

Performance Analysis of Priority Schemes for IEEE 802.11 and IEEE 802.11e Wireless LANs

Yang Xiao, *Senior Member, IEEE*

Abstract—The IEEE 802.11 distributed coordination function (DCF) enables fast installation with minimal management and maintenance costs, and is a very robust protocol for the best effort service in wireless medium. However, the current DCF is unsuitable for real-time applications. This paper studies backoff-based priority schemes for IEEE 802.11 and the emerging IEEE 802.11e standard by differentiating the minimum backoff window size, the backoff window-increasing factor, and the retransmission limit. An analytical model is proposed to derive saturation throughputs, saturation delays, and frame-dropping probabilities of different priority classes for all proposed priority schemes. Simulations are conducted to validate analytical results. The proposed priority schemes can be easily implemented, and the results from this paper are beneficial in designing good priority parameters.

Index Terms—Enhanced distributed coordination function (EDCF), IEEE 802.11e, priority, service differentiation.

I. INTRODUCTION

THE IEEE 802.11 medium access control (MAC) employs a mandatory contention-based channel access function called distributed coordination function (DCF) and an optional centrally controlled channel access function called point coordination function (PCF) [1]. The DCF adopts a carrier sense multiple access with collision avoidance with binary exponential backoff. The DCF is unsuitable for multimedia applications with quality of service (QoS) requirements. The popularity of the IEEE 802.11 market is largely due to the DCF, whereas the PCF is barely implemented in current products due to its complexity and inefficiency for normal data transmissions, even though it has a limited QoS support.

In the DCF, a station with a frame to transmit monitors the channel activities until an idle period equal to a distributed interframe space (DIFS) is detected. After sensing an idle DIFS, the station waits for a random backoff interval before transmitting. The backoff time counter is decremented in terms of slot time as long as the channel is sensed idle. The counter is stopped when a transmission is detected on the channel and reactivated when the channel is sensed idle again for more than a DIFS. The station transmits its frame when the backoff time reaches zero. At each transmission, the backoff time is uniformly chosen in the range $[0, CW - 1]$, where CW is the current backoff window size. At the very first transmission

attempt, CW equals the initial backoff window size CW_{\min} . After each unsuccessful transmission, CW is doubled 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 the retry limit L_{retry} . After the destination station successfully receives the frame, it transmits an acknowledgment frame (ACK) following a short interframe space (SIFS) time. If the transmitting station 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 above mechanism is called the basic access mechanism. To reduce the hidden station problem, an optional four-way data transmission mechanism called request-to-send (RTS)/clear-to-send (CTS) is also defined in the DCF.

There have been many performance studies for the original IEEE 802.11 DCF [2]–[5], [20]–[27]. Bianchi [2] proposed a simple and accurate analytical model to compute saturation throughput. Ziouva and Antonakopoulos (ZA) [3] improved Bianchi's model to derive a saturation delay. Wu *et al.* [4] improved Bianchi's model to consider retry limit. Xiao and Rosdahl [5] studied the maximum throughput and the throughput limit. However, these models [2]–[5] are not for priorities. In the DCF, a station might have to wait an arbitrarily long time to send a frame so that real-time applications such as voice and video may suffer [6]. One possible solution is to provide a good priority scheme for the DCF. Simple DCF priority schemes can be easily designed with minor changes in the DCF, and they are quite effective [7]. Several priority studies have been reported in the literature for the DCF. Deng and Chang [6] proposed a priority scheme by differentiating the backoff window: the higher priority class uses the window $[0, 2^{j+1} - 1]$ and the lower priority class uses the window $[2^{j+1}, 2^{j+2} - 1]$, where j is the backoff stage. Veres *et al.* [8] proposed priority schemes by differentiating the initial backoff window size and the maximum window size. Aad and Castelluccia [9] proposed a priority scheme by differentiating interframe spaces (IFS). Pallot and Miller [10] proposed an interesting prioritized backoff time distribution mechanism in which the backoff time is chosen in the current window range with different distributions for different priorities. All the priority schemes [2]–[5] were based on simulations. Recently, Xiao [7] proposed an analytical model to evaluate backoff-based priority schemes by differentiating the initial window size, the backoff window-increasing factor, and the maximum backoff stage.

Manuscript received November 9, 2002; revised June 6, 2003; accepted May 3, 2004. The editor coordinating the review of this paper and approving it for publication is K. K. Leung.

The author is with the Computer Science Department, The University of Memphis, Memphis, TN 38152 USA (e-mail: yangxiao@ieee.org).

Digital Object Identifier 10.1109/TWC.2005.850328

To support MAC-level QoS, the IEEE 802.11 standardization committee is currently working on IEEE 802.11e [11], supporting multimedia applications. The IEEE 802.11e MAC employs a channel access function called hybrid coordination function (HCF) that includes a contention-based enhanced DCF (EDCF) and a contention-free centrally controlled HCF part. The EDCF provides a priority scheme by differentiating the interframe space, the initial window size, and the maximum window size. Mangold *et al.* [12] introduced the EDCF and provided performance studies via simulations. Xiao [13] provided a good survey on the IEEE 802.11e standard. Xiao *et al.* [14] proposed two-level QoS protection and guarantee mechanisms via simulations for the EDCF.

We can classify priority schemes into three kinds: backoff-based priority schemes [6]–[8], IFS-based priority schemes [9], [10], and hybrid priority schemes [11]–[14]. Backoff-based priority schemes are listed as follows: 1) differentiating the initial window size [7], [8]; 2) differentiating the window-increasing factor [7]; 3) differentiating the maximum backoff stage [7]; 4) differentiating the maximum window size [8]; 5) differentiating the backoff time distribution [6], [10]; and 6) combining two or more of the above schemes [7], [8]. Note that scheme 4) can be achieved by schemes 1), 2), and 3). A hybrid scheme such as the EDCF adopts both a backoff-based scheme and an IFS-based scheme. In this paper, we study backoff-based priority schemes for IEEE 802.11 DCF and IEEE 802.11e EDCF by differentiating the minimum backoff window size, the backoff window-increasing factor, the retransmission limit, and any combination. Readers may refer to [18] and [19] for IFS-based priority models.

We briefly summarize the difference of this paper from [2], [3], [7]: 1) supporting priorities; 2) supporting a finite retransmission limit (retry limit); 3) differentiating among classes with different retry limits; 4) giving new performance metrics such as frame-dropping probability; 5) considering that the backoff counter is stopped when busy; 6) not having a nonbackoff stage; and 7) providing a new and accurate delay model.

The paper is organized as follows. Section II defines the priority schemes. We propose a general analytical model in Section III. Section IV customizes the general analytical model into different priority schemes. We conduct simulation validations in Section V. Numerical results are studied in Section VI. We conclude this paper in Section VII.

II. PRIORITY SCHEMES

In the original DCF, all stations/traffic have the same priority to access the channel in a distributed manner. In this paper, we assume that traffic is classified into N priority classes $i = 0, \dots, N - 1$, and we can define priority schemes as follows.

A. Backoff-Based Priority Schemes

In the IEEE 802.11 MAC protocol, the maximum backoff window size CW_{\max} is recommended with a default value, i.e., 1024 [1], [15], [16]. The initial backoff window size CW_{\min} is recommended with a default value, i.e., 32 timeslots for IEEE 802.11b [16] and 16 timeslots for IEEE 802.11a [15]. The default long retry limit is 4 and the default short retry limit is

7. This differentiation is just for frame size. In this paper, we differentiate the retry limit based on classes.

We modify 802.11 MAC to provide a combined backoff-based priority scheme by differentiating the following three metrics for the priority i class: the initial window size $W_{i,0}$ (same as $CW_{i,\min}$), the window-increasing factor σ_i , and the retry limit $L_{i,\text{retry}}$, where σ_i is the factor by which the current window size is increased when a transmitted frame collides, and is equal to 2 in the original IEEE 802.11 DCF [1]. In this paper, we let σ_i be a real number and $\sigma_i > 1$. Assuming that the priority i class has a higher priority than the priority j class, we have $W_{i,0} \leq W_{j,0}$, $1 < \sigma_i \leq \sigma_j$, and $L_{i,\text{retry}} \leq L_{j,\text{retry}}$. Furthermore, at least one of the above inequalities must be strict. A class with a smaller metric has a better chance of accessing the channel earlier.

Many special cases can be designed by differentiating any one or two metrics among the above three metrics. We will further discuss them in later sections.

For the priority i class, let $W_{i,j}$ denote CW in the j th retry/retransmission (or the j th backoff stage) and $CW_{i,\max}$ the maximum window size. The relationships among $W_{i,j}$, $CW_{i,\max}$, $CW_{i,\min}$, σ_i , and $L_{i,\text{retry}}$ are given as

$$W_{i,j} = \begin{cases} \sigma_i^j W_{i,0}, & \text{for } j = 0, 1, \dots, m_i - 1, \\ & \text{if } L_{i,\text{retry}} > m_i \\ \sigma_i^{m_i} W_{i,0} = CW_{i,\max}, & \text{for } j = m_i, \dots, L_{i,\text{retry}}, \\ & \text{if } L_{i,\text{retry}} > m_i \\ \sigma_i^j W_{i,0}, & \text{for } j = 0, 1, \dots, L_{i,\text{retry}}, \\ & \text{if } L_{i,\text{retry}} \leq m_i \end{cases} \quad (1)$$

where $m_i = \log_{\sigma_i}(CW_{i,\max}/CW_{i,\min})$.

B. IFS-Based Priority Schemes

For the priority i class, we can modify 802.11 MAC to provide a priority scheme by differentiating the interframe space IFS $[i]$, ($i = 0, \dots, N - 1$) instead of using DIFS for all. If one class has a smaller IFS, the class's traffic has a better chance to access the channel earlier.

C. Hybrid Scheme: EDCF

The EDCF adopts eight different priorities that are further mapped into four access categories (ACs). ACs are achieved by differentiating the arbitration interframe space (AIFS), the initial window size, and the maximum window size. For the AC i ($i = 0, 1, 2, 3$), the initial backoff window size is $CW_{\min}[i]$ ($= W_{i,0}$), the maximum backoff window size is $CW_{\max}[i]$, and the AIFS is AIFS $[i]$. For $0 \leq i < j \leq 3$, we have $CW_{\min}[i] \geq CW_{\min}[j]$, $CW_{\max}[i] \geq CW_{\max}[j]$, and AIFS $[i] \geq$ AIFS $[j]$, and at least one of the above inequalities must be strict. In other words, the EDCF adopts AIFS $[i]$ ($i = 0, 1, \dots, 3$), $CW_{\min}[i]$ ($i = 0, 1, \dots, 3$), and $CW_{\max}[i]$ ($i = 0, 1, \dots, 3$) instead of DIFS, CW_{\min} , and CW_{\max} , respectively. If one class has a smaller AIFS or CW_{\min} or CW_{\max} , the class's traffic has a better chance to access the wireless medium earlier.

Four transmission queues are implemented in a station, and each queue supports one AC class, behaving roughly as a

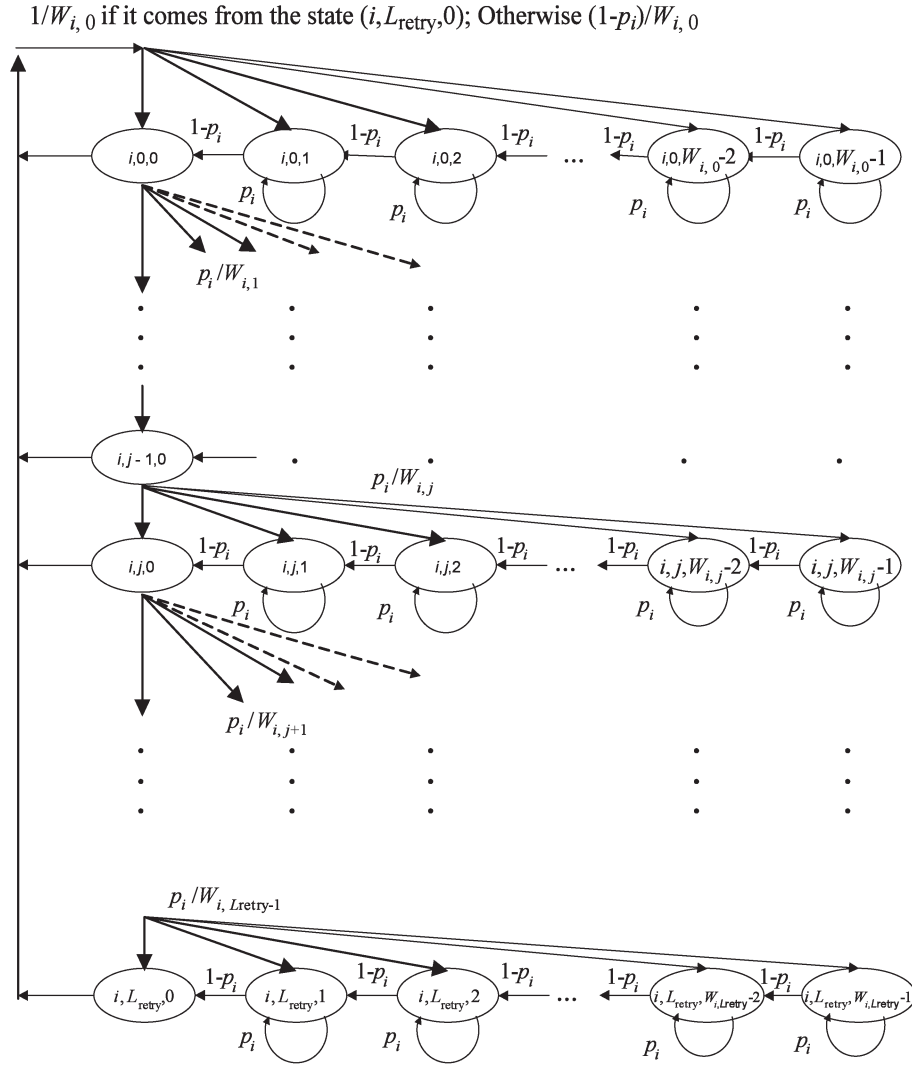


Fig. 1. State transition diagram for the priority i class.

single DCF entity in the original IEEE 802.11 MAC. It is assumed that a payload from a higher layer is labeled with a priority value, and it is pushed into the corresponding queue with the same priority value. Each queue acts as an independent MAC entity and performs the same DCF function with a different interframe space ($AIFS[i]$), a different initial window size ($CW_{min}[i]$), and a different maximum window size ($CW_{max}[i]$). Each queue has its own backoff counter ($BO[i]$) that acts independently the same way as the original DCF backoff counter introduced in the previous section. If there is more than one queue finishing the backoff at the same time, the highest priority frame is chosen to transmit by the virtual collision handler. Other lower priority frames whose backoff counters also reach zeros will increase their backoff counters with $CW_{min}[i]$ ($i = 0, 1, \dots, 3$), accordingly.

III. ANALYTICALLY MODELING MULTIPLE PRIORITIES

Based on Bianchi's model [2] and ZA's model [3], an analytical model for multiple prioritized classes of traffic under high traffic condition is proposed. We assume that each station belongs to one and only one priority class (we will lose this

assumption for the EDCF in the next section) and always has frames ready to send.

A. An Analytical Model

For a given station in the priority i class ($i = 0, \dots, N - 1$), $b(i, t)$ is defined as a random process representing the value of the backoff counter at time t , and $s(i, t)$ is defined as the random process representing the backoff stage j ($j = 0, 1, \dots, L_{i, retry}$) where $L_{i, retry}$ is the retry limit. The value of the backoff counter $b(i, t)$ is uniformly chosen in the range $(0, 1, \dots, W_{i, j} - 1)$, where $W_{i, j} = \lceil (\sigma_i)^j W_{i, 0} \rceil$. Let p_i denote the probability that a transmitted frame collides and p_i also equals to probability that a station in the backoff stage for the priority i class senses the channel busy. Similar to Bianchi's model [2], the bidimensional random process $\{s(i, t), b(i, t)\}$ is a discrete-time Markov chain. Therefore, the state of each station in the priority i class is described by $\{i, j, k\}$, where j stands for the backoff stage taking values from $(0, 1, \dots, L_{i, retry})$ and k stands for the backoff delay taking values from $(0, 1, \dots, W_{i, j} - 1)$ in timeslots. The state transition diagram for the priority i class is shown in Fig. 1.

From Fig. 1, we can observe the following several differences from Bianchi's model [2], ZA's model [3], and our old model [7]. 1) Similar to the old model, Fig. 1 adopts three variables in each state $\{i, j, k\}$ whereas the model of Bianchi and ZA adopts two variables in each state; therefore, Fig. 1 can be used to model priorities. 2) A finite retry limit $L_{i,\text{retry}}$ is modeled, whereas Bianchi's model, ZA's model, and the old model all assume an infinite retry limit. 3) Furthermore, the frame-dropping probability can also be derived. 4) Fig. 1 considers that the backoff counter is stopped as in ZA's model. 5) Both ZA's model and our old model adopt p_b , the probability that the channel is busy, whereas Fig. 1 adopts p_i , the probability that a station in the backoff stage for the priority i class senses the channel busy. Therefore, Fig. 1 is more accurate; note that 5) is only a minor aspect since all the models are approximate models, and therefore some reasonable approximations with minor differences do not influence much of the final accuracy. 6) Both ZA's model and our old model assume that after every successful transmission, a station can transmit if the medium is idle without entering the backoff stage; Fig. 1 has no "nonbackoff stage." The non-null transition probabilities in Fig. 1 are listed as follows:

$$\Pr[(i, 0, k)|(i, j, 0)] = \frac{(1 - p_i)}{W_{i,0}},$$

$$\text{for } 0 \leq k \leq W_{i,0} - 1 \text{ and } 0 \leq j < L_{i,\text{retry}}$$

$$\Pr[(i, 0, k)|(i, L_{i,\text{retry}}, 0)] = \frac{1}{W_{i,0}},$$

$$\text{for } 0 \leq k \leq W_{i,0} - 1$$

$$\Pr[(i, j, k)|(i, j, k)] = p_i,$$

$$\text{for } 1 \leq k \leq W_{i,j} - 1 \text{ and } 0 \leq j \leq L_{i,\text{retry}}$$

$$\Pr[(i, j, k)|(i, j, k + 1)] = 1 - p_i,$$

$$\text{for } 0 \leq k \leq W_{i,j} - 2 \text{ and } 0 \leq j \leq L_{i,\text{retry}}$$

$$\Pr[(i, j, k)|(i, j - 1, 0)] = \frac{p_i}{W_{i,j}},$$

$$\text{for } 0 \leq k \leq W_{i,j} - 1 \text{ and } 1 \leq j \leq L_{i,\text{retry}}.$$

Let $b_{i,j,k} = \lim_{t \rightarrow \infty} \Pr[s(i, t) = j, b(i, t) = k]$ be the stationary distribution of the Markov chain. In steady state, we can derive the following relations through chain regularities:

$$b_{i,j,0} = p_i^j b_{i,0,0} \quad 0 \leq j \leq L_{i,\text{retry}} \quad (2)$$

$$b_{i,j,k} = \frac{W_{i,j} - k}{W_{i,j}} \frac{1}{1 - p_{i,b}} b_{i,j,0} \quad 0 \leq j \leq L_{i,\text{retry}}, \quad 1 \leq k \leq W_{i,j} - 1 \quad (3)$$

and

$$\sum_{j=0}^{L_{i,\text{retry}}} \sum_{k=0}^{W_{i,j}-1} b_{i,j,k} = 1. \quad (4)$$

From (2) to (4), we have

$$b_{i,0,0} = \frac{1}{\sum_{j=0}^{L_{i,\text{retry}}} \left[1 + \frac{1}{1 - p_i} \sum_{k=1}^{W_{i,j}-1} \frac{W_{i,j}-k}{W_{i,j}} \right] p_i^j}. \quad (5)$$

Let τ_i be the probability that a station in the priority i class transmits during a generic slot time. A station transmits when its backoff counter reaches zero, i.e., the station is at any of states $\{i, j, 0\}$ ($0 \leq j \leq L_{i,\text{retry}}$)

$$\tau_i = \sum_{j=0}^{L_{i,\text{retry}}} b_{i,j,0} = b_{i,0,0} \frac{1 - p_i^{L_{i,\text{retry}}+1}}{1 - p_i}. \quad (6)$$

Let n_i ($i = 0, \dots, N - 1$) denote the number of stations in the priority i class. A transmitted frame collides when one more station also transmits during a slot time. The probability p_i that a station in the backoff stage for the priority i class senses the channel busy is given as

$$p_i = 1 - \left[\prod_{h=0}^{i-1} (1 - \tau_h)^{n_h} \right] (1 - \tau_i)^{n_i-1} \left[\prod_{h=i+1}^{N-1} (1 - \tau_h)^{n_h} \right]. \quad (7)$$

Substituting (1) and (7) to (6), we can solve unknown parameters numerically. Then, we can calculate p_i from (7). Let p_b denote the probability that the channel is busy. It happens when at least one station transmits during a slot time. Therefore, we have

$$p_b = 1 - \prod_{h=0}^{N-1} (1 - \tau_h)^{n_h}. \quad (8)$$

B. Saturation Throughput

Let $p_{s,i}$ ($i = 0, \dots, N - 1$) denote the probability that a successful transmission occurs in a slot time for the priority i class and p_s the probability that a successful transmission occurs in a slot time. We have

$$p_{s,i} = n_i \tau_i (1 - \tau_i)^{n_i-1} \prod_{h=0, h \neq i}^{N-1} (1 - \tau_h)^{n_h} \quad (9)$$

$$\begin{aligned} p_s &= \sum_{i=0}^{N-1} p_{s,i} = \sum_{i=0}^{N-1} \frac{n_i \tau_i}{1 - \tau_i} (1 - p_b) \\ &= (1 - p_b) \sum_{h=0}^{N-1} \frac{n_h \tau_h}{1 - \tau_h}. \end{aligned} \quad (10)$$

Let S_i ($i = 0, \dots, N - 1$) denote the normalized throughput for the priority i class. Let δ , $T_{E(L)}$, T_s , and T_c denote the duration of an empty slot time, the time to transmit the average payload, the average time that the channel is sensed busy because of a successful transmission, and the average time that the channel has a collision, respectively. The probability that the channel is idle for a slot time is $(1 - p_b)$, and the probability

that the channel is neither idle nor successful for a slot time is $[1 - (1 - p_b) - p_s] = (p_b - p_s)$. Therefore, we have

$$S_i = \frac{E(\text{payload transmission time in a slot time for the } i \text{ class})}{E(\text{length of a slot time})} \\ = \frac{p_{s,i} T_{E(L)}}{(1 - p_b)\delta + p_s T_s + [p_b - p_s] T_c}. \quad (11)$$

Let T_H , T_{ACK} , SIFS, L^* , and $T_{E(L^*)}$ denote the time to transmit the header (including MAC header, physical layer header, and/or tail), the time to transmit an ACK, SIFS time, the length of the longest frame in a collision, and the time to transmit a payload with length $E(L^*)$, respectively. For the basics access method, we have

$$T_s^{\text{basic}} = T_H + T_{E(L)} + \text{SIFS} + T_{ACK} + \text{DIFS} \quad (12)$$

$$T_c^{\text{basic}} = T_H + T_{E(L^*)} + \text{SIFS} + T_{ACK} + \text{DIFS}. \quad (13)$$

Please refer to [2] for calculating $T_{E(L^*)}$. Let T_{RTS} and T_{CTS} denote the time to transmit an RTS frame and a CTS frame, respectively. For the RTS/CTS access model, we have

$$T_c^{\text{RTS/CTS}} = T_{RTS} + 3 * \text{SIFS} + T_{CTS} \\ + T_H + T_{E(L)} + T_{ACK} + \text{DIFS} \quad (14)$$

$$T_c^{\text{RTS/CTS}} = T_{RTS} + \text{SIFS} + T_{ACK} + \text{DIFS}. \quad (15)$$

C. Frame-Dropping Probability

Let $P_{i,\text{drop}}$ ($i = 0, \dots, N - 1$) denote the frame-dropping probability for the priority i class. From Fig. 1, we observe that a frame can be dropped only in state $\{i, L_{\text{retry}}, 0\}$ if a collision occurs. In other words, a frame can be dropped when the retransmission counter reaches the retry limit L_{retry}

$$P_{i,\text{drop}} = p_i^{L_{i,\text{retry}}+1}. \quad (16)$$

Let $P_{i,\text{success}}$ ($i = 0, \dots, N - 1$) denote the frame-successful probability for the priority i class

$$P_{i,\text{success}} = \sum_{j=0}^{L_{i,\text{retry}}} p_i^j (1 - p_i) = 1 - p_i^{L_{i,\text{retry}}+1}. \quad (17)$$

D. Saturation Delay

Saturation delay is the average delay under the saturation condition and includes the medium access delay (due to backoff, collisions, etc.), the transmission delay, and the interframe spaces (such as SIFS). The average backoff delay depends on the value of a station's backoff counter and the duration when the counter freezes due to other transmissions. Let X_i ($i = 0, \dots, N - 1$) denote the random variable representing the total

number of backoff slots, which a frame encounters without considering the case when the counter freezes, for the priority i class. The probability that the frame is successfully transmitted after the j th retry [which is the $(j + 1)$ th transmission] is given by $p_i^j (1 - p_i) / (1 - p_i^{L_{i,\text{retry}}+1})$, which uses conditional probability on a successful transmission with probability $P_{i,\text{success}}$. The average number of backoff slots that a station needs to transmit a frame successfully at the j th retry is $\sum_{h=0}^j [(W_{i,h} - 1) / 2]$

$$E(X_i) = \sum_{j=0}^{L_{i,\text{retry}}} \frac{p_i^j (1 - p_i)}{1 - p_i^{L_{i,\text{retry}}+1}} \sum_{h=0}^j \frac{W_{i,h} - 1}{2}. \quad (18)$$

Note that from (18) only successful transmissions are considered. Let B_i ($i = 0, \dots, N - 1$) denote the random variable representing the total number of slots when the counter freezes, which a frame encounters, for the priority i class. The portion of idle slots is $(1 - p_i)$, which is used to decrease $E(X_i)$. We have

$$E(B_i) = \frac{E(X_i)}{(1 - p_i)} p_i. \quad (19)$$

We can treat $E(X_i)$ and $E(B_i)$ as the total number of idle and busy slots that the frame encounters during backoff stages, respectively. Let $E(N_{i,\text{retry}})$ denote the average number of retries for the priority i class. We have

$$E(N_{i,\text{retry}}) = \sum_{j=0}^{L_{i,\text{retry}}} \frac{j p_i^j (1 - p_i)}{1 - p_i^{L_{i,\text{retry}}+1}}. \quad (20)$$

Let D_i ($i = 0, \dots, N - 1$) denote the random variable representing the frame delay for the priority i class. Let T_o denote the time that a station has to wait when its frame transmission collides before sensing the channel again. Let $T_{ACK_timeout}$ and $T_{CTS_timeout}$ denote the duration of the ACK and CTS timeouts, respectively. Note that $E(N_{i,\text{retry}})$ is one less than the number of transmissions. The average slot lengths are δ , $\{(p_s/p_b)T_s + [(p_b - p_s)/p_b]T_c\}$, $(T_c + T_o)$, and T_s for an idle slot at states $\{i, j, k\}$ ($k > 0$), a busy slot at states $\{i, j, k\}$ ($k > 0$), a failed transmission slot for this station at states $\{i, j, 0\}$, and a successful transmission at states $\{i, j, 0\}$, respectively. We have

$$E(D_i) = E(X_i)\delta + E(B_i) \left[\frac{p_s}{p_b} T_s + \frac{(p_b - p_s)}{p_b} T_c \right] \\ + E(N_{i,\text{retry}})(T_c + T_o) + T_s \quad (21)$$

$$T_o^{\text{basic}} = \text{SIFS} + T_{ACK_timeout} \quad (22)$$

$$T_o^{\text{rts}} = \text{SIFS} + T_{CTS_timeout}. \quad (23)$$

IV. CUSTOMIZING ANALYTIC MODEL

We assume that each station belongs to one and only one priority class. We will lose this assumption in Section IV-B for the EDCF priority scheme.

A. Backoff-Based Priority Schemes

Listed above are all the requirements for the combined priority scheme using all three metrics. For those priority schemes that do not adopt all three metrics, the unused metrics will be set to constants. For example, for the priority scheme differentiating only the initial window size, we have $L_{i,\text{retry}} = L_{\text{Retry}}$ and $\sigma_i = 2$ for the priority i ($i = 0, \dots, N - 1$) class. For the priority scheme differentiating only the window-increasing factor, we have $L_{i,\text{retry}} = L_{\text{Retry}}$ and $W_{i,0} = CW_{\min}$ for the priority i ($i = 0, \dots, N - 1$) class. For the priority scheme differentiating only the retry limit, we have $\sigma_i = 2$ and $W_{i,0} = CW_{\min}$ for the priority i ($i = 0, \dots, N - 1$) class. For other priority schemes that adopt two metrics, the unused metric will be set to a constant.

B. EDCA, A Hybrid Priority Scheme

For the EDCA priority scheme, the number of ACs is four, and we do not differentiate the window-increasing factor and the retry limit. Therefore, we have $N = 4$, and for the priority i ($i = 0, \dots, 3$) class we have $\sigma_i = 2$, $L_{i,\text{retry}} = L_{\text{Retry}}$, and $W_{i,0} = CW_{\min}[i]$. Note that we only evaluate backoff-based priority schemes of the EDCA. Assuming that there are n active stations in a basic service set and each station implements all four queues, there are a total of $4n$ queue entities, which are roughly equivalent to $4n$ stations in the original IEEE 802.11 DCF. More generally, let n_i ($i = 0, \dots, 3$) denote the number of active queues in the priority i class. We assume that any queue in any station always has frames ready to send. In order to use the general model, we also assume that one of the following two assumptions will be true.

- Each station belongs to one and only one priority class and uses only one queue that belongs to that class.
- Each station has traffic of multiple priority classes. Each class traffic uses the queue that belongs to that class. If there is more than one queue finishing the backoff at the same time, we assume that the virtual collision handler will not make a choice about who will send, but it will let a collision occur in the wireless medium. With this assumption, each queue will be truly equivalent to a station in the original IEEE 802.11 DCF, and n_i , the number of active queues in the priority i class, can be treated as the number of active stations in the priority i class in the general analytical model. Note that the EDCA will perform a little better without this assumption. However, such an assumption will not influence our studies of priority schemes' characteristics.

V. VALIDATIONS

In this section, we conduct simulations to validate the proposed analytic model. The IEEE 802.11a simulation models had been developed based on the IEEE 802.11a standard [27], IEEE 802.11e draft [15], and OPNET wireless local area network simulation model 8.0 A (for IEEE 802.11b DCF). Furthermore, we adapt our simulation model with similar assumptions as those in the analytical model. The data rate is 6 Mb/s and the control rate is 6 Mb/s. The frame size is

TABLE I
SIMULATION AND ANALYTICAL RESULTS (LEGEND: ST, SATURATION THROUGHPUT; SD, SATURATION DELAYS (MICROSECOND); N, NUMBER OF ACTIVE STATIONS; S, SIMULATION; A, ANALYTICAL; E, RELATIVE ERROR)

ST	Class	S/A	$n_0 = n_1 = 10$	$n_0 = n_1 = 30$
SD	0	S	0.5161	0.4638
		A	0.520821	0.465941
		E	-0.9%	-0.45%
	1	S	0.1532	0.1008
		A	0.154653	0.101312
		E	-0.96%	-0.46%
ST	0	S	25533.74	67737.56
		A	25082.262785	67000.549513
		E	1.8%	1.1%
	1	S	88398.10	348845.78
		A	86495.210138	342341.299834
		E	2.2%	1.9%

fixed at 1024 bytes. For demonstration purposes, we adopt two priority classes, i.e., $N = 2$. All the simulation results have over 95% confidence intervals. The relative error is calculated by (Simulation Result – Analytic Result)/Analytic Result.

Table I shows the simulation versus numerical results for a combined backoff-based priority scheme when adopting the following parameters: $[\sigma_0, \sigma_1] = [1.7, 2]$, $[L_{0,\text{retry}}, L_{1,\text{retry}}] = [4, 7]$, $[W_{0,0}, W_{1,0}] = [16, 32]$, $[CW_{0,\text{max}}, CW_{1,\text{max}}] = [1024, 1024]$, and $n_0 = n_1$. As illustrated in the table, analytical and simulation results match pretty well for both saturation throughput and saturation delay.

VI. NUMERICAL RESULTS

We adopt IEEE 802.11a as an example. The IEEE 802.11a parameters can be found in [5] and [15], are used to calculate $T_H + T_{E(L)}$ accurately [15]. Both the data rate and the control rate are 6 Mb/s. The frame size is fixed at 1024 bytes. For demonstration purposes, we adopt two priority classes, i.e., $N = 2$. However, our proposed model is very general so that we can design many levels of priorities.

A. Effectiveness of a Combined Backoff-Based Priority Scheme

Figs. 2–4 have the following parameters: $[\sigma_0, \sigma_1] = [1.7, 2]$, $[L_{0,\text{retry}}, L_{1,\text{retry}}] = [4, 7]$, $[W_{0,0}, W_{1,0}] = [16, 32]$, $[CW_{0,\text{max}}, CW_{1,\text{max}}] = [1024, 1024]$, and $n_0 = n_1$. Figs. 2–4 show saturation throughputs, saturation delays, and frame-dropping probabilities, respectively, for two priority classes over the number of stations (n_0 or n_1). As illustrated in the figures, class 0 has a much better saturation throughput (delay) than class 1. Fig. 4 shows that as the number of stations increases the frame-dropping probability increases. Furthermore, since class 0 has a smaller retry limit, it has a higher frame-dropping probability. These figures indicate that the combined backoff-based priority scheme is very effective.

B. Effects of the Initial Window Size

Figs. 5 and 6 have the following parameters: $\sigma_0 = \sigma_1 = 2$, $L_{0,\text{retry}} = L_{1,\text{retry}} = 7$, $W_{0,0} = 16$, $CW_{0,\text{max}} = CW_{1,\text{max}} =$

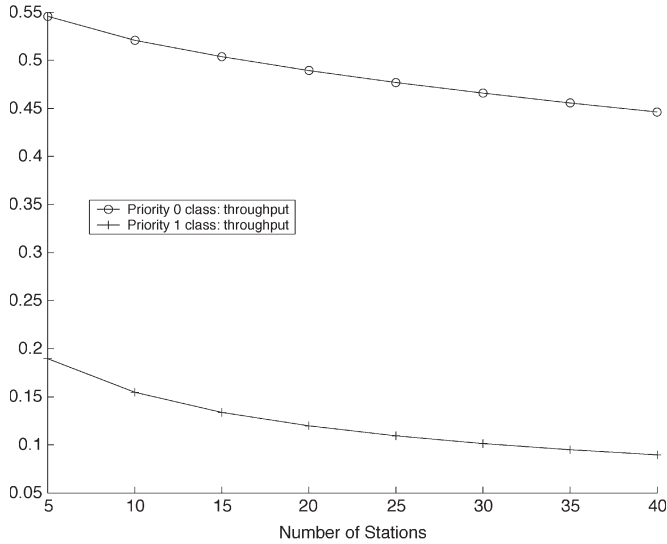


Fig. 2. Saturation throughputs.

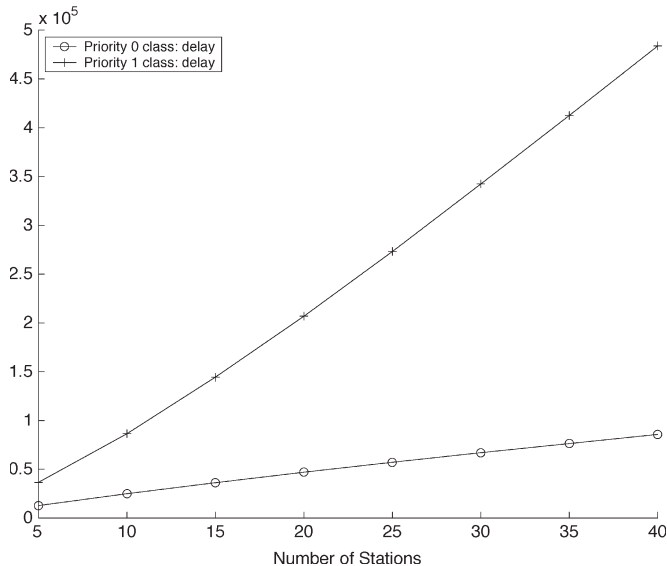


Fig. 3. Saturation delay (microsecond).

1024, and $n_0 = n_1 = 10$. Figs. 5 and 6 show saturation throughputs (saturation delays) over the minimum window size of class 1 $W_{1,0}$ that changes from 16 to 56. As illustrated in Figs. 5 and 6, when $W_{1,0} = 16$, saturation throughputs (saturation delays) are the same for both classes. As $W_{1,0}$ increases, the saturation throughput of class 1 decreases, the saturation throughput of class 0 increases, the saturation delay of class 0 decreases, and the saturation delay of class 1 increases. An interesting observation in Figs. 5 and 6 is that the throughputs/delays of class 0 and class 1 are symmetrical along a line parallel to the x -axis. This phenomenon indicates that class 0 can steal bandwidth from class 1 as the initial window size of class 1 ($W_{1,0}$) increases, whereas the total throughput of all classes increases a little. The reasons for this are as follows. As the initial window size of class 1 increases, stations in class 1 will delay accessing the channel so that the saturation throughput of class 1 will decrease and the saturation delay of class 1 will increase. Furthermore, the collision probability

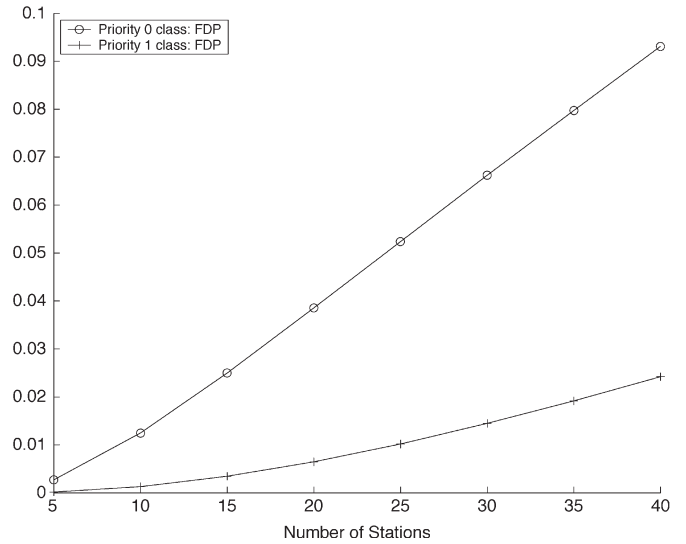


Fig. 4. Frame-dropping probability.

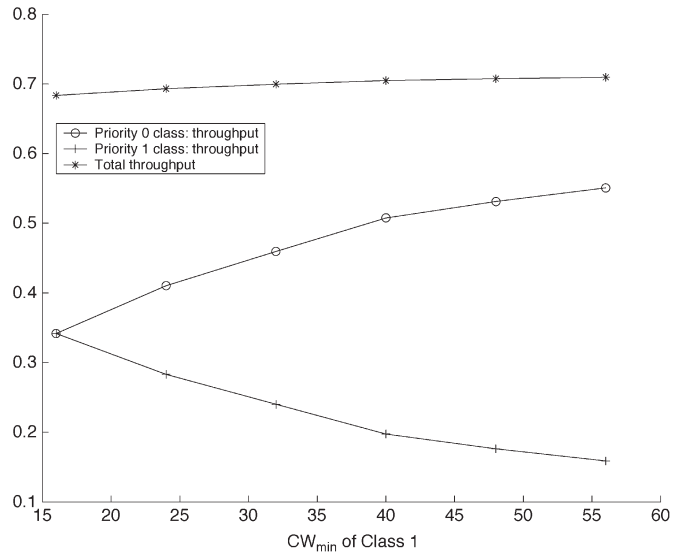


Fig. 5. Saturation throughputs (differentiating the initial window size).

of class 0 will decrease so that the saturation throughput of class 0 will increase and the saturation delay of class 0 will decrease. Therefore, differentiating the initial window size has the function of both reducing collisions and providing priorities. In fact, the total throughput increases a little bit with the same reason as follows. The overall system performance can sometimes be improved by using a contention protocol that assigns different access probabilities to different stations [17].

C. Effects of the Backoff Window-Increasing Factor

Figs. 7 and 8 have the following parameters: $\sigma_0 = 2$, $L_{0, \text{retry}} = L_{1, \text{retry}} = 7$, $W_{0,0} = W_{1,0} = 16$, $CW_{0, \text{max}} = CW_{1, \text{max}} = 1024$, and $n_0 = n_1 = 10$. Figs. 7 and 8 show saturation throughputs and saturation delays, respectively, over the backoff window-increasing factor of class 1, σ_1 that changes from 1.6 to 2.6. As illustrated in the figures, when $\sigma_1 < \sigma_0$, class 1 has better throughput and delay than class 0;

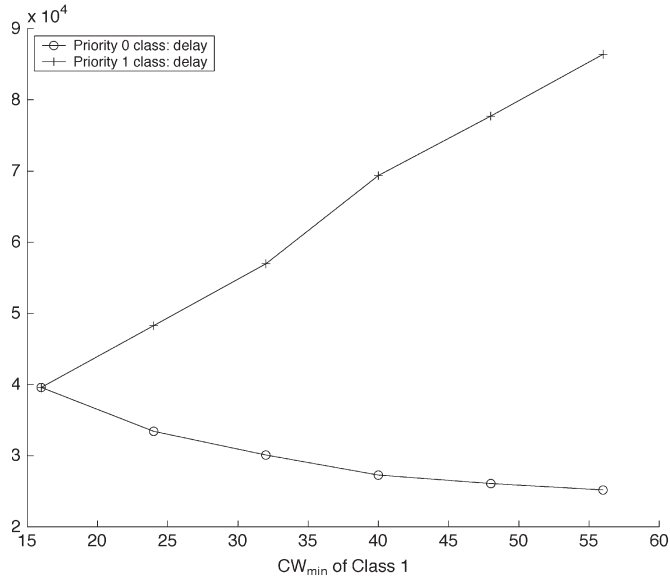


Fig. 6. Saturation delay (microsecond) (differentiating the initial window size).

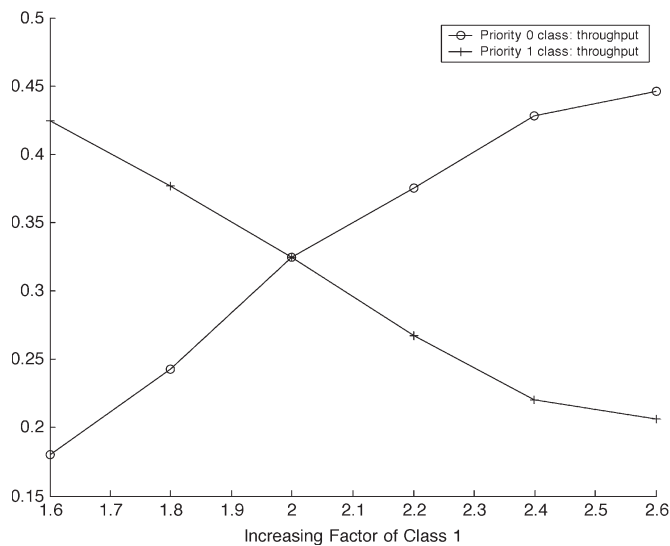


Fig. 7. Saturation throughput by differentiating the backoff window-increasing factor.

when $\sigma_1 = \sigma_0 = 2$, they have the same throughput and delay; when $\sigma_1 > \sigma_0$, class 0 has better throughput and delay than class 1. Figs. 7 and 8 also have the same effect as Figs. 5 and 6, i.e., that class 0 can steal bandwidth from class 1 with a similar reason and that the total throughput does not decrease. Differentiating the backoff window-increasing factor also has the function of both reducing collisions and providing priorities.

D. Effects of the Retry Limit

Figs. 9–11 have the following parameters: $\sigma_0 = \sigma_1 = 2$, $L_{0, \text{retry}} = 8$, $W_{0,0} = W_{1,0} = 16$, $CW_{0, \text{max}} = CW_{1, \text{max}} = 1024$, and $n_0 = n_1 = 10$. Figs. 9–11 show saturation throughputs, saturation delays, and frame-dropping probabilities, respectively, over the retry limit of class 1, $L_{1, \text{retry}}$, which changes from 0 to 8. As illustrated in the figures, class 1 has bet-

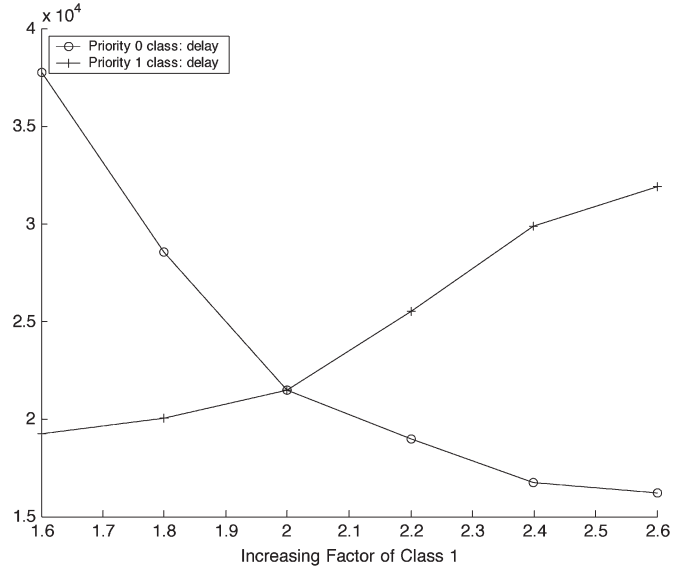


Fig. 8. Saturation delay (microsecond) by differentiating the backoff window-increasing factor.

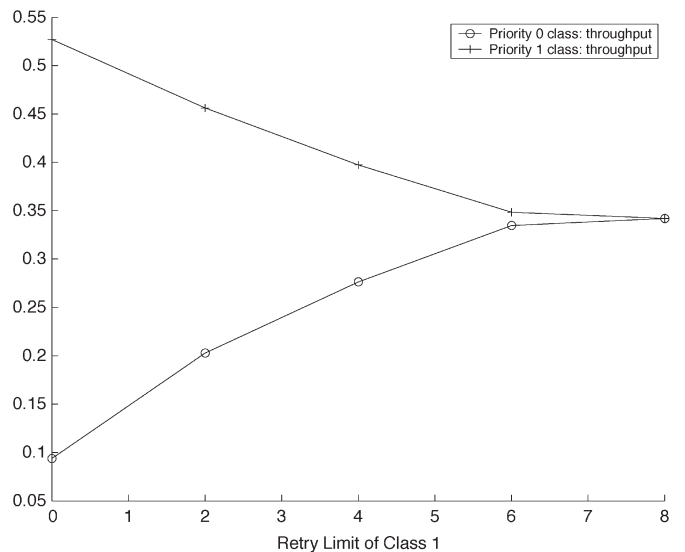


Fig. 9. Saturation throughput by differentiating the retry limit.

ter throughput and delay than class 0 when $L_{1, \text{retry}} < L_{0, \text{retry}}$, and they have the same throughput and delay when $L_{1, \text{retry}} = L_{0, \text{retry}}$. Fig. 9 also has the same effect as Figs. 5 and 7, i.e., that class 0 can steal bandwidth from class 1 with a similar reason and that the total throughput does not decrease. Differentiating the retry limit also has the function of both reducing collisions and providing priorities. Fig. 11 shows that class 1 has a higher frame-dropping probability when $L_{1, \text{retry}} < L_{0, \text{retry}}$.

For real-time multimedia traffic with sensitive delay requirements, retransmitted frames may be too late to be useful and a smaller retry limit is appropriate, whereas some nonreal-time transmissions may need a larger retry limit to enhance reliable transmissions. Fig. 11 shows that with a smaller retry limit, both throughput and delay can be improved with a higher frame-dropping probability. This is good for real-time delay-sensitive applications that can survive with frame loss, such as voice traffic.

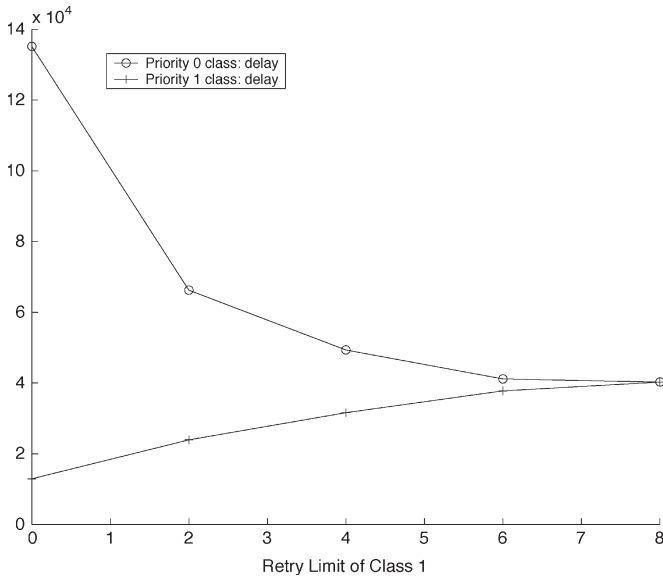


Fig. 10. Saturation delay (microsecond) by differentiating the retry limit.

VII. CONCLUSION

In this paper, we studied backoff-based priority schemes for both the original IEEE 802.11 and the emerging IEEE 802.11e. An improved analytical model was proposed to study all the proposed priority schemes in terms of saturation throughput, saturation delay, and frame-dropping probability. Our study shows the following results.

- For all three backoff-based metrics (the initial window size, the retry limit, and the backoff window-increasing factor), one class can steal bandwidth from another if the latter one increases the metric value and the total throughput increases a little. This fact indicates that they are good metrics. The reason that the total throughput increases a little is the same reason as follows: the overall system performance can sometimes be improved by using a contention protocol that assigns different access probabilities to different stations [17]. Three metrics can be implemented altogether since this does not make hardware implementation any more difficult.
- For real-time multimedia traffic with sensitive delay requirements, retransmitted frames may be too late to be useful and a smaller retry limit is appropriate, whereas some nonreal-time transmissions may need a larger retry limit to enhance reliable transmissions. Our results show that with a smaller retry limit, both throughput and delay can be improved with a higher frame-dropping probability. This is good for real-time delay-sensitive applications that can survive with frame loss, such as voice traffic.
- A backoff-based metric has the function of both reducing collisions and providing priorities.
- Simulation results match analytical results very well.

The proposed analytical priority model is very general. It might be useful to evaluate other priority schemes. The results of this paper are also beneficial in designing good prioritized QoS parameters.

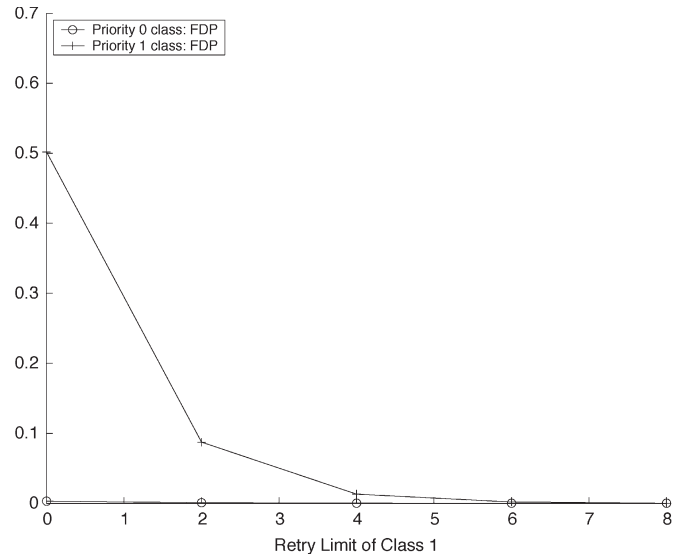


Fig. 11. Frame-dropping probability by differentiating the retry limit.

ACKNOWLEDGMENT

The author would like to thank Dr. G. Bianchi for pointing out an error in the original delay model, Dr. K. K. Leung for his helpful suggestions, and three anonymous reviewers whose valuable comments significantly improved the manuscript.

REFERENCES

- [1] IEEE 802.11 WG, *Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification*, 1999.
- [2] 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.
- [3] E. Zouva 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.
- [4] H. Wu, Y. Peng, K. Long, S. Cheng, and J. Ma, "Performance of reliable transport protocol over IEEE 802.11 Wireless LANs: Analysis and enhancement," in *IEEE Information Communications (INFOCOM)*, New York, 2002, pp. 599–607.
- [5] 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.
- [6] D.-J. Deng and R.-S. Chang, "A priority scheme for IEEE 802.11 DCF access method," *IEICE Trans. Commun.*, vol. E82-B, no. 1, pp. 96–102, Jan. 1999.
- [7] Y. Xiao, "A simple and effective priority scheme for IEEE 802.11," *IEEE Commun. Lett.*, vol. 7, no. 2, pp. 70–72, Feb. 2003.
- [8] A. Veres, A. T. Campbell, M. Barry, and L.-H. Sun, "Supporting differentiation in wireless packet networks using distributed control," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 10, pp. 2081–2093, Oct. 2001.
- [9] I. Aad and C. Castelluccia, "Differentiation mechanisms for IEEE 802.11," in *IEEE Information Communications (INFOCOM)*, Anchorage, AK, 2001, pp. 209–218.
- [10] X. Pallot and L. E. Miller, "Implementing message priority policies over an 802.11 based mobile ad hoc network," in *IEEE Military Communications Conf. (MILCOM)*, McLean, VA, 2001, pp. 860–864.
- [11] IEEE 802.11e WG, *Medium Access Control (MAC) Enhancements for Quality of Service*, IEEE 802.11e/D2.0, Nov. 2001.
- [12] S. Mangold, S. Choi, P. May, O. Kein, G. Hiertz, and L. Stibor, "IEEE 802.11e Wireless LAN for quality of service," in *European Wireless*, Florence, Italy, 2002, pp. 32–39.
- [13] Y. Xiao, "IEEE 802.11e: A QoS provisioning at the MAC layer," *IEEE Wireless Commun.*, vol. 11, no. 3, pp. 72–79, Jun. 2004.
- [14] Y. Xiao, H. Li, and S. Choi, "Protection and guarantee for voice and video traffic in IEEE 802.11e Wireless LANs," in *Proc. IEEE Information Communications (INFOCOM)*, Hong Kong, 2004, pp. 2153–2163.

- [15] IEEE 802.11a WG, *Part 11: High-Speed Physical Layer in the 5 GHz Band*, 1999.
- [16] IEEE 802.11b WG, *Part 11: High-Speed Physical Layer Extension in the 2.4 GHz Band*, 1999.
- [17] A. S. Tanenbaum, *Computer Networks*, 4th ed. New Jersey: Prentice-Hall, 2002.
- [18] G. Bianchi and I. Tinnirello, "Analysis of priority mechanisms based on differentiated inter frame spacing in CSMA-CA," in *IEEE Vehicular Technology Conf. (VTC) Fall*, Orlando, FL, 2003, pp. 1401–1405.
- [19] J. Zhao, Z. Guo, Q. Zhang, and W. Zhu, "Performance study of MAC for service differentiation in IEEE 802.11," in *IEEE Global Telecommunications Conf. (GLOBECOM)*, Taipei, Taiwan, 2002, pp. 778–782.
- [20] T. S. Ho and K. C. Chen, "Performance evaluation and enhancement of the CSMA/CA MAC protocol for 802.11 wireless LAN's," in *Proc. IEEE Personal Indoor and Mobile Radio Communications (PIMRC)*, Taipei, Taiwan, Oct. 1996, pp. 392–396.
- [21] B. Bing and R. Subramanianb, "A novel technique for quantitative performance evaluation of wireless LANs," *Comput. Commun.*, vol. 21, no. 9, pp. 833–838, Jul. 1998.
- [22] H. S. Chhaya and S. Gupta, "Performance modeling of asynchronous data transfer methods of IEEE 802.11 MAC protocol," *Wireless Netw.*, vol. 3, no. 3, pp. 217–234, May 1997.
- [23] K. C. Huang and K.-C. Chen, "Interference analysis of nonpersistent CSMA with hidden terminals in multicell wireless data networks," in *Proc. IEEE Personal Indoor and Mobile Radio Communications (PIMRC)*, Toronto, Canada, Sep. 1995, pp. 907–911.
- [24] F. Cali, 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. 785–790, Dec. 2000.
- [25] —, "IEEE 802.11 protocol: Design and performance evaluation of an adaptive backoff mechanism," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 19, pp. 1774–1786, Sep. 2000.
- [26] Y. C. Tay and K. C. Chua, "A capacity analysis for the IEEE 802.11 MAC protocol," *Wireless Netw.*, vol. 7, no. 2, pp. 159–171, Mar./Apr. 2001.
- [27] Y. Xiao and J. Rosdahl, "A performance analysis of IEEE 802.11a Wireless LAN," in *Proc. Systemics, Cybernetics, and Informatics (SCI)*, Orlando, FL, 2002, pp. 243–248.



Yang Xiao (S'98–M'01–SM'04) was a Voting Member of the IEEE 802.11 Working Group from 2001 to 2004. His research areas are wireless networks and mobile computing including wireless local area networks (LANs), wireless personal area networks (PANs), wireless metropolitan area networks (MANs), wireless wide area networks (WANs) (cellular networks), and ad hoc and sensor networks. His research interests are security/reliable communications, medium access control, mobility/location/paging managements, cache access and re-

placement policies, quality of service, energy efficiency, and routing in wireless networks and mobile computing.

Dr. Xiao currently serves as the Editor-in-Chief for the *International Journal of Security and Networks (IJSN)* and for the *International Journal of Sensor Networks (IJSNet)*. He serves as an Associate Editor or on the editorial boards for six journals. He serves as a Lead/Sole Guest Editor for five journal special issues. He served as a Technical Program Vice-Chair for two conferences, a symposium Co-Chair for two conferences, a Co-Editor for five edited books, a Transactions on Professional Communication (TPC) Member for more than 50 conferences/symposia/workshops, and a Reviewer for many journals, conferences, and funding agencies. He has served as a Panelist for the National Science Foundation (NSF) Cyber-Trust in 2005.