Performance Behavior of IEEE 802.11 Distributed Coordination Function

J.S. Vardakas, M.K. Sidiropoulos and M.D. Logothetis
Wire Communications Laboratory,
Department of Electrical & Computer Engineering,
University of Patras,
265 04 Patras, Greece
Emails: jvardakas@wcl.ee.upatras.gr, m.sidiropoulos@ieee.org and m-logo@wcl.ee.upatras.gr

Abstract- The IEEE 802.11 protocol is the dominating standard for Wireless Local Area Networks (WLANs) and has generated much interest in the investigation and improvement of its performance. The IEEE 802.11 Medium Access Control (MAC) is mainly based on the Distributed Coordination Function (DCF) with two access mechanisms: Basic and Request To Send/Clear To Send (RTS/CTS). DCF uses a Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) protocol in order to resolve contention among wireless stations, and to verify successful transmissions. In this paper we present an extensive investigation of the performance of the IEEE 802.11 MAC protocol, with respect to throughput and delay. For the protocol analysis we propose a new model, which describes the protocol’s behavior in a great extent by incorporating and extending existing models. We also present a detailed analysis of the end-to-end delay through the study of the MAC delay and the queueing delay. We use the Z-transform of backoff duration to get the mean value, the variance and the probability distribution of the MAC delay. For the queueing analysis, first we consider an M/G/1 queue in order to provide a first look in the queueing delay. Second, we modify the input process of the queue so that the packet arrival process is described by an ON-OFF model which expresses the bursty nature of traffic. In our investigations we assume data rates of 1, 5.5 and 11 Mbps to highlight the effect of the bit rate on network performance for both access mechanisms. We validate the throughput and delay analysis by simulating the DCF, while we compare our models with existing models based on their results. The accuracy of our analyses was found to be quite satisfactory.

Keywords: CSMA/CA, IEEE 802.11b, Distributed Coordination Function, Throughput, Delay, ON-OFF model, Z-transform.

I. INTRODUCTION

Wireless access networks are experiencing a tremendous growth in the past few years, and therefore the performance evaluation of the supported protocols remains challenging. The IEEE has standardized PHY (physical layer) and MAC
through the IEEE 802.11 workgroup. The related protocol defines two medium access methods, the widely used DCF and the optional Point Coordination Function (PCF). The DCF uses the CSMA/CA protocol to allow contended access to the wireless medium under binary exponential backoff rules [1]. When using CSMA/CA, each station wishing to take control of the medium has to sense whether the channel is idle or not; if it is not idle, the station defers its transmission to a random time interval (in this paper, the terms medium and channel are used interchangeably). Upon each collision notified by the absence of an acknowledgment (ACK) frame, the bound of random time interval (contention window) is increased in order for a retransmission to be scheduled. The IEEE 802.11 protocol defines that a packet has a finite number of retransmission attempts to be transmitted. If the retransmission attempts of a packet exceeds this predefined number (the retry limit), the packet is dropped.

Several studies appear in the literature investigating the performance of the IEEE 802.11 protocol. Bianchi in [2] proposes a Markov process to demonstrate a simple and tractable analytical model for the saturation throughput of the WLAN, under ideal channel conditions. However, the Bianchi’s model does not take into account basic parameters of the IEEE 802.11 standard: the backoff suspension and finite retry attempts (retry limit). In [3], Wu et al. extends Bianchi’s model to include finite packet retry limits, but still ignoring the backoff suspension issue. In [4], Ziouva proposes a Markov chain model that introduces an additional transition state to the models of Bianchi and Wu, while using a new probability (denoting that the channel is busy), in order to confront the backoff suspension case. The Ziouva’s model lacks validation through simulation. As far as the additional state of the Markovian model is concerned, it is rather extraordinary, because it permits transmissions consecutively, neglecting the fact that a new backoff procedure must commence after a successful transmission. Needless to say, this case is not included in the IEEE 802.11 standard, while it causes an unfair use of the wireless channel [5]. Another model that considers the freezing of the backoff counter and the finite retry limits is found in [6] by Foh. This model does not permit transmissions consecutively; instead, it takes into account that the medium access probability actually depends on whether the previous period is idle or busy. The latter assumption necessitates three parameters for the system state description and makes the model more complicated.

In this paper, we further extend the aforementioned models of Wu and Ziouva and propose a new Markov chain model by assuming that the backoff procedure is suspended when the medium is sensed busy. We also consider finite retry limits. The new Markov chain model describes the IEEE 802.11 protocol in a greater extent than the existing related models, because it incorporates the basic parameters of the IEEE standards in a realistic way, as the transition diagram of the proposed Markov chain proves (section III). We make the same basic assumptions as in [6], while, in addition, we take into consideration the difference between the number of Contention Window (CW) sizes and the maximum backoff stage. Moreover, our model is simpler than that of Foh ([6]), since it uses only two parameters to describe each Markov chain state, and it is easier to implement. The new model leads to throughput analysis, which is
validated through simulation. The proposed model is compared with the Wu’s and Foh’s models, in respect of the saturation throughput. As the comparison proves, the throughput results of the proposed new model are closer to the simulation results than those of the Wu’s model, while it has a performance close to that of the Foh’s model.

Furthermore, we present an end-to-end packet delay analysis of the IEEE 802.11b standard, where both the medium access delay and the queueing delay are considered. As a first approach, we calculate the mean MAC delay following the analysis of [5]. Then, we proceed to a more detailed analysis. Based on the packet collision probability, we calculate the mean and the variance of the MAC delay by getting the Z-transform of the backoff duration according to [7]. Moreover, we determine the Probability Distribution Function (PDF) of the MAC delay through the Lattice-Poisson algorithm [8]. The mean and the variance of the MAC delay provide a coarse estimation of the MAC delay, while the PDF provides a fine estimation of the MAC delay. Having determined the mean and the variance of the MAC delay, we can calculate the mean queueing delay by considering approximate queueing service models. A queue is assumed for each station with a common server. First, we provide results for the M/G/1 queue [9], because of its simplicity, in order to obtain a first look in the queueing delay. Second, we consider a more realistic case, where the queue’s input process is described by an ON-OFF model, in order to highlight the bursty nature of traffic at the packet level. In these queueing models the common server is the wireless channel, while the service time is the MAC delay. Finally, the end-to-end delay is the sum of MAC and queueing delay. The accuracy of our end-to-end delay analysis is evaluated through simulation and was found to be quite satisfactory. Especially for the MAC delay, our analytical results are compared with those of [5].

This paper is organized as follows: Section II gives a brief overview of the DCF access method. Section III presents the proposed mathematical model for the transmission probability determination. In section IV, we provide the channel throughput under saturation, based on the transmission probability. We proceed to the end-to-end delay analysis in section V; subsection A contains the MAC delay analysis, while subsection B contains the queueing delay analysis. Section VI is the evaluation section and is divided in two subsections, A and B, for the throughput analysis and the delay analysis, respectively. Finally, we conclude in section VII.

II. OVERVIEW OF THE DCF

The DCF operation supports best effort asynchronous data transfer while using two access schemes, the Basic mode and the RTS/CTS mode, as defined in IEEE 802.11 standard [1]. Each station, wishing to transmit a packet, senses the medium to ascertain that there is no transmission in progress. If the medium is idle the station transmits its packet, otherwise the transmission is postponed until the medium is sensed free for a time interval greater than Distributed Inter-Frame Space (DIFS). The station defers the transmission, after elapsing the DIFS time, for an additional time,
which is randomly selected; after the additional time has elapsed, the station is permitted to transmit its packet. As far as
the verification of a successful data reception is concerned, it is done by the reception of an ACK packet from the
destination station, after a Short InterFrame Space (SIFS) time interval from the reception of the data packet. If an ACK
packet is not detected by the source station a retransmission is scheduled.

The randomly selected time interval that a station has to wait for before commencing transmission is known as
backoff interval and is defined by a value of a backoff counter. This value is uniformly chosen in the range \((0, W_i-1)\),
where \(W_i-1\) is known as Contention Window \((CW)\) of backoff stage \(i\), where \(i \in [0, m]\), while \(m\) represents the station’s
retry count. The \(CW\) is an integer chosen in the range \((CW_{\text{min}}, CW_{\text{max}})\), where \(CW_{\text{min}}\) and \(CW_{\text{max}}\) are determined by the
characteristics of the PHY layer. At the first transmission attempt, \(CW\) equals to \(CW_{\text{min}}\). After each unsuccessful
transmission \(W_i\) is doubled up to a maximum value of \(2^m W_0\), where \(m'\) is the number of different \(CW\) sizes. Note also
that: \(W_0 = (CW_{\text{min}} + 1)\) and \(2^m W_0 = (CW_{\text{max}} + 1)\); therefore (1):

\[
\begin{align*}
W_i &= 2^i W_0 & i \leq m' \\
W_i &= 2^m W_0 & i > m'
\end{align*}
\]

The IEEE 802.11b standard [10] specifies that \(m\) could be larger or smaller than \(m'\), depending on the employed
access scheme. For the Basic access scheme we have \(m = 7\) and \(m' = 5\), whereas for the RTS/CTS access scheme we
have \(m = 4\) and \(m' = 5\). Therefore, according to (1), \(CW_{\text{min}}\) and \(CW_{\text{max}}\) are equal to 31 and 1023, respectively; these
values are adequate for the Direct Sequence Spread Spectrum (DSSS).

If the medium is sensed idle for a time interval greater than DIFS, then the backoff counter decrements as long as the
medium remains idle. When another station starts transmitting a packet during the backoff procedure, the backoff
counter is frozen to its current value (backoff suspension) and resumes when the medium is sensed idle for a time
greater than DIFS interval.

In the RTS/CTS mode, the sender transmits a short RTS packet prior to the data packet, and the receiver responds
with a CTS packet, after a SIFS time interval. If the sender detects the CTS packet, it enters the first backoff stage after
a SIFS time interval. By the end of the backoff procedure, the transmission of the data packet is initialized. An ACK
packet is sent by the receiver in order to verify the successful reception of the data packet. The IEEE 802.11 in
RTS/CTS mode partly overcomes the hidden node problem.
III. MATHEMATICAL MODEL

We consider a WLAN that consists of \(n\) stations where each station has always a packet available for transmission in its transmission queue (this is called saturated station). Moreover, at each transmission attempt, regardless of the number of retransmissions suffered, each packet collides with probability \(p\), which is constant and independent of the number of the collisions that the packet has suffered in the past. This key assumption is the same as in [2], [3] and [4]. Furthermore, as in [4], we assume that \(p_b\) stands for the probability that the channel is busy. This probability is independent of the backoff procedure; that is, independent not only from the backoff stage (number of retransmissions), but also from the value of the backoff counter. In what follows, the term time-slot refers to a constant value \(\sigma\), whereas the term system time-slot refers to a variable time interval. Note that the system time-slot equals to the time-slot \(\sigma\), only if there is no medium activity on the channel (idle). Otherwise, the system time-slot is the time demanded, either to complete a successful transmission, or to perform a failed transmission.

Following the modelling and analysis proposed in [2], let \(\{s(t), b(t)\}\) be a bi-dimensional, discrete time Markov chain, which is shown in Fig. 1. Here \(s(t)\) is the stochastic process which represents the backoff stage \(i\) at time-slot \(t\), and \(b(t)\) is the stochastic process which represents the backoff counter (with different \(CW\) size) for a given station at time-slot \(t\).

The state transition diagram of the Markov chain model shown in Fig.1 has the following one-step transmission probabilities:

1. The backoff counter decrements when the station senses the channel idle:

\[
P[i, k | i, k + 1] = 1 - p_b, \quad k \in [0, W_i - 2], \quad i \in [0, m]
\]

2. The backoff counter freezes when the channel is busy:

\[
P[i, k | i, k] = p_b, \quad k \in [1, W_i - 1], \quad i \in [0, m]
\]

3. The station enters backoff stage 0, if it detects a successful transmission of its current frame:

\[
P[0, k | i, 0] = \frac{1 - p}{W_0}, \quad k \in [0, W_0 - 1], \quad i \in [0, m - 1]
\]

4. The station enters into the next backoff stage and chooses a new value for the backoff counter after an unsuccessful transmission:

\[
P[1, k | 0, k] = p_b, \quad k \in [1, W_0 - 1], \quad i \in [0, m - 1]
\]
The station reaches the last backoff stage and returns to the initial backoff stage after a successful or unsuccessful transmission:

\[ P[i, k | i-1, 0] = \frac{2}{W_i}, \quad k \in [0, W_i - 1] \quad i \in [1, m] \]

5. The station reaches the last backoff stage and returns to the initial backoff stage after a successful or unsuccessful transmission:

\[ P[0, k | m, 0] = \frac{1}{W_0}, \quad k \in [0, W_0 - 1] \]

Let \( b_{i,k} = \lim_{t \to \infty} P[s(t) = i, b(t) = k] \) be the stationary distribution of the Markov chain, with \( i, k \) integers and \( k \in [0, W_i - 1], \quad i \in [0, m] \). In steady state, the following relations are derived from the rate balance equations of the state transition diagram of Fig. 1:

(2) obtained by assuming local balance, i.e. rate up = rate down,

\[
b_{i-1,0} \cdot p = b_{i,0} \iff b_{i,0} = p^i b_{0,0}, \quad i \in [0, m]
\]

(3) obtained by assuming global balance, i.e. rate in = rate out.

\[
b_{i,k} = \begin{cases} \frac{W_i - k}{W_i (1 - p)}, & \text{if } i = 0 \\ \frac{1}{W_i (1 - p)} & 0 < i \leq m, \quad 0 < k \leq W_i - 1 \end{cases}
\]

Thus, using (2), (3) and the normalization condition for the stationary distribution,

\[
\sum_{i=0}^{m} \sum_{k=0}^{W_i-1} b_{i,k} = 1
\]

we can express the value \( b_{0,0} \) as shown in (5).
The value of $b_{0,0}$ is expressed as a function of the collision probability $p$ and the probability $p_b$ that the channel is busy. The probability that in a system time-slot at least one of $n$ remaining stations transmits is approximated by (2):

$$p = 1 - (1 - \tau)^{n-1}$$  \hspace{1cm} (6)

Furthermore, the channel is detected busy when at least one of the $n$-1 stations transmits during a system time-slot. Note also that a station remains with probability $p_b$ at state $(i,k)$, $k \geq 1$, when at least one of $n$-1 remaining stations transmit. Therefore, the values of $p_b$ and $p$ coincide:

$$p_b = \sum_{i=1}^{n-1} \left[ \binom{n-1}{i} \cdot \tau^i \cdot (1 - \tau)^{n-1-i} \right] = 1 - (1 - \tau)^{n-1}$$  \hspace{1cm} (7)

Using (5), (6) and (7) we can calculate the probability $\tau$ that a station transmits in a randomly selected time-slot:

$$\tau = \sum_{i=0}^{m} b_{i,0} = \frac{1 - p^{m+1}}{1 - p} \cdot b_{0,0}$$  \hspace{1cm} (8)

### IV. THROUGHPUT ANALYSIS

In order to calculate the normalized system throughput $S$, we use the ratio (2),

$$S = \frac{E\left[\text{Payload information in a time-slot}\right]}{E\left[\text{Length in a time-slot}\right]} = \frac{P_s P_s E[P]}{(1 - P_s) \cdot \sigma + P_s P_s T_s + (1 - P_s) P_s T_c}$$  \hspace{1cm} (9)

where $E[P]$ is the average packet length, and $T_s, T_c$ are the average time that the channel is sensed busy due to a successful or unsuccessful transmission, respectively. The probability $P_s$ that an occurring packet transmission is
successful, and the probability $P_{s}$, that there is at least one transmission in a randomly selected system time-slot, are respectively given by ([2]):

\[
P_{s} = n \frac{\tau (1 - \tau)^{n-1}}{1 - (1 - \tau)^{n}} \tag{10}
\]

\[
P_{s} = 1 - (1 - \tau)^{n} \tag{11}
\]

The average length of a system time-slot in (9) is calculated considering that a system time-slot: (a) is empty with probability $(1 - P_{s})$, (b) includes a successful transmission with probability $P_{p}P_{s}$ and (c) includes a collision with probability $(1 - P_{p})P_{s}$.

The values of $T_{s}$ and $T_{c}$ depend on the access mechanism, considering the ACK and CTS timeout effect ([11]):

\[
\begin{align*}
T_{s}^{\text{Bas}} &= \text{DIFS} + H + T_{D} + \text{SIFS} + T_{\text{ACK}} + 2 \cdot \delta \\
T_{c}^{\text{Bas}} &= \text{DIFS} + H + T_{D} + \text{SIFS} + T_{\text{ACK}} + 2 \cdot \delta
\end{align*} \tag{12}
\]

\[
\begin{align*}
T_{s}^{\text{RTS}} &= \text{DIFS} + H + T_{\text{RTS}} + T_{\text{CTS}} + T_{D} + \\
&+ 3 \cdot \text{SIFS} + T_{\text{ACK}} + 4 \cdot \delta \\
T_{c}^{\text{RTS}} &= \text{DIFS} + T_{\text{RTS}} + \text{SIFS} + T_{\text{CTS}} + 2 \cdot \delta
\end{align*} \tag{13}
\]

where $T_{D}$, $T_{\text{ACK}}$, $T_{\text{RTS}}$ and $T_{\text{CTS}}$ is the time required to transmit the data packet, ACK, RTS and CTS respectively, while $\delta$ is the propagation delay and $H = \text{MAC} + \text{PHY}$ is the required time to transmit the packet header.

V. END-TO-END DELAY ANALYSIS

The majority of analytical work on the performance of IEEE 802.11 ([4], [5], [12], [12] and [13]) focuses on calculating the mean delay of the medium access; little attention has been drawn to the queueing delay investigation ([15]). Herein, we study the average end-to-end packet delay, which can be calculated as the sum of the average MAC delay, $E[M]$, and the average queueing delay $E[Q]$. Such a sum is possible due to the linear characteristics of the mean values.
A. Medium Access Delay Analysis

Firstly, we demonstrate the analysis for calculating the average MAC delay, $E[M]$, according to the model of [5]. The $E[M]$ equals to the time interval from the time point that the packet is ready for transmission at the head of the station’s queue, until the verification of the successful transmission. It is given by ([5]):

$$E[M] = E[X] \cdot E[slot]$$ (14)

where $E[X]$ is the average number of the system time-slots required for a successful transmission and $E[slot]$ is the average length of the system time-slots, given by the denominator of (9). The value of $E[X]$ can be calculated by multiplying the average number of the system time-slots $d_i$ that a packet delays in each backoff stage, by the probability $q_i$ for the packet to reach this backoff stage, provided that this packet is not discarded. Therefore ([5]):

$$d_i = \frac{1}{W_i} \cdot \sum_{x=0}^{W_i} x = \frac{W_i + 1}{2}, \quad i \in [0,m]$$ (15)

and

$$q_i = \frac{p^i - p^{i+1}}{1 - p^{i+1}}, \quad i \in [0,m]$$ (16)

Hence, $E[X]$ is given by:

$$E[X] = \sum_{i=0}^{m} d_i \cdot q_i = \sum_{i=0}^{m} \frac{(W_i + 1) \cdot (p^i - p^{i+1})}{2(1 - p^{i+1})}$$ (17)

We have calculated the average MAC delay according to [5], as a reference point. In what follows, we propose a new analysis for calculating the mean, the variance and the PDF of the MAC delay. The mean and the variance are obtained by using the Z-transform of the backoff duration ([7]), given that the time is considered discrete because it is measured in time-slots, as follows.

The interruption of the backoff period is a result of two different events: the collision of two or more stations with probability $p$ and the transmission of only one station other than the tagged one, with probability:
\[ p' = \binom{n-1}{1} \cdot \tau \cdot (1-\tau)^{n-2} \quad (18) \]

In order to calculate the backoff duration, we envision the Binary Exponential Backoff (BEB) Algorithm as a function of two coordinates \((x, y)\), where \(x \in [0, m]\) represents the backoff stage and \(y \in [0, W' - 1]\) represents the value of the backoff counter at the backoff stage \(x\), as shown in Fig. 2. The decrement of the backoff counter occurs when the channel is not interrupted, while the station stays at the same state as a result of two different events. Therefore, the Probability Generating Function (PGF) of each state is given by:

\[
D_{x,y}(Z) = \frac{(1-p)Z^y}{1-(p' S(Z) + (p-p') C(Z))} \quad (19)
\]

where \(S(Z)\) and \(C(Z)\) are the Z-transforms of the transmission period and the collision period, and are respectively given by:

\[
S(Z) = Z^{T_s} \quad (20)
\]

\[
C(Z) = Z^{T_c} \quad (21)
\]

Nevertheless, the backoff duration is not doubled after \(m'\) times, but stays at the same value for the remaining backoff stages:

\[
D_x(Z) = \begin{cases} 0 & 0 \leq x \leq m' \\ \frac{\sum_{y=0}^{W'} D_{x,y}(Z)}{D_{m'}(Z)} & m' < x \leq m \end{cases} \quad (22)
\]

Thus, for each \(x\) the Z-transform of the backoff duration is given by:

\[
BD(Z) = (1-p) \cdot S(Z) \sum_{i=0}^{m} \left( (p \cdot C(Z))^i \cdot \prod_{j=0}^{x} D_j(Z) \right) + (p \cdot C(Z))^{m+1} \cdot \prod_{j=0}^{m} D_j(z) \quad (23)
\]
The first term of the second part of (23) indicates the transmission delay multiplied by the delay encountered in the previous $x$ and $y$ stages, while the second term accounts for the delay of the dropping packet, which has encountered in all $x$ collisions.

Having determined the Z-transform of the backoff duration, we derive the mean value $E[M]$ and the variance $\text{Var}[M]$ of the MAC delay by taking the derivative of (23), with respect to $Z$:

$$E[M] = BD'(Z)\big|_{Z=1}$$

$$\text{Var}^2(M) = (BD'(Z)\big|_{Z=1})^2 + BD'(Z)\big|_{Z=1} - (BD'(Z)\big|_{Z=1})^2$$

In order to calculate the PDF of the MAC delay, we need the inverse transform of the Z-transform of the backoff duration [16]. To this end we use the expression:

$$BD(Z) = \sum_{k=0}^{\infty} d_k Z^k$$

The goal is to calculate $d_k$, which expresses the PDF of the backoff duration. A method that gives the inverse Z-transform with a predefined error bound is the Lattice-Poisson Algorithm [8], which is valid for $|d_k| \leq 1$. Since $d_k$ is the PDF, we obtain:

$$d_k = \frac{1}{2\pi r^k} \sum_{b=1}^{2k} (-1)^b \text{Re}(BD(re^{j2\pi k}))$$

where $\text{Re}(BD(Z))$ stands for the real part of the complex $BD(Z)$.

Eq. (27) is derived by integrating $BD(Z)$ over a circle with radius $r$, where $0< r <1$. For practical reasons we suppose that the predefined approximation error is $r^{2k}$, therefore to have accuracy $10^{-\gamma}$ we let $r = 10^{-\gamma/2k}$ [8].
B. Queueing Analysis

We proceed with the queueing analysis, by considering two different queueing systems, both with the wireless medium as common server and both having infinite capacity. The first is a simple M/G/1 queueing system; in the second system we modify the M/G/1 so that the input process is described by an ON-OFF model.

In M/G/1, the service time has the average value of the average MAC delay, $E[M]$. By applying the corresponding formula for the mean queueing delay, $E[Q]$, we obtain [9]:

$$E[Q] = \frac{A \cdot E[M]}{2 \cdot (1 - A)} \cdot \varepsilon$$

(28)

where $A$ is the offered traffic-load and $\varepsilon$ is the form factor of the holding time distribution, which equals to ([9]):

$$\varepsilon = 1 + \frac{E[M]^2}{\text{var}[M]}$$

(29)

where var$[M]$ is the variance of the service time distribution and is given by the square root of (25).

For the second queueing system, we consider that each node is a discrete-time queueing system, with infinite buffer capacity and a single server representing the wireless channel. The consideration of discrete time is possible due to the fact that according to the IEEE 802.11 all events take place in time units (e.g. the backoff counter decreases one unit per time-slot). The arrival process is described by a two-state Markov chain, that is, the arrivals are modeled by an ON-OFF model, as it is shown in Fig.3. The parameters $q_1$ and $q_2$ are the probabilities that the Markov chain remains in states ON and OFF, respectively. Calls arrive to the queue with mean arrival rate $\lambda$. When the system is in state ON and a new call arrives in the next time-slot, the system remains in state ON; this happens with probability $q_1$. When the system is in state OFF and a new call arrives, the system passes to state ON; this happens with probability $1-q_2$.

The probability of two consecutive arrivals in two consecutive time-slots is denoted by $f_1$ and equals to $q_1$. The probability that the system jumps to state OFF just after an arrival and then returns to the state ON during the next time-slot is $f_2 = (1-q_1)(1-q_2)$. Thereupon, the general expression of the inter-arrival distribution, $f_n$, is ([17]):

$$f_n = \begin{cases} q_1 & \text{if } n = 1 \\ (1-q_1)q_2^{n-1}(1-q_2) & \text{if } n > 1 \end{cases}$$

(30)
where \( n \) denotes the number of time-slots between two consecutive arrivals.

The probability generating function (PGF) of \( f_n \) is denoted by \( F(z) \) and is given by ([17]):

\[
F(z) = q_1z + (1 - q_1)(1 - q_2) \frac{z^2}{1 - q_2z}
\]  

(31)

The mean inter-arrival time can be calculated as:

\[
\frac{dF(z)}{dz} \bigg|_{z=1} = \frac{2 - q_1 - q_2}{1 - q_2} = \frac{1}{\lambda}
\]  

(32)

The stationary queue size distribution \( \pi_m \) is in the following form ([17]):

\[
\pi_m = \begin{cases} 
1 - \xi & m = 0 \\
(1 - \rho)\rho^{m-1} & m \geq 1
\end{cases}
\]  

(33)

where \( \xi \) is a constant and \( \rho \) is the unique root of the equation:

\[
z = F(\mu z + (1 - \mu))
\]  

(34)

where \( \mu \) is the mean service time of the system and equals to the mean MAC delay. Using (33) and (34) we find that:

\[
\rho = \frac{1 - \mu}{\mu} \left[ 1 - \frac{1}{\mu(1 - q_1 - q_2) + q_2} \right]
\]  

(35)

By equating the mean probability of the arrival to the mean probability of departure (i.e. by assuming local balance), we find that:

\[
\lambda = \mu(1 - \pi_0) \Leftrightarrow \pi_0 = 1 - \frac{\lambda}{\mu}
\]  

(36)
The mean queue size, \( E[S] \), and the mean waiting time, \( E[Q] \), are calculated through the Little’s theorem and (35), as follows:

\[
E[S] = \sum_{i=1}^{\infty} l_{i} \pi_{i} = \frac{\xi}{1 - \rho} = \frac{\lambda}{\mu(1 - \rho)}
\]  

(37)

\[
E[Q] = \frac{E[S]}{\lambda} = \frac{1}{\mu(1 - \rho)}
\]  

(38)

Finally, the average end-to-end packet delay \( E[D] \) is given by the summation:

\[
E[D] = E[M] + E[Q]
\]  

(39)

VI. Evaluation

We consider a WLAN topology consisting of an Access Point (AP), placed in the center of an area of 100m x 100m, and of \( n \) stations placed on a circle of radius \( R=50m \) around the AP. In order to pass over the problem of hidden terminals, we assume that all stations have Line of Sight. All stations send traffic at such a rate that ensures saturated conditions (the transmission queues remain nonempty). To assess the performance behavior of this WLAN, we apply the proposed mathematical models, while the WLAN is simulated, for comparison and model’s evaluation. We mostly use the NS-2 simulator [18], except for the end-to-end delay analysis when an ON-OFF model is assumed for the packet arrival process, where the OPNET simulator is used [19]. Unless otherwise specified, the values of the parameters used both in the simulation and the analysis, come from the values of the DSSS parameters found in [20]; they are summarized in Table I. Note that the analytical model is independent of the specific values of the parameters, so that it can be implemented for different PHY layers. All simulation results have been obtained with five replications and with different seed numbers each time; from these five runs, mean values have been calculated with a 95% confidence interval [21]. Since the resultant reliability ranges were found to be very small, only the mean values are shown in the figures.
A. Throughput Results

First, the variation of the saturation throughput against the number of the contenting stations is presented in Fig. 4, for data rate of 1 Mbps and for both Basic and RTS/CTS access mechanisms; the saturation throughput was determined by three models: a) the proposed new model in this paper, b) the Wu’s model, and c) the Foh’s model, as well as by simulation. As one can observe in Fig. 4, the new model and Foh’s model are closer to the simulation results than the Wu’s model for both access mechanisms. This is due to the fact that the new model and Foh’s model take into account the backoff suspension. The new model and Foh’s model have almost the same performance. More precisely, the Foh’s model performs slightly better, but the new model has the advantage of an easy implementation. As it was anticipated, the RTS/CTS mode gives a greater throughput than the Basic mode for all number of stations. This is because in RTS/CTS mode the collisions occur only among the small sized control packets; therefore, the bandwidth wasted on collisions in RTS/CTS mode is much less than the bandwidth wasted at collisions in the Basic mode.

The saturation throughput is also assessed by the new model and Foh’s model, for data rates at 5.5 and 11 Mbps, shown in Fig. 5 and 6, respectively. In both cases, our analysis (new model) is validated by the simulation. A similar performance between the new model and Foh’s model is observed, like in Fig. 4. It is worth noticing that when the data rate increases, the Basic mode is preferable than the RTS/CTS mode; this is because the RTS/CTS control packets are traveling across the network at the lower rate of 1 Mbps. Figures 4, 5 and 6 also show that the saturation throughput decreases as the data rate increases. This is attributed to the fact that the packet transmission time is reduced with the increase of the data rate, contrary to the collisions which increase and to the DIFS, SIFS and time-slot intervals which remain invariable.

The saturation throughput depends on the size of the transmitted packets, so it is evident that the effect of this parameter must be investigated. In particular, Fig. 7 and 8 illustrate the throughput as it is determined by the new model against packet size for data rates of 1, 5.5 and 11 Mbps for a medium network size (n=20), for both access mechanisms respectively. Interestingly, in all curves the throughput increases for higher packet sizes, a situation that can be explained by the fact that larger packets corresponds to greater periods of contention, therefore lower collision probability. Comparison of Fig. 7 and 8 clearly reveals that the Basic access method outperforms the RTS/CTS method for any packet size at 5.5 and 11 Mbps, whereas the opposite is observed only for small packet sizes (< 3 Kbytes) at 1 Mbps. Similar figures (not shown for correspondence) for small and large network sizes (n=5 and n=50, respectively) depicts that the saturation throughput decreases as the network size increases, due to the multiple number of collisions. Furthermore, higher data rates renders lower throughput, a result that is in concordance with Fig. 4, 5 and 6. Consequently, the selection of the value of packet size should be attentive, according to the size of the network and the data rate employed.
Another parameter that is examined is the minimum value of the Contention Window (\(CW_{\text{min}}\)), whose effect on the saturation throughput (determined by the new model) for a medium sized network (\(n=20\)) can be monitored in Fig. 9, 10 and 11 for 1, 5.5 and 11 Mbps respectively. These figures show that all curves have an increment up to a certain value of \(CW_{\text{min}}\); after this point, they become descending. The explanation for this behaviour is as follows. For small values of \(CW_{\text{min}}\), the collision probability is high; if, however, the contention window is significantly large, the long time spent in the backoff procedure degrades the performance of the WLAN. Similar curves are obtained for various network sizes. Carefully examining Fig. 9, 10 and 11 allows identifying the preferable value of \(CW_{\text{min}}\) according to the selected data rate and network size, for optimum performance.

B. Delay Results

We evaluate our delay analysis by considering initially the MAC delay analysis and afterwards the end-to-end delay analysis. For the MAC delay analysis, we consider two application examples. In the first example, a WLAN is considered with the parameters of Table I, but the packet length is 1023 bytes \((pl=8184 \text{ bits})\), as well as \(m=6\) and \(m'=5\), according to the corresponding example found in [5]. The selection of these parameters is made in order to compare the results of the new MAC delay analysis with the verified analytical results presented in [5]. The WLAN utilizes the basic access mechanism and 1 Mbps data rate. In the second example, a WLAN with the parameters of Table I is considered, while the parameter \(m'=7\) and \(m'=5\). In Table II, we comparatively present our analytical results for the MAC delay (obtained through equations (18) to (25)), together with the results obtained by applying the equations (14) to (17) (as in [5]), for both application examples. More precisely, the first column of Table II contains the number of stations; the second and fifth columns contain the mean MAC delay of the existing model, while the third and sixth column contain the mean MAC delay of the new MAC delay model. The new MAC delay model determines the variance of the MAC delay which is presented in the fourth and seventh columns. The results of Table II indicate that the two models produce almost the same mean MAC delay (while the variance is given by the new model only).

A first approximation of the end-to-end delay analysis is obtained by using the new MAC delay analysis and the analysis of the M/G/1 queueing system. The system is simulated through the NS-2 so that the packet arrival process is Poisson. This has only a theoretical value and offers a first look into the end-to-end delay. The average end-to-end delay versus the number of stations for both the Basic and RTS/CTS access mechanisms is depicted in Fig. 12 and 13, respectively, for data rates 1, 5.5 and 11 Mbps. The offered traffic load to each queue is \(A=0.8 \text{ erl}\). The curves in both figures indicate that the delay increases as the number of contending stations increases; this is observed in all data rates employed. In large sized networks packets suffer more collisions, the stations choose higher backoff stages, the size of
the contention window increases and therefore the stations experience longer delays. Moreover, the increase of the data rate affects constructively the delay, due to the fact that the time needed for data transmission is reduced.

Comparison of Fig. 12 and 13 also reveals that the delay observed in the RTS/CTS mode is lower than the delay observed in the Basic mode only at the data rate of 1 Mbps; the opposite behavior is observed at 5.5 and 11 Mbps. In RTS/CTS mode the RTS and CTS control packets, which are transmitted at 1 Mbps irrespective of the data rate, introduce additional overhead. This fact may not make a difference at 1 Mbps, whereas in higher data rates it influences the value of the delay. Consequently, the performance of the Basic mode seems to be inadequate only at 1 Mbps, and therefore the Basic mode should be used at higher data rates in order for the performance to be improved.

A better approximation of the end-to-end delay analysis is obtained by using the new MAC delay analysis and the analysis of a queueing system with the more realistic packet arrival process of the ON-OFF model. The offered traffic load at the ON state is \( A = 0.8 \) erl. The probabilities that the arrival process is at the ON and at the OFF state are \( q_1=0.9 \) and \( q_2=0.1 \), respectively. In Fig. 14 and 15, the analytical and simulation end-to-end delay results are illustrated, for the basic and RTS/CTS access mechanisms, respectively, for data rates of 1, 5.5 and 11 Mbps. The same impact of the data rates on the end-to-end delay as in Fig. 13 and 14 is observed. As Fig. 14 and 15 reveal, the analytical results are very close to the simulation results. It is worth noticing that if we use the ON-OFF model with \( q_1=1 \) and \( q_2=0 \) (i.e. the system is always at the ON state), we obtain almost the same results as those in Fig. 12 and 13. This consistency also validates our analysis.

VII. CONCLUSION

In summary, the performance of the IEEE 802.11 standard is extensively investigated with the aid of a new Markov chain model. The new model describes the IEEE 802.11 protocol in a relatively simple way and in a greater extent than existing models; it incorporates the important parameters of backoff suspension and finite retry limits, in a realistic way, as it is depicted in the proposed Markov chain. We provide analysis of the saturation throughput and the average end-to-end delay (i.e. including not only MAC delay but also queueing delay analysis). The accuracy of the proposed new model was validated by comparing it with the Wu’s model, the Foh’s model and simulation results, with respect to saturation throughput. The new model has a performance close to that of the Foh’s model, while it has the advantage of an easy implementation. Both models perform better than the Wu’s model. Besides, the presented throughput results in this paper aim at deriving useful guidelines for the proper selection of the values of the parameters of the WLAN (e.g. data rate, \( CW_{\text{min}} \), packet length), so as to ensure optimum network performance. As far as the accuracy of our delay analysis is concerned, it is validated through simulation for the end-to-end delay, while, for the mean MAC delay, our results are compared with existing results, well validated through simulation, and were found to be quite satisfactory. In
addition to the mean MAC delay, we provide results for the variation of the MAC delay. We study the effect of the data rates on the average end-to-end packet delay. The average end-to-end delay results, for both Basic and RTS/CTS access mechanisms, show that the Basic mechanism outperforms the RTS/CTS mechanism in higher transmission rates for the data. This conclusion also results from the saturation throughput investigation. Having considered an ON-OFF model to describe the bursty nature of the packet arrival process, in our future work we shall consider a more comprehensive model, like the MMPP/G/1, assuming in addition not only infinite but also finite queue size.

ACKNOWLEDGMENT

Work supported by the research program Caratheodory of the Research Committee of the University of Patras, and the project IENEΛ-No.03ΕΔ420, which is funded in 75% by the European Union – European Social Fund and in 25% by the Greek State – General Secretariat for Research and Technology. The authors would like to thank Dr. V. Vitsas (TEI Thessaloniki, Greece) for his helpful comments and Mr. I. Papapanagiotou (ECE, North Carolina State University, USA) for his support on the OPNET simulator.
REFERENCES


Figure 1: The state transition diagram for the Markov chain model.
Figure 2: Depiction of backoff stages.
Figure 3: State transition diagram of the ON-OFF arrival process
<table>
<thead>
<tr>
<th><strong>PARAMETERS FOR MAC AND DSSS PHY LAYER.</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Packet Payload</strong></td>
</tr>
<tr>
<td><strong>Channel Bit Rate</strong></td>
</tr>
<tr>
<td><strong>MAC Header</strong></td>
</tr>
<tr>
<td><strong>PHY Header</strong></td>
</tr>
<tr>
<td><strong>Propagation Delay</strong></td>
</tr>
<tr>
<td><strong>Time-slot (σ)</strong></td>
</tr>
<tr>
<td><strong>DIFS</strong></td>
</tr>
<tr>
<td><strong>SIFS</strong></td>
</tr>
<tr>
<td><strong>ACK</strong></td>
</tr>
<tr>
<td><strong>CTS</strong></td>
</tr>
<tr>
<td><strong>RTS</strong></td>
</tr>
</tbody>
</table>
Figure 4: Simulation versus analysis at 1 Mbps bit rate for both access mechanisms.
Figure 5: Simulation versus analysis at 5.5 Mbps bit rate for both access mechanisms.
Figure 6: Simulation versus analysis at 11 Mbps bit rate for both access mechanisms.
Figure 7: Saturation throughput versus packet size for the Basic access mechanism.
Figure 8: Saturation throughput versus packet size for the RTS/CTS access mechanism.
Figure 9: Saturation throughput versus the minimum contention window for 1 Mbps data rate.
Figure 10: Saturation throughput versus the minimum contention window for 5.5 Mbps data rate
Figure 11: Saturation throughput versus the minimum contention window for 11 Mbps data rate
<table>
<thead>
<tr>
<th>Number of stations</th>
<th>1st example (m=6, pl=8184 bits)</th>
<th>2nd example (m=7, pl=8224 bits)</th>
<th>1st example (m=6, pl=8184 bits)</th>
<th>2nd example (m=7, pl=8224 bits)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Existing model</td>
<td>New MAC delay model</td>
<td>Existing model</td>
<td>New MAC delay model</td>
</tr>
<tr>
<td></td>
<td>Mean (sec)</td>
<td>Mean (sec)</td>
<td>Variance</td>
<td>Mean (sec)</td>
</tr>
<tr>
<td>5</td>
<td>0.0499</td>
<td>0.0491</td>
<td>0.005693</td>
<td>0.0502</td>
</tr>
<tr>
<td>10</td>
<td>0.1068</td>
<td>0.1045</td>
<td>0.00792</td>
<td>0.1077</td>
</tr>
<tr>
<td>15</td>
<td>0.1660</td>
<td>0.1644</td>
<td>0.01331</td>
<td>0.1686</td>
</tr>
<tr>
<td>20</td>
<td>0.2257</td>
<td>0.2234</td>
<td>0.01752</td>
<td>0.2311</td>
</tr>
<tr>
<td>25</td>
<td>0.2852</td>
<td>0.2823</td>
<td>0.02235</td>
<td>0.2942</td>
</tr>
<tr>
<td>30</td>
<td>0.3442</td>
<td>0.3407</td>
<td>0.02852</td>
<td>0.3575</td>
</tr>
<tr>
<td>35</td>
<td>0.4026</td>
<td>0.3985</td>
<td>0.03551</td>
<td>0.4127</td>
</tr>
<tr>
<td>40</td>
<td>0.4602</td>
<td>0.4556</td>
<td>0.04489</td>
<td>0.4841</td>
</tr>
<tr>
<td>45</td>
<td>0.5272</td>
<td>0.5220</td>
<td>0.05735</td>
<td>0.5471</td>
</tr>
<tr>
<td>50</td>
<td>0.5934</td>
<td>0.5907</td>
<td>0.06821</td>
<td>0.6098</td>
</tr>
</tbody>
</table>
Figure 12: Average end-to-end delay versus the number of stations for the Basic access mode and the M/G/1 queueing model.
Figure 13: Average end-to-end delay versus the number of stations for the RTS/CTS access mode and the M/G/1 queueing model.
Figure 14: Average end-to-end delay versus the number of stations for the Basic access mode and the ON-OFF queueing model.
Figure 15: Average end-to-end delay versus the number of stations for RTS/CTS access mode and the ON-OFF queueing model.