Throughput and Delay Performance of IEEE 802.11e Enhanced Distributed Channel Access (EDCA) Under Saturation Condition

Ching-Ling Huang and Wanjiun Liao, *Senior Member, IEEE*

***Abstract*— In this paper, we analyze the saturation performance of IEEE 802.11e Enhanced Distributed Channel Access (EDCA), which provides contention-based differentiated channel access for frames of different priorities in wireless LANs. With EDAC, Quality of Service (QoS) support is provided with up to four access categories (ACs) in each station. Each AC behaves as an independent backoff entity. The priority among ACs is then determined by AC-specific parameters, called the EDCA parameter set. The behavior of the backoff entity of each AC is modeled as a two-state Markov chain, which is extended from Bianchi's model to capture the features of EDCA. The differences of our model from existing work for 802.11e EDCA include: (i) virtual collisions among different ACs in an EDCA station are modeled, thus more accurately capturing the behavior of EDCA; (ii) the influence of using different arbitrary inter-frame spaces (AIFS) for different ACs on saturation performance are considered; (iii) delay and delay jitter are derived, in addition to saturation throughput. The analytical model is validated via ns-2 simulations. The results show that our analytical model can accurately describe the behavior of IEEE 802.11e EDCA.**

***Index Terms*— Delay, EDCA, IEEE 802.11e, saturation performance, throughput.**

I. INTRODUCTION

IEEE 802.11 Wireless Local Area Network (WLAN) with the Distributed Coordination Function (DCF) [1] is the dominant wireless medium to access the Internet. With DCF, mobile stations contend for the access to the channel. All stations operate independently and share the channel bandwidth equally. IEEE 802.11e [2] is the supplementary standard of 802.11 Medium Access Control (MAC) to provide QoS for different kinds of applications. In 802.11e, QoS is supported with a new access method called the Hybrid Coordination Function (HCF). In HCF, two medium access mechanisms are defined: controlled channel access and contention-based channel access. In this paper, we focus only on the HCF contention-based channel access mechanism, referred to as Enhanced Distributed Channel Access (EDCA).

Manuscript received December 8, 2004; revised November 10, 2005 and March 21, 2006; accepted May 4, 2006. The associate editor coordinating the review of this paper and approving it for publication was K. K. Leung. This work was supported in part by the National Science Council (NSC), Taiwan, under a Center Excellence Grant NSC94-2752-E-002-006-PAE, and in part by NSC under Grant Number NSC94-2213-E-002-047.

Ching-Ling Huang was with the Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan. He is now with the Dept. of Civil and Environmental Engineering, University of California Berkeley, Berkeley, CA 94530 USA (e-mail: huang@calccit.org).

Wanjiun Liao is with the Department of Electrical Engineering and the Graduate Institute of Communication Engineering, National Taiwan University, Taipei, Taiwan (e-mail: wjliao@ntu.edu.tw).

Digital Object Identifier 10.1109/TWC.2007.04796.

A. Legacy 802.11 DCF

IEEE 802.11 DCF is based on Carrier-Sense Multiple Access with Collision Avoidance (CSMA/CA), which introduces the "Inter Frame Space" (IFS) and the backoff interval to avoid collisions. Each station wishing to transmit monitors the wireless channel. When the channel becomes idle, the station waits for a period of DCF IFS (DIFS) and a backoff process before a transmission. If the channel is still idle, the station starts a transmission; otherwise, the transmission is deferred. Once the station enters the backoff process, a backoff interval is selected. The selected value is uniformly distributed in the interval $[0, CW-1]$, where CW (i.e., Contention Window) falls between the minimum (CW_{min}) and maximum (CW_{max}) contention windows. The value of the backoff interval is decremented by one at each time slot on an idle medium. When the channel is busy, the backoff process is frozen, and resumed after the channel becomes idle again. Eventually, the backoff interval reaches zero, and the station starts transmission. A collision occurs when more than one station transmits in the same slot. In this case, the CWs of the collided stations are doubled until CW_{max} is reached. CW is reset to CW_{min} after each successful transmission.

In DCF, the "Request to Send (RTS) / Clear to Send (CTS)" access mechanism is optional for packet transmissions. A station wishing to transmit must first obtain the channel access right based on the operation of the basic access mode described above. The sender station then transmits a special short frame called RTS frame, and the receiver, on receiving an RTS frame, responds with a CTS frame after an SIFS period. The sender starts transmitting a data packet after it has received the CTS frame correctly; the receiver, on receipt of the data, sends an ACK in response after another SIFS to finish the transmission. The frames RTS and CTS carry information about the duration of channel occupancy. Other stations receiving these frames set their Network Allocation Vector (NAV) according to the duration recorded in the frames and stop sensing the channel until the end of the duration. After that, they sense the channel again. If the channel is idle for a DIFS period, they will apply the backoff procedure and compete for the channel again.

B. 802.11e EDCA

The EDCA of 802.11e is an enhanced version of 802.11 DCF for priority-based QoS support. With EDCA, a station can implement up to four access categories (ACs), corresponding to voice, video, best effort, and background traffic, respectively. Each AC is associated with one backoff entity and some AC-specific parameters called the EDCA parameter set composed of Arbitrary Inter-Frame Space Number ($AIFSN[AC]$), minimum contention window ($CW_{min}[AC]$), and maximum contention window ($CW_{max}[AC]$). $AIFSN[AC]$ is used to determine the duration of Arbitrary IFS ($AIFS[AC]$) according to

$$AIFS[AC] = SIFS + AIFSN[AC] \times aslotTime,$$

where $AIFSN[AC] >= 2$, and $aslotTime$ is the duration of one slot. Since the value of $AIFSN[AC]$ is at least 2, the earliest access time for an EDCA station is after a DIFS. The backoff entities are prioritized according to the values of their EDCA parameter sets. The smaller the $AIFS[AC]$ or $CW_{min}[AC]$, the higher the priority in medium access, and thus, the higher the throughput. The backoff interval for an AC in EDCA is randomly selected from $[1, CW]$, instead of $[0, CW-1]$ as in DCF. Another EDCA feature different from DCF is that the backoff counter will be decreased by one slot before the end of AIFS as shown in Fig. 1. This feature of EDCA may cause problems in applications with coexisting legacy DCF and EDCA systems [19].

The operation of 802.11e EDCA is described as follows. Each data frame from the higher layer arrives in the MAC layer with a specific priority value. Then, each frame is mapped into an AC based on the specified priority. The values of the EDCA parameter set for each AC are announced periodically by the AP via beacon frames. Each AC behaves as a single enhanced DCF contending entity, and the corresponding queue has its own AIFS, backoff interval, and contention window (CW). After each unsuccessful transmission attempt, the contention window is doubled until a retry limit or the maximum contention window is reached. The collision is handled in a virtual manner. That is, the highest priority frame among the colliding frames is chosen and transmitted, and the others perform a backoff with an enlarged CW value. Note that there is no priority among EDCA stations; different EDCA stations have to compete for channel access with equal opportunity.

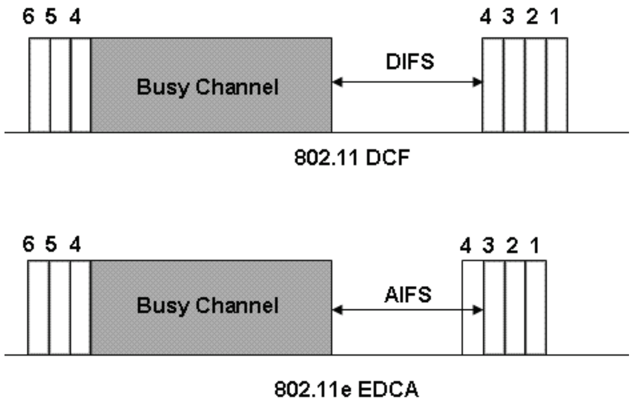

Fig. 1. Comparison of Backoff Counter Decrement behavior for DCF and EDCA.

C. Related Work

Most existing work on evaluating the performance of IEEE 802.11e QoS in the literature (e.g., [3-11]) are based on simulations (e.g., [3-6]). The analytical studies for IEEE 802.11e are mainly extended from Bianchi's model [12] (e.g., [7-11]). Bianchi's model is based on a two-state Markov chain to calculate the saturation throughput of 802.11 DCF in an ideal physical environment. In [7-9], only the saturation throughputs of 802.11e are analyzed. In [10], the saturation throughput and saturation delay are derived. However, that model does not consider the effect of virtual collisions among ACs inside an EDCA station. Thus, it cannot accurately calculate the MAC-layer delays of different ACs for each station. In [11], the behavior of differentiated CW and AIFS is considered, but the backoff procedure for different ACs with different AIFSs is not addressed.

D. Problem Specification

In this paper, we analyze the saturation performance of IEEE 802.11e EDCA throughput, access with delay, and delay jitter. The saturation throughput is defined as the maximum throughput that the system can achieve in stable conditions. The access delay is the interval from when a head-of-line data frame at the sender starts contending for the channel to when the data reaches the MAC layer of the receiver. The jitter is referred to as the standard deviation of the access delay described above. The Markov chain for the backoff entity of each AC in this paper is also extended from Bianchi's model [3] to accommodate the QoS features of 802.11e EDCA. The proposed analytical model differs from existing work in that: (i) we consider virtual collisions among different ACs inside each EDCA station in addition to external collisions among stations. Thus, our model can capture the behavior of 802.11e EDCA more accurately. (ii) We consider the impact of high-priority ACs' AIFSs on low priority ones. Thus, the probability that the backoff counter for each AC can be decreased by one in each time slot may not be identical or equal to one. (iii) We model the effect of the frame retry limit, i.e. the maximum number of retransmissions each frame can experience before being dropped. After this limit is reached, the retrying data frame is discarded. (iv) In addition to modeling the saturation throughput as in existing work, we also model the delay and delay jitter performance. Note that we do not consider channel errors in our model. Such considerations can be easily extended from existing work such as [13, 14].

The rest of the paper is organized as follows. In Section II, the saturation throughput of 802.11e EDCA is analyzed. In Section III, the delay and the jitter are analyzed. In Section IV, the analytical model is verified with network simulator ns-2. Finally, the paper is concluded is Section IV.

II. Throughput Analysis of 802.11e EDCA

In this section, we analyze the saturation throughput of IEEE 802.11e EDCA. Notations used in the analysis are summarized in Table I. To reach the throughput limit, each AC is assumed always backlogged, i.e., there is at least one data frame in each AC's queue ready to be sent. We also assume that the system operates in an ideal physical environment, i.e., no frame errors, the hidden terminal effect, and the capture effect in our model. For convenience, we denote ACs from the highest priority to the lowest priority by subscripts 0, 1, 2, and 3 in the analysis.

In [12], the DCF model is simplified by assuming the backoff decrement probability is one, which does not always hold for DCF since the decrement probability depends on whether there are other stations with backoff counters equal to zero. However, in EDCA, the backoff counter will always be decremented by one slot before the end of AIFS. Therefore, the basic Markov chain model presented in [12] is still valid for each AC and is further modified for the analysis as follows. The two-state Markov Chain for the EDCA backoff entity of an AC is shown in Fig. 2. The state $(s(t), b(t))$ is defined as follows. $s(t)$ is the backoff stage of a request packet at time t, defined as the number of collisions the request packet has suffered; $b(t)$ is the backoff counter at time t. The packet is sent whenever the backoff counter becomes zero regardless of the backoff stage. In DCF, the backoff counter is chosen in the range $(0, W_i - 1)$, but in 802.11e EDCA, it is chosen in $(1, W_i)$. The effect of the frame retry limit is considered, i.e., after $M+f$ transmission failures, the frame will be discarded. Here f is defined as the difference between M and the frame retry limit. Finally, we also consider the impact of different AIFSs of different ACs on the probability that the backoff counter of each AC can be decreased by one (denoted by PT).

The state transition probabilities for each AC in the proposed Markov model are expressed as follows.

$$\begin{cases} P(i,k|i,k+1) = PT, k \in (0, W_i - 1), i \in (0, M+f) \\ P(i,k|i,k) = 1 - PT, k \in (1, W_i), i \in (0, M+f) \\ P(0,k|i,0) = (1 - PC)/W_0, k \in (1, W_0), i \in (0, M+f-1) \\ P(i,k|i-1,0) = PC/W_i, k \in (1, W_i), i \in (1, M+f) \\ P(0,k|M+f,0) = 1/W_0, k \in (1, W_0) \end{cases} \quad (1)$$

where M is the maximum number of times the contention window may be doubled, $M+f$ is the frame retry limit of the AC, W_0 is the minimum contention window size, W_i is the window size at backoff stage i (i.e. $W_i = W_0 \times 2^i$ for $i \in (0, M-1)$ and $W_i = W_M$ for $i \in (M, M+f-1)$), PT is the probability that the backoff counter is decreased by one in a given time slot, and PC is the collision probability at each transmission attempt.

The first equation in (1) accounts for the fact that, at the beginning of each slot time, the backoff counter is decreased by one with probability PT. The second equation says that the backoff counter is not successfully decreased with probability $1 - PT$ due to the effect of different AIFSs of different ACs. Recall that, in EDCA, different ACs may wait for intervals of different lengths (i.e., $AIFS[AC]$), before the channel can be accessed again or the backoff procedure is resumed. A lower priority AC may need to wait longer, because the wireless channel may be used again by a higher priority AC with a relatively shorter AIFS. This leads to that this low-priority AC is very likely to be frozen again before it can successfully finish waiting for its AIFS. The third equation corresponds to

TABLE I

NOTATIONS USED IN THE ANALYSIS

Notation	Definition
M_i	Maximum number of times for doubling the contention window size after a transmission failure for access class i
$W_{0,i}$	Minimum contention window size for access class i
$W_{k,i}$	Contention window size at backoff stage k for access class i
$M_i + f_i$	Frame retry limit of class i for an EDCA station, i.e. retry counter, for access class i
PC_i	Collision probability of class i for an EDCA station
PI_i	Probability that a class collides with higher priority access classes inside the EDCA station
PO	Probability that class i collides with other EDCA stations while accessing the channel
$PIDLE_i$	Probability that class i senses the channel is idle during backoff state in the Markov Chain model
$diff_i$	Number of time slots between AIFSmin and AIFSi for access class i
PT_i	Probability that backoff counter can be decreased by one for access class i; this probability accounts for the effect of different AIFSs among access classes
τ_i	Probability of transmission attempt for access class i observed inside an EDCA station
σ_i	Probability of transmission attempt for access class i observed outside an EDCA station
N	Total number of EDCA stations in the system
PTR	Probability of at least one EDCA station transmits in the considered time slot
PS_i	Probability of a transmission of class i is successful in the considered time slot
PFC	Probability of a transmission attempt fails due to collisions in the considered time slot
S_i	Average saturation throughput of access class i in the system
$E[L_i]$	Average packet length of access class i
$aslotTime$	Length of one slot time
t_{s_i}	Average time of a successful transmission for access class i
t_c	Average time of a collision involved in the system
t_H	Transmission time of PHY and MAC headers
$t_{E[P_i]}$	Transmission time of the average payload size for class i
$t_{E[P_i^*]}$	Average time interval of the longest packet size involved in a collision in the system
t_{SIFS}	Time interval of SIFS
t_{AIFS_i}	Time interval of AIFS for access class i
t_{ACK}	Transmission time of an ACK massage
δ	Propagation delay
$B_{i,k}$	Random variable representing the backoff counter in stage k of access class i before transmission attempt
X_i	Random variable representing the total backoff counter of class i before the successful transmission or before reaching the frame retry limit
K_i	Average deferring time spent in every backoff decreasing attempt for access class i while current channel state is busy
Z_i	Average transition time of backoff counter decreasing successfully for access class i
AD_i	Access delay of a packet frame for class i
D_i	Total delay of a packet frame for class i, including the queueing delay, the channel access delay and the transmission delay
Y_i	Random variable representing the number of collisions with other EDCA stations while accessing the channel before the successful transmission for class i
Q_i	Waiting time spent in the queue of access class i before trying to access the wireless medium
T_i	Service time for access class i in a queueing system
F_i	Number of data frames in the queue of access class i
λ_i	Arrival rate of traffic in access class i
α_i	Variance of inter-arrival time of traffic in access class i

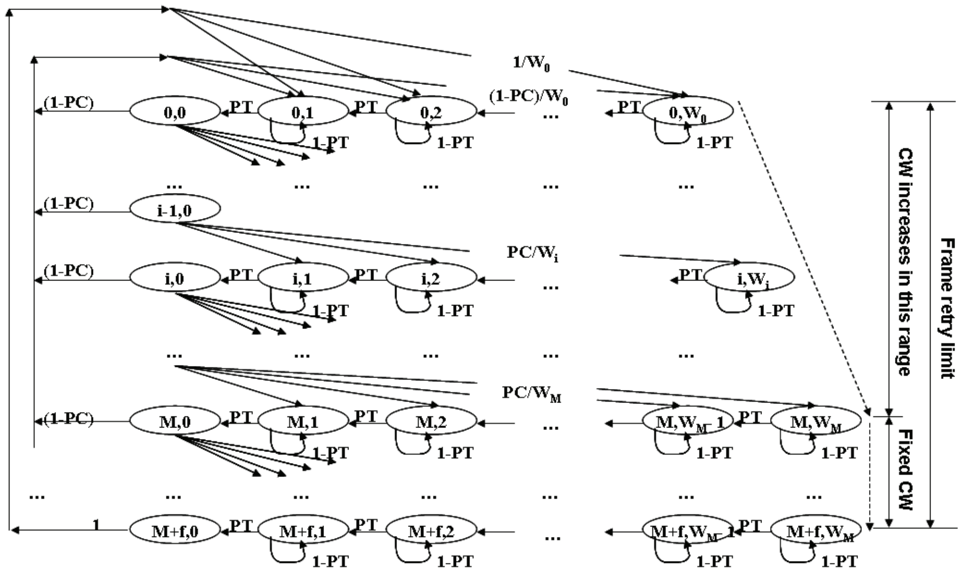

Fig. 2. Markov chain model for a Single AC inside the EDCA station.

when a new data frame starts at backoff stage 0, and thus the backoff counter is initially uniformly chosen in $(1, W_0)$. The fourth equation describes when an unsuccessful transmission occurs at backoff stage $i-1$, the backoff stage increases, and the new backoff value is uniformly chosen in $(1, W_i)$. Note that W_i will not be increased but fixed at W_M after M transmission retries. The last equation says that once the backoff stage reaches the retry limit $M+f$, the frame is discarded after transmission failure and the next data frame in the queue is served. Thus, the state will start another backoff procedure with probability one.

The stationary distribution of the states in this model is defined as $b_{i,k} = lim_{t\to\infty} P(s(t)=i, b(t)=k)$. From the transition probabilities in (1) and the fact that the sum of all states in the Markov model equals one, the limiting probability of state $b_{0,0}$ is obtained in (2). Since a transmission occurs whenever the backoff counter becomes zero, the transmission probability for an AC can be expressed by $\tau = \sum_{i=0}^{M+f} b_{i,0} = b_{0,0}\sum_{i=0}^{M+f} PC^i$, as in (3). Let τ_i denote the transmission probability of $AC[i]$ in an EDCA station, $i = 0,1,2,3$. See equation following (3) on top of next page, where PC_i is the collision probability of $AC[i]$, PT_i is the probability that the backoff counter for $AC[i]$ can be decreased by one, M_i is the maximum number of times $AC[i]$ can double its contention window, $M_i + f_i$ is $AC[i]$'s frame retry limit, and $W_{0,i}$ is the minimum contention window size of $AC[i]$.

Collisions may occur among different ACs in the same EDCA station (i.e., virtual collisions), and collisions may also take place among different EDCA stations (i.e., external collisions). Let PI_i denote the probability of virtual collisions for $AC[i]$, and PO be the probability of external collisions in the system. Hence, the collision probability of $AC[i]$ (i.e., PC_i) can be expressed by

$$PC_i = PI_i + (1 - PI_i)PO. \tag{4}$$

The probability of virtual collisions PI_i can be expressed as follows, considering that each AC will collide only with higher priority ACs in the same station.

$$\begin{cases} PI_0 = 0 \\ PI_1 = \tau_0 \\ PI_2 = 1-(1-\tau_0)(1-\tau_1) \\ PI_3 = 1-(1-\tau_0)(1-\tau_1)(1-\tau_2) \end{cases} \tag{5}$$

Let σ_i, $i = 0,1,2,3$, denote the transmission probability of $AC[i]$ for an EDCA station. Thus,

$$\begin{cases} \sigma_0 = \tau_0 \\ \sigma_1 = \tau_1(1-PI_1) = \tau_1(1-\tau_0) \\ \sigma_2 = \tau_1(1-PI_2) = \tau_2(1-\tau_0)(1-\tau_1) \\ \sigma_3 = \tau_1(1-PI_3) = \tau_3(1-\tau_0)(1-\tau_1)(1-\tau_2) \end{cases} \tag{6}$$

and the total transmission probability for an EDCA station is $\sigma_{total} = \sum_{i=0}^{3}\sigma_i$. With σ_{total}, the probability of external collisions PO can be expressed by

$$PO = 1 - (1-\sigma_{total})^{N-1}, \tag{7}$$

where N is the total number of EDCA stations in the system.

After waiting for the wireless medium to become idle for $AIFS_i$, $AC[i]$ will start decreasing its counter value. Let PT_i denote the probability that the backoff counter of $AC[i]$ can be successfully decreased by one in a given time slot, i.e., when there are no transmissions initiated by other stations or other higher priority ACs inside the same station in the period between $AIFS_{min}$ (i.e., DIFS) and $AIFS_i$. Note that, since $AC[i]$ will freeze its counter value until the channel has been idle for $AIFS_i$, PT_i will be zero before $AC[i]$ starts to count down. Since we attempt to understand the behavior of the saturated channel condition, we assume that each $AC[i]$ in all wireless stations can perceive the same collision probability as described in [12]. Let $diff_i$ denote the differences in the number of time slots between $AIFS_{min}$ and $AIFS_i$, i.e., $diff_i = \frac{AIFS_i - AIFS_{min}}{aslotTime} \cong \frac{AIFS_i - DIFS}{aslotTime}$. The relationship between PT_i, $AIFS_i$, and $diff_i$ is illustrated in Fig. 3. With $diff_i$, PT_i can be expressed by

$$\begin{cases} PT_0 = 0, \ before\ AIFS_0; \\ PT_0 = 1, \ after\ AIFS_0. \end{cases}$$

$$b_{0,0} = \frac{2PT}{\frac{(1+2PT)\times(1-PC^{M+f+1})+W_0\times(2PC)^M\times(1-PC^{f+1})}{1-PC} + \frac{W_0\times(1-(2PC)^M)}{1-2PC}} \quad (2)$$

$$\tau = \frac{2PT\times(1-PC^{M+f+1})}{(1+2PC)\times(1-PC^{M+f+1})+W_0\times(2PC)^M\times(1-PC^{f+1}-\frac{1-PC}{1-2PC})+\frac{1-PC}{1-2PC}} \quad (3)$$

$$\tau_i = \frac{2PT_i\times(1-PC_i^{M_i+f_i+1})}{(1+2PC_i)\times(1-PC_i^{M_i+f_i+1})+W_{0,i}\times(2PC_i)^{M_i}\times(1-PC_i^{f_i+1}-\frac{1-PC_i}{1-2PC_i})+\frac{1-PC_i}{1-2PC_i}}$$

$$\begin{cases} PT_1=0,\ before\ AIFS_1; \\ PT_1=((1-\tau_0)(1-\sigma_0)^{N-1})^{diff_1-diff_0},\ after\ AIFS_1. \end{cases}$$
$$\begin{cases} PT_2=0,\ before\ AIFS_2; \\ PT_2=((1-\tau_0)(1-\sigma_0)^{N-1})^{diff_1-diff_0}\times \\ (\prod_{i=0}^{1}(1-\tau_i)(1-\sigma_i)^{N-1})^{diff_2-diff_1},\ after\ AIFS_2. \end{cases}$$
$$\begin{cases} PT_3=0,\ before\ AIFS_3; \\ PT_3=((1-\tau_0)(1-\sigma_0)^{N-1})^{diff_1-diff_0}\times \\ (\prod_{i=0}^{1}(1-\tau_i)(1-\sigma_i)^{N-1})^{diff_2-diff_1}\times \\ (\prod_{i=0}^{2}(1-\tau_i)(1-\sigma_i)^{N-1})^{diff_3-diff_2},\ after\ AIFS_3. \end{cases} \quad (8)$$

Based on PI_i, PT_i, and PO, both τ_i and σ_i, $i = 0,1,2,3$, can be derived accordingly with numerical techniques.

We are now ready to derive the saturation throughput for each AC in the system. Let PTR denote the probability that at least one station transmits in the considered time slot, PS_i be the probability that a transmission attempt of $AC[i]$ is successful given that there is at least one station transmitting in the considered time slot, and PFC denote the probability that a transmission attempt fails due to a collision given that there is at least one station transmitting in the considered time slot. By definition,

$$\begin{cases} PTR = 1-(1-\sigma_{total})^N; \\ PS_i = \frac{N\times\sigma_i\times(1-\sigma_{total})^{N-1}}{PTR}, i=0,1,2,3; \\ PFC = \frac{1-(1-\sigma_{total})^N - N\times\sigma_{total}\times(1-\sigma_{total})^{N-1}}{PTR}. \end{cases} \quad (9)$$

Let S_i denote the average throughput of $AC[i]$ in the system. Thus,

$$S_i = \frac{PS_i\times PTR\times E[L_i]}{(1-PTR)\times aslotTime+\sum_{j=0}^{3} PTR\times PS_j\times t_{s_j}+PTR\times PFC\times t_c} = \frac{E[L_i]}{\frac{1-\sigma_{total}}{N\times\sigma_i}\times aslotTime+\sum_{j=0}^{3}\frac{\sigma_j}{\sigma_i}\times t_{s_j}+\frac{PFC}{PS_i}\times t_c}, \quad (10)$$

where $E[L_i]$ is the mean packet size of $AC[i]$, and $\frac{PFC}{PS_i} = \frac{1-(1-\sigma_{total})^N-N\times\sigma_{total}\times(1-\sigma_{total})^{N-1}}{N\times\sigma_i\times(1-\sigma_{total})^{N-1}}$. The first term of the denominator in (10) corresponds to an idle slot, the second term accounts for a successful transmission, and the third term is for a collision. The expressions t_c and t_{s_i} for $AC[i]$ can be derived based on two access modes.

A. Basic Access Mode

$$\begin{cases} t_{s_i} = t_H + t_{E[P_i]} + t_{SIFS} + \delta + t_{ACK} + t_{AIFS_i} + \delta \\ t_c = t_H + t_{E[P_i^*]} + t_{AIFS_i} + \delta \end{cases} \quad (11)$$

where t_H is the transmission time periods of the frame header, t_{SIFS} and t_{AIFS_i} are the time periods of an SIFS and an AIFS, respectively, t_{ACK} is the transmission time of an ACK frame, $t_{E[P_i]}$ is the transmission time of the average payload for $AC[i]$, $t_{E[P_i^*]}$ is the transmission time of the largest payload involved in a collision, and δ is the propagation delay.

B. RTS/CTS Access Mode

$$\begin{cases} t_{s_i} = t_{RTS} + t_{SIFS} + \delta + t_{CTS} + t_{SIFS} + \delta \\ +t_H + t_{E[P_i]} + t_{SIFS} + \delta + t_{ACK} + t_{AIFS_i} + \delta \\ t_c = t_{RTS} + t_{AIFS_i} + \delta \end{cases} \quad (12)$$

Note that t_c is the same for every AC in the RTS/CTS mode since only RTS frames are involved in collisions. Substituting t_s and t_c into (10), we obtain the average throughput for each AC accordingly.

III. Access Delay and Jitter Analysis of 802.11e EDCA

In this section, we analyze the access delay and the delay jitter for channel access in 802.11e EDCA. To simplify the analysis, we assume no data frame is discarded due to exceeding the frame retry limit as in [15].

Let $B_{i,k}$, $k \in (0, M_i+f_i)$, denote the random variable representing the backoff counter at stage k for $AC[i]$ before a transmission attempt, and is uniformly distributed in $(1, W_{i,k})$. The expectation of $B_{i,k}$ can be expressed by $E[B_{i,k}] = \sum_{j=1}^{W_{i,k}} \frac{j}{W_{i,k}} = \frac{W_{i,k}+1}{2}$. Let X_i be the random variable representing the total backoff counter of a successful transmission for $AC[i]$. Since $B_{i,k}, k \in (0, M_i+f_i)$, are mutually independent random variables, X_i can be expressed by $X_i = \sum_{k=0}^{M_i+f_i} B_{i,k}\times(PC_i)^k$. The expectation of X_i then can be expressed by $E[X_i] = \sum_{k=0}^{M_i+f_i} E[B_{i,k}]\times(PC_i)^k$. Substituting $E[B_{i,k}]$ into the expectation and the variance of X_i, we obtain

$$E[X_i] = \sum_{k=0}^{M_i-1}\frac{W_{i,k}+1}{2}\times {PC_i}^k + \sum_{k=M_i}^{M_i+f_i}\frac{W_{i,M_i}+1}{2} \times {PC_i}^k = \frac{W_{0,i}}{2}\times[\frac{1-(2PC_i)^{M_i}}{1-2PC_i} + \frac{(2PC_i)^{M_i}\times(1-PC_i^{f_i+1})}{1-PC_i}] + \frac{1-PC_i^{M_i+f_i+1}}{2(1-PC_i)}. \quad (13)$$

Since X_i is the number of decrements in the backoff counter for $AC[i]$ before a successful transmission, from Fig. 2 we can observe that the behavior of going through several backoff stages (due to transmission failures) until first success may be approximated by a geometric random variable. To simplify

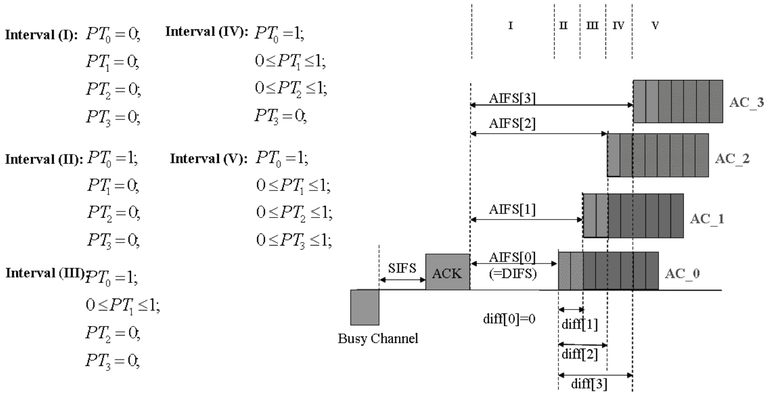

Fig. 3. Illustration of PT_i, $AIFS_i$, and $diff_i$: An example.

the calculation, we assume that X_i follows the geometric distribution to derive the variance. Thus,

$$Var[X_i] = E[X_i](E[X_i]-1). \tag{14}$$

This assumption will be verified in the simulation section.

Let $PIDLE_i$ represent the probability that $AC[i]$ senses the channel idle, i.e., the probability that no other EDCA station or other AC in the same EDCA station is currently using the channel. Thus, $PIDLE_i = (1-\sigma_{total})^{N-1}\prod_{j\neq i}(1-\tau_j)$. Let K_i be the time spent in every backoff attempt for $AC[i]$ given that the current channel state is busy. Thus,

$$\begin{aligned} K_i \cong & \textstyle\sum_{n=0}^{diff_i-1} n\times aslotTime\times(1-PIDLE_i)\times(PIDLE_i)^n + \\ & \textstyle\sum_{j=0}^{3} PS_j\times t_{s_j} + PFC\times t_c \\ = & [\frac{PIDLE_i\times(1-PIDLE_i^{diff_i-1})}{1-PIDLE_i} - (diff_i-1)\times PIDLE_i^{diff_i}] \\ & \times aslotTime + \textstyle\sum_{j=0}^{3} PS_j\times t_{s_j} + PFC\times t_c, \end{aligned} \tag{15}$$

where t_{s_i} is the average time of a successful transmission for $AC[i]$, and t_c is the average time of a collision. Note that a lower priority AC needs to wait longer than those ACs of higher priorities, thus the channel may become busy after the low-priority AC has waited longer than a DIFS but shorter than its AIFS. The first term in (15) is the time duration when $AC[i]$ tries to finish waiting for its AIFS but still fails due to the channel used by other high-priority ACs. The second (respectively, the third) term corresponds to the "frozen time" caused by the channel busy period used by other ACs due to a successful transmission (respectively, a collision).

Let Z_i denote the average transition time of a decrease in the backoff counter for $AC[i]$. Thus, Z_i can be expressed with K_i as follows.

$$\begin{aligned} Z_i \cong & (1-PTR)\times aslotTime + PTR\times[(AIFS_i - AIFS_{min}) + \\ & \textstyle\sum_{j=0}^{3} PS_j\times t_{s_j} + PFC\times t_c + \sum_{n=1}^{\infty} n\times K_i\times PT_i\times(1-PT_i)^n] \\ = & (1-PTR)\times aslotTime + PTR\times[(AIFS_i - AIFS_{min}) + \\ & \textstyle\sum_{j=0}^{3} PS_j\times t_{s_j} + PFC\times t_c + \frac{K_i\times(1-PT_i)}{PT_i}]. \end{aligned} \tag{16}$$

The first part in (16) corresponds to the situation that the channel is idle before the transition, and thus the state transition time takes only one $aslotTime$. The second part describes the situation that the channel was previously occupied and just being released. Thus every AC either resumes the backoff countdown or starts a new backoff procedure after waiting for its AIFS. An AC may fail to finish its AIFS several times before a success, i.e. counted down to one, as shown in Fig. 4. Thus the last term in the second part (i.e., $\sum_{n=1}^{\infty} n\times K_i\times PT_i\times(1-PT_i)^n$) accounts for the total time for an AC to successfully finish waiting for its AIFS.

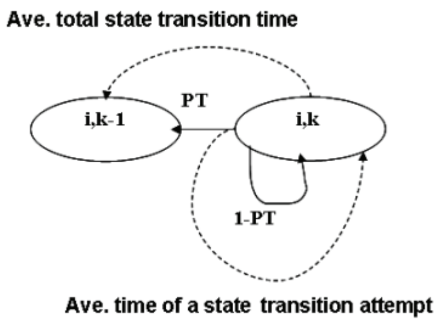

Fig. 4. Illustration of average state transition time.

The access delay AD_i for $AC[i]$ then can be expressed by

$$AD_i = X_i\times Z_i + Y_i\times\frac{(1-PI_i)\times PO}{PC_i}\times t_c. \tag{17}$$

The first term in (17) corresponds to the total time spent in the backoff procedure. The second term represents the total time due to external collisions. Y_i of the second term represents the number of collisions before a successful transmission for $AC[i]$. According to the Markov model in Fig. 2, we obtain

$$\begin{aligned} E[Y_i] &= \textstyle\sum_{k=0}^{M_i+f_i-1} k\times(1-PC_i)(PC_i)^k; \\ E[Y_i^2] &= \textstyle\sum_{k=0}^{M_i+f_i-1} k^2\times(1-PC_i)(PC_i)^k; \\ Var[Y_i] &= E[Y_i^2]-E^2[Y_i]. \end{aligned}$$

Recall that there are two kinds of collisions for each AC. One is virtual collisions with higher priority ACs in the same EDCA station, with probability PI_i. The other is external collisions with other EDCA stations for channel accessing, with probability $(1-PI_i)\times PO$. Assume that the scheduler entity of the EDCA station resolves virtual collisions within negligible time. Thus approximately only $\frac{(1-PI_i)\times PO}{PC_i}$ of those collisions are external collisions.

Taking expectation and standard deviation of AD_i, we obtain the average delay and the jitter as $E[AD_i]$ and $\sqrt{Var[AD_i]}$, respectively, i.e.,

$$E[AD_i] \cong E[X_i]\times Z_i + E[Y_i]\times\frac{(1-PI_i)\times PO}{PC_i}\times t_c; \tag{18}$$

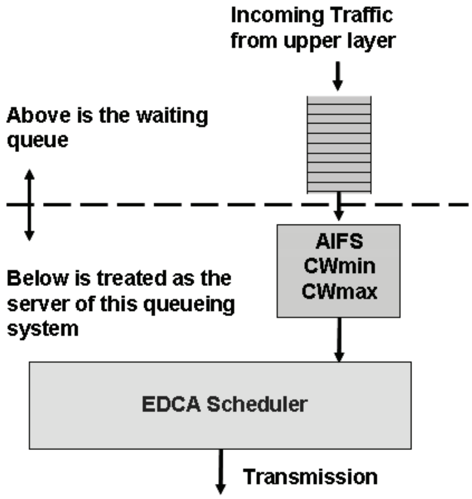

Fig. 5. Illustration of treating each AC as a queueing system.

$$Var[AD_i] \cong Var[X_i] \times Z_i^2 + Var[Y_i] \times (\frac{(1-PI_i) \times PO}{PC_i} \times t_c)^2. \quad (19)$$

We can also derive the mean waiting time in the queue for each data received from the upper layer to each AC, as shown in Fig. 5. Based on (18), we can express the end-to-end delay of a packet transmission for $AC[i]$ by $E[D_i] \cong E[Q_i] + E[X_i] \times Z_i + E[Y_i] \times \frac{(1-PI_i)\times PO}{PC_i} \times t_C + t_{s_i}$, where Q_i is the waiting time spent in the queue of $AC[i]$ before the packet tries to access the wireless medium, and t_{s_i} corresponds to the average time of a successful data transmission for $AC[i]$. To derive $E[Q_i]$, we consider two arrival traffic patterns.

A. Poisson Arrival Traffic

Each AC in an EDCA station can be modeled as an M/G/1 queueing system if the traffic into the waiting queue follows Poisson arrivals. Thus, the backoff procedure along with transmission mechanism of each AC is the server of the M/G/1 system. The mean and the variance of the service time T_i in the queueing system then can be expressed as below.

$$E[T_i] \cong E[X_i] \times Z_i + E[Y_i] \times \frac{(1-PI_i)\times PO}{PC_i} \times t_c + t_{s_i}\ ; \quad Var[T_i] \cong Var[X_i] \times Z_i^2 + Var[Y_i] \times (\frac{(1-PI_i)\times PO}{PC_i} \times t_c)^2. \quad (20)$$

With the P-K formula for an M/G/1 system in [16] and [17], the average waiting time in the queue $E[Q_i]$ can be expressed by (21), where λ_i is the Poisson arrival rate of traffic to $AC[i]$. Here we assume that λ_i is large enough so that $AC[i]$ is always backlogged.

$$E[Q_i] = \frac{\lambda_i \times E[T_i^2]}{2(1-\lambda_i E[T_i])} = \frac{\lambda_i (Var[T_i] + E^2[T_i])}{2(1-\lambda_i \times E[T_i])}. \quad (21)$$

The expected number of data frames $E[F_i]$ in the queue of $AC[i]$ can be derived by

$$E[F_i] = \lambda_i \times E[Q_i] = \frac{\lambda_i^2 (Var[T_i] + E^2[T_i])}{2(1-\lambda_i \times E[T_i])}.$$

With average data frame size $E[L_i]$, the average queue size of $AC[i]$ is expressed by $E[L_i] \times E[F_i]$.

B. Non-Poisson Arrival Traffic

For non-Poisson arrivals, each AC can be modeled as a G/G/1 system, from which the upper and lower bounds of average waiting time spent in the queue can still be derived. Let λ_i be arrival rate and α_i be the variance of inter-arrival time for the traffic entering $AC[i]$. Here we assume that λ_i is large enough such that $AC[i]$ is always backlogged. With the upper and lower bounds of a G/G/1 system, the range of the average waiting time in the queue of $AC[i]$ can be derived with these two characteristic parameters of arrival traffics: (λ_i, α_i).

$$\frac{\lambda_i Var[T_i] + E[T_i](\lambda_i E[T_i] - 2)}{2(1-\lambda_i E[T_i])} \le E[Q_i] \le \frac{\lambda_i(\alpha_i + Var[T_i])}{2(1-\lambda_i E[T_i])} \quad (22)$$

Note that $\lambda_i \times E[T_i]$ is the server utilization of a G/G/1 system. When $\lambda_i \times E[T_i]$ approaches one, i.e. when the traffic load is heavy or the wireless medium is very busy, the waiting time in the queue will grow dramatically. From (22), the range of the mean end-to-end delay $E[D_i]$ can also be derived. Then the expected range of the number of data frames $E[F_i]$ in the queue of $AC[i]$ can be expressed by

$$\frac{\lambda_i^2 Var[T_i] + E[T_i](\lambda_i^2 E[T_i] - 2\lambda_i)}{2(1-\lambda_i E[T_i])} \le E[F_i] \le \frac{\lambda_i^2(\alpha_i + Var[T_i])}{2(1-\lambda_i E[T_i])}.$$

Again, with the average data frame size $E[L_i]$, the average queue size of $AC[i]$ can be expressed by $E[L_i] \times E[F_i]$.

For constant bit rate traffic, such as voice and media streaming, the characteristic parameters of those arrival traffic should simply be $(\lambda_i, \alpha_i = 0)$. Thus, (22) can further be re-expressed by

$$\begin{cases} A - B \le E[Q_i] \le A; \\ A = \frac{\lambda_i \times Var[T_i]}{2(1-\lambda_i \times E[T_i])}; \\ B = \frac{E[T_i] \times (2-\lambda_i \times E[T_i])}{2(1-\lambda_i \times E[T_i])} = E[T_i] \times (1 + \frac{\lambda_i \times E[T_i]}{2(1-\lambda_i \times E[T_i])}). \end{cases} \quad (23)$$

Note that $\lambda_i \times E[T_i]$ is the utilization of the server. In order to make the system stable, the value of $\lambda_i \times E[T_i]$ must be confined between 0 and 1. Here 0 corresponds to an idle server, and 1 means that the server is fully utilized. Apparently, when $\lambda_i \times E[T_i]$ is close to 1, $E[T_i] \times (1 + \frac{\lambda_i \times E[T_i]}{2(1-\lambda_i \times E[T_i])}) \cong E[T_i]$, which means B is small as compared to the total delay spent in the queue. Therefore, when the traffic load is very heavy, $E[Q_i] \cong A = \frac{\lambda_i \times Var[T_i]}{2(1-\lambda_i \times E[T_i])}$ for constant bit rate traffic with arrival rate λ_i. This result can be very useful, since each AC is always busy with sending frames in our previous assumption, i.e. the utilization of each AC is close to 1.

IV. Performance Evaluation

In this section, the analytical model derived in the previous section is validated with the ns-2 EDCA implementation created by G. Chesson and A. Singla [18]. The data frames of all ACs are fixed at 256 bytes, and the frame retry limit is fixed at 7. For each AC, CBR UDP traffic is generated at a rate of 1Mbps. Each station keeps sending packets to the AP to achieve the saturation throughput. There is no traffic going between any two neighboring stations, and no traffic from an AP to other stations except ACK frames. The simulation parameters are listed in Table II. We simulate two access modes: basic mode and RTS/CTS mode. In each mode, we measure the following items: (i) Aggregate throughput for all ACs: the overall throughput of each AC in the system. (ii)

TABLE II

PARAMETER SETTINGS USED IN THE SIMULATION

General Parameters	Value
Data Transmission Rate	24Mbps
Control Message Transmission Rate	6Mbps
Bytes per OFDM symbol	27
An idle slot time	9 μsec
SIFS time	16 μsec
DIFS=SIFS+2 slot	34 μsec
Propagation delay	1 μsec
Maximum station number	25
RTS frame length	20 octets
CTS frame length	14 octets
Data payload length	256 octets
ACK frame length	14 octets
MAC sublayer overhead	28 octets
PHY layer overhead	20 μsec
OFDM symbol interval	4 μsec
CBR sending interval of each AC	1 $msec$
$AC_{0,1,2,3}$ retry limit	7
First Set of EDCA parameters	**Value**
$AC_0(CW_{min}, CW_{max})$	(15,31)
$AFIS_0$	SIFS + 2 slots
$AC_1(CW_{min}, CW_{max})$	(31,63)
$AFIS_1$	SIFS + 3 slots
$AC_2(CW_{min}, CW_{max})$	(31,127)
$AFIS_2$	SIFS + 4 slots
$AC_3(CW_{min}, CW_{max})$	(63,255)
$AFIS_3$	SIFS + 4 slots
Second Set of EDCA parameters	**Value**
$AC_0(CW_{min}, CW_{max})$	(31,63)
$AFIS_0$	SIFS + 2 slots
$AC_1(CW_{min}, CW_{max})$	(63,127)
$AFIS_1$	SIFS + 2 slots
$AC_2(CW_{min}, CW_{max})$	(127,255)
$AFIS_2$	SIFS + 3 slots
$AC_3(CW_{min}, CW_{max})$	(127,255)
$AFIS_3$	SIFS + 7 slots

MAC-to-MAC delay of each AC: the delay from the MAC layer of the sender to the MAC layer of the receiver. (iii) MAC-to-MAC jitter of each AC: the standard deviation of MAC-to-MAC delay of each AC. (iv) End-to-end delay of each AC: the delay from the UDP layer of the sender to the UDP of the receiver, including both queueing delay and the MAC-to-MAC delay.

Figs. 6 to 9 plot the analytical curves (solid lines) and the simulation curves (dashed lines) based on the two sets of different EDCA parameter combinations described in Table II. In each figure, the sub-figure(a) represents the results obtained from the first set of parameter settings, and the sub-figure(b) represents the results from the second set of parameter settings. We can see that all the analytical curves match the simulation ones very well. Thus, the total backoff counter of a successful transmission for an AC, which is assumed to be a geometric distribution (eq (14)), can be verified. Note that due to space limitations, we only include the results of the RTS/CTS mode. The results of the basic access mode are consistent with those included in this paper.

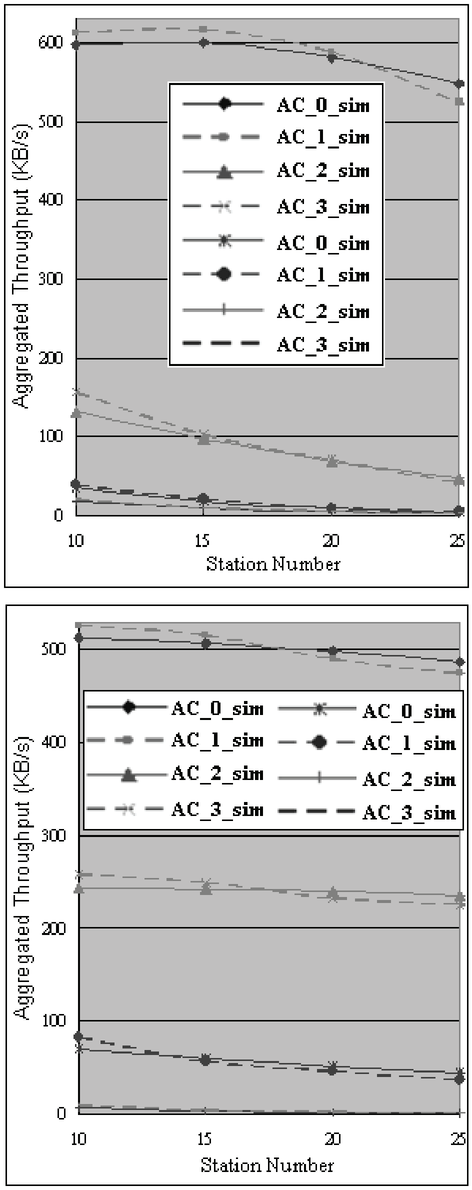

Fig. 6. Aggregate throughput of each AC from first parameter settings (left) and second parameter settings (right).

V. CONCLUSION

In this paper, we have developed the performance of EDCA in terms of the saturation throughput, the delay, and the jitter for IEEE 802.11e WLANs. The contributions of this paper are summarized as follows. (i) We consider virtual collisions among different ACs inside each EDCA station in addition to external collisions among stations. Thus, our model can capture the behavior of 802.11e EDCA more accurately than existing work. (ii) We consider the impact of high-priority ACs' AIFSs on low priority ones. Thus, the probability that the backoff counter of each AC can be decreased by one in each time slot may not be identical or equal to one. (iii) We model the effect of the frame retry limit, i.e., the maximum number of retransmissions each frame can experience before being dropped. After this limit is reached, the retrying data frame is discarded. (iv) In addition to modeling the saturation throughput as in existing work, we also model the delay and delay jitter performance. Simulations based on ns-2 are conducted to verify the analytical model. The results show that the analytical curves match the simulation results very well, proving that our analytical model is very accurate in describing the behavior of IEEE 802.11e EDCA.

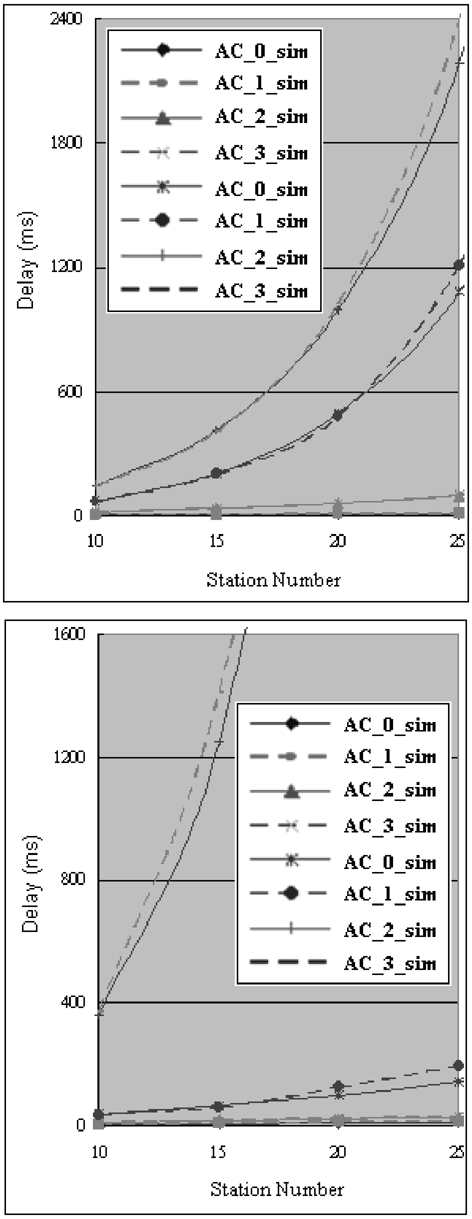

Fig. 7. MAC-to-MAC delay of each AC from first parameter settings (left) and second parameter settings (right).

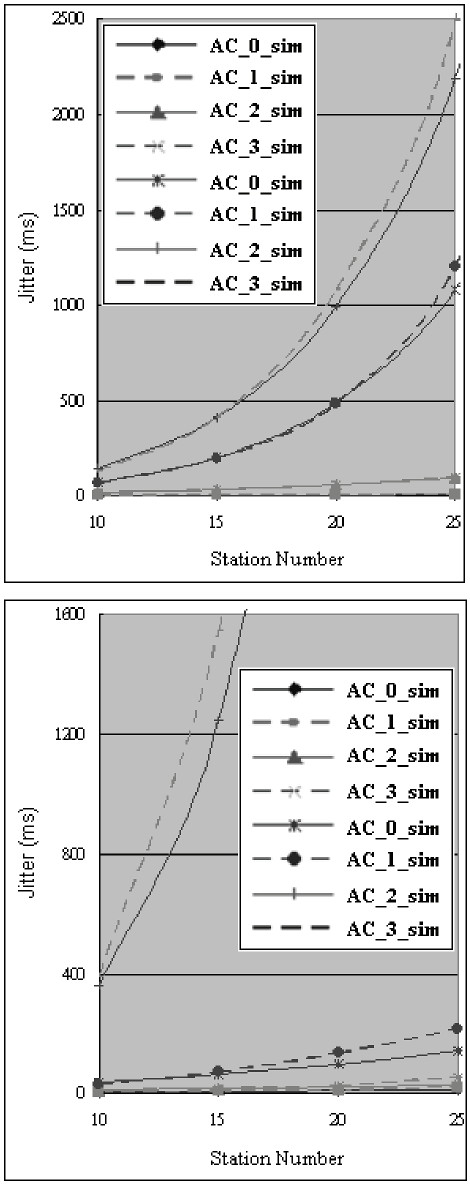

Fig. 8. MAC-to-MAC jitter of each AC from first parameter settings (left) and second parameter settings (right).

REFERENCES

[1] IEEE 802.11, "Wireless LAN medium access control (MAC) and physical layer (PHY) specifications," IEEE Standard, Aug. 1999.

[2] IEEE 802.11 WG, IEEE 802.11e/D5.0, "Draft supplement to standard for telecommunications and information exchange between systems-LAN/MAN specific requirements-Part 11: Wireless medium access control (MAC) and physical layer (PHY) specifications: Medium access control (MAC) enhancements for quality of service (QoS)," Aug. 2003.

[3] A. Grilo and M. Nunes, "Performance evaluation of IEEE 802.11e," in *Proc. IEEE PIMRC*, Sep. 2002, vol. 1, pp. 511-517.

[4] D. He and C. Q. Shen, "Simulation study of IEEE 802.11e EDCF," in *Proc. IEEE VTC-Spring*, Apr. 2003, vol. 1, pp. 685-689.

[5] S. Choi, J. Pavon, S. N. Shankar, and S. Mangold, "IEEE 802.11e EDCF performance evaluation," in *Proc. IEEE ICC*, May 2003, vol. 2, pp. 1151-1156.

[6] J. Pavon and S. N. Shankar, "Impact of frame size, number of stations, and mobility on the throughput performance of IEEE 802.11e," *IEEE WCNC*, Mar. 2004, vol. 2, pp. 2789-2795.

[7] H. Zhu and I. Chlamtac, "An analytical model for IEEE 802.11e EDCF differential services," in *Proc. ICCCN*, Oct. 2003, pp. 163-168.

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

[9] P. Ferre, A. Doufexi, A. Nix, and D. Bull, "Throughput analysis of IEEE 802.11 and IEEE 802.11e MAC," in *Proc. IEEE WCNC*, Mar. 2004, vol. 2, pp. 783-788.

[10] X. Yang, "Enhanced DCF of IEEE 802.11e to support QoS," in *Proc. IEEE WCNC*, 2003.

[11] S. Mangold, S. Choi, G. R. Hiertz, O. Klein, and B. Walke, "Analysis of IEEE 802.11e for QoS support in wireless LANs," *IEEE Wireless Commun. Mag.*, vol. 10, pp. 40-50, Dec. 2003.

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

[13] J. Gosteau, M. Kamoun, S. Simoens, and P. Pellati, "Analytical developments on QoS enhancements provided by IEEE 802.11 EDCA," in *Proc. IEEE ICC*, June 2004, vol. 7, pp. 4197-4201.

[14] P. Chatzimisios, A. C. Boucouvalas, and V. Vitsas, "Performance analysis of IEEE 802.11 DCF in presence of rransmission errors," in *Proc. IEEE ICC*, June 2004, vol. 7, pp. 3854-3858.

[15] ———, "IEEE 802.11 packet delay-A finite retry limit analysis," in *Proc. IEEE Globecom*, Dec. 2003, vol. 2, pp. 950-954.

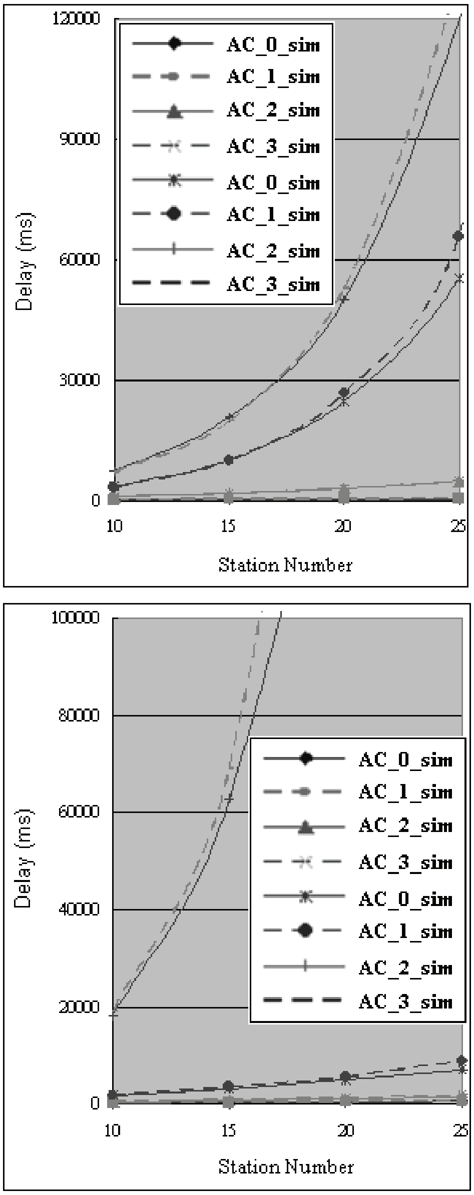

Fig. 9. End-to-End delay of each AC from first parameter settings (left) and second parameter settings (right).

[16] D. Bertsekas and R. Gallager, *Data Networks*. Upper Saddle River, NJ: Prentice Hall PTR, 1992.

[17] D. Gross and C. M. Harris, *Fundamentals of Queueing Theory*. New York: John Wiley and Sons, Inc, 1998.

[18] ns-2. [Online.] Available: `http://www-sop.inria.fr/planete/qni/Research.html`

[19] G. Bianchi, I. Tinnirello, and L. Scalia, "Understanding 802.11e contention-based prioritization mechanisms and their coexistence with legacy 802.11 stations," *IEEE Network*, vol. 19, no. 4, pp. 28-34, July/Aug. 2005.

Ching-Ling Huang received his BS ('99) and MS ('04) degrees from Electrical Engineering, National Taiwan University. From 2005, Ching-Ling is working at CCIT(California Center for Innovative Transporation) on Berkeley campus. His current projects includes BHL(Berkeley Highway Lab) video traffic monitoring, the WiMax backhaul network, HICOMP(Highway Congestion Monitoring Program) software architecture design, and Travel Time estimation on CMS(Changeable Message Signs). Ching-Ling is going to pursuit his PhD degree at Civil System, U.C. Berkeley from Fall'06. His research areas cover wireless communication, vehicular network, VII(Vehicle-Infrastrcutrue Integration), and realted ITS (Intellient Transportation System) issues.

Wanjiun Liao received her BS and MS degrees from National Chiao Tung University, Taiwan, in 1990 and 1992, respectively, and the Ph.D. degree in Electrical Engineering from the University of Southern California, Los Angeles, California, USA, in 1997. She joined the Department of Electrical Engineering, National Taiwan University (NTU), Taipei, Taiwan, as an Assistant Professor in 1997. Since August 2005, she has been a full Professor. Her research interests include wireless networks, multimedia networks, and broadband access networks.

Dr. Liao is actively involved in the international research community. She is currently an Associate Editor of IIEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and IEEE TRANSACTIONS ON MULTIMEDIA. She was a tutorial co-chair of IEEE INFOCOM 2004, and the TPC vice chair of IEEE Globecom 2005 (Autonomous Network Symposium). Dr. Liao has received many research awards. Two papers she co-authored with her students received the Best Student Paper Award of the First IEEE International Conferences on Multimedia and Expo (ICME) in 2000, and the Best Paper Award of the First International Conference on Communication, Circuits and Systems in 2002. She was a recipient of K. T. Li Young Research Award honored by ACM Taipei chapter 2003for her research achievement. Dr. Liao was elected as one of Ten Distinguished Young Women in Taiwan in 2000 and was listed in the Marquis Who's Who in Science and Engineering, and the Contemporary Who's Who in 2003.