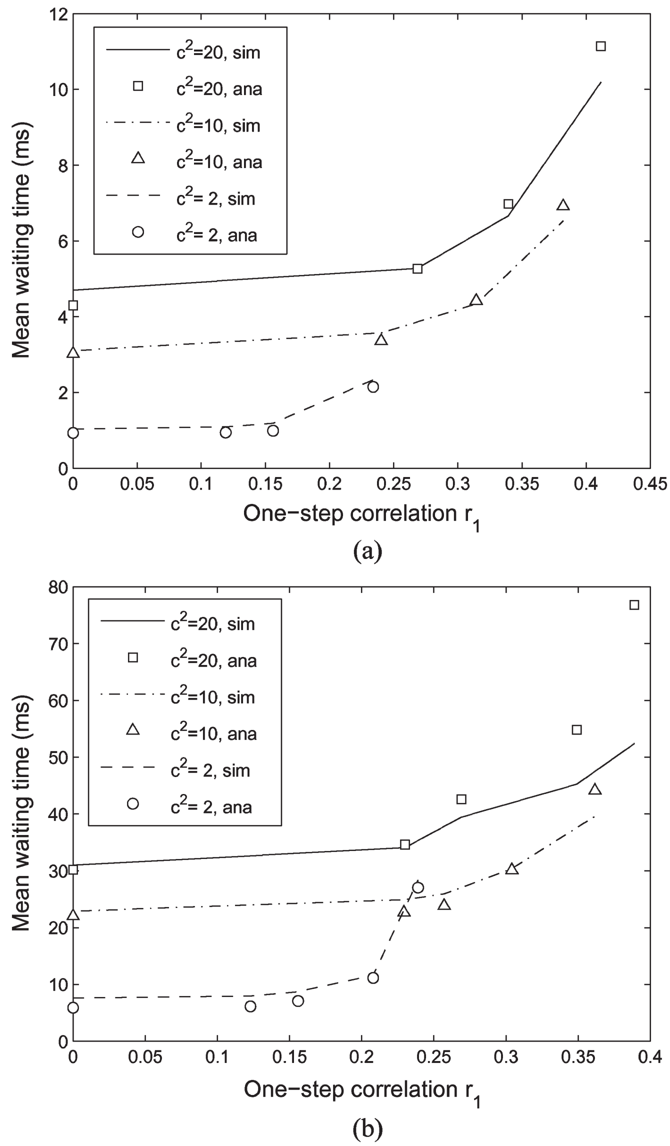


Fig. 3. Impact of burstiness and correlation in interarrival times to the mean waiting time of AC_1 in different traffic loads, with $M = 2$. (a) Low traffic load $\lambda_a = 0.3$. (b) High traffic load $\lambda_a = 0.6$.

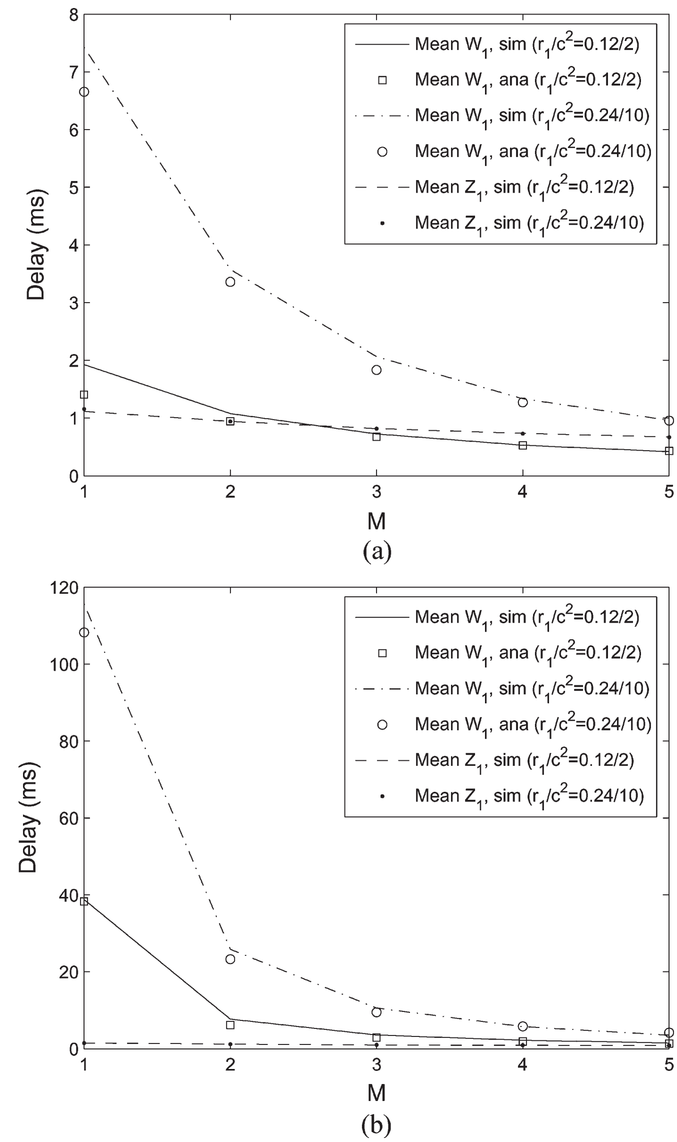


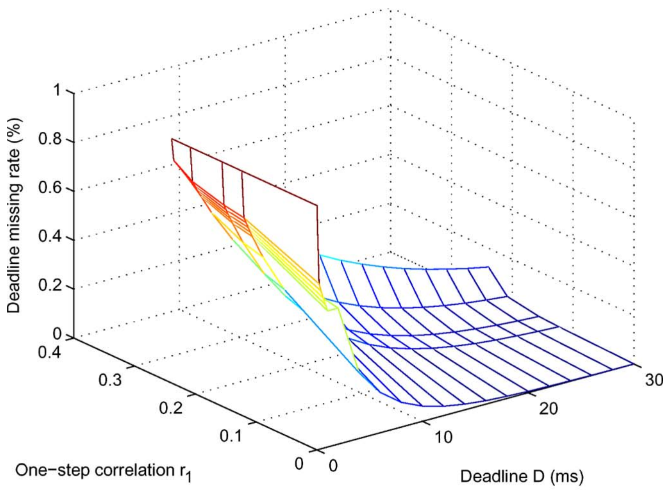
Fig. 4. Impact of AIFS on mean waiting time with $N_1 = N_2 = 5$, where $M = \mathrm{AIFS}N_2 - \mathrm{AIFS}N_1$. (a) Low traffic load $\lambda = 0.3$. (b) High traffic load $\lambda = 0.6$.

M represents the difference between AIFS_1 and AIFS_2, and a larger M provides more protection to AC_1 transmissions. We compare the mean waiting time W_1 of AC_1 resulting from two scenarios, i.e., low correlated/bursty interarrivals (i.e., $r_1 = 0.12, c^2 = 2$) and high correlated/bursty interarrivals (i.e., $r_1 = 0.24, c^2 = 10$). We also report the mean service time Z_1 of AC_1, obtained from simulations, to demonstrate the effect of AIFS differentiation.

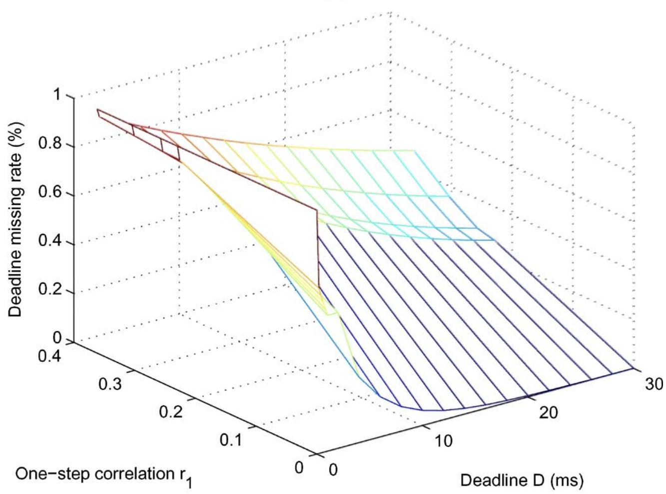
We have the following observations. 1) The descending trend in both figures shows that, although setting a larger M can help to reduce the mean waiting time of AC_1, the achieved gain is most significant when M is increased from 1 to 2, and its strength is reduced for a larger M. On the other hand, the results in [14] have shown that increasing M could remarkably degrade the throughput of low-priority AC_2 stations, whereas the increase of AC_1's throughput is minor. Hence, this gross observation suggests that a conservative setting of AIFS should be considered in differentiating the TXOP of high-priority traffic from low-priority traffic. 2) A traffic with higher bursty/correlation levels is more sensitive to AIFS differentiation. Take the low traffic load case in Fig. 4(a), for example. W_1 drops by about 87% from $M = 1$ to $M = 5$ for the high correlated/bursty scenario, whereas the reduction is by about 78% for the low correlated/bursty scenario. 3) The effect of AIFS is magnified for a higher traffic load. Consider the highly correlated/bursty interarrivals case for instance. Increasing M from 1 to 2 yields about a 52% decrease in mean waiting time W_1 for $\lambda_a = 0.3$ [Fig. 4(a)], whereas it is about 78% for $\lambda_a = 0.6$ [Fig. 4(b)]. The above observations indicate that dynamically changing the contention parameters should be beneficial to improving the QoS provisioning for multimedia traffic using the PCA protocol.

D. Potential Application

Finally, we present a potential application of our analysis. For a multimedia traffic sensitive to delay, the deadline missing



(a)



(b)

Fig. 5. Dead line rate (DMR) versus one-step correlation (r_1) for different degrees of burstiness, with $N_1 = N_2 = 5$, $M = 2$, and $\lambda_a = 0.6$. (a) Low bursty traffic $c^2 = 2$. (b) High bursty traffic $c^2 = 10$.

ratio (DMR) is a useful temporal metric in characterizing the QoS provisioning. DMR is defined as the probability that the frame waiting time in the MAC buffer exceeds a predefined deadline, i.e., $\mathbb{P}[W > D]$. A direct computation of this tail probability is generally difficult since the exact waiting time distribution may not exist in explicit form. Alternatively, we can adopt the approximation of (39), as suggested in [6], to obtain the analytical value of DMR, by taking advantage of its adequate accuracy in most cases, as we have discussed above. Here, we fix $M = 2$ and $\lambda_a = 0.6$, and vary the degrees of burstiness and correlation.

The results are shown in Fig. 5(a) and (b) for $c^2 = 2$ and $c^2 = 10$, respectively. For the low bursty traffic $c^2 = 2$, we can see that the DMR quickly drops for D less than 10 ms, and the tail becomes quite flat for larger D, since the correlation r_1 has a minor impact on the mean waiting time when the traffic load is light. For a highly bursty traffic, as shown in Fig. 5(b), not only the DMR is higher compared to that of a low correlated one, but the correlation r_1 also has a dramatic impact on the DMR. Furthermore, if we compare the DMR curve for $r_1 = 0$ in both figures, we can find that they are nearly the same. However, as the correlation r_1 increases, the DMR surface of the high-burstiness traffic $(c^2 = 10)$ is clearly different from that of low-burstiness traffic $(c^2 = 2)$. For the high-burstiness traffic, a deadline (say $D = 10$), which is sufficient to ensure low DMR for the low correlated interarrivals, is not applicable to the high correlated interarrivals, where additional protections, such as smaller minimum CW and longer TXOP, may be cooperatively used with AIFS to ensure a desired low DMR.

VII. Conclusion

We have presented a simple yet accurate model for the performance study of the distributed PCA protocol in the WiMedia MAC specification. We have focused on the interrelation between the AIFS mechanism specified in PCA and the burstiness/correlation properties in the multimedia traffic. While the burstiness/correlation in the interarrivals has been neglected in most studies, we have shown their significant impact on the mean frame waiting time. We derive the PGF of service time distribution and model the multimedia traffic as an MMPP process, which is able to capture the bursty/correlated characteristics of interarrivals. The mean frame waiting time is obtained using queueing analysis, where we consider several approximate systems, including the exact MMPP arrival process and its *GI* counterpart, combined with the exponential and Gamma service time approximations. The asymptotically heavy traffic approximation is also considered.

Although none of these methods is clearly the best in all cases, the Q^{Γ}_{MMPP} and Q^{M}_{MMPP} approximations provide reasonable accuracy and adequately reflect the impact of burstiness/correlation in interarrivals. Q^{M}_{PMRQ} achieves an accuracy similar those of the other two methods (Q^{Γ}_{MMPP} and Q^{M}_{MMPP}) in certain cases, and its notable benefit in easing the computational burden deserves further study.

It is demonstrated that the effect of AIFS tends to be magnified when the traffic load is high or the interarrivals are highly bursty and correlated. The burstiness has a significant impact on the queueing performance, and the correlation has a stronger impact on highly bursty traffic. We have also presented a potential application of our analysis in QoS provisioning for real-time traffic. Our analysis has suggested that dynamically adjusting the contention parameters in response to the traffic characteristics and the network condition may need to be considered to support multimedia traffic with a stringent delay requirement.

Appendix
Derivation of $G_U(z)$

To derive the PGF of the pre-backoff period $G_U(z)$, consider the fact that when the tagged AC_2 station is backoff in zone 2, the backoff procedure is interrupted if any other stations transmit, which occurs with probability $1 - \gamma$, where $\gamma = (1 - \tau_1)^{N_1}(1 - \tau_2)^{N_2 - 1}$. Hence, when the tagged station experiences x_i backoffs in stage i, there are $(1 - \gamma)\sum_{i=1}^{C-1} x_i$ interruptions or *segments*, providing a total of C transmission trials before one successful transmission. In each segment, the backoff counter can be decremented only when zone 1 is idle with probability θ_2. With probability $1-\theta_2$, one or more AC_1 stations may transmit in any of the slots in zone 1, and the current uncompleted zone 1 is immediately ended due to this transmission. Therefore, for each backoff segment of the tagged AC_2 station, there are a number of "pre-backoff waiting" periods, which are denoted Q, preceding the "pure" backoff stage. While Q itself is a geometric random variable with parameter θ_2, to simplify the analysis, we let $Q = 1/\theta_2$. As a result, the overall number of pre-backoff waiting periods, which is denoted $N_U(C)$, is equal to $N_U(C) = (1-\gamma)/\theta_2 \sum_{i=1}^{C-1} x_i$, and its PGF is given by

$$G_{N_U(C)}(z) = \mathbb{E}\left[z^{\frac{1-\gamma}{\theta_2}\sum_{j=0}^{C-1} x_j}\right] = \prod_{j=0}^{C-1} G_{x_j}\left(z^{(1-\gamma)/\theta_2}\right). \quad (53)$$

To compute the length of a pre-backoff waiting period, we use an argument similar to computing the length of a generic slot. Define η as an i.i.d. random variable representing the length of the pre-backoff waiting period. If the first interrupted slot in zone 1 is the $\tilde{N}$th slot, the length of the pre-backoff waiting period equals $\eta = (\tilde{N}-1)\Delta + T_s$. Therefore

$$G_\eta(z) = G_{\tilde{N}}(z^\Delta) z^{-\Delta} G_{T_s}(z) \quad (54)$$

where

$$G_{\tilde{N}}(z) = \sum_{k=1}^{M} z^{\tilde{N}} P[\tilde{N}=k] = (1-p_{I,1}) \sum_{k=1}^{M} z^k (p_{I,1})^{k-1} \quad (55)$$

and $p_{I,1}$ is given by (17). The total pre-backoff period $U = \sum_{i=1}^{N_U(C)} \eta_i$ has the PGF given by

$$\begin{aligned} G_U(z) &= (1-p)\sum_{c=1}^{m} p^{c-1} \mathbb{E}\left[z^{\sum_{i=1}^{N_U(c)} \eta_i}\right] + p^m z^{m\Delta} \\ &= (1-p)\sum_{c=1}^{m} p^{c-1} G_{N_U(c)}\left(G_\eta(z)\right) + (pz)^m z^\Delta. \end{aligned} \quad (56)$$

Acknowledgment

The authors would like to thank Prof. B. Balcioglu for his helpful suggestions in obtaining the numerical result for PMRQ approximation.

References

[1] L. Yang and G. B. Giannakis, "Ultra-wideband communications: An idea whose time has come," *IEEE Signal Process. Mag.*, vol. 21, no. 6, pp. 26–54, Nov. 2004.

[2] X. Shen, W. Zhuang, H. Jiang, and J. Cai, "Medium access control in ultra-wideband wireless networks," *IEEE Trans. Veh. Technol.*, vol. 54, no. 5, pp. 1663–1677, Sep. 2005.

[3] K.-H. Liu, L. Cai, and X. Shen, "Multi-class QoS scheduling in UWB networks using discrete stochastic optimization," *IEEE Trans. Veh. Technol.*, vol. 57, no. 2, pp. 1176–1187, Mar. 2008.

[4] *High Rate Ultra Wideband PHY and MAC Standard*, ECMA Int. ECMA-368, Dec. 2005. [Online]. Available: www.ecma-international.org/publications/standards/Ecma-368.htm

[5] Q. Wu, Y. Xiong, H. Wu, Z. Guo, X.-G. Xia, Q. Zhang, and Z. Li, "Performance evaluation of the beacon period contraction algorithm in UWB MBOA MAC," *IEEE Commun. Lett.*, vol. 9, no. 10, pp. 933–935, Oct. 2005.

[6] D. L. Jagerman, B. Balcioglu, T. Altiok, and B. Melamed, "Mean waiting time approximations in the G/G/1 queue," *Queueing Syst.*, vol. 46, no. 3/4, pp. 481–506, Mar./Apr. 2004.

[7] C. Hu, H. Kim, J. Hou, D. Chi, and S. S. Nandagopalan, "Provisioning quality controlled medium access in ultrawideband-operated WPANs," in *Proc. IEEE INFOCOM*, Barcelona, Spain, Apr. 23–29, 2006, pp. 1–11.

[8] J. W. Robinson and T. S. Randhawa, "Saturation throughput analysis of IEEE 802.11e enhanced distributed coordination function," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 5, pp. 917–928, Jun. 2004.

[9] Z.-N. Kong, D. H. K. Tsang, B. Bensaou, and D. Gao, "Performance analysis of IEEE 802.11e contention-based channel access," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 10, pp. 2095–2106, Dec. 2004.

[10] Y. Xiao, "Performance analysis of priority schemes for IEEE 802.11 and IEEE 802.11e wireless LANs," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1506–1515, Jul. 2005.

[11] J. Hui and M. Devetsikiotis, "A unified model for the performance analysis of IEEE 802.11e EDCA," *IEEE Trans. Commun.*, vol. 53, no. 9, pp. 1498–1510, Sep. 2005.

[12] X. Chen, H. Zhai, X. Tian, and Y. Fang, "Supporting QoS in IEEE 802.11e wireless LANs," *IEEE Trans. Wireless Commun.*, vol. 5, no. 8, pp. 2217–2227, Aug. 2006.

[13] Y. Lin and V. Wong, "Saturation throughput of IEEE 802.11e EDCA based on mean value analysis," in *Proc. IEEE WCNC*, Las Vegas, NV, Apr. 2006, pp. 475–480.

[14] X. Ling, K.-H. Liu, Y. Cheng, X. Shen, and J. W. Mark, "A novel performance model for distributed prioritized MAC protocols," in *Proc. IEEE GLOBECOM*, 2007, pp. 4692–4696.

[15] O. Tickoo and B. Sikdar, "Queueing analysis and delay mitigation in IEEE 802.11 random access MAC based wireless networks," in *Proc. IEEE INFOCOM*, Hong Kong, 2004, pp. 1404–1413.

[16] C. G. Park, H. S. Jung, D. H. Han, "Queueing analysis of IEEE 802.11 MAC protocol in wireless LAN," in *Proc. ICN/ICONS/MCL*, Morne, Mauritius Island, Apr. 2006, pp. 139.

[17] K. M. Khalil and Y. S. Sun, "The effect of bursty traffic on the performance of local area networks," in *Proc. IEEE GLOBECOM*, Orlando, FL, Dec. 1992, pp. 597–603.

[18] C. J. Chang, C. H. Lin, D. S. Guan, and R. G. Cheng, "Design of a power-spectrum-based ATM connection admission controller for multimedia communications," *IEEE Trans. Ind. Electron.*, vol. 45, no. 1, pp. 52–59, Feb. 1998.

[19] S. Molnar and G. Miklos, "On burst and correlation structure of teletraffic models," in *Proc. 5th IFIP Workshop Perform. Model. Evaluation ATM Netw.*, Ilkley, U. K., Jul. 1997.

[20] I. W. Habib and T. N. Saadawi, "Multimedia traffic characteristics in broadband networks," *IEEE Commun. Mag.*, vol. 30, no. 7, pp. 48–54, Jul. 1992.

[21] D. P. Heyman and D. Lucantoni, "Modeling multiple IP traffic streams with rate limits," *IEEE/ACM Trans. Netw.*, vol. 11, no. 6, pp. 948–958, Dec. 2003.

[22] S. Shah-Heydari and T. Le-Ngoc, "MMPP models for multimedia traffic," *Telecommun. Syst.*, vol. 15, no. 3/4, pp. 273–293, Dec. 2000.

[23] A. Heindl, "Decomposition of general queueing networks with MMPP inputs and customer losses," *Perform. Eval.*, vol. 51, no. 2–4, pp. 117–136, Feb. 2003.

[24] S. A. Qahtani and S. A. Dhahran, "A simulation-based comparison of multimedia traffic prioritization schemes for high-performance input-queued packet switches," *J. Comput. Sci.*, vol. 2, no. 4, pp. 347–354, 2006.

[25] Y. Zheng, K. Lu, D. Wu, and Y. Fang, "Performance analysis of IEEE 802.11 DCF in imperfect channels," *IEEE Trans. Veh. Technol.*, vol. 55, no. 5, pp. 1648–1656, Sep. 2006.

[26] Y. Cheng, X. Ling, W. Song, L. X. Cai, W. Zhuang, and X. Shen, "A cross-layer approach for WLAN voice capacity planning," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 4, pp. 678–688, May 2007.

[27] A. Zanella and F. De Pellegrini, "Statistical characterization of the service time in saturated IEEE 802.11 networks," *IEEE Commun. Lett.*, vol. 9, no. 3, pp. 225–227, Mar. 2005.

[28] G. L. Choudhury and D. M. Lucantoni, "Numerical computation of the moments of a probability distribution from its transform," *Oper. Res.*, vol. 44, no. 2, pp. 368–381, Mar./Apr. 1996.

[29] J. Abate, G. L. Choudhury, and W. Whitt, "An introduction to numerical transform inversion and its application to probability models," *Int. Series Oper. Res. Manage. Sci.*, vol. 24, pp. 257–324, 2000.

[30] W. Fischer and K. Meier-Hellstern, "The Markov-modulated Poisson process (MMPP) cookbook," *Perform. Eval.*, vol. 18, no. 2, pp. 149–171, 1993.

[31] D. Gross and C. M. Harris, *Fundamentals of Queueing Theory*. New York: Wiley, 1985.

[32] A. E. Eckberg, “Generalized peakedness of teletraffic processes,” in *Proc. 10th ITC*, Montreal, QC, Canada, 1983. Session 4.4b.

[33] A. Ekbal and J. M. Cioffi, “Effect of wireless channel process on queueing delay-approximate analysis using peakedness function,” in *Proc. IEEE ICC*, Seoul, Korea, May 2005, pp. 468–472.

[34] S. Q. Li and C. L. Hwang, “On the convergence of traffic measurement and queueing analysis: A statistical-matching and queueing (SMAQ) tool,” *IEEE/ACM Trans. Netw.*, vol. 5, no. 1, pp. 95–110, Feb. 1997.

[35] L. Muscariello, M. Mellia, M. Meo, M. Ajmone Marsan, and R. Lo Cigno, “Markov models of internet traffic and a new hierarchical MMPP model,” *Comput. Commun.*, vol. 28, no. 16, pp. 1835–1851, Oct. 2005.

[36] Z. Cui and A. A. Nilsson, “The impact of correlation on delay performance of high speed networks,” in *Proc. 26th Southeastern Symp. Syst. Theory*, Athens, OH, Mar. 2004, pp. 371–374.

[37] K. Sriram and W. Whitt, “Characterizing superposition arrival processes in packet multiplexers for voice and data,” *IEEE J. Sel. Areas Commun.*, vol. SAC-4, no. 6, pp. 833–846, Sep. 1986.

[38] J. She, F. Hou, and P.-H. Ho, “An application-driven MAC-layer buffer management with active dropping for real-time video streaming in 802.16 networks,” in *Proc. IEEE Int. Conf. AINA*, Niagara Falls, ON, Canada, May 21–23, 2007, pp. 451–458.

Kuang-Hao Liu received the B.S. degree in applied mathematics from National Chiao Tung University, Hsinchu, Taiwan, R.O.C., in 1998 and the M.S. degree in electrical engineering from National Chung Hsing University, Taichung, Taiwan, in 2000. He is currently working toward the Ph.D. degree in electrical and computer engineering at the University of Waterloo, Waterloo, ON, Canada.

From 2000 to 2002, he was a Software Engineer with Siemens Telecom, System Ltd., Taipei, Taiwan. His research interests include UWB communications in personal area networks, resource allocation problems, and performance analysis for wireless communication protocols.

Xinhua Ling received the B.Eng. degree in radio engineering from Southeast University, Nanjing, China, in 1993, the M.Eng. degree in electrical engineering from the National University of Singapore, in 2001, and the Ph.D. degree in electrical and computer engineering from the University of Waterloo (UW), Waterloo, ON, Canada, in 2007.

He is currently a Post-Doctoral Fellow with UW. His general research interests are in the areas of WLAN; WPAN; mesh, *ad hoc*, cellular, and WiMAX networks; and their Internet working, focusing on protocol design and performance analysis.

Xuemin (Sherman) Shen (M’97–SM’02) received the B.Sc. degree in electrical engineering from Dalian Maritime University, Dalian, China, in 1982 and the M.Sc. and Ph.D. degrees in electrical engineering from Rutgers University, Camden, NJ, 1987 and 1990, respectively.

He is currently with the Department of Electrical and Computer Engineering, University of Waterloo (UW), Waterloo, ON, Canada, where he is a Professor and the Associate Chair for graduate studies. He is the coauthor of three books and has published more than 300 papers and book chapters in wireless communications and networks, control, and filtering. His research focuses on mobility and resource management in interconnected wireless/wired networks, UWB wireless communications systems, wireless security, and *ad hoc* and sensor networks.

Dr. Shen is a registered Professional Engineer in the province of Ontario. He has served as the Technical Program Chair for many conferences, including IEEE Globecom’07. He has also served as Editor/Associate Editor/Guest Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, *Computer Networks*, *ACM/Wireless Networks*, the *IEEE Journal of Selected Areas in Communications*, IEEE WIRELESS COMMUNICATIONS, IEEE COMMUNICATIONS MAGAZINE, etc. He received the Outstanding Performance Award from the University of Waterloo in 2002 and 2004 for outstanding contributions to teaching, scholarship, and service, and the Premier’s Research Excellence Award (PREA) in 2003 from the Province of Ontario for demonstrated excellence of scientific and academic contributions.

Jon W. Mark (M’62–SM’80–F’88–LF’03) received the B.A.Sc. degree in electrical engineering from the University of Toronto, Toronto, ON, Canada, in 1962 and the M.Eng. and Ph.D. degrees in electrical engineering from McMaster University, Hamilton, ON, in 1968 and 1970, respectively.

From 1962 to 1970, he was an Engineer and then a Senior Engineer with Westinghouse Canada Ltd., Hamilton, where he conducted research on sonar signal processing and submarine detection. Since 1970, he has been with the Department of Electrical Engineering (now Electrical and Computer Engineering), University of Waterloo (UW), Waterloo, ON, where he is currently a Distinguished Professor Emeritus. He was promoted to full Professor in 1978 and served as Department Chairman from July 1984 to June 1990. In 1996, he established the Centre for Wireless Communications (CWC), UW, and has since been serving as the founding Director. He was on sabbatical leave at the IBM Thomas Watson Research Center, Yorktown Heights, NY, as a Visiting Research Scientist (1976–1977); AT&T Bell Laboratories, Murray Hill, NJ, as a Resident Consultant (1982–1983); the Laboratoire MASI, Université Pierre et Marie Curie, Paris, France, as an Invited Professor (1990–1991); and the Department of Electrical Engineering, National University of Singapore, as a Visiting Professor (1994–1995). He is a coauthor of *Wireless Communications and Networking* (Prentice-Hall, 2003). His current research interests are in wireless communications and wireless/wireline interworking, particularly in the areas of resource management, mobility management, cross-layer design, and end-to-end information delivery with QoS provisioning.

Dr. Mark has served as a member of a number of editorial boards, including the IEEE TRANSACTIONS ON COMMUNICATIONS, *ACM/Baltzer Wireless Networks, Telecommunication Systems*, etc. He was a member of the Inter-Society Steering Committee of the IEEE/ACM TRANSACTIONS ON NETWORKING from 1992 to 2003 and a member of the IEEE COMSOC Awards Committee during 1995–1998.