A Markovian Framework for Performance Evaluation of IEEE 802.11

Chuan Heng Foh, *Member, IEEE,* Moshe Zukerman, *Fellow, IEEE,* and Juki Wirawan Tantra, *Member, IEEE*

***Abstract*— A new approach for modeling and performance analysis of the IEEE 802.11 medium access control (MAC) protocol is presented. The approach is based on the so-called system approximation technique, where the protocol service time distribution of the IEEE 802.11 MAC protocol is studied and approximated by an appropriate phase-type distribution, leading to the construction of a versatile queueing model which is amenable to analysis and, at the same time, general enough to allow for bursty arrival process as well as key statistical characteristics of the protocol operations. The versatility of the model is demonstrated by considering Markov modulated and on/off arrival processes as well as various data frame size distributions. The accuracy of the analytical results is verified by simulation.**

***Index Terms*— Computer network performance, IEEE 802.11, wireless LAN.**

I. Introduction

TO support development of an efficient and robust network in a wireless environment, the IEEE 802.11 working group has chosen the *carrier sense multiple access with collision avoidance* (CSMA/CA) protocol as the standard protocol for wireless *local area networks* (LANs) [1]. The CSMA/CA protocol is a collision-based random access protocol whereby, in the case of a collision, each mobile station executes the *binary exponential backoff* (BEB) retransmission algorithm to resolve the collision and maintain the stability of the CSMA/CA channel.

The standardization of the IEEE 802.11 *medium access control* (MAC) protocol has triggered research on its performance evaluation and enhancements [2]-[8]. Performance evaluation of the IEEE 802.11 MAC protocol with all its details and under realistic traffic conditions has been considered difficult. Therefore, many analyses have assumed simpler traffic conditions such as Poisson sources with fixed-size data frames and/or simplifications of the protocol operations, for example, the simplification of the actual retransmission algorithm used in IEEE 802.11 [4],[5].

Bianchi [2] has analyzed the IEEE 802.11 MAC protocol capturing the protocol operation in great details. His performance evaluation assumes saturation traffic whereby all stations are saturated, namely, stations always have data frames to transmit. Since in the actual operation, persistent saturation continues only during a short time period, it is also of interest to evaluate the performance of IEEE 802.11 under statistical traffic conditions.

Accepted
IEEE T Wireless Communications 2007

In [9], we have presented a pioneering work of performance modeling of IEEE 802.11 under statistical traffic conditions using queueing analysis. In our queueing analytical approach, an IEEE 802.11 WLAN is modeled as a single server queue (SSQ), where its service process is fitted [9]. Later, several other methods that use queueing analysis with approximated service processes have been published [10]-[12].

Beside the queueing analytical approach, some recent works propose the extension of Bianchi's Markov Chain model for the protocol performance evaluation under non-saturation traffic condition [13]-[15]. This approach introduces one or more additional states into Bianchi's Markov Chain to model the idle state of a station. This approach assumes that after a successful transmission by a station, it generates a new packet based on a Bernoulli process over uneven timeslots. This assumption imposes a limitation on the model by considering only effect of first order statistics. In contrast, our queueing analytical approach with the fitting strategy can account for second order effects which are more appropriate for the busty traffic that occurs in practice.

In this paper, we revise and substantially extend our previous work in [9]. Focusing on the analysis of the IEEE 802.11 protocol service time distribution, we obtain close form results for some of its important statistical characteristics, which were only demonstrated via simulation in [9]. The derivation of the service time distribution function of the IEEE 802.11 protocol and its mean and variance expressed in simple closed form expressions are useful for the performance study of higher layer protocols operating in wireless LANs [16]. Besides, we analyze the performance of the popular IEEE 802.11b [17] WLANs reflecting the latest development in WLANs, as opposed to [9] that focuses on the low speed frequency hopping spread spectrum (FHSS) WLAN. Moreover, we present and discuss the protocol performance under bursty traffic.

Our approach is based on the so-called *system approximation* technique [18]. Defining the IEEE 802.11 service time to be the time between the start of two consecutive packet transmissions observed on the IEEE 802.11 common channel, we first study the statistical characteristics of the IEEE 802.11 service time, and we fit a phase-type (PH) distribution to it. We observe that the IEEE 802.11 service time distribution can be approximated by a particular PH distribution using the right parameter set. Using the right PH distribution for the protocol service time and assuming a certain bursty packet arrival process, we are able to construct an SSQ, that models the MAC protocol. Here we demonstrate by simulations that our approach which includes the PH approximation of the IEEE 802.11 service time and the SSQ modeling provides an accurate means for evaluating the mean delay performance of the IEEE 802.11 protocol. It is well known that Poisson is not an accurate model for packet traffic. It is therefore desirable to analyze the protocol under a more realistic traffic model. In this paper, we use the versatile *Markovian arrival process* (MAP) [19] which is known to capture statistical characteristics of bursty traffic while maintaining the Markovian property.

A Markovian framework is finally developed describing a queueing system based on the IEEE 802.11 MAC protocol under a certain network and traffic conditions. This framework provides means to obtain accurate performance results. Examples of new delay performance results of IEEE 802.11 that can now be obtained using the new approach include: (i) performance under bursty traffic - Markov Modulated Poisson Process (MMPP); (ii) performance under on/off sources; (iii) and performance under various data frame size distributions. In addition, this technique also enables us to study some complex MAC protocols, such as the out-of-band signaling MAC protocol for IEEE 802.11 WLANs [20].

In the next section, the operation of the IEEE 802.11 MAC protocol is revisited. In Section III, we study the characteristics of IEEE 802.11 service process and describe the Markovian framework for the performance analysis of IEEE 802.11. In Section IV, we obtain analytical results for performance evaluation for IEEE 802.11 under several traffic conditions.

II. The IEEE 802.11 MAC Protocol

According to the *distributed coordination function* (DCF) of the IEEE 802.11 protocol, stations accessing the channel use the *basic access method*, and also the optional *four-way handshaking access method* with an additional Request-To-Send/Clear-To-Send (RTS/CTS) message exchange (see Fig. 1). Under the basic access method, when a station is ready for a new data frame transmission, it first senses the channel status. If the channel is found to be busy, the station defers its transmission and continues to sense the channel until it is idle. After the channel is idle for a specified period of time called the DCF interframe space (DIFS) period, the station chooses a random number as a backoff timer. The time immediately after the DIFS period is slotted. As shown in Fig. 1, the timeslot duration is at least the time required for a station to detect an idle channel, plus the time required to switch from listening to transmitting mode. The backoff timer is decreased by one for each idle slot, stopped if the channel is sensed busy, and then reactivated if the channel is idle again and remains idle for more than a DIFS time period. When the backoff timer reaches zero, the data frame is transmitted.

The choice of the random number for the backoff timer is based on the binary exponential backoff algorithm. A station chooses an integer between 0 and $CW - 1$ randomly with equal probability. The contention window (CW) is set to the minimum backoff contention window, W, for every new

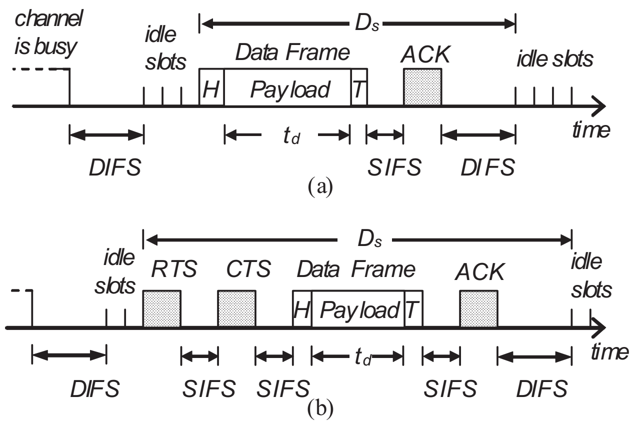


Fig. 1. The IEEE 802.11 access methods: (a) Basic access method. (b) Four-way handshaking access method.

data frame transmission. For each unsuccessful transmission attempt, the station increases its backoff stage count and doubles its CW, until the station reaches the predefined maximum backoff stage, M, or its CW reaches the maximum contention window, $CWmax$. At that stage, CW ceases to increase. Hence given that the initial backoff stage is zero, then $CWmax = 2^M W$. To determine whether a data frame transmission is successful, after its completion, a positive acknowledgement (ACK) is transmitted by the receiver. ACK is transmitted after a short interframe space (SIFS) period upon receiving the entire data frame successfully. If ACK is not detected within an SIFS period after the completion of the data frame transmission, the transmission is assumed to be unsuccessful, and a retransmission is required.

In the four-way handshaking access method, an additional operation is introduced on top of the basic access method before a data frame transmission. When the backoff timer of a station reaches zero, the station first transmits an RTS frame to request a transmission right. Upon receiving the RTS frame, the receiver replies with a CTS frame after a SIFS period. Once the RTS/CTS information is exchanged successfully, the sender transmits its data frame.

After a successful transmission, a station must perform a compulsory DIFS deference and backoff even if it has no queued data frame in its local buffer. This process is often known as "post-backoff", it ensures that consecutive transmissions are separated by at least one backoff interval. If a new data frame is generated after the post-backoff procedure, the data frame may be transmitted immediately if the channel has been sensed idle for a period longer than the DIFS.

III. The Markovian Framework Approach

We first study the service characteristics of the IEEE 802.11 MAC protocol given a certain number of stations. Next, we obtain exact analytical results for the mean and variance of the service time of the IEEE 802.11 MAC protocol. Observing that the service time distribution can be approximated by a particular PH distribution, parameter fitting is performed to identify an appropriate PH distribution describing the IEEE 802.11 MAC protocol service time distribution. A Markovian framework is then developed for the performance evaluation

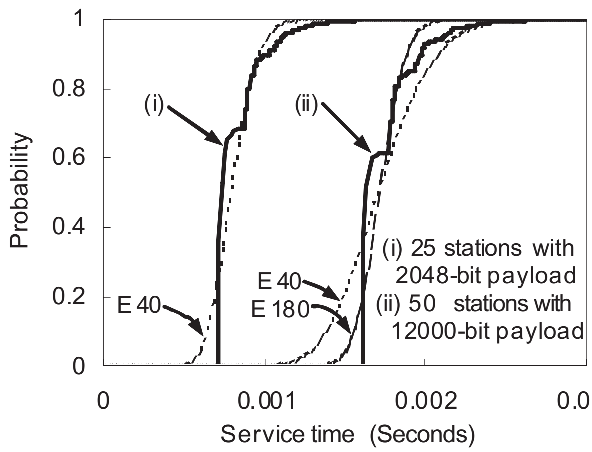


Fig. 2. Cumulative distributions of the service time of IEEE 802.11b with the four-way handshaking access method from (2).

of the IEEE 802.11 MAC protocol under a Markovian arrival process.

A. Study of the Statistical Characteristics of the IEEE 802.11 MAC Protocol Service Process

We consider a single hop IEEE 802.11 WLAN consisting of N stations. Each station in the WLAN can communicate with others directly. We further consider, as in [2], that all stations in the WLAN are *saturated* such that whenever a station transmits its data frame successfully, the station is ready for another new data frame transmission immediately after its previous data frame transmission. This traffic condition is known as *saturation* [2].

The focus of the study here is the service process of the protocol under saturation. We are particularly interested in the average time for the IEEE 802.11 MAC protocol to serve a data frame, as well as the variance of such service time, under saturation.

Bianchi has derived the saturation throughput formula for IEEE 802.11 in [2]. A few important performance quantities are also derived in [2], namely the probabilities, P_i, P_s and P_c, that a particular timeslot on the channel contains no transmission (idle), one transmission (a successful transmission) and two or more transmissions (a collision), respectively. Given N saturated stations in a WLAN, these quantities can be determined by the following equation set

$$\begin{array}{l} P_i = (1-\tau)^N \\ P_s = N\tau(1-\tau)^{N-1} \\ P_c = 1 - P_s - P_i \end{array} \quad (1)$$

with

$$\left\{ \begin{array}{l} \tau = \frac{2(1-2p)}{(1-2p)(W+1)+pW\left(1-(2p)^M\right)} \\ p = 1-(1-\tau)^{N-1} \end{array} \right.$$

where W denotes the minimum contention window, M denotes the maximum backoff stage.

As in [2] we assume that in steady state, under saturation condition, an event occurring in a particular timeslot is independent of the events occurred in previous timeslots. Let θ_N be a discrete random variable representing the IEEE 802.11 service time of the four-way handshaking method. Let D_s be a discrete random variable representing the duration of a data frame transmission. To be precise, D_s is the sum of the payload transmission duration and the transmission overheads (see Fig. 1). Since θ_N also depends on D_s, we further define θ_{N,d_s} to be θ_N given that $D_s = d_s$. Knowing that a timeslot at any arbitrary time may carry zero, one, and two or more transmissions with probabilities P_i, P_s, and P_c respectively, then given a certain data frame transmission duration, d_s, the probability that a service period is shorter than or equal to a particular time period t, is

$$\Pr\{\theta_{N,d_s} \le t\} = \sum_{n_i=0}^{\infty}\sum_{n_c=0}^{\infty} \phi(n_i, n_c, t) \quad (2)$$

where

$$\phi(n_i, n_c, t) = \left\{ \begin{array}{ll} \binom{n_i+n_c}{n_i} P_i^{n_i} P_c^{n_c} P_s, & d_s + n_i d_i + n_c d_c \le t \\ 0, & \text{otherwise} \end{array} \right.$$

and $\binom{n_i+n_c}{n_i}$ is the binomial coefficient, d_i and d_c are the duration of an idle period and a collision respectively.

TABLE I
IEEE 802.11b Protocol Parameters

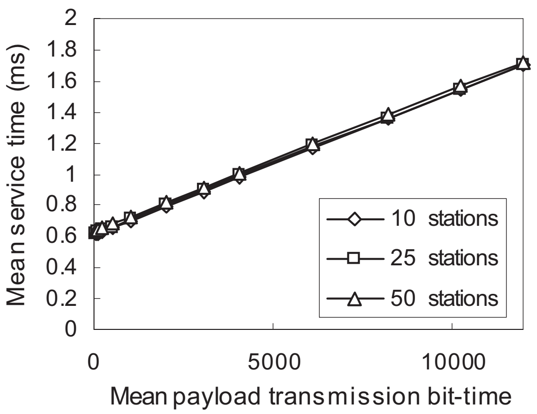
Parameters for IEEE 802.11b	**Value**
Channel data rate	11 Mb/s
Idle slot duration, d_i	20 μs
Collision slot duration (basic)*, d_c^{bas}	$t_d + 171.727\mu$s
Successful data frame transmission duration (basic)*, D_s^{bas}	$t_d + 288.909\mu$s
Collision slot duration (rts/cts)*, d_c^{rts}	161.545μs
Successful data frame transmission duration (rts/cts)*, D_s^{rts}	$t_d + 527.636\mu$s
Backoff parameters	$W = 32$, $M = 5$
*The duration is inclusive of signal propagation delay of $1\mu s$ after a transmission; see also [2] for details.	

We adopt the IEEE 802.11b physical layer [17] with the protocol parameters given in Table I for our study. We further consider constant payload sizes, that is, D_s is assumed to be a constant. From (2), we plot the cumulative distribution of the service time in Fig. 2. Two cases are illustrated: (i) 25 saturated stations transmitting 2048-bit payloads, and (ii) 50 saturated stations transmitting 12000-bit payloads. The dotted lines shown in Fig. 2 correspond to the equivalent Erlang distribution functions that will be discussed in the next subsection. One immediate observation from the results is the small variance of the service time distributions for both cases.

To study the statistical characteristics of θ_N, we first derive $E[\theta_N]$ by conditioning and unconditioning on the channel slot type, that is

$$\begin{aligned} E[\theta_N] &= E[D_s] P_s + (d_i + E[\theta_N]) P_i + (d_c + E[\theta_N]) P_c \\ &= E[D_s] + \frac{P_i d_i + P_c d_c}{P_s}. \end{aligned} \quad (3)$$

Having obtained the mean service time for one data frame transmission under the saturation condition, the service rate of the protocol, $\mu(N)$, is $(E[\theta_N])^{-1}$. Furthermore, define t_d to be the mean payload transmission time, the throughput of the protocol can also be easily computed by $t_d \cdot \mu(N)$, which agrees with (13) presented in [2].



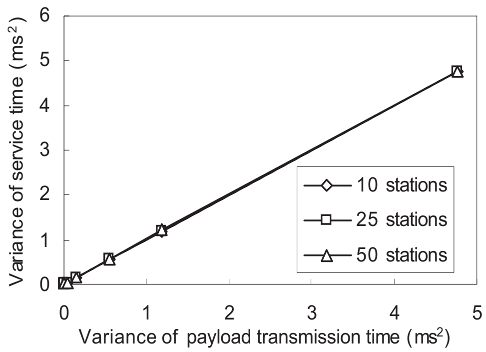


Fig. 3. Mean and variance of service time versus payload transmission bit-time for 10, 25 and 50 saturated stations.

The relationship between the $E[\theta_N]$ and t_d is presented in Fig. 3. It is found that the mean service time of IEEE 802.11 depends mainly on the mean payload transmission time measured in bit-time. Different number of stations represented by the three curves appear to have no impact on the mean service time as the curves overlap each other.

We now turn our attention to the variance of θ_N. We first define A to be the first event to occur on the channel after a successful transmission. There are three possible events for A, which are: $A = 0$ denoting an idle channel, $A = 1$ denoting a successful transmission and $A = 2$ denoting a collision. We notice that $Var(D_s)$ only depends on the variance of the payload size since the header/trailer size remains unchanged during the operation. We use the condition variance formula to derive the variance of θ_N,

$$Var\left(\theta_N\right) = E\left[Var\left(\theta_N\left|A\right.\right)\right] + Var\left(E\left[\theta_N\left|A\right.\right]\right). \quad (4)$$

We notice that θ_N can be divided into two independent random variables: the first is the duration of A, denoted τ_A, and the second is the remaining time, ψ_A. Therefore, revoking the independent of τ_A and ψ_A, we have $Var(\theta_N) = Var(\tau_A) + Var(\psi_A)$. For the cases of $A = 0$ and $A = 2$, we have $Var(\tau_A) = 0$ since τ_A is constant, thus $\tau_0 = d_i$ and $\tau_2 = d_c$, respectively. The remaining time is itself a service time due to its memoryless property, that is, $\psi_A = \theta_N$, hence $Var(\theta_N|A) = Var(\theta_N)$. For $A = 1$, the appearance of the data frame transmission on the channel ends the service time, hence $Var(\theta_N|A = 1) = Var(D_s)$. Therefore we obtain the following relationships

$$\begin{cases} Var\left(\theta_N\left|A = 0\right.\right) = Var\left(\theta_N\right) \\ Var\left(\theta_N\left|A = 1\right.\right) = Var\left(D_s\right) \\ Var\left(\theta_N\left|A = 2\right.\right) = Var\left(\theta_N\right). \end{cases}$$

From the above formula set, we obtain

$$E\left[Var\left(\theta_N\left|A\right.\right)\right] = \left(P_c + P_i\right)Var\left(\theta_N\right) + P_s Var\left(D_s\right). \quad (5)$$

Similarly, we have

$$\begin{cases} E\left[\theta_N\left|A = 0\right.\right] = d_i + E\left[\theta_N\right] \\ E\left[\theta_N\left|A = 1\right.\right] = E\left[D_s\right] \\ E\left[\theta_N\left|A = 2\right.\right] = d_c + E\left[\theta_N\right] \end{cases}$$

thus we obtain

$$\begin{aligned} Var\left(E\left[\theta_N\left|A\right.\right]\right) &= E\left[\left(E\left[\theta_N\left|A\right.\right] - E\left[\theta_N\right]\right)^2\right] \\ &= P_i d_i^2 + P_c d_c^2 + \frac{\left(P_i d_i + P_c d_c\right)^2}{P_s}. \quad (6) \end{aligned}$$

By substituting (5) into (4), we have

$$Var\left(\theta_N\right) = \frac{Var\left(E\left[\theta_N\left|A\right.\right]\right)}{P_s} + Var\left(D_s\right) \quad (7)$$

where $Var\left(E\left[\theta_N|A\right]\right)$ is given by (6).

We use the same approach to derive the variance of the protocol service time for the basic access method. Let θ_N^{bas} and d_c^{bas} be the protocol service time and the collision period of the basic access method, respectively. The only difference between the two access methods for this study is that the collision period of the basic access method is not constant, instead, it comprises a fixed overhead plus a data frame transmission duration which may be variable. Hence we have $Var(d_c^{bas}) = Var(D_s)$, which gives $Var(\theta_N^{bas}|A = 2) = Var(\theta_N^{bas}) + Var(D_s)$. Using this modification, we obtain

$$Var\left(\theta_N^{bas}\right) = \frac{Var\left(E\left[\theta_N^{bas}\left|A\right.\right]\right)}{P_s} + \left(1 + \frac{P_c}{P_s}\right)Var\left(D_s\right) \quad (8)$$

where, similar to (6),

$$Var\left(E\left[\theta_N^{bas}\left|A\right.\right]\right) = P_i d_i^2 + P_c\left(d_c^{bas}\right)^2 + \frac{\left(P_i d_i + P_c d_c^{bas}\right)^2}{P_s}.$$

From (7)-(8), we notice a similar phenomenon that both the variances of θ_N and θ_N^{bas} are mainly influenced by the variance of the data frame transmission time, especially for θ_N^{bas}. Focusing on the four-way handshaking method, to evaluate the effect of the data frame size distribution on the variance of θ_N, we plot, in Fig. 3, $Var(\theta_N)$ versus $Var(D_s)$ for various N. The results show that $Var(\theta_N)$ changes according to $Var(D_s)$ rather than N, as various values of N produce similar curves. This suggests that the $Var(\theta_N)$ is mainly affected by $Var(D_s)$.

To further study and demonstrate the effect of data frame size distribution, we show $E[\theta_N]$ and $Var(\theta_N)$ versus the number of stations in Fig. 4 given a fixed data frame size with payload of 2048 bits. This plot demonstrates that when the payload size is fixed, the changes in both mean and variance of the service time with different numbers of stations are insignificant, as compared to changes in the mean and the variance of the data frame size presented in Fig. 3. The

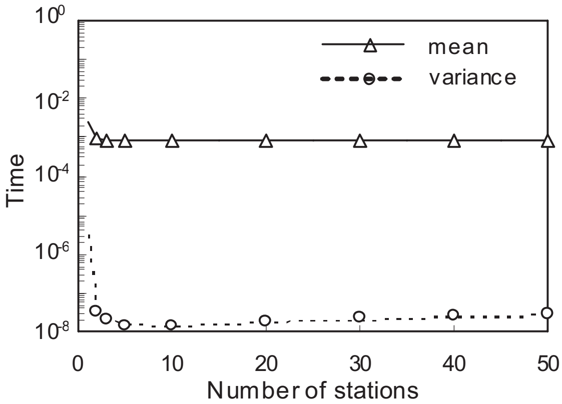


Fig. 4. Mean and variance of service time versus number of saturated stations for constant data frame size.

results again suggest that the payload size distribution dictates the protocol service time distribution. The numerical results (shown in lines) presented in Fig. 4 are also validated by simulation results (shown in symbols) to demonstrate the accuracy of our analysis.

Using the above results, together with the study in [2] that has suggested a memoryless process for IEEE 802.11 service process, we find a PH distribution, of which its mean and variance match those of the service time distribution of the IEEE 802.11 given a particular data frame size distribution. After the fitting of the mean and variance, we use the PH distribution to approximate the IEEE 802.11 service time distribution.

B. Markovian Framework for IEEE 802.11 MAC Protocol

We consider that there are k stations in a network. Each station may be in an active or an idle mode. A station is said to be idle when it has no packet in its local buffer for transmission, otherwise, it is said to be in active mode. If an active station always finds packets in its local buffer for transmission even after a successful transmission, then the active station is saturated and never turns idle. Bianchi models the situation when all k stations are always active, which is the saturation load condition. To evaluate the performance under statistical traffic where the number of active stations statistically varies among the values $1, 2, \ldots, k$, we first repeat the calculations of Bianchi's model for each i, $i = 1, 2, \ldots, k$. We then determine the steady state probability that the number of active stations is i by considering and analyzing the associated continuous time Markov chain single server queue (CTMC-SSQ) model and solving the resulted set of steady state equations. Finally, we compute the overall performance of the system using weighted average of the performances for each i, where the weights are the steady state probabilities that the number of active stations is i.

For the purpose of illustrating the idea, we consider a simple example of a CTMC-SSQ, the *state dependent* M/M/1/k (SD-M/M/1/k) queue, shown in Fig. 5. The queue size of the SD-M/M/1/k queue is analogous to the number of active stations. Having a single server in the queue represents the fact that in the IEEE 802.11 network, only one station is served at

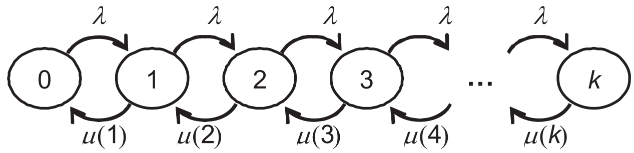


Fig. 5. The state dependent service rates M/M/1/k queue for the analysis of the IEEE 802.11 protocol.

a time. The rate, denoted $\mu(i)$, that a station is served is a function of the number of active stations, denoted i. Therefore, we consider state dependent service rate in our queue. The Poisson arrival in the SD-M/M/1/k queue is analogous to the Poisson process associated with points in time in which stations switch from idle to active mode, the rate is denoted λ. This simple example assumes a Poisson arrival process and Markovian state dependent service.

The results provided in the previous subsection suggest the modeling of the IEEE 802.11 service time distribution using an appropriate PH distribution. Incorporating this into our model, we construct an equivalent PH-based CTMC-SSQ to describe IEEE 802.11 WLANs under more general arrival processes. Since the service rate of the protocol also depends on the number of active stations, the service rate in our model is state dependent. The service rate at each state can be estimated by (3).

The use of a CTMC-SSQ as the model for IEEE 802.11 WLANs performance analysis relies on the assumption that with any arrival or departure event triggering a change in the system state, the system has renewed its service process partially or totally depending on the model with the appropriate number of active stations. This is in fact different from the actual protocol operation, since in the actual operation, a station does not reset its backoff timer and contention window when an arrival or a departure event occurs. However, based on our observation in the previous subsection, which is consistent with the earlier suggestion by [2], in steady state, the protocol service process mostly depends on the data frame size distribution than other factors. In other words, the actual values of backoff timer and the contention window of each station, at a point in time when an arrival or a departure event occurs, do not have a significant effect on the overall protocol performance in steady state. Therefore, the protocol service process after an arrival or a departure event can be assumed to be in steady state and to follow a particular PH process. This allows us to construct an equivalent CTMC-SSQ system that has similar performance behavior to the IEEE 802.11 MAC protocol.

To further extend the model for more realistic performance study, we incorporate the versatile *Markovian arrival process* (MAP) into our CTMC-SSQ system. There are two reasons why the MAP is considered. Firstly, the MAP can be used to model bursty traffic (e.g. MMPP). Secondly, the MAP maintains Markovian property of the system allowing the system to be amenable to analysis.

By evaluating results shown in Fig. 3, particularly when $Var(D_s)$ is zero, the variance of the IEEE 802.11 service time is very small compared to its mean service time. This finding supports the consideration of the Erlang distribution as an ac-

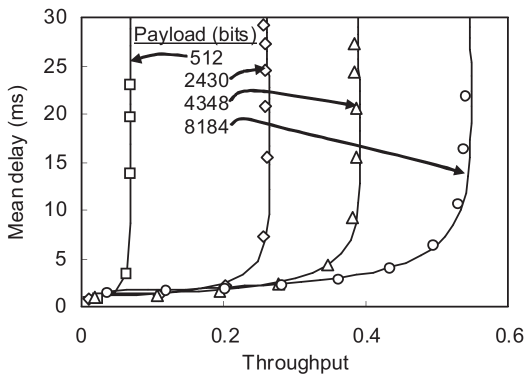


Fig. 6. Delay performance of the IEEE 802.11b MAC protocol with the four-way handshaking access method.

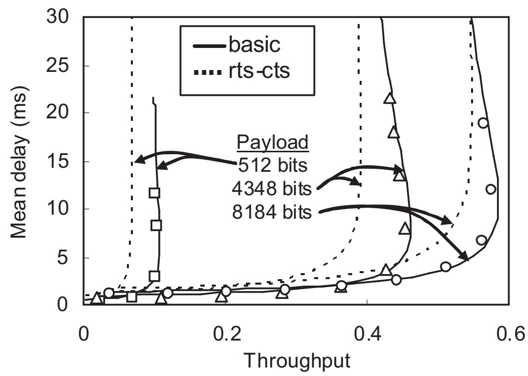


Fig. 7. Delay performance of the IEEE 802.11b MAC protocol with the basic access method.

curate model for the IEEE 802.11 service time distribution for fixed data frame size distribution. According to the numerical results shown in Fig. 4, a fixed payload size of 2048 bits gives a mean service time of around 8×10^{-4}s and a variance of service time of around $1.6 \times 10^{-8}\text{s}^2$. This corresponds to an Erlang distribution with 40 exponential phases, E_{40}. A visual comparison between the actual distribution and the equivalent E_{40} is illustrated in Fig. 2. We show the case of 25 saturated stations labeled as (i), and the equivalent E_{40} drawn with a dotted line. In one extreme example, not shown in Fig. 4, a large fixed payload size of 12 kbits, gives a mean and a variance service time of around 1.7×10^{-3}s and $1.6 \times 10^{-8}\text{s}^2$, respectively, corresponding to E_{180}. The distribution of the latter example for the case of 50 saturated stations is plotted in Fig. 2, labeled as (ii), along with the equivalent E_{180} depicted by the dotted lines. Another dotted line partially overlapping E_{180} is the equivalent E_{40}, which is illustrated to be also a possible choice for modeling the IEEE 802.11b service time distribution.

Since our work here focuses on the common performance measure of MAC protocol–the mean delay, which can be computed from the mean queue size using Little's formula, let us consider, for the sake of argument, the M/G/1 queue. In such a queue, the mean queue size, Q, is

$$Q = \frac{2\rho - \rho^2 + \rho^2 \left(\frac{\sigma_s}{s}\right)^2}{2\left(1-\rho\right)},$$

where ρ, σ_s and s are the utilization, the standard deviation and the mean of the service time, respectively. This result suggests a negligible contribution to the mean queue size from the term (σ_s/s) when it is very small. Taking the results presented in Fig. 4 as an example, the term (σ_s/s) produces values of 0.158 and 0.074 when the service distribution is approximated by E_{40} and E_{180} respectively, and both cases give similar results for Q given any valid ρ. Hence, E_{40} and E_{180} are almost equally good for modeling the service time distribution of various fixed payload sizes, as also illustrated visually in Fig. 2, and therefore, we choose E_{40}.

For variable data frame sizes, a particular PH distribution will be identified to model the data frame size distribution and the corresponding IEEE 802.11 service time distribution.

IV. Performance Study of IEEE 802.11

In this section, we present several performance studies of the IEEE 802.11 MAC protocol under various traffic conditions using our proposed framework. The protocol parameters in this study are based on the IEEE 802.11b standards (see also Table I for some important parameters). Simulation is used to validate our proposed framework. Apart from the protocol operations considered in [2], our simulator further implements the post-backoff and its related procedure as well as the accurate backoff freezing operation discussed in [8]. When these additional implementations are disabled in our simulator, under saturation traffic condition, our simulator reproduces results matching that of [2] with the FHSS settings. For any of the presented simulation results, the radius of the 95% confidence interval is less than 1% of the mean.

A. Delay Performance Under ON/OFF Sources

Here we focus on the delay performance of a single hop IEEE 802.11 WLAN consisting of a finite number of stations modeled as on/off sources. Each station can be in one of two states: an idle state and an active state. A station is active when its local buffer carries a data frame ready for transmission or the data frame is being transmitted. The station transits to an idle state when the data frame is successfully transmitted. After entering an idle state, the station remains in that state for an exponentially distributed time period and then it generates a new data frame and hence becomes active again.

We consider a network of k stations, each generates a fixed-size data frame. Based on our approach, the IEEE 802.11 service time distribution of a fixed-size data frame can be approximated by an E_{40}. To model the arrival process, a state-dependent Markovian process is used. This results in a state-dependent M/E_{40}/1/k system for the purpose of performance analysis. The parameters used for the protocol are given in Table I. The four-way handshaking access method is considered here. The numerical results computed by Successive Over-Relaxation (shown in solid lines) and the simulation results (shown in symbols) are compared in Fig. 6.

Fig. 6 demonstrates the effect of the data frame size on the delay performance. The benefit of using a larger data frame is clearly demonstrated as the protocol offers a higher achievable throughput level for a larger data frame size. In

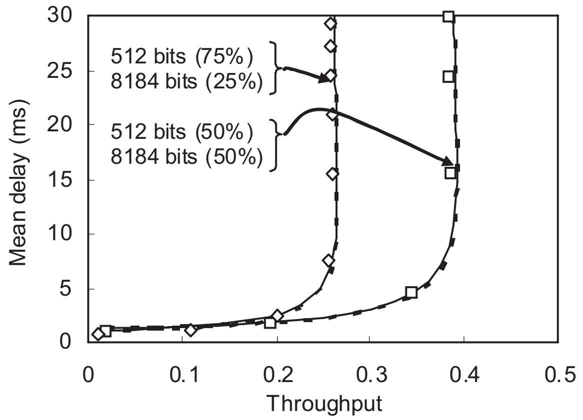


Fig. 8. Delay performance of the IEEE 802.11b MAC protocol for dual-fixed data frame sizes the four-way handshaking access method.

Fig. 7, we further compare the previous results with those obtained for the basic access method. For the case of 512-bit payload with the basic access method, the network becomes saturated with throughput at around 0.1. Given 50 saturated stations, using Little's formula to calculate the mean delay under the saturation [21], we get the mean delay to be around 23 ms, which is why the mean delay for 512-bit payload does not continue to increase beyond 23 ms.

A comparison between the four-way handshaking and the basic access methods indicates that basic access method offers higher achievable throughput than that of the four-way handshaking access method for the same data frame size. A different result was reported in our earlier work in [9] where the four-way handshaking access method offered higher achievable throughput on a low speed FHSS channel. This is mainly because the four-way handshaking access method leads to higher protocol efficiency when transmission time of its payload is much longer than that of its transmission overhead, which is the case of FHSS. However, when the payload transmission time is shortened by 11 times on an IEEE 802.11b channel, with only slight decrease in transmission overhead compared to that of FHSS, relatively high wastage due to transmission overhead is resulted. As for the basic access method, lacking the additional RTS/CTS helps maintain low transmission overhead on a high speed channel; hence for the considered payload sizes in Fig. 7, the basic access method offers better performance in terms of mean delay and maximum achievable throughput.

B. Delay Performance Under a Dual Data Frame Size Case

To further investigate the effect of data frame size on delay, we now consider a case with dual fixed data frames. This assumption is said to be more realistic than the one based on fixed data frames [22], and it effectively increases the variance of the data frame size distribution. In particular, we consider two possible data frame payload sizes: 512 bits and 8184 bits. Two cases are assumed here. In the first, 75% (25%) of the payloads are of 512 (8184) bits long, and in the second case, half of the payloads are of 512 bits long and the other half are of 8184 bits long.

Using our approach, to model the dual data frame size distribution, we construct a state-dependent M/PH/1/k system of which the PH service time distribution is a Hyper-Erlangian distribution [23]. The balance equations for this system are provided in the Appendix. Numerical (shown in lines) and simulation (shown in symbols) results are plotted in Fig. 8. Moreover, we include the numerical results of Fig. 6 (depicted by dotted lines in Fig. 8) to compare the delay performance between the case of fixed data frame size and the case of dual fixed data frame size, maintaining the same mean value for both cases.

As seen in Fig. 8, an excellent agreement between the numerical and simulation results has again been achieved. This also indicates the robustness and versatility of the approach where the model is not limited to a certain data frame distribution. Our interest here is also the investigation of the effect of the data frame distribution on the delay performance. Surprisingly, the increase in variance in payload distribution has only a minor effect on the mean delay, as the mean delay of the payloads of dual size appears to be just slightly higher than that of the fixed size.

C. Delay Performance Under Various Data Frame Size Distributions

Here we study the mean delay performance for the following four payload size distributions: (i) fixed, (ii) dual-fixed, (iii) geometric and (iv) dual-geometric. The dual-geometric random variable, X, has the probability density function

$$\Pr\{X = x\} = \sum_{b=1}^{2} \alpha_b (1-q_b)^{x-1} q_b, \; x = 1, 2, \ldots$$

and

$$\begin{aligned} E[X] &= \alpha_1 q_1^{-1} + \alpha_2 q_2^{-1}, \\ Var(X) &= 2\alpha_1 q_1^{-2} + 2\alpha_2 q_2^{-2} - (E[X])^2. \end{aligned}$$

Considering the traffic arrival process in the previous subsection, the performance analysis for the case of the fixed and dual-fixed payload size distributions can be performed using M/E_{40}/1/50 and M/PH/1/50 systems respectively. For the payload size having a geometrically distributed function, M/M/1/50 is used instead. This is because as we have learned in the previous subsections, the IEEE 802.11 protocol service time is highly dominated by the data frame transmission time, thus exponential service time, which is the continuous-time version of its geometric counterpart, is an appropriate choice. For dual-geometric payload size distribution, M/PH/1/k is used where the service time distribution follows a hyper-exponential distribution.

The delay performances of the IEEE 802.11 MAC protocol under the four different payload size distributions are presented in Fig. 9. Firstly, we notice from the figure a good match between the numerical (shown in lines) and the simulation (shown in symbols) results. Secondly, the mean delay performance is more sensitive to the distribution function than just the variance. For example, at a system throughput of 0.3, the order of the mean delays from high to low is the dual-geometric, the geometric, followed by the dual-fixed, and finally the fixed payload sizes, despite the fact that the variance

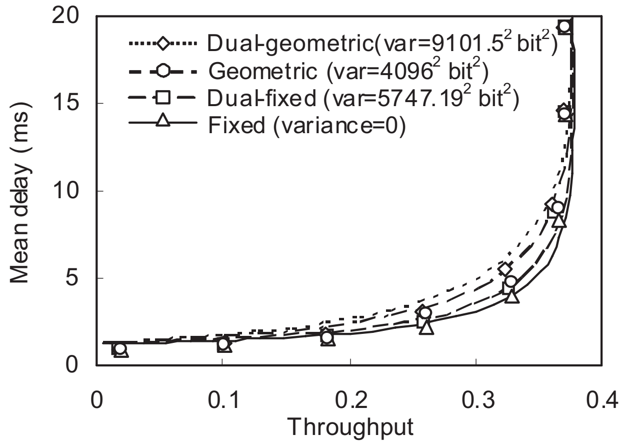


Fig. 9. Delay performance of the IEEE 802.11b MAC protocol for various data frame size distributions with mean payload size of 4096 bits.

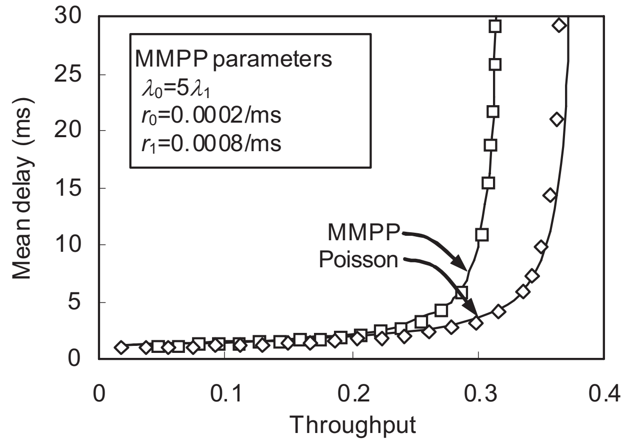


Fig. 10. Delay performance of the IEEE 802.11b MAC protocol under MMPP arrival process.

of the dual-fixed is higher than that of the geometric payload sizes. Hence identifying the actual distribution of payload sizes is essential to accurately study the delay performance of the IEEE 802.11b MAC protocol.

D. The Delay Performance Under MMPP Arrival Model

A study of data traffic [25] has shown that data traffic generally exhibits long range dependent (LRD) behavior, however, a later study in [26] has observed that the LAN traffic loses its self-similarity property when the time-scale is in the order of days. Hence in this study, the bursty MMPP with appropriate parameters (which is short range dependent [27]) is used as our LAN traffic model. However, justification and fitting of MMPP parameters to LAN traffic is out of the scope of this paper, and the reader is referred to [28],[29], for information on these topics.

The system we use to model IEEE 802.11 under an arrival process that follows MMPP is an MMPP/E_{40}/1/k system with state-dependent service rate. While the arrival rate is not state-dependent, arrivals are truncated at k, meaning when all k stations are active, there can be no further arrivals. This assumption makes the model more realistic in a LAN environment than one based on the assumption of an infinite number of stations. Numerical results (shown in solid lines) of delay performance of IEEE 802.11 under MMPP are presented in Fig. 10. An excellent match between the numerical and the simulation results (shown in symbols) has again confirmed the accuracy of the modeling approach. The payload sizes are fixed: each is 4096-bit long.

It is also our interest to evaluate the effect of burstiness on the delay performance of IEEE 802.11. To this aim, we include an equivalent Poisson arrival process for comparison. As can be seen, the results obviously show a clear impact of traffic burstiness. While IEEE 802.11 achieves as high as 37% throughput level before the delay exceeds 30 ms for the case of memoryless traffic, when MMPP is considered, no more than 32% throughput can be achieved. This suggests that traffic burstiness adversely affects the performance of IEEE 802.11, and that the assumption of Poisson arrival traffic overestimates the performance of IEEE 802.11 under bursty traffic.

V. Conclusion

Aiming to provide analytical means for performance evaluation of the IEEE 802.11 under statistical traffic, we have initially focused on the analysis of the service process of IEEE 802.11 under saturation traffic conditions. This has led to the derivation of closed form expressions for the mean and variance of the service time of the IEEE 802.11 MAC protocol. By studying the statistical characteristics of the IEEE 802.11 protocol service time, we have demonstrated that it can accurately be approximated by a PH distribution for the purpose of performance evaluation.

Furthermore, we have developed a new analytical framework based on the system approximation technique to study the performance of IEEE 802.11 under various realistic traffic conditions. The proposed framework is based on modeling the IEEE 802.11 service time distribution by a phase type distribution (in particular, Erlang or hyper-exponential), and allows for a wide range of Markovian models to describe the traffic behavior of the sources. This framework enables studies of IEEE 802.11 performance under many realistic network situations that are difficult to study using other known techniques. The agreement between numerical and simulation results, demonstrated for a wide range of models and scenarios, has confirmed the robustness and versatility of our new approach.

From the numerical results, the delay performance of the IEEE 802.11 MAC protocol is presented for various network scenarios and load conditions; of special interest are the insensitivity of the average MAC delay to the data frame size distribution, and even more so to its variance. However, the average MAC delay is demonstrated to be sensitive to traffic burstiness.

Appendix A
The M/PH/1/k System With Hyper-Erlangian Service Process

Recall that the system state corresponds to the number of active stations in an IEEE 802.11 single hop WLAN. Let k be the number of stations in the network, and p_n be the probability that the system is in state n, equivalently, there are n out of k active stations in the system, with $0 \leq n \leq k$, the mean arrival rate, $\overline{\lambda}$, observed by the system is

$$\overline{\lambda} = \sum_{n=0}^{k} \left(\lambda(n) \cdot p_i \right) \tag{A.1}$$

and the throughput, ρ, is given by

$$\rho = \overline{\lambda} \cdot d_t \tag{A.2}$$

where $\lambda(n)$ is the mean arrival rate of system state n, and d_t is the mean transmission time of the payload in a data frame.

Since packet arrival processes of stations are statistically identical and since each station transmits a new data frame after an exponential random time provided that it has just completed a data frame transmission, the overall arrival rate of a particular state n can be expressed as

$$\lambda(n) = \lambda_{ind} \cdot (k-n), \; n = 0, 1, \ldots, k \tag{A.3}$$

where λ_{ind} is the individual arrival rate of a station.

To facilitate the dual-fixed payload size distribution, we construct an M/PH/1/k system where the service time distribution is Hyper-Erlangian with j exponential stages. Two service rates are required in this system, which are $\mu_b(n)$ where $b = 1, 2$. Moreover, we define α_b to be the probability that the service rate of a data frame is $\mu_b(n)$. Appropriate settings to the system parameters reduce the system to some special cases, such as (i) setting $j = 1$ reduces the system to the hyper-exponential payload size distribution case; (ii) setting $\alpha_1 = 1$ and $\alpha_2 = 0$ reduces the system to an M/E$_j$/1/k system; and (iii) combination of the above two settings reduces the system to an M/M/1/k system.

The balance equations for such an M/PH/1/k system are

$$\begin{aligned}
0 =& -\lambda(0)p_0 + j\mu_1(1)p_{1,1,1} + j\mu_2(1)p_{1,1,2}, \\
0 =& -\left(\lambda(1) + j\mu_b(1)\right) p_{1,j,b} + \alpha_b\lambda(0)p_0 + \alpha_b j\mu_1(2)p_{2,1,1} \\
& + \alpha_b j\mu_2(2)p_{2,1,2}, (b = 1, 2) \\
0 =& -\left(\lambda(1) + j\mu_b(1)\right) p_{1,i,b} + j\mu_b(1)p_{1,i+1,b}, \\
& (i = 1, 2, \ldots, j-1; \; b = 1, 2) \\
0 =& -\left(\lambda(n) + j\mu_b(n)\right) p_{n,j,b} + \lambda(n-1)p_{n-1,j,b} \\
& + \alpha_b j\mu_1(n+1)p_{n+1,1,1} + \alpha_b j\mu_2(n+1)p_{n+1,1,2}, \\
& (n = 2, 3, \ldots, k-1; \; b = 1, 2) \\
0 =& -\left(\lambda(n) + j\mu_b(n)\right) p_{n,i,b} + \lambda(n-1)p_{n-1,i,b} \\
& + j\mu_b(n)p_{n,i+1,b}, \\
& (n = 2, 3, \ldots, k-1; \; i = 1, 2, \ldots, j-1; \; b = 1, 2) \\
0 =& -j\mu_b(k)p_{k,j,b} + \lambda(k-1)p_{k-1,j,b}, (b = 1, 2) \\
0 =& -j\mu_b(k)p_{k,i,b} + \lambda(k-1)p_{k-1,i,b} + j\mu_b(k)p_{k,i+1,b}, \\
& (i = 1, 2, \ldots, j-1; \; b = 1, 2)
\end{aligned}$$

with $p_n = \sum_{i=1}^{j} \left(p_{n,i,0} + p_{n,i,1}\right)$ where $n = 1, 2, \ldots, k$, and $\sum_{n=0}^{k} p_n = 1$.

The quantity $j \cdot \mu_b(n)$ represents the service rate of state $\{n, i, b\}$ for $n \geq 1$; p_0 is the probability that the system is in state 0, and $p_{n,i,b}$ is the probability that the system is in state $\{n, i, b\}$. State $\{n, i, b\}$ represents the state where there are n active stations and the current service process has i remaining phases to serve a station carrying type b data frame size.

The mean arrival rate and the throughput of the system can be computed by (A.1)-(A.3) given a particular λ_{ind} value. The mean delay can be obtained by $D = (\overline{\lambda})^{-1} \sum_{n=0}^{k} (np_n)$.

Appendix B
The MMPP/E$_j$/1/k System

The steady state balance equations for the MMPP/E$_j$/1/k queue are

$$\begin{aligned}
0 =& -\left(\lambda_m + r_m\right) p_{0,m} + j\mu(1)p_{1,m,1} + r_{m+1}p_{0,m+1}, \\
0 =& -\left(\lambda_m + j\mu(1) + r_m\right) p_{1,m,j} + \lambda_m p_{0,m} \\
& + j\mu(2)p_{2,m,1} + r_{m+1}p_{1,m+1,j}, \\
0 =& -\left(\lambda_m + j\mu(1) + r_m\right) p_{1,m,i} + j\mu(1)p_{1,m,i+1} \\
& + r_{m+1}p_{1,m+1,i}, (i = 1, 2, \ldots, j-1) \\
0 =& -\left(\lambda_m + j\mu(n) + r_m\right) p_{n,m,j} + j\mu(n+1)p_{n+1,m,1} \\
& + \lambda_m p_{n-1,m,j} + r_{m+1}p_{n,m+1,j}, (n = 2, 3, \ldots, k-1) \\
0 =& -\left(\lambda_m + j\mu(n) + r_m\right) p_{n,m,i} + j\mu(n)p_{n,m,i+1} \\
& + \lambda_m p_{n-1,m,i} + r_{m+1}p_{n,m+1,i}, \\
& (n = 2, 3, \ldots, k-1; \; i = 1, 2, \ldots, j-1) \\
0 =& -\left(j\mu(k) + r_m\right) p_{k,m,j} + \lambda_m p_{k-1,m,j} + r_m p_{k,m+1,j}, \\
& (n = 2, 3, \ldots, k-1) \\
0 =& -\left(j\mu(k) + r_m\right) p_{k,m,i} + j\mu(k)p_{k,m,i+1} + \lambda_m p_{k-1,m,i} \\
& + r_{m+1}p_{k,m+1,i}, (i = 1, 2, \ldots, j-1)
\end{aligned}$$

with

$$\begin{aligned}
& m+1 = \begin{cases} 1, & \text{if } m = 0 \\ 0, & \text{if } m = 1 \end{cases} \\
& p_0 = p_{0,0} + p_{0,1} \\
& p_n = \textstyle\sum_{i=1}^{j} \left(p_{n,0,i} + p_{n,1,i}\right), \quad n = 1, 2, \ldots, k
\end{aligned}$$

and

$$\sum_{n=0}^{k} p_n = 1$$

where $m = 0$ or 1 is the mode of MMPP, λ_m is the mean arrival rate in mode m of MMPP, $\mu(n)$ is the service rate of a protocol for n saturated stations. State $\{n, m, i\}$ represents the state where there are n (non-zero) active stations, the current service process has i remaining phases to serve a station, and the current arrival mode of MMPP is m. For zero active station, the system is either in state $\{0, 0\}$ when $m = 0$, or state $\{0, 1\}$ when $m = 1$. In this model, the mean arrival rate observed by the system is $\overline{\lambda} = \sum_{m=0}^{1} (\lambda_m p_{0,m}) + \sum_{n=1}^{k-1} \sum_{i=1}^{j} \sum_{m=0}^{1} (\lambda_m p_{n,m,i})$.

References

[1] "IEEE Standard for Information Technology - Part 11: Wireless LAN MAC and PHY Specifications," IEEE Std 802.11-1999, 1999.

[2] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," *IEEE J. Select. Areas Commun.*, vol. 18, pp. 535-547, Mar. 2000.

[3] C. H. Foh and M. Zukerman, "Performance evaluation of IEEE 802.11," in *Proc. IEEE VTC Fall*, May 2001, pp. 841-845.

[4] T. S. Ho and K. C. Chen, "Performance analysis of IEEE 802.11 CSMA/CA medium access control protocol," in *Proc. IEEE PIMRC*, Oct. 1996, pp. 407-411.

[5] F. Cali, M. Conti, and E. Gregori, "IEEE 802.11 protocol: Design and performance evaluation of an adaptive backoff mechanism," *IEEE J. Select. Areas Commun.*, vol. 18, pp. 1774-1786, Sep. 2000.

[6] A. Grilo and M. Nunes, "Performance evaluation of IEEE 802.11e," in *Proc. IEEE PIMRC*, Sep. 2002, pp. 511-517.

[7] A. Grilo, M. Macedo, and M. Nunes, "A scheduling algorithm for Qos support in IEEE 802.11 networks," *IEEE Wireless Commun.*, vol. 10, no. 3, pp. 36-43, June 2003.

[8] C. H. Foh and J. W. Tantra, "Comments on IEEE 802.11 saturation throughput analysis with freezing of backoff counters," *IEEE Commun. Lett.*, vol. 9, no. 2, pp. 130-132, Feb. 2005.

[9] C. H. Foh and M. Zukerman, "Performance analysis of the IEEE 802.11 MAC protocol," in *Proc. European Wireless*, Feb. 2002, pp. 184-190.

[10] G. Zeng, H. Zhu, and I. Chlamtac, "A novel queueing model for IEEE 802.11 WLANs," in *Proc. 1st Annual Wireless Networking Symposium*, Oct. 2003.

[11] O. Tickoo and B. Sikdar, "Queueing analysis and delay mitigation in IEEE 802.11 random access MAC based wireless networks," in *Proc. IEEE Infocom*, Mar. 2004, vol. 2, pp. 1404-1413.

[12] H. Zhai, Y. Kwon, and Y. Fang, "Performance analysis of IEEE 802.11 MAC protocols in wireless LANs," *Wireless Commun. Mobile Comput.*, vol. 4, pp. 917-931, Dec. 2004.

[13] E. Ziouva and T. Antonakopoulos, "The IEEE 802.11 distributed coordination function in small-scale ad-hoc wireless LANs," *Intl. J. Wireless Inform. Networks*, vol. 10, no.1, pp. 1-15, Jan. 2003.

[14] F. Alizadeh-Shabdiz and S. Subramaniam, "A finite load analytical model for the IEEE 802.11 distributed coordinated function MAC," in *Proc. Workshop Modeling Optimization Mobile Ad Hoc Wireless Networks*, Mar. 2003, pp. 321-322.

[15] B. Li and R. Battiti, "Analysis of the IEEE 802.11 DCF with service differentiation support in non-saturation conditions," in *Proc. International Workshop Quality Future Internet Services*, Sep. 2004, pp. 64-73.

[16] D. Barman, I. Matta, E. Altman, and R. El Azouzi, "TCP optimization through FEC, ARQ and transmission power tradeoffs," in *Proc. WWIC*, Feb. 2004, pp. 87-98.

[17] "IEEE Standard for Information Technology - Part 11: Higher-Speed Physical Layer Extension in the 2.4 GHz Band," IEEE Std 802.11b-1999, 1999.

[18] D. Gross and C. M. Harris, *Fundamentals of Queueing Theory*, 3rd ed. New York: Wiley, 1998.

[19] D. M. Lucantoni, K. S. Meier-Hellstern, and M. F. Neuts, "A single-server queue with server vacations and a class of non-renewal arrival processes," *Advances in Applied Probability*, vol. 22, pp. 676-705, Sep. 1990.

[20] J. W. Tantra, C. H. Foh, G. Bianchi, and I. Tinnirello, "Performance analysis of the out-of-band signaling scheme for high speed wireless LANs," in *Proc. IEEE Globecom*, Dec. 2004, vol. 5, pp. 3002-3006.

[21] C. H. Foh and M. Zukerman, "Performance comparison of CSMA/RI and CSMA/CD with BEB," in *Proc. IEEE ICC*, June 2001, pp. 2670-2675.

[22] J. F. Shoch and J. A. Hupp, "Measured performance of an ethernet local network," *Commun. ACM*, vol. 23, no. 12, pp. 711-721, Dec. 1980.

[23] Y. Fang, "Hyper-Erlang distribution model and its application in wireless mobile networks," *Wireless Networks*, vol. 7, no. 3, pp. 211-219, May 2001.

[24] M. F. Neuts, *Matrix-Geometric Solutions in Stochastic Models - An Algorithmic Approach*. New York: Dover Publications Inc., 1981.

[25] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, "On the self-similar nature of ethernet traffic (extended version)," *IEEE/ACM Trans. Networking*, vol. 2, no. 1, pp. 1-15, Feb. 1994.

[26] A. Erramilli, O. Narayan, and W. Willinger, "Fractal Queueing Models," in *Frontiers in Queueing*, J. H. Dshalalow, ed., pp. 245-269. New York: CRC Press, Inc., 1997.

[27] C. H. Foh, B. Meini, B. Wydrowski, and M. Zukerman, "Modeling and performance evaluation of GPRS," in *Proc. IEEE VTC*, May 2001, pp. 2108-2112.

[28] K. S. Meier-Hellstern, "A fitting algorithm for Markov modulated Poisson processes having two arrival rates," in *European J. Operational Research*, vol. 29, no. 3, pp. 370-377, 1987.

[29] T. Yoshihara, S. Kasahara, and Y. Takahashi, "Practical time-scale fitting of self-similar traffic with Markov-modulated Poisson process," *Telecommun. Syst.*, vol. 17, pp. 185-211, May 2001.