# A Unified Model for the Performance Analysis of IEEE 802.11e EDCA

Jie Hui, *Student Member, IEEE,* and Michael Devetsikiotis, *Senior Member, IEEE*

***Abstract*—Rapid deployment of IEEE 802.11 wireless local area networks (WLANs) and their increasing quality of service (QoS) requirements motivate extensive performance evaluations of the upcoming 802.11e QoS-aware enhanced distributed coordination function (EDCA). Most of the analytical studies up-to-date have been based on one of the three major performance models in legacy distributed coordination function analysis, requiring a large degree of complexity in solving multidimensional Markov chains. Here, we expose the common guiding principle behind these three seemingly different models. Subsequently, by abstracting, unifying, and extending this common principle, we propose a new unified performance model and analysis method to study the saturation throughput and delay performance of EDCA, under the assumption of a finite number of stations and ideal channel conditions in a single-hop WLAN. This unified model combines the strengths of all three models, and thus, is easy to understand and apply; on the other hand, it helps increase the understanding of the existing performance analysis. Despite its appealing simplicity, our unified model and analysis are validated very well by simulation results. Ultimately, by means of the proposed model, we are able to precisely evaluate the differentiation effects of EDCA parameters on WLAN performance in very broad settings, a feature which is essential for network design.**

***Index Terms*—Enhanced distributed coordination function (EDCA), IEEE 802.11, performance analysis, quality of service (QoS), wireless local area network (WLAN).**

## NOMENCLATURE

| | |
|---|---|
| AC | Access category. |
| AP | Access point. |
| ACK | Acknowledgment. |
| AIFS | Arbitration Inter Frame Space (802.11e). |
| AIFSN | Arbitration Inter Frame Space Number (802.11e). |
| BEB | Binary exponential backoff. |
| BSP | Backoff subperiods. |
| CA | Collision avoidance. |
| CD | Collision detection. |
| CI | Confidence interval. |
| CSMA | Carrier-sense multiple access. |
| CTS | Clear to send. |
| CW | Contention window. |
| CWmin | Contention window minimum. |
| CWmax | Contention window maximum. |
| DCF | Distributed coordination function. |
| EB | Exponential backoff. |
| EIFS | Extended interframe space. |
| EDCA | Enhanced DCF (802.11e). |
| HOL | Head of line. |
| LAN | Local area network. |
| MAC | Medium access control. |
| QAP | QoS access point. |
| QoS | Quality of service. |
| RTS | Request to send. |
| SIFS | Short interframe space. |
| TC | Traffic category (802.11e). |
| TXOP | Transmission opportunity. |
| UP | User priority. |
| WLAN | Wireless local area network. |
| $c_{ij}$ | Slot collision probability of a station with AC $= 4-i$ in BSP$_j$. |
| $d_i$ | AIFS $[4-i]$ in unit of slots. |
| $\Delta_j$ | Difference between $d_{j+1}$ and $d_j$. |
| $D$ | Delay period. |
| $l_j$ | Conditional average backoff delay in BSP$_j$. |
| $m_i$ | Maximum backoff stage of a station with AC $= 4-i$. |
| $n_i$ | Number of stations with AC $= 4-i$. |
| $p_{ij}$ | Slot transmission probability of a station with AC $= 4-i$ in BSP$_j$. |
| $P$ | Slot transmission probability matrix. |
| $p_{\mathrm{tr}_j}$ | System slot transmission probability in BSP$_j$. |
| $P_{\mathrm{tr}}$ | System slot transmission probability matrix. |
| $p_{\Delta_j}$ | Probability that transmission begins in BSP$_j$. |
| $P_\Delta$ | System BSP transmission probability matrix |
| $P_s$ | System successful transmission probability. |
| $p_{\mathrm{suc}_{ij}}$ | Successful transmission probability of a station with AC $= 4-i$ in BSP$_j$. |
| $p_{s_i}$ | Successful transmission probability of a station with AC $= 4-i$. |
| $S$ | Saturation throughput of the system. |
| $S_i$ | Saturation throughput of a station with AC $= 4-i$. |
| $T$ | Payload transmission time in unit of slots. |
| $T_c$ | Collision time in unit of slots. |
| TP | Transmission period. |
| $T_s$ | Successful transmission time in unit of slots. |
| $U$ | Successful payload transmission time in a cycle. |
| $W_i$ | CWmin $[4-i]+1$. |

Paper approved by M. Zorzi, the Editor for Multiple Access of the IEEE Communications Society. Manuscript received February 13, 2004; revised August 19, 2004 and February 2, 2005. This work was supported in part by the Center for Advanced Computing and Communication under Grant 04-08. This paper was presented in part at the Global Telecommunications Conference (Globecom), Dallas, TX, December 2004 [1].

The authors are with the Electrical and Computer Engineering Department, North Carolina State University, Raleigh, NC 27695 USA (e-mail: jhui@ncsu.ed; mdevets@ncsu.edu).

Digital Object Identifier 10.1109/TCOMM.2005.855013

## I. INTRODUCTION AND MOTIVATION

THROUGHPUT and delay analysis of contention-based random multiple-access techniques, especially CSMA and its variations, has long been a research focus in packet networks since the 1970s. Accompanying the standardization and rapid deployment of IEEE 802.11 WLANs in the 1990s, the performance analysis of its contention-based DCF MAC access function [2], a CSMA with CA (CSMA/CA) scheme with slotted BEB, has been studied extensively by analytical or numerical means in recent years. Among those analytical studies, three major performance models have been proposed in parallel in order to analyze the saturation throughput/capacity performance. Assuming a constant collision probability for each station, Bianchi [3], [4] proposed a Markov chain to approximately model the behavior of CSMS/CA/BEB DCF, found the equilibrium packet-transmission probability in a generic slot time by solving the Markov chain, and finally obtained the saturation throughput by applying regenerative analysis to a generic slot time; Calì [5], [6] analyzed a $p$-persistent variant of DCF, with persistence factor $p$ derived from the CW in DCF, then found the capacity similarly using renewal theory; Tay [7] used instead an *average value* mathematical model in order to calculate the packet-collision probability, and solved the maximum throughput in terms of collision probability. A variation of Bianchi's model was proposed by Wu *et al.* in [8] for the further consideration of retry limits.

Driven by the rapid growth of WLAN traffic volume and the different needs of applications, the IEEE 802.11 Task Group E has been working for several years to enhance the current best-effort 802.11 MAC to support a QoS-aware WLAN. EDCA, one of the main and mandatory schemes in 802.11e [9], parameterizes the DCF CSMA/CA scheme with prioritized EB to achieve differentiated QoS. In recent years, the performance of EDCA has been explored by means of not only simulation [10]–[18], but also analytical evaluations [18]–[29]. Most of the EDCA analytical studies are based on the modifications of DCF analysis mentioned above: the work in [18] and [22] extends and parameterizes Calì's $p$-persistent DCF to accommodate different classes; with the exception of [21], the others all modify or extend Bianchi's Markov chain model [4] to accommodate the differentiation of AIFS and/or CW. [19], [23], [26], and [27] analyze the differentiation effects of only CW, while in the others, the differentiation effects of both AIFS and CW are considered. [25] and [24] vary transition rates on top of the original Markov chains, while [20] and [29] enlarge the original bidimensional Markov chain to tridimensional, and [28] enlarges it even to multidimensional. Other than multidimensional Markov chains, [21] provides a new rigorous analytical approach to model AIFS-based priority mechanisms, but also with the weakness of high complexity.

To achieve more successful embedding of QoS-aware MAC in network schemes, such as call-admission control and scheduling schemes, a performance model/analysis needs to be easier to understand and apply. The EDCA analysis mentioned above, which can evaluate differentiation effects of all EDCA parameters, all require high complexity.

Motivated by this need to *simplify* and to *unify*, we thoroughly reexamine the foundations of these EDCA analyses, including Bianchi's seminal Markov model, Calì's $p$-persistent CSMA model, and Tay's average-value model for DCF analysis. We establish and expose a common guiding principle behind these seemingly different models: all of them assume saturation traffic conditions and homogeneous slots accessed with the same probability. The probability can be either the constant transmission probability (Calì) or constant collision probability (Bianchi and Tay), and they can be converted to each other. For example, the persistence factor $p$ in Calì's model derived from CW using an iterative algorithm, actually is the transmission probability calculated from Bianchi's Markov chain model, and can be converted from the collision probability in Tay's model, which was solved by average-value analysis. In other words, both Bianchi's and Tay's models imply a constant transmission probability, which agrees with the constant transmission probability resulting from a geometrically distributed backoff interval in Calì's model. This homogeneity, in turn, guarantees the application of regenerative analysis.

The differences among these methods consist of rather technical side issues, concerning diverse naming for transmission probability, different methods of finding the probabilities, and varied choices of renewal cycles. For instance, the renewal cycle in Calì's model was explicitly marked to be the time between two adjacent successful transmissions; implicitly, Bianchi picked a generic slot, and Tay chose time between two transmissions, to be their renewal cycles. However, in the common assumption of slot homogeneity, these different renewal cycles are all valid, and can be transformed to each other.

Keeping these commonalities in mind, in this paper, we borrow strengths from these three models and compose a new unified performance model for EDCA analysis, without involving large complexity. We still use renewal theory to formulate the throughput performance, but assume constant packet-transmission probabilities for different stations in different periods of time to account for differentiation effects of both AIFS and CWmin/CWmax in EDCA. The secondary assumption is that each packet-transmission probability only depends on a unique collision probability. These transmission probabilities consist of a *2-D* persistence-factor matrix, resulting in a $p$-persistent-like CSMA/CA performance model. We denote this set of transmission probabilities by matrix $\mathbf{P}$, and thus name the unified model the ***generalized*** $\mathbf{P}$***-persistent CSMA/CA model*** (called in the following $\mathbf{P}$-persistent, for short). In other words, we adopt a $p$-persistent-like system in which persistence factors are time-dependent, while all of the other previous models used static persistence factors.

In order to solve for those transmission probabilities or the persistence factors in matrix $\mathbf{P}$, we provide extensions to both Bianchi's Markov chain analysis and Tay's mean value analysis. This unified model, on one hand, reduces the complexity of Markov chains and is easy to apply; on the other hand, it increases the understanding of several efforts in the past on saturation throughput analysis of 802.11 MAC, and allows better understanding of the system behavior by exploiting the time-dependent persistence factors. Finally, the accuracy of the model and the analysis is well validated by simulation results.

The remainder of this paper is organized as follows. First, we briefly review DCF and EDCA mechanisms in Section II. Second, in Section III, we propose our unified performance model, a $\mathbf{P}$-persistent CSMA/CA, for EDCA. How to derive $\mathbf{P}$, the key factor of the model, is illustrated in Section III-C. Applying regenerative analysis to this model with knowledge of $\mathbf{P}$, we calculate the saturation throughput performance and delay performance of 802.11e EDCA separately in Sections IV and V, in the assumption of a finite number of stations and ideal

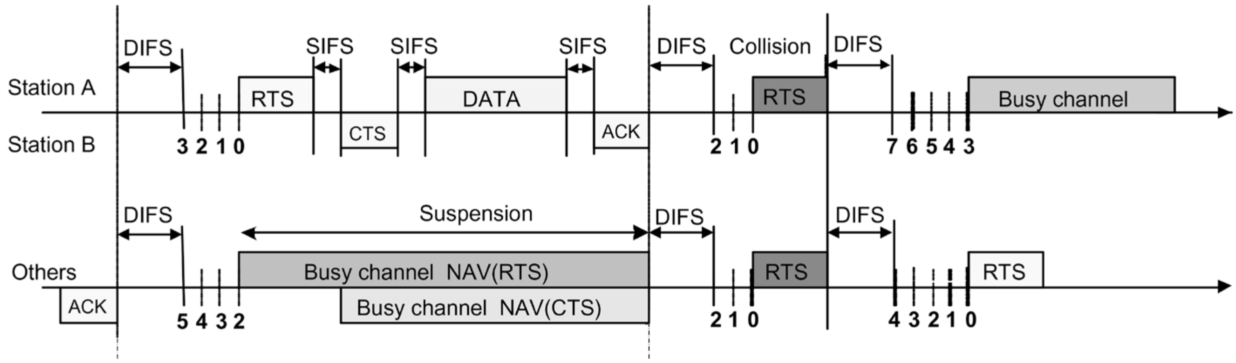



Fig. 1. DCF access procedure.

channel conditions in a single-hop WLAN. We then establish the accuracy of this model by simulation results, and study the differentiation effects of EDCA parameters in Section VII. Finally, in Section VIII, we present our conclusions and describe our future directions.

## II. Protocol Description of DCF and EDCA

A legacy DCF WS performs CSMA/CA with the following BEB procedures [2] to access the wireless medium (Fig. 1).

**Defer:** A station with a pending packet has to wait for the channel to be idle for the duration of a different interframe space (DIFS) before the transmission, in order to give priority access to polling or control messages. If the channel is sensed busy during this period, the station has to wait for another idle DIFS after the channel is idle again, and then performs a random backoff. **Backoff:** Then, the station has to wait for an additional random backoff time, which is uniformly distributed between 0 and CWmin slots. If the channel is sensed to be busy during this period, the station suspends backoff until the channel is idle for DIFS again. **Handshaking:** If the packet size is bigger than a threshold, a two-way handshaking procedure is performed to further reduce the data collision probability, including an RTS and a CTS packet. **Data Transmission:** If the above procedures are successful, the data packet will be transmitted. **Confirmation:** Then, the station awaits an ACK from the destination for confirmation. **Collision and Retransmission:** If more than one station begins their transmissions at the same time, collision happens. The collided station will defer, backoff, and retransmit with a new CW size $(\text{CW}_{\text{new}} = \text{CW}_{\text{old}} * 2 + 1)$ until CWmax is reached, and then stays unchanged at CWmax. If the retransmission attempts reach a retry limit, the packet will be discarded.

To provide differentiated channel access, EDCA [9] supports up to four access categories (ACs) in one QoS station for packets belonging to eight user priorities (UPs) or frame types. AC values of 0, 1, 2, and 3 represent best effort, background, video, and voice AC, respectively. The mapping between UP or frame type and AC can be found in [9].

Compared with the equal access of DCF contentions by using the same DIFS, CWmin, and CWmax, EDCA offers differentiated access through EDCA parameter set AIFS[AC], CWmin[AC], CWmax[AC], and TXOPlimit[AC] for a corresponding AC (AC $= 0, 1, 2, 3$). AIFS[AC] is determined by AIFS[AC] = SIFS + AIFSN[AC]$*T_{\text{slot}}$, where AIFSN[AC] is an integer indicating the number of slots after a SIFS duration a station should defer before either invoking a backoff or starting a transmission. The minimum value for AIFSN is two for backoff entities of non-AP stations and one for back entities of QAP. TXOP is a new scheme to improve the efficiency of the protocol. A backoff entity can transmit multiple packets within one TXOP, of which the maximum length is TXOPlimit[AC].

## III. Unified Performance Model

### A. Assumptions and Configurations

*1) Assumptions:* According to the latest draft of the 802.11e standard, we make the following assumptions about the EDCA random-access system which is going to be analyzed in the following section.

- Finite number of stations contend in a single-hop network.
- Each station has one access entity which belongs to one of the four ACs, and the number of stations deploying an AC is $N$[AC].
- Ideal channel conditions: Propagation delay is zero, there is no channel error, no hidden node problem, and no capture effects.
- Synchronized and slotted system: The time immediately following a busy medium and an idle AIFS[AC] is slotted, and a station is allowed only to transmit at the beginning of each slot time $(T_{\text{slot}})$. Here we assume the slot boundaries in all stations are synchronized. This assumption, together with ideal channel conditions, implies a synchronous start of frame transmissions.
- Saturation traffic: All stations always have constantly backlogged queues.
- Constant successful transmission time $T_s$ slots, including constant payload transmission time $T$ slots,[1] overhead, and control-message time: $T =$ PacketSize/DataRate$/T_{\text{slot}}$, $T_s = ($PHY and MAC header + RTS + CTS + ACK +3* SIFS$)/T_{\text{slot}} + T$.
- Constant collision time $T_c =$ RTS$/T_{\text{slot}}$ slots. In this paper, we assume RTS/CTS access mode. However, the analysis can be easily applied to the basic access mechanism without RTS/CTS with the modification of $T_c$.
- TXOPlimit[AC] $= 0$ for all ACs. We assume only one data frame is transmitted per EDCA TXOP.
- Infinite retry limit for a frame after collision, as assumed in [4].[2]

[1]Not affecting the validity of the model, we assume a constant packet payload size in this paper; otherwise, $T$ will be a function of the average payload size.

[2]The inclusion of a finite retry limit in the model is straightforward [8] and is not discussed in this paper.

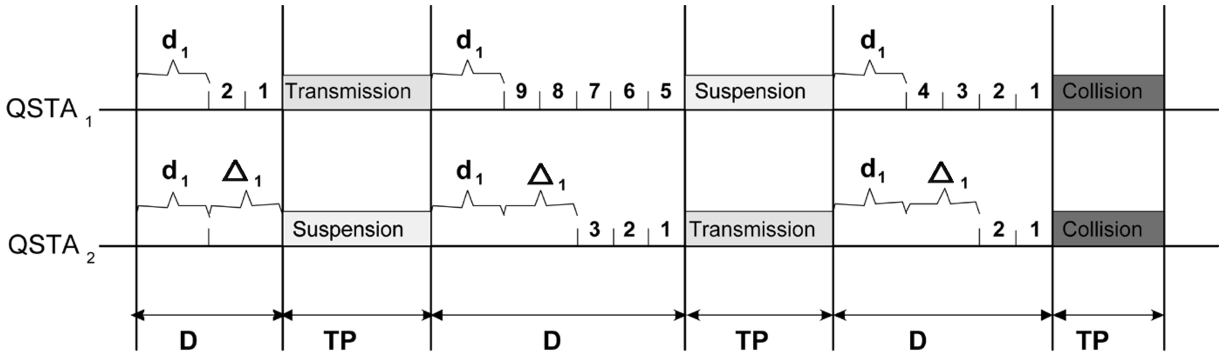



Fig. 2. EDCA channel states with two stations ($d_1 = 2, d_2 = 4$).

*2) EDCA Parameter Sets:* We also define the priority EDCA parameter set for a corresponding AC as AIFS[AC], CWmin[AC], and CWmax[AC]. For the ease of mathematic expressions in this paper, we transform the standard EDCA parameters into the following: $d_i = \text{AIFS}\,[4-i]/T_{\text{slot}}, W_i = \text{CWmin}[4-i]+1, m_i = \log_2^{(\text{CWmax}[4-i]+1)/(\text{CWmin}[4-i]+1)}$, and $n_i = N[4-i]$.

In other words, there are $n_1$ stations using voice AC (AC $= 3), \ldots, n_4$ stations using best-effort AC (AC $= 0$). $d_i$ can be interpreted as the length of AIFS$[4-i]$ measured in slots. From [9], we know AIFS[0] $\geq$ AIFS[1] $\geq$ AIFS[2] $\geq$ AIFS[3]. Therefore, $d_1 \leq d_2 \leq d_3 \leq d_4$. $W_i$ is just an even number larger than the CWmin size by one, and $m_i$ is the maximum backoff stage for AC $= 4-i$.

*3) Backoff Range:* There are different conventions regarding the inclusion of the bounds in the range of the backoff slots. The uncertainties and changes in 802.11e drafts during the standardization process also reflect this.

In this paper, we assume the range to be [1,CW[AC]] inclusive. The reason why the backoff counter starts from one is because of the backoff suspension during a busy channel. For example, two backoff entities A and B contend for channel access. Entity A initiates a frame exchange at a particular slot, and then B will defer from channel access upon detecting channel busy and suspend the decreasing of its backoff counter. After transmission, Entity A will randomly pick up a new backoff slot from zero to CW, and B will resume its backoff function upon detecting channel idle again. Before A can transmit again, B has to count down at least one more slot. This means the minimum backoff slot for A should not be zero. Therefore, a lower bound of one is a must for the backoff range. Reference [18] also explains why the backoff counter should start from one instead of zero.

### B. Unified Performance Model

If several EDCA stations contend for a radio channel with the configurations that we assumed above, we will observe on the time axis an alternate sequence of idle periods (consisting of defer time and backoff slots) and TPs (successful or unsuccessful). An idle period or a delay period $(D)$ and a following TP compose a cycle. An example of a channel state with two stations contending for it is shown in Fig. 2. To study the throughput and delay performance, as most of the classical CSMA analysis did, we need to find out the average length of the delay period, the TP, and the useful message-transmission time using regenerative analysis.

However, the packet scheduling governed by EDCA BEB is not memoryless, since it depends on the transmission history (e.g., how many retransmissions the HOL packet has suffered). Therefore, those cycles are not, strictly speaking, *regenerative* cycles, and the average length of the delay periods and TPs, which depend on the backoff algorithm, cannot be calculated easily.

Since we cannot directly study the performance of EDCA system, we are going to construct a performance model to approximate the behavior of the EDCA backoff algorithm, and then study the performance of this model instead. If the model is sufficiently appropriate based on good approximations, we should find good consistency between real system performances (from simulation) and model performances (from analysis).

Next, we construct a performance model which abstracts, unifies, and extends the common guiding principle behind the three previous DCF performance models by assuming constant transmission probabilities, varying according to different stations and differing periods of time. This key assumption, together with the memoryless packet scheduling regulated by this ***unified model*** as follows, makes the renewal analysis of throughput and delay performance possible.

- First, we divide the possible random delay period into four BSPs, as shown in Fig. 3. The $j$th BSP is defined as the period of time between AIFS$[4-j]$ and AIFS$[3-j]$ for $j = 1, 2, 3$; and the fourth BSP is defined as the period of time greater than AIFS[4]. Thus, the length of BSP$_j$ under the unit of slots is $\Delta_j = d_{j+1} - dj$ for $j = 1, 2, 3$.
- Second, we define the *slot transmission probability* $p_{ij}$ as the transmission probability of a station belonging to AC$(4-i)$ in a slot boundary within the BSP$_j$, i.e., when the channel is idle, all $n_i$ stations of AC$(4-i)$ will transmit in a slot within $\Delta_j$ with probability $p_{ij}$, or postpone the transmission by one slot with probability $1 - p_{ij}$, where $i = 1, 2, 3, 4$, and $j = 1, 2, 3, 4$.
- Third, there are two possibilities for transmission. If a station transmits and succeeds, the other stations have to wait for $T_s$ slots until the transmission finishes, and then repeat the same procedures. Otherwise, if more than one station tries to grab the channel in the same slot, a collision happens. All stations have to wait for $T_c$ slots until the collision is detected, and then repeat the above contention procedures.

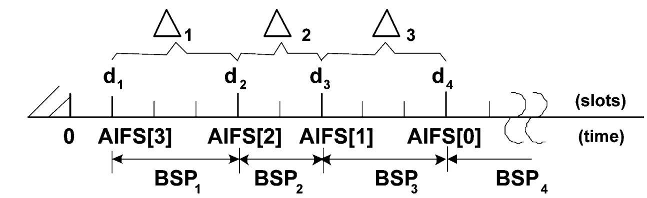



Fig. 3. Random delay model.

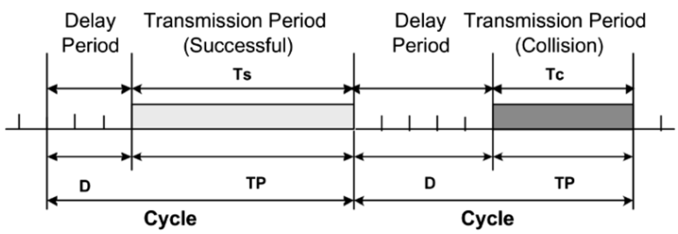



Fig. 4. Channel model.

We call this model the generalized $\mathbf{P}$-persistent CSMA/CA, because all of the slot-transmission probabilities $(p_{ij})$ can actually be seen as persistence factors, assuming each station is using a classical $p$-persistent CSMA/CA. In this case, the persistence factors vary in different time periods, also according to different ACs. We insert all the persistence factors into a matrix, and name it the *slot-transmission probability matrix* $\mathbf{P}$. As we can see, $\mathbf{P}$ is a $4 \times 4$ upper triangular matrix, since $i$ and $j$ are both from 1 to 4, and stations of AC$(4-i)$ cannot transmit in BSP$_j$

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ 0 & p_{22} & p_{23} & p_{24} \\ 0 & 0 & p_{33} & p_{34} \\ 0 & 0 & 0 & p_{44} \end{bmatrix}.$$

In this model, the packet scheduling is *memoryless*, because the transmission probabilities are independent, and there are always $\sum_{i=1}^{4} n_i$ packets waiting in the beginning of a cycle (saturation traffic condition). Therefore, the time in which a transmission ends are *renewal points*. Fig. 4 shows the renewal cycles of the channel states. We denote by random variable $D$ the duration of the random delay period,[3] and by random variable TP the duration of a TP, which varies according to the failure or the success of the transmission.

Using the classical renewal property, we can calculate the saturation throughput $S$ as the ratio of expected successful transmission time over the expected cycle length as in

$$S = \frac{E(U)}{E(D) + E(\mathrm{TP})} \tag{1}$$

where $U$ represents the successful payload-transmission time during a cycle. Furthermore, access delay can be directly derived from throughput using the relationship AccessDelay $= T/S$.

## C. Calculation of $\mathbf{P}$

In order to calculate the throughput and delay performance for the model, we first need to find the key factors, namely, the slot-transmission probabilities in matrix $\mathbf{P}$ from the EDCA backoff algorithm. We use the following three steps to calculate them.

*1) Represent Slot-Collision Probabilities $c_{ij}$ as a Function of $p_{ij}$:* A transmission happens with probability $p_{ij}$ for a station of AC$(4-i)$ during a slot in BSP$_j$. It may succeed or collide. The slot-collision probability $c_{ij}$ is one minus the probability that all other stations do not transmit, including the other $(n_i - 1)$ stations of the same AC, and all the other stations of different ACs

$$c_{ij} = 1 - \frac{\prod_{k=1}^{4} (1 - p_{kj})^{n_k}}{1 - p_{ij}}, \quad i, j = 1, 2, 3, 4. \tag{2}$$

*2) Represent Slot-Transmission Probabilities $p_{ij}$ as a Function of $c_{ij}$:* Besides the key assumption, we further assume that the $c_{ij}$'s are constant and independent of the backoff stage (secondary assumption), as assumed in other EB analyses [4], [7], [27], [30]. Thus, we can represent $p_{ij}$ as a function of a unique $c_{ij}$ using two methods here: mean value analysis, extended from [7], and Markov chain analysis, extended from [4].

*a) Mean Value Analysis:* Given collision probabilities $c_{ij}$, the number of transmissions for a station with AC$= 4-i$ to transmit a packet successfully in BSP$_j$ is geometrically distributed with parameter $(1 - c_{ij})$. When the backoff stage is $r$, the CW size is updated to be $2^r * W_i - 1$ for $r = 0, 1, \ldots, m_i$. Since we only consider the contentions in BSP$_j$, which begins from the end of BSP$_{j-1}$, the average backoff slots in BSP$_j$ is $(2^r * W_i - \sum_{k=1}^{j-1} \Delta_k / 2)$, where $\sum_{k=1}^{j-1} \Delta_k$ represents the summation of all the past BSPs.

Subsequently, we can calculate the expectation of the number of backoff slots for a station of AC$= 4-i$ in BSP$_j$ conditioning on backoff stage $r$

$$\begin{aligned} \overline{w_{ij}} &= (1 - c_{ij}) \frac{W_i - 1 - \sum_{k=1}^{j-1} \Delta_k}{2} \\ &\quad + + c_{ij}(1 - c_{ij}) \frac{2 * W_i - 1 - \sum_{k=1}^{j-1} \Delta_k}{2} + \cdots \\ &\quad + (c_{ij})^{m_i} (1 - c_{ij}) \frac{2^{m_i} * W_i - 1 - \sum_{k=1}^{j-1} \Delta_k}{2} \\ &\quad + (c_{i,j})^{m_i + 1} \frac{2^{m_i} * W_i - 1 - \sum_{k=1}^{j-1} \Delta_k}{2} \\ &= \frac{1 - c_{ij} - c_{ij} * (2 * c_{ij})^{m_i}}{1 - 2 * c_{ij}} * \frac{W_i}{2} - 1 + \sum_{k=1}^{\frac{j-1}{\Delta_k}} 2. \end{aligned}$$

Therefore $p_{ij}$, the slot-transmission probability of a station with AC$= 4-i$ in BSP$_j$, is the reciprocal of the average number of backoff slots plus one, as in (3), where $i = 1, \ldots, j$ and $j = 1, 2, 3, 4$. The other stations with AC values less than $4-j$ transmit in this $\Delta_j$ period with a probability of zero, implying $p_{ij} = 0$, when $i > j$

$$\begin{aligned} p_{ij} &= \frac{1}{\overline{w_{ij}} + 1} \\ &= \frac{2}{W_i - \sum_{k=1}^{j-1} \Delta_k + c_{ij} * (W_i - 1) * \frac{1 - (2 * c_{ij})^{m_i}}{1 - 2 * c_{ij}}}. \end{aligned} \tag{3}$$

*b) Markov Chain Analysis:* For each station AC$= 4-i$ in BSP$_j$, we let $r_{ij}(t)$ represent the stochastic process of backoff stage $(0, \ldots, m_i)$, and let $b_{ij}(t)$ be the stochastic process of backoff timer $(1, 2, \ldots, W_i^{(m_i)} - 1)$ at time $t$. We can model $\{r_{ij}(t), b_{ij}(t)\}$ as a 2-D embedded Markov chain, as shown in

[3]The delay model is shown in Fig. 3.

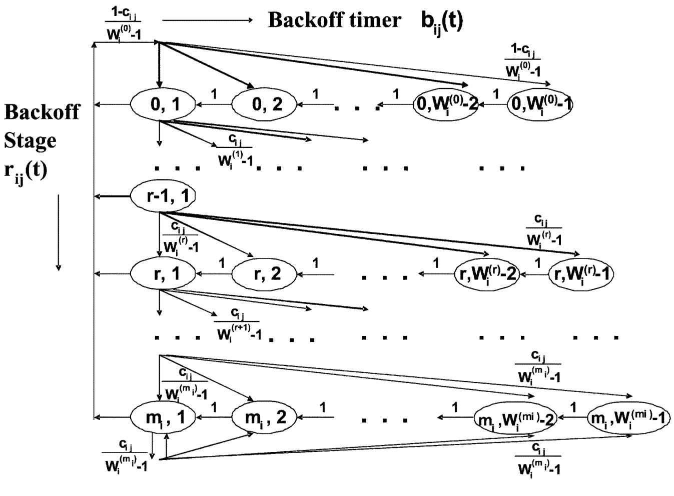



Fig. 5. Markov chain of $\{r_{ij}(t), b_{ij}(t)\}$ for a station AC= $4-i$ in BSP$_j$.

Fig. 5.[4] Similar to the previous method, stations are not using the original CW sizes, but the equivalent *smaller* CW sizes $W_i^{(r)}$, subtracting all *past* BSPs from the CWs, $W_i^{(r)} = 2^r * W_i - \sum_{k=1}^{j-1}\Delta_k$, for backoff stage $r = 0, 1, \ldots, m_i$.

In this Markov chain $\{r_{ij}(t), b_{ij}(t)\}$, the only non-null one-step transition probabilities are[5]

$$\begin{cases} P_{ij}\{r,k|r,k+1\} = 1, & r \in [0, m_i], \\ & k \in \left[1, W_i^{(r)} - 2\right] \\ P_{ij}\{0,k|r,1\} = \frac{1-c_{ij}}{W_i^{(0)}-1}, & r \in [0, m_i], \\ & k \in \left[1, W_i^{(r)} - 1\right] \\ P_{ij}\{r,k|r-1,1\} = \frac{c_{ij}}{W_i^{(r)}-1}, & r \in [1, m_i], \\ & k \in \left[1, W_i^{(r)} - 1\right] \\ P_{ij}\{m_i,k|m_i,1\} = \frac{c_{ij}}{W_i^{(m_i)}-1}, & k \in \left[1, W_i^{(m_i)} - 1\right]. \end{cases}$$

Whenever the backoff timer of a station is one, the station is going to transmit in the next slot. As stated in [27], for the consideration of backoff suspension stage and the consistency of Markovian states of different stations, we combine the backoff timer state 0 with state 1, and combine states before and during suspension to be a new state. Therefore, the slot-transmission probability $p_{ij}$ is just the summation of the stationary state probabilities of states with backoff timer value of one. Let $b_{ij}(r,k) = \lim_{t->\infty} P\{r_{ij}(t) = r, b_{ij}(t) = k\}$ be the stationary distribution of the chain. Therefore, for $i = 1, \ldots, j,\ j = 1, 2, 3, 4$, we have

$$\begin{aligned} p_{ij} &= \sum_{r=1}^{m_i} b_{ij}(r,1) \\ &= \frac{2}{W_i - \sum_{k=1}^{j-1}\Delta_k + c_{ij} * (W_i - 1) * \dfrac{1-(2*c_{ij})^{m_i}}{1-2*c_{ij}}}. \end{aligned} \tag{4}$$

[4]In total, there are ten Markov chains for different $i$'s and $j$s, since $i = 1, \ldots, j$ and $j = 1, 2, 3, 4$.

[5]We adopt the short notation $P_{ij}\{k_1, l_1|k_0, l_0\} = P_{ij}\{r_{ij}(t+1) = k_1, b_{ij}(t+1) = l_1 | r_{ij}(t) = k_0, b_{ij}(t) = l_0\}$.

As we can see, the result is the same as what was derived using the mean value analysis above. The derivation is similar to that in [4], and is thus not discussed here in detail.

*3) Solve $c_{ij}$ and $p_{ij}$:* From the last two steps, we know that $c_{ij}$ is a function of $n_i$, and $p_{1j}, \ldots, p_{jj}, p_{ij}$ is a function of $c_{ij}, W_i$, and $m_i$. Since $n_i, W_i$, and $m_i$ are all known, by solving the $2j$-dimensional nonlinear equations composed of equations of $c_{1j}, \ldots, c_{jj}$ from (2) and equations of $p_{1j}, \ldots, p_{jj}$ from (3) or (4), we can solve the values of $p_{1j}, p_{2j}, \ldots$, and $p_{jj}$. Repeating this procedure for all of the BSPs ($j = 1, 2, 3, 4$), we can obtain the transmission probability matrix $\mathbf{P}$.

## IV. Saturation Throughput Analysis

By knowing $\mathbf{P}$ and then applying probabilistic analysis to the $\mathbf{P}$-persistent CSMA/CA performance model, we can calculate the average random delay $E(D)$, the average successful transmission time $E(U)$, and the average TP $E(\text{TP})$, and then obtain the generalized saturation throughput performance for 802.11 EDCA.

### A. Average Random Delay $E(D)$

Before formulating $E(D)$, let us define two more transmission matrices besides $\mathbf{P}$.

*System Slot-Transmission Probability Matrix $P_{\text{tr}}$:* We already know that $p_{ij}$ represents the transmission probability of a *station* AC= $4-i$ at a *slot* boundary in BSP$_j$. For the whole system, it is possible that no one transmits in a slot. Then, we define the probability of the *system* to transmit at a *slot* boundary in BSP$_j$ as *system slot-transmission probability* $p_{\text{tr}_j}$. It is related to $p_{ij}$ by $p_{\text{tr}_j} = 1 - \prod_{i=1}^{j}(1-p_{ij})^{n_i}$. Integrating all $p_{\text{tr}_j}$ for $j = 1, 2, 3, 4$ into a matrix, we denote by $P_{\text{tr}}$ the *system slot-transmission probability matrix*

$$P_{\text{tr}} = [p_{\text{tr}_1}\ p_{\text{tr}_2}\ p_{\text{tr}_3}\ p_{\text{tr}_4}].$$

*System BSP Transmission-Probability Matrix $P_\Delta$:* A BSP consists of many slots. In each slot, the system can be in transmission state with probability $p_{\text{tr}_j}$. We define another matrix $P_\Delta$ as the *system BSP transmission-probability matrix*, $P_\Delta = [p_{\Delta_1}\ p_{\Delta_2}\ p_{\Delta_3}\ p_{\Delta_4}]$. The $j$th elements of the matrix $p_{\Delta_j}$ represent the probability that transmission begins in BSP$_j$. Thus, it is the product of the probability that no transmission happens in the past BSPs and the probability that at least one transmission happens in BSP$_j$

$$p_{\Delta_j} = [1-(1-p_{\text{tr}j})^{\Delta_j}] \prod_{k=1}^{j-1}(1-p_{\text{tr}k})^{\Delta_k}, \quad j = 1, 2, 3$$

$$p_{\Delta_4} = \prod_{k=1}^{3-1}(1-p_{\text{tr}k})^{\Delta_k}.$$

Then, we can calculate $E(D)$ using the following theorem.

*Theorem:* The average random delay can be computed by conditioning on the BSP in which transmission begins as

$$E(D) = d_1 + L \times P_\Delta^T = d_1 + \sum_{j=1}^{4} l_j \times p_{\Delta_j}$$

where $L$ is the *conditional average backoff delay matrix* $L = [l_1\ l_2\ l_3\ l_4]$, and $l_j$ represents the average backoff delay if transmission begins in BSP$_j$, with $l_j = (1/p_{\text{tr}_j}),\ j = 1, 2, 3, 4$.

*Proof:* The random delay $D$ is a random variable. It can take any value from the set $[d_1 + 1, \ldots, d_2, \ldots, d_3, \ldots, d_4, \ldots, \infty)$.[6] For given system slot-transmission probabilities, the tail-distribution function of $D$ can be expressed as

$$\Pr(k) = \Pr[D \geq k] = \begin{cases} (1-p_{\mathrm{tr}_1})^{k-d_1-1}, & d_1+1 \leq k \leq d_2 \\ (1-p_{\mathrm{tr}_2})^{k-d_2-1}(1-p_{\mathrm{tr}1})^{\Delta_1}, & d_2+1 \leq k \leq d_3 \\ (1-p_{\mathrm{tr}_3})^{k-d_3-1}\prod_{j=1}^{2}(1-p_{\mathrm{tr}j})^{\Delta_j}, & d_3+1 \leq k \leq d_4 \\ (1-p_{\mathrm{tr}_4})^{k-d_4-1}\prod_{j=1}^{3}(1-p_{\mathrm{tr}j})^{\Delta_j}, & d_4+1 \leq k < \infty. \end{cases}$$

Then, the expected delay can be represented as a function of $\Pr(k)$ as

$$\begin{aligned} E(D) &= \sum_{k=d_1+1}^{\infty} k[\Pr(k)(k) - \Pr(k+1)] \\ &= d_1 + 1 + \sum_{k=d_1+2}^{\infty} \Pr(k) \\ &= d_1 + \sum_{k=d_1+1}^{\infty} \Pr(k). \end{aligned}$$

By inserting the values of $\Pr(k)$, we obtain

$$\begin{aligned} E(D) = d_1 &+ \sum_{k=d_1+1}^{d_2} (1-p_{\mathrm{tr}_1})^{k-d_1-1} \\ &+ \sum_{k=d_2+1}^{d_3} (1-p_{\mathrm{tr}_2})^{k-d_2-1}(1-p_{\mathrm{tr}_1})^{\Delta_1} \\ &+ \sum_{k=d_3+1}^{d_4} (1-p_{\mathrm{tr}_3})^{k-d_3-1}\prod_{j=1}^{2}(1-p_{\mathrm{tr}_j})^{\Delta_j} \\ &+ \sum_{k=d_4+1}^{\infty} (1-p_{\mathrm{tr}_4})^{k-d_4-1}\prod_{j=1}^{3}(1-p_{\mathrm{tr}j})^{\Delta_j} \\ = d_1 &+ \frac{1}{p_{\mathrm{tr}_1}}[1-(1-p_{\mathrm{tr}1})^{\Delta_1}] \\ &+ \frac{1}{p_{\mathrm{tr}_2}}[1-(1-p_{\mathrm{tr}_2})^{\Delta_2}](1-p_{\mathrm{tr}_1})^{\Delta_1} \\ &+ \frac{1}{p_{\mathrm{tr}_3}}[1-(1-p_{\mathrm{tr}_3})^{\Delta_3}]\prod_{j=1}^{2}(1-p_{\mathrm{tr}_j})^{\Delta_j} \\ &+ \frac{1}{p_{\mathrm{tr}_4}}\prod_{j=1}^{3}(1-p_{\mathrm{tr}_j})^{\Delta_j} \\ = d_1 &+ \sum_{j=1}^{4} l_j \times p_{\Delta_j} = d_1 + L \times P_{\Delta}^{T} \end{aligned}$$

where $l_j, p_{\Delta_j}$ for $j = 1, 2, 3, 4$ are defined as above. ■

### B. Average Successful TP $E(U)$

A transmission can be successful or not. In order to calculate the average successful TP $E(U)$, we first need to determine the probability for a transmission to succeed.

[6] Since the minimum value of backoff slots is 1, as explained in Section III-A.3, the minimum random delay is $d_1 + 1$.

We define $P_s$ as the successful transmission probability for the whole system. That is, the probability for a certain transmission in a cycle to be successful. It is the summation of the successful transmission probabilities of each station $P_s = \sum_{i=1}^{4} n_i * p_{s_i}$, where we denote by $p_{s_i}$ the probability that a transmission is successful for one of the $n_i$ stations with AC equal to $4 - i$.

We introduce another quantity $p_{\mathrm{suc}_{ij}}$, which represents the probability that the transmission is successful for a station AC$= 4 - i$ in BSP$_j$, given that there is at least one transmission in BSP$_j$. It is related to $p_{ij}$ and $p_{\mathrm{tr}_j}$ by $p_{\mathrm{suc}_{ij}} = (p_{ij})/(1-p_{ij}) \times (1-p_{\mathrm{tr}_j})/(p_{\mathrm{tr}_j})$.

Similar to the previous section, we can calculate $p_{s_i}$ conditioning on the BSP in which the transmission begins: $p_{s_i} = \sum_{j=1}^{4} p_{\mathrm{suc}_{ij}} * p_{\Delta_j}$.

Thus, $E(U)$, the expected time during a cycle that the channel is used without a collision is the product of payload transmission time and the probability of a successful transmission

$$E(U) = P_s \times T.$$

### C. Mean TP $E(TP)$

By conditioning on the success of the transmission, we can calculate the mean TP as

$$E(\mathrm{TP}) = P_s T_s + (1 - P_s)T_c.$$

### D. Saturation Throughput

*System Saturation Throughput $S$:* Finally, we can obtain the total saturation throughput for the system by plugging $E(U), E(D)$ and $E(\mathrm{TP})$ into (1) as

$$S = \frac{P_s \times T}{d_1 + \sum_{j=1}^{4} l_j * p_{\Delta_j} + P_s T_s + (1-P_s)T_c}. \quad (5)$$

*Station Saturation Throughput $S_i$:* The saturation throughput for a station of AC$= 4 - i$ is the total throughput multiplied by $p_{s_i}/P_s$ for $j = 1, 2, \ldots, N$ to yield

$$S_i = \frac{p_{s_i} \times T}{d_1 + \sum_{j=1}^{4} l_j * p_{\Delta_j} + P_s T_s + (1-P_s)T_c}. \quad (6)$$

*Saturation Throughput Ratio:* The $n_i$ stations belonging to the same AC$(4 - i)$ receive the same amount of saturation throughput. The saturation throughput ratio among stations of different ACs can be expressed as the ratio of successful transmission probabilities

$$S_1 : S_2 : S_3 : S_4 = p_{s_1} : p_{s_2} : p_{s_3} : p_{s_4}.$$

## V. Access Delay Analysis

We define the access delay as the time duration from the packet becoming HOL on the sender's side, until it receives ACK from the receiver. The contributors to access delay include *medium access delay*, which is the time between the packet becoming HOL and its beginning transmission. It further includes the backoff time, deferring periods, and retransmission time due to collisions. The second contributor to delay is *successful transmission delay*, which is the transmission time of useful data and overhead.

The calculation of access delay is straightforward following the throughput analysis. For one of the $n_i$ stations with AC= $4-i(i=1,2,3,4)$ , we have

$$\text{AccDelay}_i = \frac{T_{\text{cycle}}}{p_{s_i}} = \frac{T}{S_i}. \quad (7)$$

The mean access delay is given by the packet transmission time $T$ divided by the throughput. Similarly, for the whole system, the mean access delay is given by

$$\text{AccDelay} = \frac{T_{\text{cycle}}}{P_s} = \frac{T}{S}. \quad (8)$$

## VI. Special Cases of EDCA

From the unified model, we can easily analyze the throughput of legacy 802.11 DCF and other special cases. The results totally agree with those in [4] and [27], in which the analysis is carried out in different ways.

### A. Legacy 802.11 DCF

In legacy 802.11 DCF, all stations use the same MAC parameters, including AIFS= $d * T_{\text{slot}}$, CWmin, and CWmax, meaning that there is only one AC and one BSP. If we assume that there are total $n$ stations and let $W = \text{CWmin} + 1$ and $m = \log_2^{(\text{CWmax}+1)/(\text{CWmin}+1)}$, in a slot within BSP, all $n$ stations transmit with a same probability $p$ and succeed with a same successful probability $p_{\text{suc}}$ to yield

$$\mathbf{P} = [p], \quad P_{\text{suc}} = [p_{\text{suc}}].$$

By solving the nonlinear equations, $p = (2)/(W + c*(W-1)*(1-(2*c)^m)/(1-2*c))$ and $c = 1-(1-p)^{n-1}, p$ can be found, and thus $p_{\text{suc}} = (p(1-p)^{n-1})/(1-(1-p)^n)$.

From $\mathbf{P}$, we can derive $P_{\text{tr}} = [p_{\text{tr}}]$, where $p_{\text{tr}} = 1-(1-p)^n; P_\Delta = [p_\Delta]$, $p_\Delta = 1$, and the average random delay $E(D) = d + (1)/(p_{\text{tr}})$.

Subsequently, each station receives the same saturation throughput. The normalized system saturation throughput is given by the equation shown at the bottom of the page, where $P_s = n * p_{\text{suc}} = (np(1-p)^{n-1})/(1-(1-p)^n)$. This expression is the same as that in [27], which was derived using the concept of generic slot introduced in [4].

### B. EDCA (Same AIFS, Different CW)

In this special case, only the CW sizes are used to differentiate the services for stations of different ACs, while the AIFS= $d * T_{\text{slot}}$ are kept the same for all stations, thus resulting in different $W_i = \text{CWmin}[4-i]+1$ and $m_i = \log_2^{(\text{CW}_{802\max}[4-i]+1)/(\text{CWmin}[4-i]+1)}$. We also assume there are $n_i$ stations for AC of $4-i$, where $i=1,2,3,4$. This differentiation scheme is very commonly used practically.

Stations only compete in $\text{BSP}_4$ with a length of $\Delta_4$, since $\Delta_1 = \Delta_2 = \Delta_3 = 0$. Similarly, only the fourth columns of matrixes $\mathbf{P}$ and $P_{\text{suc}}$ are nonzero with different values

$$\mathbf{P} = \begin{bmatrix} 0 & 0 & 0 & p_1 \\ 0 & 0 & 0 & p_2 \\ 0 & 0 & 0 & p_3 \\ 0 & 0 & 0 & p_4 \end{bmatrix}, \quad P_{\text{suc}} = \begin{bmatrix} 0 & 0 & 0 & p_{\text{suc}1} \\ 0 & 0 & 0 & p_{\text{suc}2} \\ 0 & 0 & 0 & p_{\text{suc}3} \\ 0 & 0 & 0 & p_{\text{suc}4} \end{bmatrix}.$$

From $\mathbf{P}$, we can derive $P_{\text{tr}} = [0\ 0\ 0\ p_{\text{tr}}]$, where $p_{\text{tr}} = 1 - \prod_{i=1}^4 (1-p_i)^{n_i}; P_\Delta = [0\ 0\ 0\ p_\Delta]$, where $p_\Delta = 1$. $p_1, p_2, p_3, p_4$ can be solved from the nonlinear $p_i = (2)/(W_i + c_i*(W_i-1)*(1-(2*c_i)^{m_i})/(1-2*c_i))$ and $c_i = 1 - (\prod_{k=1}^4 (1-p_k))/((1-p_i))$. The successful probability of a station of AC= $4-i$ can be computed as $p_{s_i} = (p_i/1-p_i) * (1-p_{\text{tr}}/p_{\text{tr}})$, and the total $P_s = \sum_{i=1}^4 n_i * p_{s_i}$.

Similarly, the average random delay and the total throughput take the same value as those in the general case. The saturation throughput ratio among stations of different ACs is expressed as $S_1 : S_2 : S-3 : S_4 = (p_1)/(1-p_1) : (p_2)/(1-p_2) : (p_3)/(1-p_3) : (p_4)/(1-p_4)$, also the same as derived in [27].

## VII. Simulation Validation and Discussion

### A. Simulation Model

We programmed a discrete-event simulation of a single-hop static WLAN. For each station, a traffic generator feeds traffic packets into a MAC queue. The packet stays in the MAC queue until it reaches HOL and wins the contention of the medium access. After the packet departs from MAC queue, it is transmitted by a transmitter. The receiver in the destination station accepts the packet and sends it to a sink after collecting statistics.

*1) Traffic Model:* The traffic generator generates packets according to the distribution of packet interarrival time and packet size. The distribution of packet interarrival time can be any distribution in our simulator. Because we study the network under saturation traffic in this paper, we use constant packet interarrival time and set the traffic rate larger than the capacity of the network in order to make it saturated. We use in our simulation a constant payload size of 1500 B.

*2) Radio Channel Model:* As assumed in the analysis, we implement an ideal radio channel in the simulation. This means that the propagation delay is zero, and there is no channel error and no exposed or hidden node problem.

*3) MAC Implementation:* In order to minimize the number of events and speed up the simulation, we did not simulate the behavior of backoff entity in each station individually. Instead,

$$S = \frac{P_s \times T}{d + \frac{1}{p_{\text{tr}}} + P_s T_s + (1-P_s) T_c} = \frac{p_{\text{tr}} \times P_s \times T}{(1-p_{\text{tr}})*1 + p_{\text{tr}} P_s (T_s + d + 1) + p_{\text{tr}} (1-P_s)(T_c + d + 1)}$$

TABLE I
802.11 MAC/PHY SIMULATION PARAMETERS

| Parameter | Value | Parameter | Value |
|---|---|---|---|
| RTS | 0.352 ms | SIFS | 0.01 ms |
| CTS | 0.304 ms | PHY/MAC header | 0.328 ms |
| ACK | 0.304 ms | Tslot | 0.02 ms |

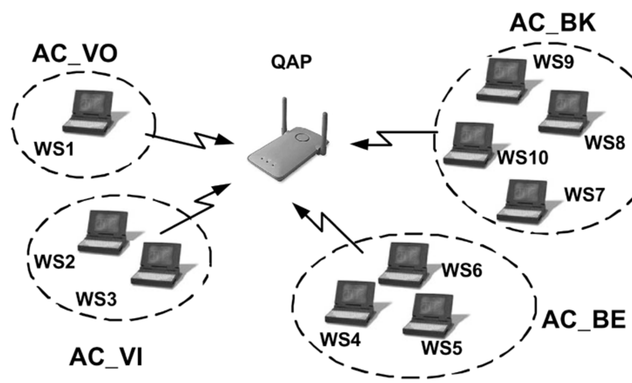



Fig. 6. Simulation scenario for experiment 1.

we simulated the behavior of the 802.11e EDCA medium access as a virtual scheduler.[7] For example, the scheduler observes how many packets are in HOL position and then compares their AIFS[AC + backoff] values. It tells the stations with the smallest value to transmit the packet, tells the other station to hold the packet and retransmit, and then advances the simulation time by AIFS[AC + backoff] of the winner station plus the packet transmission time. If only one station has the smallest delay value, the packet transmission is a success and packet transmission time is $T_s$; if more than one station has the same smallest delay value, collision happens and the packet transmission time is $T_c$.

The parameters of 802.11 MAC and PHY deployed in the simulation, as well the comparative analysis, are shown in Table I.

## B. Simulation Results

*1) Simulation Validation: Experiments 1–3:* The WLAN scenario we simulated in experiments 1–3 is shown in Fig. 6. In this WLAN, ten stations send traffic to QAP. Each station deploys one backoff entity of one AC to contend for the channel. Among these ten wireless stations (WS), there is one backoff entity per AC_VO (WS1), two backoff entities per AC_VI (WS2 and WS3), three per AC_BE (WS4, WS5, and WS6), and four per AC_BK (WS7, WS8, WS9, and WS10).

The EDCA parameter sets of the three experiments are (see Table II) as follows.

- Experiment 1: default setting from draft [9].
- Experiment 2: vary AIFS from the default setting.
- Experiment 3: vary CW from the default setting.

In each experiment, we simulate ten scenarios for this WLAN; progressively, from scenario 1 to scenario 10, we add WS1 to WS10 to the system one at a time. Then, we collect the aggregate saturation throughput and access delay for each AC and the whole system and compare the results from simulation

[7]An ideal channel condition and the synchronized system make this simulation method feasible.

TABLE II
802.11e EDCA PARAMETER SETS FOR EXPERIMENTS 1–3

| Experiment | AC | CWmin | CWmax | AIFSN |
|---|---|---|---|---|
| 1 | AC_BK | 31 | 1023 | 7 |
| (default) | AC_BE | 31 | 1023 | 3 |
| | AC_VI | 15 | 31 | 2 |
| | AC_VO | 7 | 15 | 2 |
| 2 | AC_BK | 31 | 1023 | 7 |
| | AC_BE | 31 | 1023 | 5 |
| | AC_VI | 31 | 1023 | 3 |
| | AC_VO | 31 | 1023 | 2 |
| 3 | AC_BK | 31 | 1023 | 2 |
| | AC_BE | 31 | 1023 | 2 |
| | AC_VI | 15 | 31 | 2 |
| | AC_VO | 7 | 15 | 2 |

with the results from our analysis in Figs. 7–9. Lines represent analytical results, while the markers near the lines represent the corresponding simulation ones. We use subscript "A" to represent analysis and "S" to represent simulation in the legend. Each simulation result is averaged from 20 simulation replications, and each simulation replication lasts for 1 000 000 transmission cycles. The 95% CI is shown as a bar around each simulation result. For most of the results, the CIs are too small to be visible except for the last four for access delay of AC_BK in each figure.

The agreement between analysis and simulation is remarkable. The only exception is the access delay for AC_BK in Fig. 7. The reason for this is that the throughput for a station of AC_BK is very small (of the order of $10^{-3}$) and the access delay for an AC_BK station is very large (of the order of $10^3$). In our analysis, the access delay is inversely proportional to the throughput. Therefore, a tiny difference between simulation and analysis result in throughput will result in a huge gap between access delay simulation and analysis results.

From experiment 1 (Fig. 7), we can also gain another insight: Although only one AC_VO station competes with two AC_VI stations, three AC_BE stations and four AC_BK stations, the maximum (saturation) throughput per AC_VO pumped into the network is still the largest, the maximum (saturation) throughput per AC_BK pumped into the network is the smallest and nearly zero. The reason for this is that the default EDCA parameter set differentiates among the four ACs very effectively through the combined effects of AIFS, CWmin, and CWmax. AC_BK stations are almost starved in this experiment due to the long AIFS and big CWmin and CWmax, compared to the very small AIFS and CW of AC_VO.

By varying the EDCA parameter set from the standard default values, we can study the individual differentiation effects of AIFS and CW separately. In experiment 2, we keep CWmin and CWmax the same for all ACs and just vary the AIFSN by two units between ACs; in experiment 3, we keep AIFSN all the same but vary CWmin and CWmax. Figs. 8 and 9 show the results from these two experiments. The agreement between simulation and analysis is still very good. Besides that, by comparing these three figures (Figs. 7–9), we can tell that the combined effects surely are greater than each individual one. However, among them, the relationship is not simply additive. The

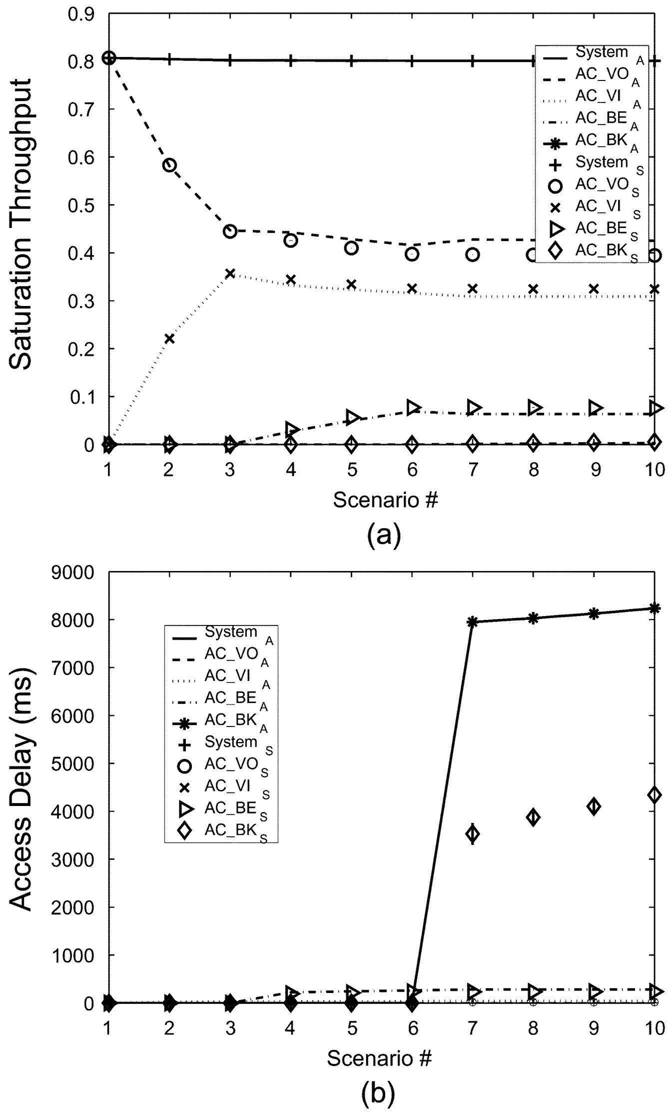



Fig. 7. Experiment 1: performance using default EDCA parameter set.

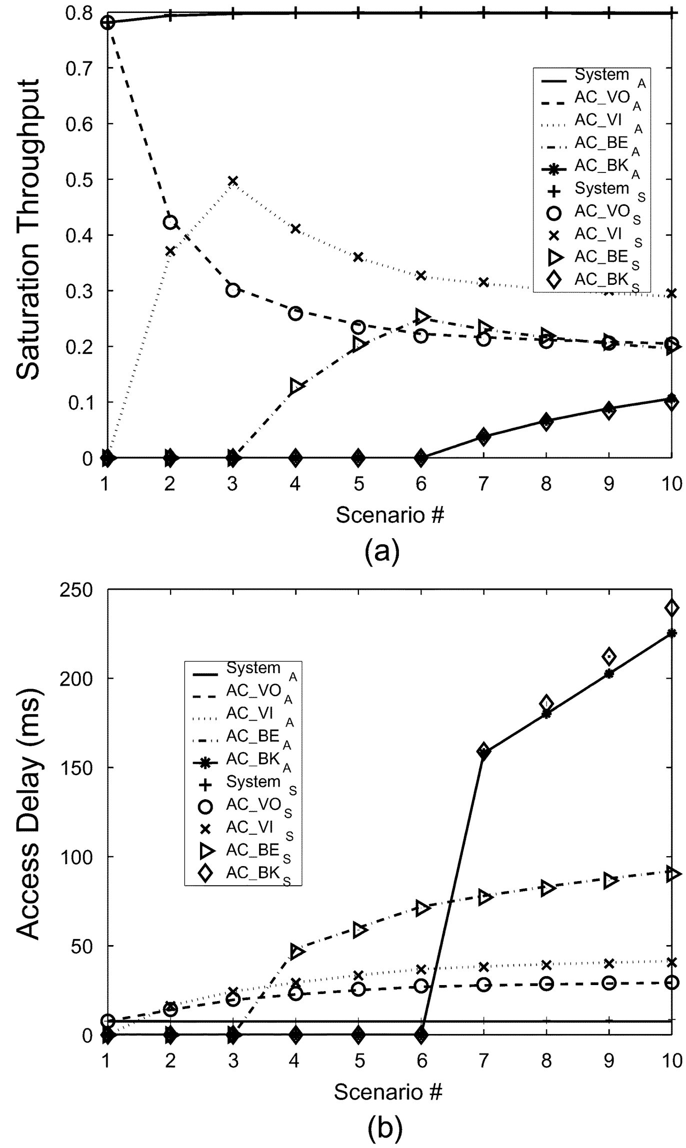



Fig. 8. Experiment 2: performance by varying only AIFS.

difference between AC_VO and AC_VI is mainly due to the nonoverlapping CW ranges. However, the starvation of AC_BK is not due to each individual but is due to the combined effects of both AIFS and CW differentiation.

*2) Differentiation Effects: Experiment 4–6:* Another important factor not considered in the above experiments is the number of stations per AC. Realizing the correlation among the differentiation effects of all EDCA parameters, we believe that a formal sensitivity analysis over the whole response surfaces will be ideal to carry out a thorough study of parameter effects. However, such a study is beyond the scope of this paper.

In the following, we perform three more experiments of specific settings and attempt to gain more understanding about the differentiation effects of AIFS and CWmin and make some further observations. To filter out the effect of the number of stations and CWmax, we fix the number of stations for each AC to be equal to 3 and $\mathrm{CWmax} = (\mathrm{CWmin} + 1)^5 - 1$. The EDCA parameter sets are (see Table III) as follows.

- Experiment 4: only differentiate AIFS but keep CWmin and CWmax equal to 31 and 1023, respectively. The AIFSN differences increase from 0 to 4.
- Experiment 5: only differentiate CWmin and CWmax but keep the AIFS equal to 2. The CWmin difference increases from 0 to 4.
- Experiment 6: differentiate both AIFS and CWmin, CWmax by combining experiments 1 and 2.

From the result comparison in Fig. 10, we can clearly see that a larger difference in AIFS or CW result in a larger difference in throughput and delay performance among different ACs. Furthermore, for this specific base setting $\{\mathrm{AIFSN}, \mathrm{CWmin}, \mathrm{CWmax}\} = \{2, 31, 1023\}$, we can infer the following.

- AIFSN has a larger differentiation effect on the performance than CW for the same variation from 1 to 4.
- The combined differentiation effects of AIFSN and CW are bigger than both individual ones.
- By keeping all other parameters unchanged, a larger AIFS or CW results in a lower throughput and a longer access delay.

*3) Special Case With Very Small CWmin: Experiment 7:* For some special settings with very small CW values (like

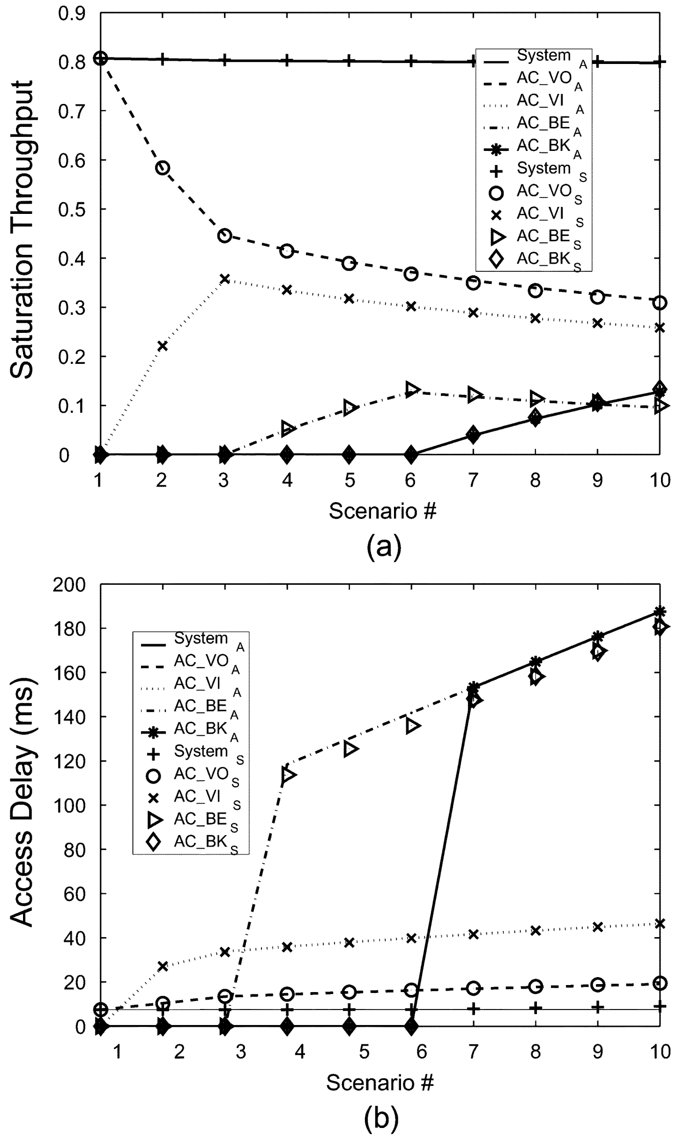



Fig. 9. Experiment 3: performance by varying only CW.

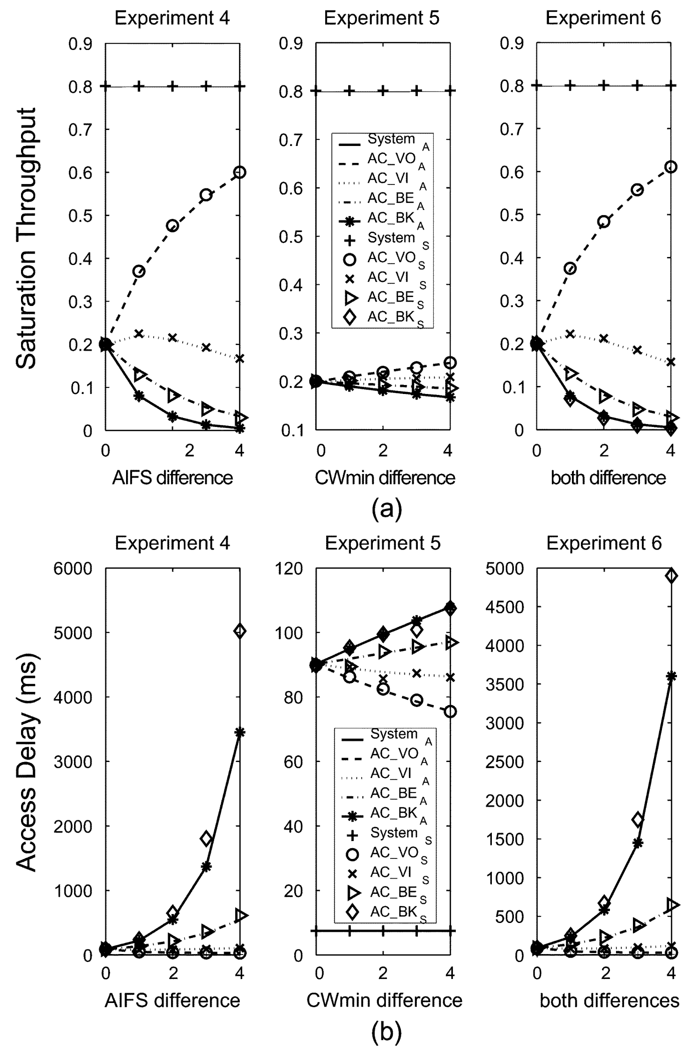



Fig. 10. Experiments 4–6: comparison of differentiation effects.

TABLE III
802.11e EDCA PARAMETER SETS FOR EXPERIMENTS 4–6

| Exp | AC | CWmin | AIFSN |
|---|---|---|---|
| 4 | AC_VO | 31 | 2, 2, 2, 2, 2 |
| | AC_VI | 31 | 2, 3, 4, 5, 6 |
| | AC_BE | 31 | 2, 4, 6, 8,10 |
| | AC_BK | 31 | 2, 5, 8,11,14 |
| 5 | AC_VO | 31,31,31,31,31 | 2 |
| | AC_VI | 31,32,33,34,35 | 2 |
| | AC_BE | 31,33,35,37,39 | 2 |
| | AC_BK | 31,34,37,40,43 | 2 |
| 6 | AC_BK | 31,31,31,31,31 | 2, 2, 2, 2, 2, |
| | AC_BE | 31,32,33,34,35 | 2, 3, 4, 5, 6, |
| | AC_VI | 31,33,35,37,39 | 2, 4, 6, 8,10 |
| | AC_VO | 31,34,37,40,43 | 2, 5, 8,11,14 |

TABLE IV
802.11e EDCA PARAMETER SETS FOR EXPERIMENT 7

| Experiment | AC | CWmin | CWmax | AIFSN |
|---|---|---|---|---|
| 7 | AC_BK | 3,3,3,3,3 | 3,3,3,3,3 | 2 |
| | AC_BE | 3,4,5,6,7 | 3,4,5,6,7 | 2 |
| | AC_VI | 3,5,7,9,11 | 3,5,7,9,11 | 2 |
| | AC_VO | 3,6,9,12,14 | 3,6,9,12,14 | 2 |

CWmin $= 1, 2, 3$), the CWmin variations can show dramatic effects in performance differentiations. Also, a bigger CW may result in a higher throughput and a shorter access delay. The reason is that a very small CW value causes a high collision probability; therefore, an increase from the this small CW can help reduce the collision probability, thus resulting in a better performance. Experiment 7 (see Table IV) explores such a special case. The results in Fig. 11 support the above observations.

From the above experiments and discussions, we surmise that the differentiation effectiveness around different parameter settings can vary from or even contradict each other. The number of stations, AIFS, CWmin, and CWmax for each AC all affect the throughput and delay performance, and they are also correlated with each other. Therefore, as mentioned earlier, a formal sensitivity analysis has to be conducted to complete a more thorough study of the parameter effectiveness. This will be left as our future work.

*4) Sensitivity With Respect to the Number of Stations: Experiment 8:* To study how our model performs with the changing number of stations, we conduct experiment 8: the EDCA parameter set is as in experiment 1, and the number of stations per

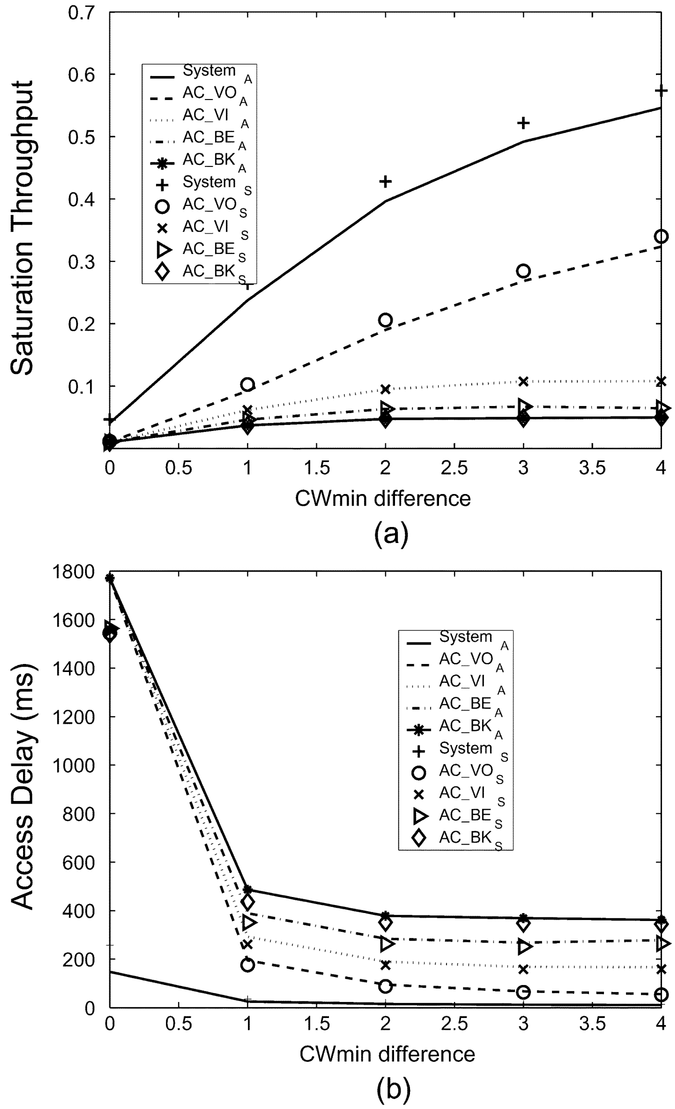



Fig. 11. Experiment 7: special case of very small CWmin.

AC increases from 1 to 10 $(N(0) = N(1) = N(2) = N(3) = 1, 2, \ldots, 10)$.

From the saturation throughput results in Fig. 12, we observe that the predictive capability of our model remains high even with an increasing number of stations. Furthermore, another effect is that the saturation throughput drops due to more frequent collisions, as the number of stations increases (greater than two per AC).

Comparing to the previous performance evaluations about both AIFS- and CW-based differentiation approaches, and although we do not offer a rigorous sensitivity analysis, our work does provide the evaluation of four ACs with the standard 802.11e EDCA parameter setting. Most previous work only evaluates a WLAN with two or three classes. The good agreement between simulation and our analysis proves the accuracy of our model and provides a useful tool for network design and analysis.

## VIII. Summary

EDCA is introduced in 802.11e for QoS improvements over legacy 802.11 DCF. The understanding of how the EDCA parameters affect the performance of WLAN is a crucial prerequisite for the design of any QoS scheme using EDCA.

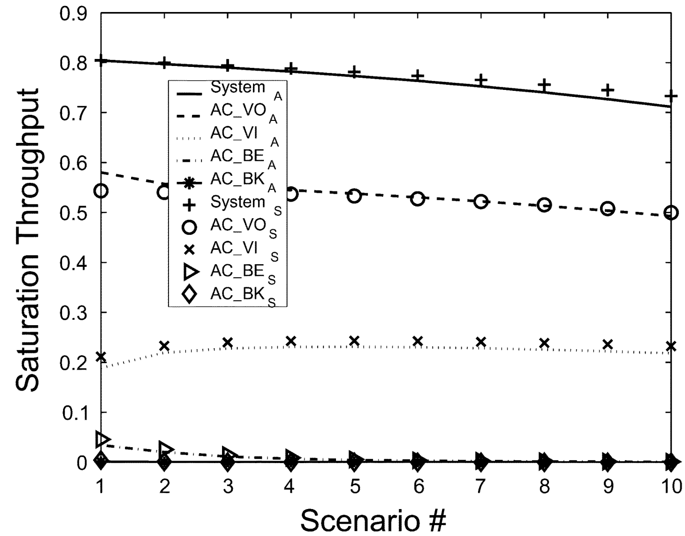



Fig. 12. Experiment 8: sensitivity with respect to the number of stations.

Our main contributions in this paper are threefold. First, we abstract and unify a common guiding principle behind three major performance models, thus increasing the understanding and applicability of these efforts. Second, we propose a unified performance model and analysis method for 802.11e EDCA by taking elements of all three models, while maintaining their common principles. In our model, the memory effects of backoff counter and backoff stage are still accounted for by using a bidimensional-state Markov Chain as in [4] or mean value analysis as in [7]; in a novel manner, in order to account for the effect of different AIFS values, we did not introduce further dimension(s) to the state space as in [20], [23], [29], and [28], but we used multiple bidimensional chains or multiple average value analysis in separate BSPs under the main assumption of time-dependent $p$-persistence behavior. This new model is easy to apply by reducing the complexity of Markov chains and offering an alternative mean value analysis method to compute persistence factors. For another aspect, this model also allows better understanding of the system behavior by exploiting the concept of BSPs and by using the persistence factors matrix $\mathbf{P}$. Third, simulation results validate our model and analysis, showing that our model will be a helpful tool for 802.11e network designers.

All of the analyses and simulations in this paper are performed with the assumption of ideal channel conditions, saturation traffic, and a single-hop network environment. The study of throughput and delay performance of EDCA with a more realistic wireless channel model, different traffic models, and even in multihop networks are included in our future plans. Another area of research can be to find closed-form solutions for a simplified performance model and then perform a formal sensitivity analysis of each parameter, finally providing a dynamic turning of parameters according to the desired QoS level.

## References

[1] J. Hui and M. Devetsikiotis, "Performance analysis of IEEE 802.11e EDCA by a unified model," in *Proc. Globecom*, Dec. 2004, pp. 754–759.
[2] *IEEE 802.11 Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, Aug. 1999.

[3] G. Bianchi, "IEEE 802.11 saturation throughput analysis," *IEEE Commun. Lett.*, vol. 2, no. 12, pp. 318–320, Dec. 1998.

[4] ——, "Performance analysis of the IEEE 802.11 distributed coordination function," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 3, pp. 535–547, Mar. 2000.

[5] F. Cali, M. Conti, and E. Gregori, "IEEE 802.11 wireless LAN: Capacity analysis and protocol enhancement," in *Proc. INFOCOM*, 1998, pp. 142–149.

[6] ——, "Dynamic tuning of the IEEE 802.11 protocol to achieve a theoretical throughput limit," in *IEEE/ACM Trans. Netw.*, vol. 18, Dec. 2000, pp. 785–799.

[7] Y. C. Tay and K. C. Chua, "A capacity analysis for the IEEE 802.11 MAC protocol," *Wireless Netw.*, pp. 159–171, Jul. 2001.

[8] H. Wu, Y. Peng, K. Long, S. Cheng, and J. Ma, "Performance of reliable transport protocol over IEEE 802.11 wireless LAN: Analysis and enhancement," in *Proc. INFOCOM*, Jun. 2002, pp. 599–607.

[9] *Draft Supplement to IEEE Standard 802.11-1999: Medium Access Control (MAC) Enhancements for Quality of Service (QoS)*, IEEE 802.11 WG, IEEE 802.11e D6.0, Nov. 2003.

[10] A. Imad and C. Claude, "Differentiation mechanisms for IEEE 802.11," in *Proc. IEEE INFOCOM*, Anchorage, AK, Apr. 2001, pp. 209–218.

[11] A. Lindgren, A. Almquist, and O. Schelén, "Evaluation of quality of service schemes for IEEE 802.11 wireless LANs," in *Proc. 26th Annu. IEEE Conf. Local Computer Netw.*, Nov. 2001, pp. 348–351.

[12] ——, "Quality of service schemes for IEEE 802.11 wireless LANs: An evaluation," *ACM Mobile Netw. Appl.*, pp. 223–235, Jun. 2003.

[13] S. Mangold, S. Choi, P. May, O. Klein, G. Hiertz, and L. Stibor, "IEEE 802.11e wireless LAN for quality of service (invited paper)," in *Proc. Eur. Wireless*, vol. 1, Florence, Italy, Feb. 2002, pp. 32–39.

[14] S. Choi, J. D. Prado, N. S. Shankar, and S. Mangold, "IEEE 802.11e contention-based channel access (EDCF) performance evaluation," in *Proc. ICC*, May 2003, pp. 1151–1156.

[15] H. L. Truong and G. Vannuccini, "The IEEE 802.11e MAC for quality of service in wireless LANs," in *Proc. SSGRR*, 2003, [CD-ROM].

[16] D. He and C. Q. Shen, "Simulation study of IEEE 802.11e EDCF," in *Proc. 57th IEEE Semiannu. VTC*, Apr. 2003, pp. 685–689.

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

[18] S. Mangold, "Analysis of IEEE 802.11e and application of game models for support of quality-of-service in coexisting wireless networks," Ph.D. dissertation, Aachen Univ., Aachen, Germany, Jun. 2003.

[19] D. Qiao and K. G. Shin, "Achieving efficient channel utilization and weighted fairness for data communications in IEEE 802.11 WLAN under the DCF," in *Proc. IWQoS*, May 2002, pp. 227–236.

[20] J. Zhao, Z. Guo, Q. Zhang, and W. Zhu, "Performance study of MAC for service differentiation in IEEE 802.11," in *Proc. Globecom*, Nov. 2002, pp. 778–782.

[21] G. Bianchi and I. Tinnirello, "Analysis of priority mechanisms based on differentiated inter frame spacing in CSMA-CA," in *Proc. VTC*, Fall 2003, pp. 1401–1405.

[22] Y. Ge and J. Hou, "An analytical model for service differentiation in IEEE 802.11," in *Proc. ICC*, May 2003, pp. 1157–1162.

[23] Y. Xiao, "Backoff-based priority schemes for IEEE 802.11," in *Proc. ICC*, May 2003, pp. 1568–1572.

[24] ——, "Enhanced DCF of IEEE 802.11e to support QoS," in *Proc. WCNC*, 2003, pp. 1291–1296.

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

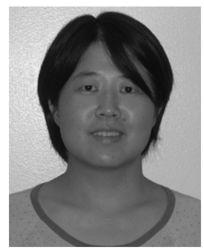
[26] J. He, L. Zheng, Z. Yang, and C. T. Chou, "Performance analysis and service differentiation in IEEE 802.11 WLAN," in *Proc. 28th Int. Conf. Local Computer Netw.*, Oct. 2003, pp. 691–697.

[27] J. Hui and M. Devetsikiotis, "Designing improved MAC packet scheduler for 802.11e WLAN," in *Proc. Globecom*, Dec. 2003, pp. 184–189.

[28] K. Xu, Q. Wang, and H. Hassanein, "Performance analysis of differentiated QoS supported by IEEE 802.11e enhanced distributed coordination function (EDCF) in WLAN," in *Proc. Globecom*, Dec. 2003, pp. 1048–1053.

[29] R. Battiti and B. Li, "Supporting service differentiation with enhancements of the IEEE 802.11 MAC protocol: Models and analysis," Univ. Trento, Trento, Italy, Tech. Rep., May 2003.

[30] B. J. Kwak, N. O. Song, and L. E. Miller, "Analysis of the stability and performance of exponential backoff," in *Proc. WCNC*, 2003, pp. 1754–1759.

**Jie Hui** (S'02) was born in Shanxi, China. She received the B.S. degree in electrical engineering from Xi'an Jiaotong University, Xi'an, China, in 1996, and the M.S. degree in electrical engineering from Institute of Space Electronic Equipment, Shanghai, China, in 1999. She is currently working toward the Ph.D. degree at North Carolina State University, Raleigh.

From 2000 to July 2001, she was with the 3G R&D Center, ZTE Corporation, Shenzhen, China, as a System and Software Engineer. Her research interests include QoS provisioning for multimedia over WiFi, medium access control, cross-layer control and optimization, modeling, performance evaluation, and efficient simulation of wireless networks.

**Michael Devetsikiotis** (S'85–M'94–SM'03) was born in Thessaloniki, Greece. He received the Diploma degree in electrical engineering from the Aristotle University of Thessaloniki, Thessaloniki, Greece, in 1988, and the M.Sc. and Ph.D. degrees in electrical engineering from North Carolina State University, Raleigh, in 1990 and 1993, respectively.

In October 1993 he joined the Broad-band Networks Laboratory, Carleton University, Ottawa, ON, Canada, as a Research Associate. He became an Adjunct Professor in the Systems and Computer Engineering Department, Carleton University, in 1995, an Assistant Professor in 1996, and an Associate Professor in 1998. Since November 2000, he has been with the Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, as an Associate Professor. His research work has been in the areas of telecommunication systems modeling, performance evaluation, and efficient simulation, traffic characterization and management, and optimization techniques applied to the analysis and design of communication systems. His present focus is on the performance of information networks as they become larger in size, and more complex in topology and workload.

Dr. Devetsikiotis currently serves as the chairman of the IEEE Communication Society Committee on Communication Systems Integration and Modeling (CSIM), as an Associate Editor for the *ACM Transactions on Modeling and Computer Simulation* (TOMACS), and as an Associate Editor for the IEEE COMMUNICATIONS LETTERS. He has cochaired the High-Speed Networks symposium under IEEE International Communications Conference (ICC) 2004 in Paris, France, and the QoS, Reliability and Performance Evaluation symposium, under IEEE ICC 2005, in Seoul, Korea.