

Refinements on IEEE 802.11 Distributed Coordination Function Modeling Approaches

Ilenia Tinnirello, Giuseppe Bianchi, and Yang Xiao, *Senior Member, IEEE*

Abstract—With the popularity of the IEEE 802.11 standards, many analytical saturation throughput studies for the distributed coordination function (DCF) have been reported. In this paper, we outline a number of issues and criticalities raised by previously proposed models. In particular, a careful look at backoff counter decrement rules allows us to conclude that, under saturation conditions, the slot immediately following a successful transmission can be accessed only by the station (STA) that has successfully transmitted in the previous channel access. Moreover, due to the specific acknowledgment (ACK) timeout setting adopted in the standard, the slot immediately following a collision cannot be accessed by any STA. Thus, the hypothesis of uncorrelation between consecutive channel slots and statistical homogeneity is not generally true. We propose a new backoff decrement model that retains the simplicity of traditional DCF models while being able to take into account such a correlation, and we compare the accuracy of our model with that of previously proposed approaches.

Index Terms—Backoff freezing, distributed coordination function (DCF), performance evaluation.

I. INTRODUCTION

OVER THE PAST few years, the networking world has witnessed an impressive deployment of products based on the IEEE 802.11 standard for wireless local area networks (WLANs). IEEE 802.11 represents the adaptation, to the wireless media, of the Ethernet medium access control (MAC) protocol. As such, it inherits many of the advantages of the Ethernet, particularly in terms of protocol simplicity and fully distributed operation. To cope with the wireless medium specificities, the IEEE 802.11 standard employs a carrier sense multiple access with collision avoidance (CSMA/CA) MAC protocol with binary exponential backoff called distributed coordination function (DCF). An optional centralized scheme called point coordination function has been specified in the IEEE 802.11 standard, although its success in terms of support in commercial products and adoption in deployed WLANs is marginal. The original 802.11 standard [1] has subsequently been amended with quality-of-service extensions, which have been integrated into the 2007 version of the standard.

Manuscript received May 24, 2008; revised November 21, 2008 and April 10, 2009. First published August 4, 2009; current version published March 19, 2010. The review of this paper was coordinated by Prof. C. Lin.

I. Tinnirello is with the University of Palermo, 90128 Palermo, Italy (e-mail: ilenia.tinnirello@tti.unipa.it).

G. Bianchi is with the School of Engineering, University of Roma Tor Vergata, 00133 Rome, Italy (e-mail: giuseppe.bianchi@uniroma2.it).

Y. Xiao is with the Department of Computer Science, The University of Alabama, Tuscaloosa, AL 35487-0290 USA (e-mail: yangxiao@ieee.org).

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

Digital Object Identifier 10.1109/TVT.2009.2029118

Starting in the mid 1990's, many papers have studied DCF performance via simulation and/or simplified analytical means [2]–[5]. Some of these literature studies [2], [4] derive the optimal setting of the backoff contention window to achieve a maximum throughput and describe an adaptive DCF operation devised to let the system operate close to the maximum throughput values. Reference [6] quantified how the maximum achievable throughput is bounded by the protocol overhead and tends to converge toward an upper bound, regardless of the available transmission rate at the physical layer.

A simple but accurate modeling approach for evaluating the saturation throughput performance of the IEEE 802.11 DCF was proposed in [7] and further detailed in [8]. Since then, a number of papers have enhanced this basic model by either adapting it to different backoff freezing assumptions [9], [10] or accounting for supplementary modeling details, such as finite retransmission attempts [11], k -ary exponential backoff and multiple traffic classes [12], error-prone channel conditions [13], channel capture [14], and hidden terminals [15]. In addition to the throughput analysis, a companion derivation of the average delay performance is described in [9] for the case of infinite retransmission attempts, whereas a derivation in the case of finite retry limits is given in [16].

The model initially proposed in [8], as well as many of its extensions, implicitly assumes that the backoff counter is deterministically decremented at the beginning of a slot time. As thoroughly discussed in Section III, this assumption does not conform to the original DCF specification [1]. To face this issue, the original model has been corrected in [9] by introducing the concept of backoff decrement probability. The correction has been addressed in many subsequent studies [17]–[20], although Foh and Tantra [10] and Chatzimisios *et al.* [21] proved that the new modeling approach has lower accuracy than the original approach.

The goal of this paper is threefold. First, we clarify the implications of the actual backoff rules on the channel access process by enlightening the presence of *anomalous* slots in which the channel access probability is much lower than the average probability. Second, we present an improved analytical framework, which thoroughly accounts for the backoff freezing details of the DCF specification while retaining the original simplicity of [8] and its basic assumption about the statistical independence of consecutive model slots. The approach is very different from the other solutions that appeared in the literature, e.g., [10], which account for anomalous slots by introducing temporal correlations among consecutive channel slots. Finally, we also justify why the backoff freezing model proposed in [9] does not provide accurate results and why the analysis

versus simulation comparison in [8] remains accurate, despite the imprecise modeling of the backoff counter decrement.

The rest of this paper is organized as follows: The DCF function is briefly reviewed in Section II. A detailed discussion on the presence of anomalous slots is carried out in Section III, and the modeling extensions needed to account for such an anomaly are presented in Section IV. Section V presents throughput and delay analysis. A comparison between different DCF modeling approaches is presented in Section VI. We finally conclude this paper in Section VII.

II. DISTRIBUTED COORDINATION FUNCTION

The IEEE 802.11 DCF follows the “listen-before-talk” baseline principles of the Carrier Sense Multiple Access mechanisms with Collision Avoidance (CSMA/CA). A station (STA) with a frame to transmit shall invoke the carrier-sense mechanism to determine whether the wireless medium is busy or idle. In the case of a busy medium, the STA shall defer transmission until the medium is idle without interruption for a period of time equal to a distributed interframe space (DIFS). After this period, the STA shall generate a random backoff period for an additional deferral time before transmitting. The backoff period is slotted for efficiency reasons and is expressed in terms of an integer number of elementary backoff slots. Such a number, which is called the backoff counter, is decremented as long as the medium is sensed idle, “frozen” when a transmission is detected on the channel, and reactivated when the medium is sensed idle again for more than a DIFS. The STA transmits when the backoff time reaches zero.

At each transmission, the backoff time is uniformly chosen in the range $[0, CW]$, where CW is the current backoff window size. At the very first transmission attempt, CW is equal to the initial backoff window size CW_{\min} (e.g., 31 in 802.11b). After each unsuccessful transmission, $CW := 2(CW + 1) - 1$, until a maximum backoff window size value CW_{\max} is reached. Once it reaches CW_{\max} , CW shall remain at the value CW_{\max} until it is reset. CW shall be reset to CW_{\min} after every successful attempt to transmit, or the retransmission counter reaches a predefined retry limit, which is referred to as R hereinafter. (Two retry limits, i.e., a short and a long one, are defined in the standard and used for frames whose size is lower or greater than a given threshold, i.e., the $RTSTreshold$.) When the retry limit is reached, the frame is dropped.

Since the CSMA/CA does not rely on the capability of STAs to detect a collision by hearing their own transmission, a positive acknowledgment (ACK) is transmitted by the destination STA to signal the successful packet reception. The ACK is immediately transmitted at the end of the packet after a period of time called short interframe space (SIFS). As the SIFS is shorter than the DIFS, no other STA is able to detect the channel to be idle for a DIFS until the end of the ACK. If the transmitting STA does not receive the ACK within a specified ACK timeout or it detects the transmission of a different frame on the channel, it reschedules the frame transmission according to the previous backoff rules.

The previously described two-way handshaking technique for the packet transmission is called the basic access mecha-

nism. The DCF defines an optional four-way handshaking technique to be used for a packet transmission known as the request-to-send/clear-to-send (RTS/CTS) mechanism. In the RTS/CTS mechanism, before transmitting a data frame, a short RTS frame is transmitted. If the RTS frame succeeds, the receiver STA responds with a short CTS frame. Then, a data frame and an ACK frame will follow. All four frames (RTS, CTS, data, and ACK) are separated by a SIFS time. Furthermore, an STA shall not transmit within an extended interframe space (EIFS) after it determines that the medium is idle, following the reception of an error frame.

Let δ , SIFS, DIFS, EIFS, ACK_{Timeout} , CTS_{Timeout} , and T_{Ack} denote an empty slot time, a SIFS time, a DIFS time, an EIFS time, the ACK timeout, the CTS timeout, and an ACK transmission time, respectively. We have the following relations [1]: $DIFS = SIFS + 2 \times \delta$, $EIFS = SIFS + T_{\text{Ack}} + DIFS$, and¹ $ACK_{\text{Timeout}} = CTS_{\text{Timeout}} = SIFS + T_{\text{Ack}} + \delta$.

III. BACKOFF DECREMENT AND ANOMALOUS SLOTS

Different interpretations of how the backoff counter is frozen have lead in the past to different CSMA/CA models. The analytical model proposed in [8] implicitly relies on the assumption that the backoff counter is decremented at the beginning of a slot time. This implies that, at each slot time, the backoff counter is decremented, regardless of the fact that the slot is empty or contains a frame transmission (or collision). However, a closer look at the original 802.11 specification² reveals that a more conforming modeling assumption is to decrement the backoff counter at the end of a slot time. In fact, the IEEE 802.11 standard specifies in Section 9.2.5.2 that, if the medium is determined to be busy at any time during a backoff slot, then the backoff procedure is suspended.

To clarify the consequences of such a rule, assume that an STA has a backoff counter equal to b at the beginning of a slot time. If the current slot time is idle, at the end of the slot time, the backoff counter is duly decremented, and the station starts the next slot time with backoff value $b - 1$. Conversely, if the current slot time is sensed busy (because another STA starts transmitting in the considered slot), the STA freezes the backoff counter to the value b and starts the slot immediately following the busy slot with the same backoff value b . In other words, the backoff counter is decremented only during idle slots. This interpretation was recognized and first adopted in [9].

The end-of-slot backoff decrement rule, conforming to the standard, has some implications on the DCF modeling framework that are not immediately evident. Fig. 1 shows what happens when two STAs access the channel with different backoff values. In the example, at slot time t , STAs A and B, which are assumed to be synchronized by previous monitoring

¹We remark that the value of ACK_{Timeout} , although apparently not specified in the core part of the standard, is indeed specified in Appendix C (formal description of MAC operations; see details in the Trsp timer setting) as $ACK_{\text{Timeout}} = CTS_{\text{Timeout}} = aSifs + Duration(ACK) + PLCLHeader + PLCPpreamble + aSlotTime$.

²We remark that the newer 802.11 specification [22], while accounting for quality-of-service extensions in the frame of the Enhanced Distributed Coordination Access, also changes the way the backoff counter is therein decremented (see, e.g., [23] for a discussion of these new decrement rules).

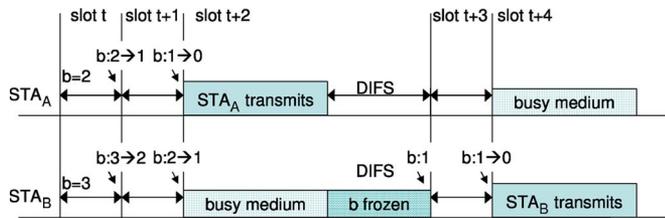


Fig. 1. First slot anomaly. The slot immediately following a transmission can be accessed only by the transmitting STA.

of the channel activity, start with a backoff counter equal to 2 and 3, respectively. Hence, we might expect them to transmit in consecutive slots, i.e., slots $t + 2$ and $t + 3$. As expected, STA A decrements the backoff counter to 0 at the end of slot $t + 1$ and thus transmits a frame in slot $t + 2$. It can then schedule the transmission for the next frame, starting from a DIFS time after the end of its transmission. An immediate transmission, in the slot referred in the figure as $t + 3$, will occur only if the STA extracts 0 as a new backoff counter. Since the transmission in slot $t + 2$ was successful, the new backoff counter will uniformly be drawn in the range $[0, CW_{\min}]$, and thus, a transmission by STA A in slot $t + 3$ will occur with known probability $1/(CW_{\min} + 1)$.

Let us now focus on STA B. If slot $t + 2$ were idle, it would have transmitted in slot $t + 3$. However, since, in slot $t + 2$, STA B sees a transmission on the channel, it freezes the backoff counter. Thus, it starts slot $t + 3$ with a backoff counter value that is still equal to 1, and (assuming slot $t + 3$ to be empty) it ultimately transmits only in slot $t + 4$. We conclude that *a slot immediately following a successful transmission cannot be used for transmissions by any other STA, except the transmitting STA*. Hence, under ideal channel conditions, such a transmission is granted to be successful as no collision may occur.

Consider now the case of collision. We assume that all STAs that listen to a collision detect a PHY-RXEND.indication, which returns an error and thus resumes the backoff process after an EIFS time from the previous channel activity.³

In this assumption, neither the transmitting STAs nor other STAs can use the first slot after an EIFS time from the previous transmission. In fact, by considering a basic rate of 1 Mb/s, the EIFS duration is equal to $ACK_{\text{Timeout}} + DIFS - \delta$. Thus, after an EIFS from a previous collision, an STA involved in the collision and waiting for the ACK timeout has to wait for a further backoff slot before extracting a new backoff value. Moreover, STAs sensing the collision resume the backoff counter to the frozen value, which is different from zero, because they were not transmitting. Hence, they cannot use the first slot after an EIFS time either. We can conclude that *the extra slot after*

³This assumption verifies whether a station listening to a collision succeeds in synchronizing, in the presence of two or more colliding preambles, and is used by the most common simulation platform (such as NS-2 or OPNET). An alternative assumption is to use a DIFS (e.g., as done in [8]), hence implying that no PHY-RXEND.indication primitive is generated by the PHY at the end of the collision, and the channel is simply interpreted as “generically” busy. We stress that the consistent usage of an EIFS (or DIFS) is a convenient, although coarse, assumption. In practice, these two interframe spaces will be case-by-case selected based on the outcome of the PHY operation in the presence of each specific channel-collision event.

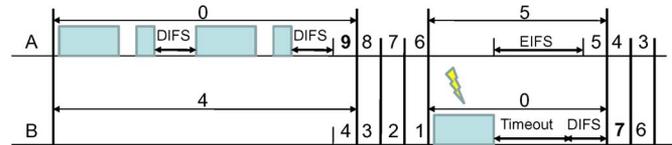


Fig. 2. Slot definition.

the end of an EIFS will not be used by any STA (either those involved in a collision or other STAs monitoring the channel).

Note that the preceding two conclusions are correct when there is no first-time-entered STA, particularly in the saturation assumption on which our paper is based. Furthermore, the case in which a first-time-entered STA accidentally transmits in the first slot is very rare. We can conclude that the slots right after a channel busy period are *anomalous* slots since the probability of transmitting and colliding in these slots is different from that in other slots.

IV. ANALYTICAL MODEL

In this section, we prove that the Markov chain presented in [8] can easily be adapted for modeling the backoff freezing and the anomalous behaviors of the channel slots following a busy period.

A. Model Time Scale

The key idea of the proposed approach is to define the discrete time scale used in the analytical model in a slightly different manner with respect to the definition given in [8]. In terms of modeling, we consider that the system time evolution is triggered whenever a nontransmitting STA decrements its backoff counter. This implies that a model slot time corresponds to either an idle backoff slot, a time interval including one or more consecutive transmissions and a final extra backoff slot, or a time interval including a collision, followed by an EIFS interval plus an extra backoff slot. The rationale that drives this definition is the modeling of backoff transitions only. Moreover, it allows maintaining a synchronized backoff countdown among all contending STAs.

These considerations are evident by analyzing Fig. 2, in which the model slots are indicated by bold lines. Consider STA A after a generic successful transmission. Since it extracts a new backoff counter equal to 0, it holds the channel for a subsequent transmission. After the second transmission, it initially sets the backoff counter to 9. Such a value is then decremented at the end of each succeeding idle slot. When the backoff counter reaches 5, it is frozen as long as STA A reveals the medium as busy. Since the frame transmitted by STA B has been corrupted and such an error has been detected by STA A, STA A waits for an EIFS time before resuming the backoff counter to the frozen value 5. Consider now STA B. During the subsequent transmissions performed by STA A, it holds the backoff counter to the frozen value 4. Such a value is resumed after a DIFS from the last transmission and decremented to 0 in the succeeding idle slots. Since STA-B transmission results in a collision, STA B waits for the ACK timeout plus a DIFS time before extracting a new backoff counter equal to 7. In the figure, the new backoff extractions are indicated by bold numbers.

Note that such extractions are included in a new model slot in the case of corrupted transmission (STA-B case) and in the previous model slot in the case of successful transmission (STA-A case).

B. Bidimensional Markov Chain

The contention process among all the STAs can easily be modeled by considering as state parameters the backoff counter, which represents the random delay introduced before accessing the channel, and the backoff stage, which represents the number of retransmissions suffered by the head-of-line (HOL) frame. Given a tagged STA, let $b(t)$ be the random process representing the value of the backoff counter at model slot time t , and let $s(t)$ be the random process representing the backoff stage j , where $j = 0, \dots, R$.

According to the new definition of the discrete time scale composed of uneven sized slots, even considering the actual backoff decrement rule, we can follow up as in [8] and assume that the backoff counter is decremented at the end of each slot time. However, the new time scale and decrement rules slightly affect the resulting bidimensional Markov chain proposed in [8]. In particular, we have to differently model stage 0, by accounting for a different number of backoff counter states and for the event that triggers the backoff stage reset (successful transmission or packet drop due to retransmission limit expiration). The rationale of these model details is discussed here.

Consider the backoff process of a target STA involved in a channel access. After a successful transmission, since the model slot is ended whenever the target STA extracts a backoff counter different from 0, the new counter value can assume only CW_{\min} different states. In fact, in the case the extracted value is 0, with probability $1/(CW_{\min} + 1)$, we do not enter into the next model slot, but the STA transmits a subsequent frame in the same slot. Since the STA enters into the next model slot only if it initially extracts a backoff counter in the range $[1, CW_{\min}]$ and since the model slot includes a last idle backoff slot, at the beginning of the next model slot, the new backoff counter will be found in the range $[0, CW_{\min} - 1]$. After a collision, when the target STA enters the backoff stage j , since the ACK_{Timeout} expiration includes a last backoff slot, thus resulting in being synchronized with the model slot evolution, the new backoff extraction is uniformly performed in the range $[0, CW_j]$. Note that stage j entered after a collision can also correspond to stage 0 whenever the maximum retransmission limit is reached. Thus, for $j > 0$, all the new extractions are performed in the range $[0, CW_j]$, whereas for $j = 0$, the new extractions can be drawn in the range $[0, CW_{\min} - 1]$ or $[0, CW_{\min}]$ according to the event that has triggered the stage-0 transition (successful transmission or packet drop, respectively).

For convenience of notation, let W_{\min} , W_{\max} , and W_j denote $CW_{\min} + 1$, $CW_{\max} + 1$, and $CW_j + 1$ in the j th retry/retransmission for the rest of this paper. We have

$$W_j = \begin{cases} 2^j W = 2^j W_{\min}, & j \in [0, m-1], R > m \\ 2^m W = 2^m W_{\min}, & j \in [m, R], R > m \\ 2^j W = 2^j W_{\min}, & j \in [0, R], R \leq m \end{cases} \quad (1)$$

where $m = \log_2 W_{\max}/W_{\min}$, and $W_{\min} = W$.

Let p denote the probability that a transmitted frame collides. The bidimensional random process $\{s(t), b(t)\}$ shown in Fig. 3 is the resulting discrete-time Markov chain under the assumptions that the probability p is independent of the backoff procedure [8] and that the backoff counter is decremented at the end of the backoff slot. Therefore, the state of each STA is described by $\{j, k\}$, where j stands for the backoff stage, and k stands for the backoff delay and takes values in $[0, W_j - 1]$.

Let $b_{j,k} = \lim_{t \rightarrow \infty} \Pr\{s(t) = j, b(t) = k\}$ be the stationary distribution of the Markov chain, and let τ be the probability that an STA transmits during a generic slot time, i.e., $\tau = \sum_{i=0}^R b_{i,0}$. For the sake of presentation, each state referring to backoff stage 0 is split into two states 0^+ and 0^- (which are plotted in two different columns), which correspond to the different events originating from the stage-0 transition. In fact, as we said, according to the event that triggers the stage-0 transition, the backoff counter belongs to different ranges. Specifically, the backoff counter belongs to the range $[0, W - 2]$ after successful transmission (which is indicated as a transition to stage 0^+) and to the range $[0, W - 1]$ after a packet drop (which is indicated as a transition to stage 0^-). The nonnull one-step transition probabilities result in⁴

$$\begin{cases} P\{i, k|i, k+1\} = 1, & k \in [0, W_i - 2], \\ & i \in [1, R] \\ P\{0^-, k|0^-, k+1\} = 1, & k \in [0, W_0 - 2] \\ P\{0^+, k|0^+, k+1\} = 1, & k \in [0, W_0 - 3] \\ P\{0^-, k|R, 0\} = p/W_0, & k \in [0, W_0 - 1] \\ P\{0^+, k|i, 0\} = (1-p)/(W_0 - 1), & k \in [0, W_0 - 2] \\ P\{i, k|i-1, 0\} = p/W_i, & k \in [0, W_i - 1] \\ & i \in [1, R]. \end{cases} \quad \forall i \quad (2)$$

Owing to the chain regularities, there is a simple relation between all the states belonging to the same row (i.e., corresponding to the same stage j)

$$\begin{cases} b_{j,k} = \frac{W_j - k}{W_j} b_{j,0}, & k \in [1, W_j - 1], \quad j > 0 \\ b_{0^-, k} = \frac{W - k}{W} b_{0^-, 0}, & k \in [1, W - 1] \\ b_{0^+, k} = \frac{W - 1 - k}{W - 1} b_{0^+, 0}, & k \in [1, W - 2] \end{cases} \quad (3)$$

where we can consider $W_0^+ = W - 1$ to include the first idle backoff slot after a successful transmission in the model slot. Given $b_{0,0} = b_{0^+,0} + b_{0^-,0}$, there is also a simple relation between the states of the first column, i.e.,

$$\begin{cases} b_{j,0} = p^j b_{0,0}, & j \in [1, R] \\ b_{0^-, 0} = p^{R+1} b_{0,0} \\ b_{0^+, 0} = (1 - p^{R+1}) b_{0,0}. \end{cases} \quad (4)$$

Since all the states are expressed as a function of the probability $b_{0,0}$, by imposing the normalization condition, we can solve the

⁴We adopt the short notation

$$P\{i_1, k_1|i_0, k_0\} = P\{s(t+1) = i_1, b(t+1) = k_1 | s(t) = i_0, b(t) = k_0\}.$$

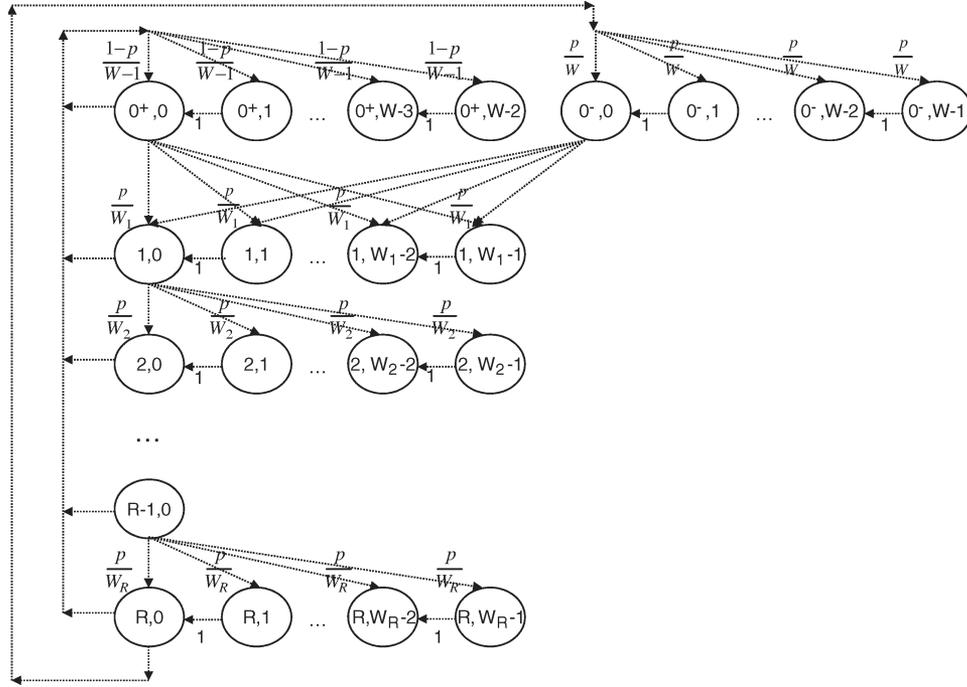


Fig. 3. State transmission diagram.

Markov chain

$$\begin{aligned}
 1 &= \sum_{j=1}^R \sum_{k=0}^{W_j-1} b_{j,k} + \sum_{k=0}^{W-1} b_{0^-,k} + \sum_{k=0}^{W-2} b_{0^+,k} \\
 &= \frac{b_{0^-,0}}{2} \left[\sum_{j=0}^R p^j (2^j W + 1) - (1 - p^{R+1}) \right]. \quad (5)
 \end{aligned}$$

We can now express the probability τ that an STA transmits in a randomly chosen slot. As any transmission occurs when the backoff counter is equal to zero, regardless of the backoff stage, it is

$$\begin{aligned}
 \tau &= \sum_{j=1}^R b_{j,0} + b_{0^-,0} + b_{0^+,0} = \frac{1 - p^{R+1}}{1 - p} b_{0^-,0} \\
 &= \frac{1}{1 + \frac{1-p}{2(1-p^{R+1})} \left[\sum_{j=0}^R p^j \cdot (2^j W - 1) - (1 - p^{R+1}) \right]}. \quad (6)
 \end{aligned}$$

Given the transmission probability τ , we can express the conditional collision probability p as the probability that a tagged STA sees a transmission originated by at least one of the contending STAs, i.e.,

$$p = 1 - (1 - \tau)^{n-1}. \quad (7)$$

Then, we can numerically solve τ and p based on (6) and (7).

C. Alternative Derivation

In this section, we provide an alternative derivation of (6) and (7), based on conditional probability arguments [24] rather

than bidimensional Markov chains. Given the discrete time scale introduced in the previous sections, let us denote by TX the event that an STA is transmitting a frame into a time slot and denote with $(s = i)$ the event that the STA is found in backoff stage $i \in (0, \dots, R)$. We are ultimately interested in the derivation of the transmission event probability $P(TX)$, which corresponds to the τ parameter defined in the previous section. Due to Bayes' theorem

$$P(TX) \frac{P(s = i|TX)}{P(TX|s = i)} = P(s = i), \quad i \in (0, \dots, R). \quad (8)$$

Since this equality holds for each $i \in (0, \dots, R)$, it also holds for the summation

$$P(TX) \sum_{i=0}^R \frac{P(s = i|TX)}{P(TX|s = i)} = \sum_{i=0}^R P(s = i) = 1. \quad (9)$$

We can thus express τ as

$$\tau = P(TX) = \frac{1}{\sum_{i=0}^R \frac{P(s=i|TX)}{P(TX|s=i)}}. \quad (10)$$

The conditional probability $P(s = i|TX)$ represents the probability that a transmitting STA is found in stage i . This probability is the steady-state distribution of a discrete-time Markov chain $s(k)$ describing the evolution of the backoff stage during the STA's transmission instants k and whose nonnull one-step transition probabilities are

$$\begin{cases} P(s(k+1) = i|s(k) = i-1) = p, & i = 1, \dots, R \\ P(s(k+1) = 0|s(k) = i) = 1-p, & i = 0, \dots, R-1 \\ P(s(k+1) = 0|s(k) = R) = 1, & i = R \end{cases} \quad (11)$$

where p is the previously defined conditional collision probability. It readily follows that $P(s = i|TX)$ is a (truncated, in the case of finite value R) geometric distribution, i.e.,

$$P(s = i|TX) = \frac{(1-p)p^i}{1-p^{R+1}}, \quad i \in (0, \dots, R). \quad (12)$$

Let us now focus on the conditional transmission probability $P(TX|s = i)$, i.e., the probability that an STA transmits while being in backoff stage i . We can envision the transmission process as the recurrence of consecutive transmission cycles, which are composed of transmission events separated by backoff times. Since we are conditioning on a given backoff stage i , we have to consider a subset of cycles corresponding to the backoff times and transmissions that originated while in stage i . Assuming independence among transmission cycles, from the renewal theory,⁵ we may conclude that the probability $P(TX|s = i)$ can be computed by dividing the average number of slots spent for transmissions in a transmission cycle (which is exactly one slot owing to the time scale adopted), with the average number of slots spent by the STA during the whole cycle. Since, according to our model slot definition, the number of slots spent for the backoff expiration is equal to the backoff extraction value, in formulas, it follows that

$$P(TX|s = i) = \frac{1}{1 + E[b_i]}, \quad i \in (0, \dots, R) \quad (13)$$

where $E[b_i]$ is the average value of the backoff counter in stage i . Because of the uniform backoff counter distribution, for $i > 0$, $E[b_i]$ is immediately given by $(W_i - 1)/2 = (2^i W - 1)/2$. For $i = 0$, we have to consider that the range of new backoff counter values is different, according to the event that triggered the stage-0 transition (frame delivery or frame drop). By accounting for the further conditioning events, we have

$$\begin{aligned} P(TX|s = 0) &= P(TX|s = 0 \wedge \text{delivery})P(\text{delivery}) \\ &\quad + P(TX|s = 0 \wedge \text{drop})P(\text{drop}) \\ &= \frac{1}{1 + (W - 2)/2}(1 - p^{R+1}) \\ &\quad + \frac{1}{1 + (W - 1)/2}p^{R+1} \\ &= \frac{1}{1 + (W - 1)/2 - (1 - p^{R+1})/2} \\ &= \frac{1}{1 + E[b_0]}. \end{aligned} \quad (14)$$

Substituting (12)–(14) into (10)

$$\begin{aligned} \tau &= \frac{1}{\sum_{i=0}^R \frac{1-p}{1-p^{R+1}} p^i (1 + E[b_i])} \\ &= \frac{1}{1 + \frac{1-p}{1-p^{R+1}} \sum_{i=0}^R p^i (2^i W - 1)/2 - \frac{1-p}{2}} \end{aligned} \quad (15)$$

which is equal to (6).

⁵Specifically, this computation can be interpreted as an application of the Long-Run Renewal rate theorem (see, e.g., W. Feller, *An introduction to probability theory and its Applications*, Vol. II, Wiley, ch. XI, pp. 368–380).

V. THROUGHPUT AND DELAY

In this section, we derive throughput and delay performance metrics, with the consideration of the first slot anomaly in Sections V-A and B, respectively.

A. Throughput

Let P_b denote the probability that the channel is busy and P_s denote the probability that a successful transmission occurs in a slot time. As in [8], it readily follows that

$$P_b = 1 - (1 - \tau)^n \quad (16)$$

$$P_s = n\tau(1 - \tau)^{n-1}. \quad (17)$$

We now recall that, according to the model slot definition provided and discussed in Section IV-A, a successful transmission may comprise multiple consecutive frames. This implies that, to compute the throughput S , we need to find appropriate values for the following: 1) the average payload size $\overline{E[P]}$ accounting for multiple frames transmitted into the same slot; 2) the average time $\overline{T_s}$ that the channel is sensed busy because of a successful transmission; and 3) the average time $\overline{T_c}$ that the channel experiences a collision. With these parameters, with δ being the duration of an empty slot time, we can express the throughput as

$$S = \frac{P_s \overline{E(P)}}{(1 - P_b)\delta + P_s \overline{T_s} + [P_b - P_s] \overline{T_c}} \quad (18)$$

which can be simplified into

$$S = \frac{P_s \overline{E(P)}}{E[slot]} = \frac{P_s \overline{E(P)}}{E[TX]}$$

where $E[slot]$ is the average model slot duration, and $E[TX]$ is the average time elapsed between two consecutive transmissions that originated in normal slots.

We now provide expressions for $\overline{E[P]}$, $\overline{T_s}$, and $\overline{T_c}$. Let T_{MPDU} , T_{ACK} , T_{RTS} , T_{CTS} , SIFS, and DIFS denote the time to transmit the MPDU (including MAC header, PHY header, and/or tail), the time to transmit an ACK, the time to transmit an RTS frame, the time to transmit a CTS frame, the SIFS time, and the DIFS time, respectively. The first slot anomaly effect described can be accounted by including an extra slot time at the end of a transmission period and noting that, for a successful STA, a subsequent collision-free transmission of a frame occurs with probability $1/W_{\text{min}}$ (i.e., the probability that the successfully transmitting STA extracts a backoff counter equal to 0; see Section III). We have

$$\overline{E[P]} = P + \sum_{k=1}^{\infty} (1/W_{\text{min}})^k P = P \frac{W_{\text{min}}}{W_{\text{min}} - 1} \quad (19)$$

$$\overline{T_s} = T_s + \sum_{k=1}^{\infty} (1/W_{\text{min}})^k T_s + \delta = T_s \frac{W_{\text{min}}}{W_{\text{min}} - 1} + \delta \quad (20)$$

where T_s is the successful transmission time of a single-frame transmission

$$T_s = T_{\text{MPDU}} + \text{SIFS} + T_{\text{ACK}} + \text{DIFS} \quad (21)$$

for the basic access mechanism, and

$$T_s = T_{RTS} + SIFS + T_{CTS} + SIFS + T_{MPDU} + SIFS + T_{ACK} + DIFS \quad (22)$$

for the RTS/CTS case.

Finally, since, according to the standard specification, a contending STA will wait for an ACK_{Timeout} greater than an EIFS, before reattempting to transmit

$$\bar{T}_c = T_c + \delta \quad (23)$$

where, for the basic model, we have

$$T_c = T_{MPDU} + SIFS + T_{ACK} + DIFS \quad (24)$$

and for the RTS/CTS model, we have

$$T_c = T_{RTS} + SIFS + T_{ACK} + DIFS. \quad (25)$$

B. Delay

Under saturation conditions, the total delay experienced by a frame is not meaningful. The time elapsing from the instant of time that the frame is inserted in the transmission buffer to the time in which it is successfully transmitted depends on how long the system has remained under saturation conditions and, consequently, how congested the transmission buffer has become. (This, in turn, depends on how greater the offered load is with respect to the saturation throughput bound.) Nevertheless, it is instructive to quantify the average access delay D , which is defined as the time elapsing between the instant of time that the frame is put into service (i.e., it becomes HOL) and the instant of time that the frame terminates a successful delivery. In the assumption of no retry limits, i.e., that all the HOL frames are ultimately delivered, this computation is straightforward. In fact, we may rely on the well-known Little's Result, which states that, for any queuing system, the average number of customers in the system is equal to the average experienced delay multiplied by the average customer departure rate. If no retry limit were considered, then

$$D = \frac{N}{S/E[P]}. \quad (26)$$

In fact, in the assumption that no frames are lost because of the retry limit (i.e., $R = \text{infinite}$), each of the N STAs is contending with an HOL frame. Moreover, $S/E[P]$ represents the throughput S measured in frames per seconds and thus represents the frame departure rate from the system. The delay computation is more elaborate when a frame is discarded after reaching a predetermined maximum number of retries R . In fact, in such a case, a correct delay computation should take into account only the frames successfully delivered at the destination while excluding the contribution of dropped frames because of the frame retry limit. (Indeed, the delay experienced by dropped frames would have no practical significance.) To determine the average delay in the finite retry case, we can still start from Little's Result, but we need to replace N in (26) with

the average number of HOL frames that will successfully be delivered. This value is lower than the number of competing STAs, as some of the competing frames will ultimately be dropped. Thus, (26) can be rewritten as follows:

$$D = \frac{N(1 - P\{LOSS\})}{S/E[P]} \quad (27)$$

where $P\{LOSS\}$ represents the probability that a randomly chosen HOL frame will ultimately be dropped. Let us now randomly pick an HOL frame among the N contending frames. Such an HOL frame can be found in any of the ($i = 0, \dots, R$) possible backoff stages. The probability that a random frame is found in backoff stage i can be computed as

$$P\{s = i\} = \sum_{k=0}^{W_i-1} b_{i,k} = \begin{cases} b_{i,0} \sum_{k=0}^{W_i-1} \frac{W_i-k}{W_i}, & i > 0 \\ p^{R+1} b_{0,0} \sum_{k=0}^{W-1} \frac{W-k}{W} + (1 - p^{R+1}) b_{0,0} \sum_{k=0}^{W-2} \frac{W-1-k}{W-1}, & i = 0. \end{cases} \quad (28)$$

By conditioning on the backoff stage i , $P\{LOSS\}$ can now be computed as

$$P\{LOSS\} = \sum_{i=0}^R P\{LOSS|s = i\} P\{s = i\}. \quad (29)$$

The average access delay expression is now found by substituting (29) into (27). In the derivation of (29), we have made use of the fact that the probability that a frame found in backoff stage i is ultimately dropped is given by the probability that it first reaches the backoff stage R (i.e., it collides for $R - i$ times), and then, it also collides during the last transmission attempt. Hence, $P\{LOSS|s = i\} = p^{R+1-i}$.

VI. COMPARISON WITH PREVIOUS MODELS AND ANALYSIS VALIDATION

In this section, we show that the approach [9] (and subsequent similar work) that models the backoff decrement through a specifically devised decrement probability leads to inaccurate results, because they do not verify the assumption of statistical independence of consecutive channel slots. The inaccuracy of model [9] was first pointed out in [21], where a comparison with simulation results and other models has been provided. However, Chatzimisios *et al.* [21] was focused on the delay analysis presented in [9], rather than on the model assumptions that yield such inaccuracy. Although some extensions, such as [10], to the model [9] explicitly deal with the correlation introduced among consecutive channel slots, we show that our approach is more effective in capturing the actual protocol behavior and providing simple results.

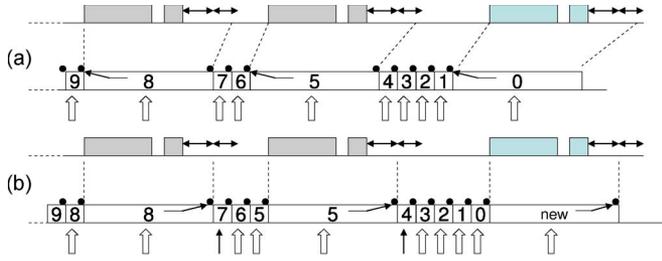


Fig. 4. Actual channel transmissions and different model descriptions. (a) Original model without backoff freezing. (b) Corrected backoff freezing model.

A. Modeling Backoff Freezing

In [9], the backoff freezing process is simply modeled by defining a backoff decrement probability that is different from 1. Specifically, since the backoff counter is decremented only during idle slots, the backoff decrement probability is set to the probability of finding an idle slot, i.e., no contending STA transmits in the current slot. Given the number n of contending STAs and the probability τ of accessing the channel, it results in $(1 - \tau)^{n-1} = 1 - p$, which is also the probability of experiencing success during a transmission. By setting the backoff decrement probability of the chain to $1 - p$, instead of 1, the average number of slots spent for the backoff expiration in stage i is equal to $1 + E[b_i]/(1 - p)$ since each backoff decrement, on average, lasts $1/(1 - p)$ model slots. Thus, the channel access probability yields [9]

$$\tau = \frac{1}{1 + \frac{1}{1-p^{R+1}} \sum_{i=0}^R p^i \frac{2^i W - 1}{2}} \quad (30)$$

which is lower than the channel access probability computed according to (6). However, the resulting model is biased by the fact that still considers all the slots as equivalent. Indeed, for the reasons discussed in Section III, the slots that immediately follow the channel activities (*anomalous* slots) are only rarely accessed (when the STA that has just transmitted gets a success and extracts a new backoff counter equal to 0). Thus, by summarizing the protocol behavior into a single average channel access probability, we overestimate the probability of accessing the *anomalous* slots and underestimate the probability of accessing all the other *normal* slots. This, in turn, corresponds to an underestimation of the collision probability since most parts of frame transmissions originated from the normal slots, in which the actual level of contention is higher than the average level.

To clarify the previous observations, Fig. 4 shows an example of channel activity that has been modeled according to the following: 1) the original model [8], without taking into account the backoff freezing, and 2) the backoff freezing model considered in [9]. The frame transmissions are represented by the long shaded boxes, which are followed by the ACK transmissions. In the example, there is no collision. Suppose to consider the backoff decrement process of a tagged STA that extracts a new backoff counter equal to 9. We observe that both the models include some approximations.

In the first case, the backoff counter state is equal to the value assumed at the beginning of the slot. Since the model

evolves at each backoff counter variation, the model time does not correspond to the actual time. In fact, after each busy slot, the backoff counters are decremented when resumed, i.e., after a DIFS or an EIFS from the previous transmission. This corresponds to considering the busy slot as one backoff slot shorter than the actual time. In our example, two busy slots occur during the backoff countdown. Thus, the tagged STA transmits after nine model slots, whose overall duration is underestimated of two backoff slots.

In the second case, the backoff counter state is equal to the value assumed at the end of the backoff slot. Since the model evolves for either a backoff counter variation or a backoff resumption, we see that the model time perfectly follows the actual time. By comparing with the previous model, we note that there is a further model slot for each channel occupancy. In our example, since two busy slots occur during the backoff countdown and since the backoff extraction is performed at the end of a busy slot, the tagged STA transmits after $9 + 3$ model slots, whose overall duration is evaluated exactly. Thus, we could conclude that such a model is more accurate than the previous model.

However, it is very easy to intuitively justify how this channel slotting leads to an underestimation of the actual collision probability experienced by the STAs. For a given observation window, let N be the number of slots modeled according to the *case-a* approach, and let A be the number of busy slots. It follows that the number of slots modeled according to the *case-b* approach is $N + A$. We can estimate the probability P_b of a busy slot occurrence as $\hat{P}_b^a = A/N$ or $\hat{P}_b^b = A/(N + A)$, according to the modeling approach indicated by the superscript. Given that n STAs are competing on the channel, we can assume that

$$\hat{P}_b = 1 - (1 - \hat{\tau})^n \simeq 1 - (1 - \hat{\tau})^{n-1} = \hat{p}$$

where \hat{p} represents an estimation of the collision probability encountered by the contending STAs. As a result, the collision probability estimated with the second model is smaller than that estimated with the first model. Note that this probability represents an average collision probability, but from our previous observations, we know that the collision probability experienced in the *anomalous* slots is equal to 0. In fact, the *anomalous* slots that follow the busy slot occurrence can be accessed by one STA only. (This fact is indicated in the figure by a thin arrow.) Given the probability p_N of colliding in the normal slots, the average collision probability results

$$\hat{p} \simeq \frac{N p_N}{N + A}$$

from which $p_N = A/N$.

In conclusion, the *case-a* approach underestimates the busy time of a backoff slot duration, but its collision probability evaluation corresponds to the value experienced in the normal slots. As most parts of the transmissions occur in the normal slots, it results in being more accurate than the *case-b* approach, which makes a rough collision-probability evaluation by still assuming that all slots have identical statistical properties.

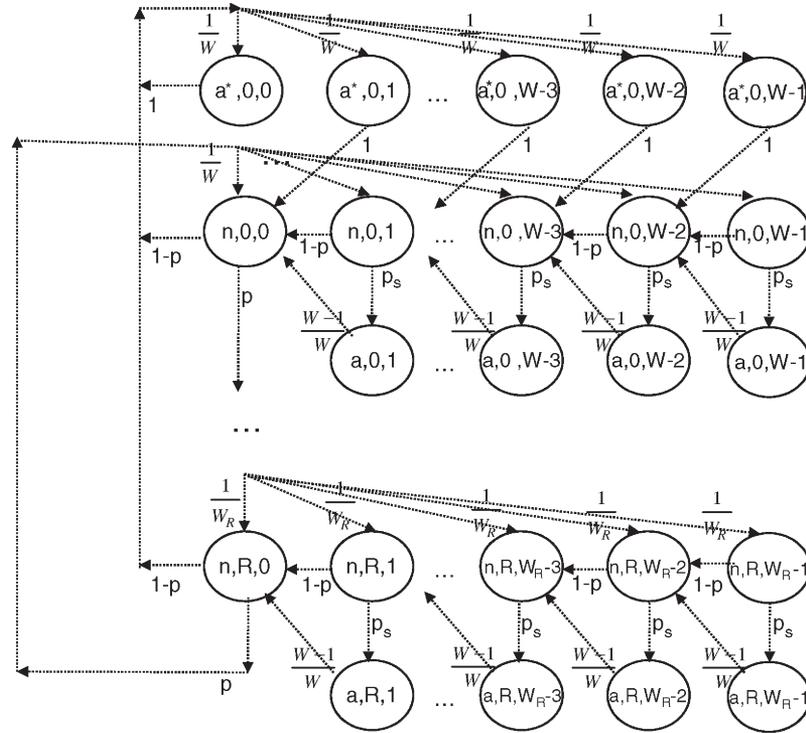


Fig. 5. Alternative tridimensional state transition diagram.

B. Modeling Backoff Freezing and Slot Differentiation

The existence of anomalous slots and nonuniform channel access probability has also been recognized in [10]. In this work, a new model is defined to correctly represent the actual backoff behavior and the correlation effects among consecutive channel slots. The behavior of a target STA is modeled in terms of a tridimensional Markov chain, in which the state is given by the status (idle or busy) of the previous model slot, the backoff stage, and the backoff counter of the target STA. Since the authors in [10] neglect ACK timeouts and EIFS times, they assume that anomalous slots follow each channel activity, regardless of the transmission outcome. The backoff decrement probability is then expressed as $1 - p_i$ or $1 - p_b$, according to the idle/busy status of the previous slot. This also corresponds to defining two different channel access probability τ_i and τ_b , which are the probability of transmitting in normal slots (i.e., slots following idle slots) or anomalous slots (i.e., slots following busy slots). Performance figures are finally evaluated by averaging the P_b and P_s probability according to the occurrence probability of normal slots (P_i) and anomalous slots ($1 - P_i$).

This approach is much more accurate than that proposed in [9]. However, it introduces other approximations. Specifically, the memory of the previous channel status is summarized into a bimodal variable (idle/busy), which does not differentiate the busy slots in which the target STA successfully transmits. After these slots, the backoff decrement probability from a generic state $(busy, 0, b)$ is equal to 1 (rather than $1 - p_b$). In fact, for $b > 0$, the slot following the successful transmission of the tagged STAs is always idle. (Neither the frozen counters nor the new backoff is equal to 0.) Conversely, the backoff

decrement probability is $1 - p_b$ whenever the transition to state $(busy, 0, b)$ is due to busy slots originated by other STAs.

By taking into account this correction and embedding the EIFS + δ time following a collision in a single transmission slot, it is possible to easily adapt the model proposed in [10] to our assumptions. Fig. 5 shows an example of a revised tridimensional model, in which we consider the channel memory in terms of anomalous and normal slots. In the model, we differentiate the anomalous slot states $(a^*, 0, b)$ following the transmission of the target STA from the anomalous slot states $(a, 0, b)$ following other transmissions. Let τ_n be the probability of transmitting in the normal slots. The model can be solved by considering the following: 1) The probability of observing a transmission in an anomalous slot is fixed to $1/W$, and 2) p_s is the probability of observing a successful transmission due to other STAs, i.e., $p_s = (n - 1)\tau_n(1 - \tau_n)^{n-2}$. It can be shown that, by solving this chain, τ_n is equal to our derivation given in (6). Clearly, solving this tridimensional chain as a function of two unknown parameters p and p_s is more complicated than solving our model. Moreover, our time scale maintains the statistical independence of consecutive slots, thus allowing easier derivation of per-slot average statistics.

C. Model Validation

To validate our new model, we used a C++ custom-made simulation tool focused on MAC-level performance analysis. We prefer this simulation tool to NS2 for many practical reasons since it allows avoiding interactions with the higher levels of the protocol stack and fully controlling and tracing the MAC protocol operations and performance figures. Moreover, the

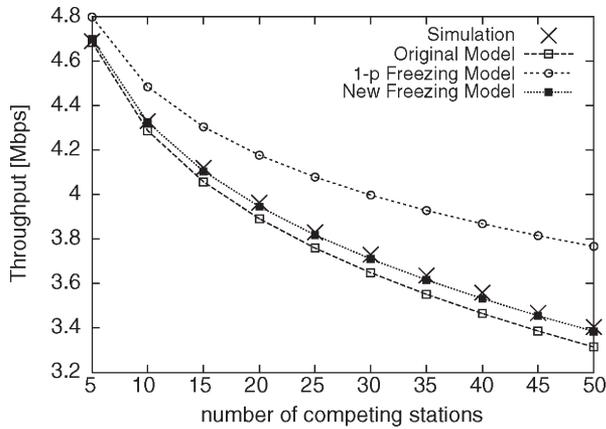


Fig. 6. Aggregated throughput versus the number of contending STAs: $CW_{\min} = 15$, $CW_{\max} = 1023$, and $R = 7$.

simulator has been obtained on the basis of the simulator used in [8] by simply extending the PHY layer with the 802.11a specifications. All the PHY and MAC parameters are compliant with the 802.11a standard definition, whereas the frame size has been set to 1500 B. Unless stated otherwise, the data rate and the basic rate are equal to 6 Mb/s.

Fig. 6 plots the aggregated throughput versus the number of contending STAs in the network. The points represent the simulation results, which are compared with three different analytical models, i.e., the original model proposed in [8] without the backoff freezing, the model that includes the $1 - p$ backoff freezing probability, and our new model. From the figure, we see that our results are very close to the simulation results. The simulation performance is slightly higher than the analytical performance. We suspect that this difference is due to some correlation effects among consecutive contentions, which are not taken into account in all the models.⁶ Indeed, the figure shows a scale of 1.6 Mb/s only. Thus, the mismatch between the simulation and analytical results is really negligible.

From the figure, we also note that the results of the model [8] are very close to our results. As the number of contending STAs increases, the accuracy of model [8] is slightly reduced. This phenomenon occurs, because model [8] neglects the anomalous slots, in which transmissions are always successful. The effects of these anomalous slots on the total throughput are more and more evident as the contention level increases. Finally, in Fig. 6, it is evident that the $1 - p$ backoff freezing approach does not provide good accuracy, particularly when the number of contending STAs is high.

To prove that the inaccuracy is due to the underestimation of the actual collision probability, Fig. 7 plots the average collision probability experienced in simulation and the average collision probability estimated by the three models. We observe that both our model and the original model provide very good accuracy, whereas the $1 - p$ freezing model is not able to correctly evaluate the collision probability. The errors can even be more dramatic if we set the retransmission limit to 0. In fact, in this

⁶This fact is confirmed in Fig. 8, in which the retry limit is equal to 0. As a result of a fixed contention window, which corresponds to a channel access probability independent of the backoff stage, the mismatch between analysis and simulations disappears.

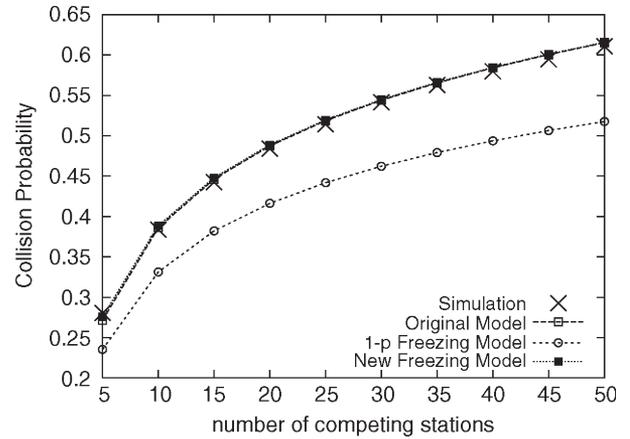


Fig. 7. Collision probability versus the number of contending STAs: $CW_{\min} = 15$, $CW_{\max} = 1023$, and $R = 7$.

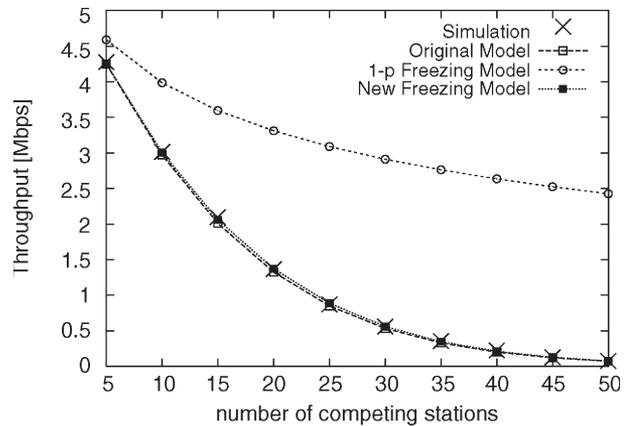


Fig. 8. Aggregated throughput versus the number of contending STAs: $CW_{\min} = 15$, $CW_{\max} = 1023$, and $R = 0$.

case, the contention window is fixed to the minimum value (which, for the 802.11a PHY, is equal to 16), and the collision probability is very severe.

Fig. 8 shows again the throughput evaluation obtained via simulation and via the three different models for the case $R = 0$. From the figure, we see that, because of the large portion of anomalous slots, the results provided by the $1 - p$ freezing model are totally wrong. For example, in the case $n = 50$, the throughput is almost zero via simulation or via the original model, but it results in being slightly less than 2.5 Mb/s via the $(1 - p)$ freezing model. The small differences between the original model [8] and our new model, which arise as the number of STA increases, are not evident in the figure because of the enlarged throughput scale (i.e., the range of $[0 : 5]$ Mb/s).

Note that the difference between the results using the original model [8] and our revised model also increases as the probability of accessing the anomalous slots increases, i.e., as the CW_{\min} value is set to smaller values. Although, in standard DCF, the probability of accessing anomalous slots is very low (1/32 after a successful 802.11b frame and 1/16 after a successful 802.11a/g frame), nowadays, it is very easy [25] to set small (nonstandard) CW_{\min} values. Table I compares the throughput results obtained at 6 Mb/s via simulation and via both the original model and our new model,

TABLE I
THROUGHPUT EVALUATIONS FOR SMALL CW_{min} VALUES FOR $n = 10$

CW_{min}	Original Model	New Model	Simulation
15	4.28 Mbps	4.32 Mbps	4.32 Mbps
7	3.94 Mbps	4.07 Mbps	4.05 Mbps
3	3.56 Mbps	3.84 Mbps	3.79 Mbps
1	3.11 Mbps	3.93 Mbps	3.83 Mbps

for $n = 10$ and different CW_{min} values. From the table, we see that the original model accuracy degrades only for $CW_{min} = 3$ or $CW_{min} = 1$.

D. Model Accuracy in Emerging PHY Layers

According to the modeling approach presented in [8] (and to all the refined models derived from it), the probability of having a successful transmission on the channel just depends on the contention window settings and the number of contending STAs. In fact, because of the slotted time scale considered in the model, the channel access performance is not affected by the actual frame duration, i.e., by the data transmission rates and frame lengths. However, the throughput evaluation also depends on the actual time that elapsed, on average, between two consecutive transmissions (i.e., $E[TX]$). This time is affected by the PHY layer parameters, such as backoff slot durations, interframe spaces, frame formats, and frame transmission times. Since the backoff countdown usually represents a small fraction of the whole $E[TX]$ interval, the throughput evaluation is not very sensitive to inaccurate predictions of these countdown intervals. Indeed, these considerations could not be valid in case of the emerging PHY layers, such as 802.11n, which significantly shorten the data and ACK transmission times. As the data-transmission rate increases, the frame durations approach the preamble durations, which are about two or three times the backoff slot duration. Under these conditions, the original DCF model [8] leads to potential inaccuracy problems, because it underestimates (of one backoff slot) the time that elapsed between two consecutive transmissions.

To evaluate the model accuracy when the frame duration is comparable to the backoff slot size, we refer to the current 802.11n specifications. In 802.11n, data rates are increased by exploiting multiple-input–multiple-output antenna systems and three further innovations on the 802.11a/g PHY layers: 1) channel bonding; 2) optional reduction of the orthogonal frequency-division multiplexing (OFDM) guard interval (GI) from 800 to 400 ns; and 3) increment of the OFDM subcarriers from 48 to 52 (in a 20-MHz channel). In the assumption of four different spatial streams and a 40-MHz bonded bandwidth, the new physical rates available with 802.11n are summarized in Table II. Obviously, for a lower number of spatial streams, the transmission rates are proportionally reduced. For example, for a single data stream, the maximum available rate is equal to 150 Mb/s. Note that, according to the GI setting, the symbol duration can be 4 or 3.6 μ s.

About the other PHY parameters, such as SlotTime duration and SIFSTime, 802.11n maintains the same settings defined in 802.11a/g. Conversely, the PPDU maximum length has significantly been increased to 65 535 octets for en-

TABLE II
AVAILABLE RATES IN THE NEW STANDARD IEEE 802.11n IN THE CASE OF FOUR SPATIAL STREAMS AND A 40-MHz CHANNEL

Modulation	Coding Rate	Carriers	Rate [Mbps]	
			GI=800ns	GI=400ns
BPSK	1/2	108	54	60
QPSK	1/2	108	108	120
QPSK	3/4	108	162	180
16-QAM	1/2	108	216	240
16-QAM	3/4	108	324	360
64-QAM	2/3	108	432	480
64-QAM	3/4	108	486	540
64-QAM	5/6	108	540	600

TABLE III
THROUGHPUT EVALUATIONS FOR HIGH TRANSMISSION RATES

4 streams - r = 600 Mbps			
n	Original Model	New Model	Simulation
5	57.96 Mbps	56.07 Mbps	56.13 Mbps
10	54.81 Mbps	53.01 Mbps	53.01 Mbps
20	50.79 Mbps	49.26 Mbps	49.19 Mbps
30	48.06 Mbps	46.70 Mbps	46.65 Mbps
1 stream - r = 150 Mbps			
n	Original Model	New Model	Simulation
5	46.27 Mbps	45.11 Mbps	45.15 Mbps
10	43.42 Mbps	42.41 Mbps	42.41 Mbps
20	40.06 Mbps	39.24 Mbps	39.20 Mbps
30	37.82 Mbps	37.13 Mbps	37.04 Mbps

abling frame aggregation. Although we do not explicitly consider the 802.11n capabilities of sending aggregated frames, in the rest of this section, we assume that ACKs are sent according to the block ACK format and transmitted at the data rate.

Table III compares the original DCF model [8] and our new model in terms of the throughput results gathered for different numbers of contending STAs at two different rates (i.e., 600 and 150 Mb/s). The packet payload is set to 1500 B. From the table, it is evident that the accuracy of model [8] depends on the employed data rate. For example, for $n = 5$, the difference between simulation and model predictions is about 1 Mb/s at 150 Mb/s (2.5%) and about 2 Mb/s at 600 Mb/s (3.5%). These errors, which could become appreciable for systems working with shorter preambles, are obviously mitigated in case of multiple frame aggregation.

Fig. 9 shows the 802.11n maximum PHY-SAP throughput performance provided by our new model and by the original DCF model, for a larger PSDU size (i.e., 8000 B). The curves have been obtained, for different antenna configurations (i.e., spatial streams), at the maximum possible transmission rate, which are 600, 300, and 150 Mb/s in the case of four, two, and one spatial stream, respectively. From the figure, we see that the results provided by the original model are quite close to the results provided by the new model.

We can also observe that the increment of the spatial streams does not correspond to a proportional increment of the throughput. In fact, the resource consumption due to the PHY header proportionally increases as the data rate increases. Moreover, using more spatial streams requires enlarging the PHY header to send more training sequence fields. The result is that the channel-utilization efficiency is always poor, despite the

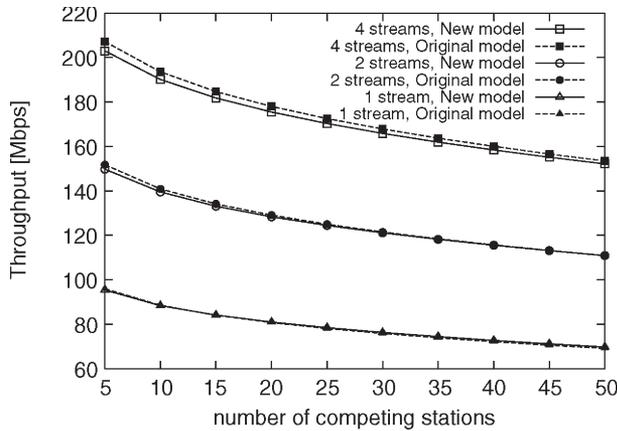


Fig. 9. Aggregated throughput versus the number of competing STAs: $P = 8000$ B.

payload size allowing aggregation of multiple packets (assuming a common size of 1500 B) and the MAC header overheads not being taken into account. For example, in the case of $n = 10$ and four spatial streams, the maximum throughput is less than 200 Mb/s, i.e., about 33% of the data rate. The situation is significantly worse in the absence of packet aggregation. In the same case where $n = 10$, a payload of 1500 B leads to an aggregated throughput of 53 Mb/s with four spatial streams (i.e., lower than 10% of the data rate) and to a throughput of 42 Mb/s with one spatial stream only. In such a case, the benefits of spatial multiplexing are negligible in comparison to the hardware complications required for enabling multiple stream transmissions. We can conclude that packet aggregation has to be considered mandatory for justifying the adoption of complex (and expensive) transceivers.

VII. CONCLUSION

The performance analysis of the IEEE 802.11 DCF has extensively been addressed in the last ten years, and several extensions have been proposed to the original Markov Chain approach proposed in [8]. However, despite the huge amount of literature work, some basic issues such as how to properly model the DCF freezing/resumption process of the backoff counter have remained open and/or improperly addressed.

In this paper, we have shown how the introduction of a backoff freezing state creates some strong correlation effects in the slots that immediately follow the channel transmissions. In fact, because of the protocol operations, these slots are available only for the STA that successfully transmitted in the previous channel access. By neglecting the anomalous behaviors of such slots, severe errors can be introduced in the DCF performance evaluation.

We have proposed a new model devised to properly capture the actual DCF protocol behavior without explicitly considering a backoff freezing state, thus retaining the original simplicity of model [8]. A comparison between previous approaches and our new model has extensively been carried out. The new proposed model has been tested against the more-challenging modeling conditions emerging in data-rate 802.11 scenarios, with specific reference to the 802.11n physical-layer specification.

REFERENCES

- [1] *Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification*, IEEE 802.11 WG, Aug. 1999.
- [2] G. Bianchi, L. Fratta, and M. Oliveri, "Performance evaluation and enhancement of the CSMA/CA MAC protocol for 802.11 wireless LANs," in *Proc. IEEE PIMRC*, 1996, pp. 392–396.
- [3] H. S. Chhaya and S. Gupta, "Performance modeling of asynchronous data transfer methods of IEEE 802.11 MAC protocol," *Wirel. Netw.*, vol. 3, no. 3, pp. 217–234, 1997.
- [4] F. Calf, M. Conti, and E. Gregori, "Dynamic tuning of the IEEE 802.11 protocol to achieve a theoretical throughput limit," *IEEE/ACM Trans. Netw.*, vol. 8, no. 6, pp. 785–799, Dec. 2000.
- [5] Y. C. Tay and K. C. Chua, "A capacity analysis for the IEEE 802.11 MAC protocol," *Wirel. Netw.*, vol. 7, no. 2, pp. 159–171, Mar./Apr. 2001.
- [6] Y. Xiao and J. Rosdahl, "Throughput and delay limits of IEEE 802.11," *IEEE Commun. Lett.*, vol. 6, no. 8, pp. 355–357, Aug. 2002.
- [7] G. Bianchi, "IEEE 802.11—Saturation throughput analysis," *IEEE Commun. Lett.*, vol. 2, no. 12, pp. 318–320, Dec. 1998.
- [8] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 3, pp. 535–547, Mar. 2000.
- [9] E. Ziouva and T. Antonakopoulos, "CSMA/CA performance under high traffic conditions: Throughput and delay analysis," *Comput. Commun.*, vol. 25, no. 3, pp. 313–321, Feb. 2002.
- [10] 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.
- [11] H. Wu, Y. Peng, K. long, S. Cheng, and J. Ma, "Performance of reliable transport protocol over IEEE 802.11 WLAN: Analysis and enhancement," in *Proc. IEEE INFOCOM*, 2002, vol. 2, pp. 599–607.
- [12] Y. Xiao, "A simple and effective priority scheme for IEEE 802.11," *IEEE Commun. Lett.*, vol. 7, no. 2, pp. 70–72, Feb. 2003.
- [13] V. M. Vishnevsky and A. I. Lyakhov, "802.11 LANs: Saturation throughput in the presence of noise," in *Proc. IFIP Netw.*, Pisa, Italy, 2002, pp. 1008–1019.
- [14] Z. Hadzi-Velkov and B. Spasenovski, "An analysis of CSMA/CA protocol with capture in wireless LANs," in *Proc. IEEE WCNC*, New Orleans, LA, Mar. 2003, vol. 2, pp. 16–20.
- [15] T.-C. Hou, L.-F. Tsao, and H.-C. Liu, "Throughput analysis of the IEEE 802.11e DCF scheme in multi-hop ad hoc networks," in *Proc. ICWN*, Jun. 2003, pp. 653–659.
- [16] P. Chatzimisios, A. C. Boucouvalas, and V. Vitsas, "IEEE 802.11 packet delay: A finite retry limit analysis," in *Proc. IEEE GLOBECOM*, San Francisco, CA, Dec. 2003, vol. 2, pp. 950–954.
- [17] M. Ergen and P. Varaiya, "Throughput analysis and admission control for IEEE 802.11a," *Mob. Netw. Appl.*, vol. 10, no. 5, pp. 705–716, Oct. 2005.
- [18] P. E. Engelstad and O. N. Osterbo, "Analysis of the total delay of IEEE 802.11e EDCA and 802.11 DCF," in *Proc. IEEE ICC*, Jun. 2006, vol. 2, pp. 552–559.
- [19] D. Y. Yang, T. J. Lee, K. Jang, J. B. Chang, and S. Choi, "Performance enhancement of multirate IEEE 802.11 WLANs with geographically scattered stations," *IEEE Trans. Mobile Comput.*, vol. 5, no. 7, pp. 906–919, Jul. 2006.
- [20] G. Sharma, A. Ganesh, and P. Key, "Performance analysis of contention based medium access control protocols," in *Proc. IEEE INFOCOM*, Apr. 2006, pp. 1–12.
- [21] P. Chatzimisios, A. C. Boucouvalas, and V. Vitsas, "IEEE 802.11 wireless LANs: Performance analysis and protocol refinement," *EURASIP J. Wireless Commun. Netw.*, vol. 2005, no. 1, pp. 67–78, Mar. 2005.
- [22] *Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification*, IEEE 802.11 WG, Jun. 2007.
- [23] G. Bianchi, L. Scalia, and I. Tinnirello, "Understanding 802.11e contention-based prioritization mechanisms and their coexistence with legacy 802.11 stations," *IEEE Netw.*, vol. 19, no. 4, pp. 28–34, Jul./Aug. 2005.
- [24] G. Bianchi and I. Tinnirello, "Remarks on IEEE 802.11 DCF performance analysis," *IEEE Commun. Lett.*, vol. 9, no. 8, pp. 765–767, Aug. 2005.
- [25] G. Bianchi, A. Di Stefano, C. Giaconia, L. Scalia, G. Terrazzino, and I. Tinnirello, "Experimental assessment of the backoff behavior of commercial IEEE 802.11b network cards," in *Proc. IEEE INFOCOM*, Anchorage, AK, 2007, pp. 1181–1189.
- [26] W. Feller, *An Introduction to Probability Theory and Its Applications*. New York: Wiley, 1971.



Ilenia Tinnirello received the Laurea degree in electronic engineering and the Ph.D. degree in communications in 2000 and 2004, respectively.

Since January 2005, she has been an Assistant Professor with the University of Palermo, Palermo, Italy. She was a Visiting Researcher with Seoul National University, Seoul, Korea, in 2004 and Nanyang Technological University, Singapore, in 2006. Her research interests include wireless networks, particularly multiple access algorithms with quality-of-service provisioning, cross-layer interactions

between access solutions and the physical layer, and mobility management and load balancing in wireless packet networks.



Giuseppe Bianchi received the Laurea degree in electrical engineering from Politecnico di Milano, Milan, Italy, in 1990 and the Specialist degree in information technology from the Center of Excellence for Research, Innovation, Education and Industrial Labs Partnership (CEFRIEL) Research Center of Milan in 1991.

Since January 2007, he has been a Full Professor of telecommunications with the School of Engineering, University of Roma Tor Vergata, Rome, Italy. Prior to this, he was a Research Consultant for

CEFRIEL during 1991–1993, an Assistant Professor with the Politecnico di Milano during 1994–1998, and an Associate Professor with the University of Palermo, Palermo, Italy during 1999–2003 and the University of Roma Tor Vergata during 2004–2006. His research interests include the design and performance evaluation of broadband networks, multiple access in wireless local area networks, and network security and privacy.



Yang Xiao (SM'04) received the B.S. and M.S. degrees from Jilin University, Changchun, China, and the M.S. and Ph.D. degrees in computer science and engineering from Wright State University, Dayton, OH.

He is currently with Department of Computer Science, The University of Alabama, Tuscaloosa. He currently serves as Editor-in-Chief for the *International Journal of Security and Networks*, the *International Journal of Sensor Networks*, and the *International Journal of Telemedicine and Applications*.

His research interests are security, telemedicine, robots, sensor networks, and wireless networks.

Dr. Xiao serves as an Associate Editor for several journals, e.g., the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY. He was a voting member of the IEEE 802.11 Working Group from 2001 to 2004.