Performance Analysis for IEEE 802.11e EDCF Service Differentiation

Hua Zhu and Imrich Chlamtac, *Fellow, IEEE*

***Abstract*—Having been originally developed as an extension of the wired local area networks, IEEE 802.11 lacks support for quality-of-service (QoS) and differential services. Since its introduction, various extensions and modifications have been studied to address this current need and the IEEE 802.11 Task Group E is responsible for developing a QoS-aware MAC protocol that considers several service differentiation mechanisms. However, the performance of service differentiation has only been evaluated by simulation. The analytical model that calculates the differential service performance corresponding to the contention parameter configuration has not been found yet. In this paper, we first briefly explain the enhanced distributed coordination function (EDCF) access method of IEEE 802.11e. We then introduce an analytical model, which can be used to calculate the traffic priority, throughput, and delay corresponding to the configuration of multiple DCF contention parameters under the saturation condition. A detailed simulation is provided to validate the proposed model. Finally, using the analytical model, we analyze the effect on service differentiation for each contention parameter. The contention parameters can be configured appropriately at each station to achieve specific needs of service differentiation for applications.**

***Index Terms*—Enhanced distributed coordination function (EDCF), IEEE 802.11, quality-of-service (QoS), service differentiation.**

I. Introduction

THE emergence of wireless computing in the 1990s has led to yet another extension of the 802 specifications. The emerging IEEE 802.11 [29] uses the standard 802 LLC protocol but provides physical layer (PHY) and MAC sublayer optimized for wireless communications. The IEEE 802.11 has been highly successful and is being considered for inclusion in third-generation (3G) cellular networks [16].

The basic access method [distributed coordination function (DCF)] in IEEE 802.11 MAC layer protocol does not provide priorities and service differentiation mechanism to guarantee an access delay bound to stations or any specific traffic stream. Due to the significant demand for the transmission of delay sensitive video/voice data, IEEE 802.11 Task Group E is working on developing a quality-of-service (QoS)-aware MAC protocol with service differentiation mechanism. IEEE 802.11e MAC protocol is considered as an extension of the previous IEEE 802.11 MAC protocol, with a new hybrid coordination function (HCF) [3]. HCF consists of two access methods, i.e., enhanced DCF (EDCF) and the polling-based [point coordination function (PCF)] scheme.

The IEEE 802.11 standard considers two network topologies: ad-hoc and infrastructure. In an ad-hoc wireless local area network (LAN), a group of mobile terminals communicate with each other in an independent basic service set (BSS) without connectivity to the wired backbone network. In an infrastructure wireless LAN, mobile terminals in a BSS communicate with the backbone network through an access point (AP). The AP is an internetworking device seamlessly integrating the wireless BSS with the wired backbone network, and it provides the connectivity between multiple BSSs to form an extended service set (ESS). A mobile terminal can roam among BSSs within one ESS without losing connectivity with the backbone network.

The optional polling-based (PCF) scheme is designed for time-bounded services in the BSS of infrastructure configuration. When PCF is enabled, the wireless channel is divided into superframes. Each superframe consists of a *contention free period* (CFP) for PCF and a *contention period* (CP) for DCF. At the beginning of CFP, the point coordinator (usually the access point) will contend for the access to the wireless medium. Once it acquires the medium, it cyclically polls stations giving them opportunity to transmit. However, PCF has certain restrictions: 1) PCF can only be used in the BSS of infrastructure configuration, while the DCF based scheme is good for BSS and ESS in both infrastructure and ad-hoc configuration; 2) PCF is highly complex and experiences substantial delay at low load, i.e., stations must always wait for polling, even in an otherwise idle system; 3) since the AP needs to contend for the channel using DCF protocol at the beginning of CFP, the effective period of contention free polling may vary; and 4) many other issues remain unsolved in PCF [12], e.g., how should the point coordinator manage the polling of a large number of interactive streams without harming the applications using DCF contention. This will be a common problem for the AP with dense node deployment. For all of the above, the DCF-based access method is being considered as primary candidate for differential services.

Given the large number of DCF contention parameters in 802.11e, such as $AIFS$, $CW_{\max}$ and $CW_{\min}$, that could be used to prioritize traffic, an analytical model to calculate the performance of the traffic prioritization corresponding to the configuration of those contention parameters is highly desirable. Although the performance of service differentiation has been evaluated by simulation [3], [25], so far there is no analytical model proposed to analyze the performance of differential service without simplifying the system by excluding the variation

Manuscript received July 10, 2003; revised January 6, 2004; accepted February 25, 2004. The editor coordinating the review of this paper and approving it for publication is Z. Zhang.

The authors are with the University of Texas at Dallas, Richardson, TX 75080 USA (e-mail: zhuhua@utdallas.edu; chlamtac@utdallas.edu).

Digital Object Identifier 10.1109/TWC.2005.847113

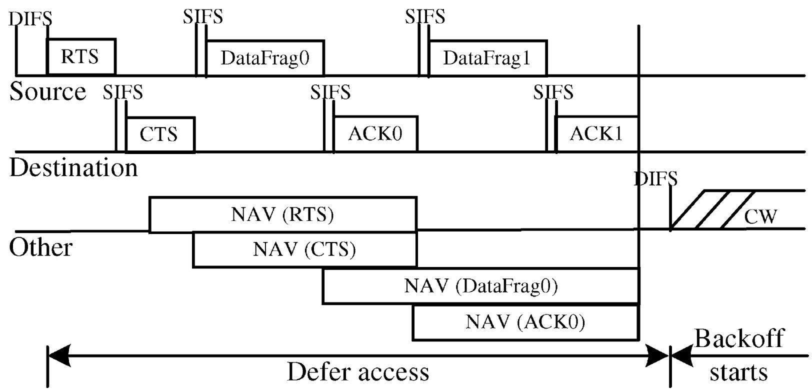


Fig. 1. CSMA/CA-RTS/CTS with fragmentation access scheme.

of certain contention parameters, and thereby a guide on setting all contention parameters in the IEEE 802.11e has not been found.

In this paper, we first give an overview of IEEE 802.11 DCF and its enhanced service differentiation version—802.11e. We then introduce an analytical model for 802.11e EDCF access method. Finally, we validate the analytical model based on the simulation result, and give a guide on setting contention parameters. The proposed analytical model can be used to calculate the traffic priority and saturation throughput. We note that while under nonsaturation conditions, the QoS requirements (throughput and/or delay) may be satisfied without a differential service mechanism, under saturation conditions, prioritized traffic and differential service become more important in order to guarantee the transmission of throughput-critical (and/or time-critical) traffic by sacrificing the transmission throughput (and/or delay) of other traffic.

The remainder of the paper is organized as follows. Section II provides an overview of IEEE 802.11 DCF scheme. Section III investigates tunable parameters in the IEEE 802.11e EDCF access method. Section IV describes the proposed EDCF analytical model. Section V validates the model by simulation and gives a guide to configure contention parameters. Finally, we conclude the paper in Section VI.

II. Overview of the IEEE 802.11 DCF Scheme

The scope of IEEE 802.11 standard is to develop a MAC sublayer and PHY specification for wireless connectivity for fixed, portable, and moving stations within a local area.

Certain MAC contention parameters, such as $SIFS$, *slot time*, the length of PHY preamble header, and channel capacity, are pre-defined based on the specific PHY. Since all other contention parameters are defined based on these pre-defined parameters, and the contention schemes are all the same for different PHYs, therefore, we are be able to analyze the IEEE 802.11 MAC without further details of the PHY.

At MAC sublayer, 802.11 includes two medium access protocols, i.e., DCF and an optional PCF. In this paper, we only consider the DCF-based access protocol. Since 802.11e EDCF inherits all the contention scheme and parameters of the original 802.11 DCF, we provide an overview of the original 802.11 DCF scheme in this section.

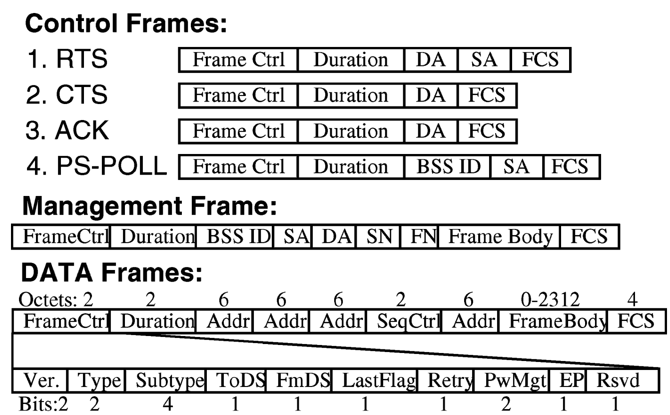


Fig. 2. IEEE 802.11 frame format.

DCF is based on carrier sense multiple access with collision avoidance (CSMA/CA). Carrier sensing multiple access with collision detection (CSMA/CD) is unable to be used because the station cannot listen to the channel for collision while transmitting. In 802.11, carrier sensing (CS) is performed at both PHY and MAC layers, i.e., physical carrier sensing and MAC layer virtual carrier sensing. If the MAC frame length (including the payload and 34 B MAC header) exceeds the *RTS_threshold*, request-to-send (RTS) and clear-to-send (CTS) are used to solve the problem of hidden terminal and capture effect. In order to further increase the channel utilization, if the payload length exceeds *Frag_threshold*, it is divided into fragments before transmitting within one contention window. As a result, if error occurs in the transmission of a specific fragment, the station does not have to wait to backoff until the whole payload is transmitted, also it does not have to transmit previous fragments that have been transmitted correctly. DCF with RTS/CTS and fragmentation is shown in Fig. 1. The range of *RTS_threshold* is from 0 to 2347 (default), while the range of *Frag_threshold* is from 256 to 2312 (default). However, vendors may choose different range for both thresholds. The IEEE 802.11 frame format is shown in Fig. 2.

Error recovery (i.e., retransmission) is always the responsibility of the station that initiates a frame exchange sequence. Many circumstances may cause an error to occur that requires the retransmission. For example, the CTS frame may not be returned after an RTS frame is transmitted. This may happen due to a collision with the transmission of another station, due to interference in the channel during the transmission of the RTS or CTS frame, or because the station receiving the RTS frame has an active virtual carrier sense condition (indicating a busy medium time period). Stations have two retry counters: *short retry count* and *long retry count*. Each packet has a single retry counter associated with it. Packets that are shorter than *RTS_threshold* are associated to the short retry count; otherwise to the long retry count. The retry counts begin at 0 and are incremented when a frame (or fragment) transmission fails. The frame will be dropped when the retry count exceeds the maximum retry limit. The short count is reset to 0 when: 1) a CTS is received in response to a transmitted RTS; 2) an ACK, which is shown in Fig. 1 and 2, is received after a non-fragmented transmission; or 3) a broadcast or multicast frame is received. The long retry count is reset to 0 when: 1) an ACK is received for a frame longer than RTS threshold; or 2) a broadcast or multicast frame is received.

In DCF, a backoff window follows the DIFS in order to minimize the probability of collision among stations. The length of window is equal to a uniformly distributed random integer multiplied by the slot time, which is a medium dependent parameter. The range of the random number is $[0, CW]$. CW is the backoff window size, which is equal to (31,63,127,255,512,1023,1023) corresponding to the retry count from 1 to 7, respectively.

For a better understanding of IEEE 802.11, the interested reader is referred to [5], [11], [18], [19], [27], and [28].

III. IEEE 802.11e EDCF Service Differentiation

IEEE 802.11 provides a best effort service, which indicates that every data packet handed over to the 802.11 interfaces receives similar treatment as other packets in terms of delivery guarantees, i.e., available bandwidth, latency, jitter, etc. In order to deliver real-time video traffic [13], which is sensitive to packet latency and effective bandwidth characteristic of the underlying network, QoS and service differentiation [1], [7], [9] have become one of the most important issues of IEEE802.11 standard.

The IEEE 802.11e recently established working group of IEEE 802.11e provides applications with QoS and service differentiation supported by priority-based contention service, i.e., the EDCF. A concept of traffic category is introduced. Each *traffic category* is associated with the predetermined contention parameters, *arbitration interframe space* $(AIFS)$, $CW_{\min}$ and $CW_{\max}$, and the *backoff persistence factor* (PF) [3], [14]. The lower $AIFS/CW_{\min}/CW_{\max}$ results in the higher probability of winning the channel contention. In EDCF, the contention window is expanded by the PF after collision. In the original DCF, the contention window is always doubled after collision $(PF = 2)$, while in EDCF PF may be a different value. For each station, up to eight traffic categories with different contention parameters can exist in parallel, thus leading to internal contention in each station. The collisions among internal contention are avoided by letting the highest priority traffic category win the contention window. Please note that the persistence factor PF in IEEE 802.11e is different from that in p-DCF [23], which is a version different from the IEEE 802.11 standard DCF in that instead of using the binary exponential backoff technique, a station determines whether to attempt transmission following an idle time of DIFS by the probability p (in other words, the backoff interval is sampled from geometric distribution with parameter p).

Besides the previous contention parameters, there are more station-based tunable parameters, which can also support differential services. They are the following.

- *RTS_threshold*: If there are a large number of terminals (not necessarily hidden terminals), decreasing *RTS_threshold* results in higher effective throughput; on the other hand, if there are only few terminals, increasing RTS_threshold results in higher effective throughput by reducing RTS/CTS overhead.
- *Frag_threshold*: In noisy areas, increasing *Frag_threshold* results in greater effective throughput, because interference corrupts only fragments, not whole frames.On the other hand, in noise-free areas, decreasing *Frag_threshold* results in greater effective throughput by reducing fragmentation acknowledgment overhead.
- *Long and short retry limit*: Higher retry limit decreases the frame drop-rate, but may throttle the data rate and throughput because of longer backoff time; while smaller retry limit increases frame drop-rate but shorten backoff time.

IV. Analytical Model

In this section, we give an analytical model of EDCF in the *saturation* condition, i.e., the transmission queue of each traffic category is assumed to be always nonempty. Performance under saturation condition, which is not always the case in practice though, gives us fundamental bounds on system throughput and delay. The proposed analytical model provides quantitative results of channel contention among prioritized traffic flows. It also gives us an insight on the different influence on service differentiation by each individual contention parameter (i.e., $AIFS$, $CW_{\min}$, $CW_{\max}$, and PF). Based on this result, one can configure traffic categories to achieve the desirable performance of service differentiation, whether moderate service differentiation or service separation between high and low traffic priorities. The proposed model extends the Markov model described in [2], [15], [22], [24], and [26] by introducing differential service parameters, as shown in Fig. 3.

Assume there are total T priority traffic categories (TC), and for each TC, the number of traffic flows is $N_t, t = 1, 2, \ldots, T$. The total number of traffic flows is

$$N = \sum_{t=1}^{T} N_t. \tag{1}$$

The nth traffic flow is associated with $AIFS_n$, $CW_{n,\min}$, $CW_{n,\max}$, pf_n, and all other parameters that have a subscript n. Define for convenience $f_n = pf_n$, $W_n = CW_{n,\min} + 1$, $W_{n,i} = (f_n)^i W_n$, and we have

$$W;_{n,m'_n} = (f_n)^{m'_n} W_n = CW_{n,\max} + 1 \tag{2}$$

$$\begin{cases} W_{n,i} = (f_n)^i W_n & i \le m'_n \\ W_{n,i} = (f_n)^{m'} W_n & m'_n < i \le m_n \end{cases} \tag{3}$$

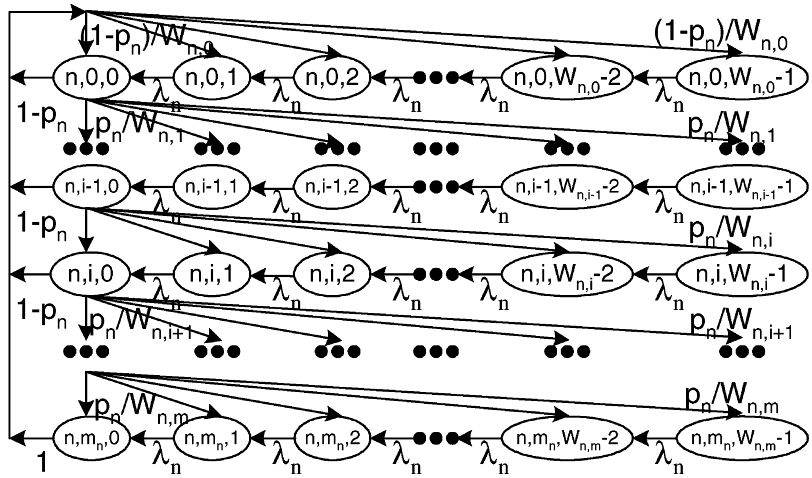


Fig. 3. Markov chain model of EDCF.

where (m_n+1) is the retry limit of the packet in the n^{th} traffic flow. So far in 802.11 short retry limit is 7, while the long retry limit is 4.

In Fig. 3, the state of the nth traffic flow is described by $\{n,i,k\}$, where n stands for the nth traffic flow and ranges [1, N], where i stands for the backoff stage (i.e., the number of retries) and ranges [0, m_n], where k stands for the backoff delay in timeslots and ranges [0, $W_{n,i}-1$].

The analysis is divided into two parts. First, we study the traffic priority τ_n (i.e., the stationary probability that a station transmits a packet in a randomly selected slot time) by a Markov model. Second, we derive the TC throughput equation by studying the events that can occur in a generic slot time.

A. TC Priority Analysis

Each traffic flow has a conditional collision probability p_n (the probability of collision seen by a packet being transmitted on channel) and the backoff state transition rate λ_n. Given a specific backoff window size $W_{n,i}$, the average backoff timer $\mathrm{E}[bk_{n,i}] = (W_{n,i}/2)$ in slot time. The average backoff window size [10] of a MAC frame is calculated as (4), located at the bottom of the page.

The overall average backoff timer of a packet in nth traffic flow is

$$E[bk_n] = \frac{E[W_n]}{2} \times slotTime. \tag{5}$$

Without loss of generality, let us assume that from the first to the last traffic flow, we have the highest to the lowest priority, i.e., $AIFS_1 = AIFS_{\min}$. We define $\delta_{n,aifs} = AIFS_n - AIFS_{\min}$. In saturation condition, the transmission queues of the highest priority traffic flows are always nonempty. The backoff timer of all other traffic flows will be paused/resumed repeatedly when it decreases. Therefore, we calculated λ_n for the nth traffic flow as follows:

$$\lambda_n = \begin{cases} \left[\frac{(E[bk_1]-\delta_{n,aifs})}{E[bk_n]}\right]^{N_1}, & E[bk_1] > \delta_{n,aifs} \\ 0, & o.w. \end{cases} \tag{6}$$

where N_1 stands for the number of traffic flows in the highest priority traffic category. In (6), if $\mathrm{E}[bk_1] \le \delta_{n,aifs}$, which means the traffic flow with the highest priority always starts transmitting before the deferment of $AIFS_n$ is completed. Therefore, the backoff timer of the nth traffic flow is always in the pause state, i.e., the nth traffic flow will not be able to win the channel contention. Readers may argue about the accuraqcy of (6). However, our numerical experiments has shown that a more accurate equation, which takes account of all traffic categories instead of only the highest priority traffic category, will generate the result with only negligible difference.

In the Markov chain of the nth traffic flow, the only nonnull one-step transition probabilities are[1]

$$\begin{cases} P\{n,i,k|n,i,k+1\} = \lambda_n, & k\in[0,W_{n,i}-2], i\in[0,m_n] \\ P\{n,0,k|n,i,0\} = \frac{(1-p_n)}{W_{n,0}}, & k\in[0,W_{n,0}-1], i\in[0,m_n-1] \\ P\{n,i,k|n,i-1,0\} = \frac{p_n}{W_{n,i}}, & k\in[0,W_{n,i}-1], i\in[1,m_n] \\ P\{n,0,k|n,m_n,0\} = \frac{1}{W_{n,0}}, & k\in[0,W_{n,0}-1] \end{cases}. \tag{7}$$

Note that $b_{n,i-1,0}p_n = b_{n,i,0}$ for $(0 \le i \le m_n)$, therefore,

$$b_{n,i,0} = p_n^i b_{n,0,0} \quad (0 \le i \le m_n). \tag{8}$$

[1]We adopt the similar short notation in [15]$P\{n,i_1,k_1|n,i_0,k_0\} = P\{s(t+1)=i_1, b(t+1)=k_1|s(t)=i_0, b(t)=k_0$ for the nth traffic flow$\}$.

$$E[W_n] = \frac{(1-p_n)W_{n,0} + p_n(1-p_n)W_{n,1} + \cdots + p_n^{m_n-1}(1-p_n)W_{n,m_n-1} + p_n^{m_n}(1-p_n)W_{n,m_n}}{(1-p_n) + p_n(1-p_n) + \cdots + p_n^{m_n-1}(1-p_n) + p_n^{m_n}(1-p_n)} \tag{4}$$

For each $k \in (0, W_{n,i-1}]$, we have

$$b_{n,i,k} = \frac{W_{n,i}-k}{W_{n,i}\lambda_n} \times \begin{cases} (1-p_n)\sum_{j=0}^{m_n-1} b_{n,j,0} + b_{n,m_n,0}, & i=0 \\ p_n b_{n,i-1,0}, & 0 < I \le m_n \end{cases} = \frac{W_{n,i}-k}{W_{n,i}\lambda_n} b_{n,i,0} \quad 0 \le i \le m_n. \tag{9}$$

By using the normalization condition for stationary distribution, we have

$$\begin{aligned} 1 &= \sum_{i=0}^{m_n}\sum_{k=0}^{W_{n,i}-1} b_{n,i,k} \\ &= \sum_{i=0}^{m_n} b_{n,i,0}\left(1+\sum_{k=1}^{W_{n,i}-1}\frac{W_{n,i}-k}{W_{n,i}\lambda_n}\right) \\ &= \sum_{i=0}^{m_n} b_{n,i,0}\left(1+\frac{W_{n,i}-1}{2\lambda_n}\right) \\ &= b_{n,0,0}\sum_{i=0}^{m_n} p_n^i\left(1+\frac{W_{n,i}-1}{2\lambda_n}\right). \end{aligned} \tag{10}$$

With all the TC parameters known, a general function of $b_{n,0,0}$ with respect to p_n can be found for a specific traffic flow, see (11), located at the bottom of page.

The traffic priority τ_n can be expressed as

$$\tau_n = \sum_{i=0}^{m_n} b_{n,i,0} = \frac{1-p_n^{m_n+1}}{1-p_n} b_{n,0,0} \quad 1 \le n \le N. \tag{12}$$

In the stationary state, a station transmits a packet with probability τ_n, so we have

$$p_n = 1 - \prod_{j=1, j\ne n}^{N} (1-\tau_j) \quad 1 \le n \le N. \tag{13}$$

Therefore, (10), (12), and (13) represent a nonlinear system with the two unknown vectors $(\tau_1, \tau_2, \ldots, \tau_n)$ and $(p_1, p_2, \ldots, p_n)$ which can be solved by numerical results. Note that we must have all τ_n's and p_n's $\in (0,1)$.

B. TC Throughput Analysis

Let P_b denote the probability that the channel is busy in a slot time is

$$P_b = 1 - \prod_{j=1}^{N}(1-\tau_j). \tag{14}$$

Let $P_{n,s}$ denote the probability that the transmission of the nth traffic flow is successful in a slot time. So we have

$$P_{n,s} = \tau_n \prod_{j=1, j\ne n}^{N}(1-\tau_j). \tag{15}$$

We assume that the traffic flows in the same traffic category have the same value of the average payload length. As we mentioned at the beginning of this section, there are T (TCs). The tth TC has N_{t} number of traffic flows with the same value of the average payload length, i.e., from the $(N_1+N_2+, \ldots, +N_{t-1}+1)$th traffic flow to the $(N_1+N_2+, \ldots, +N_{t-1}+N_t)$th traffic flow.

Let $P_{t,s}$ denote the overall probability that the transmission of the tth TC is successful in a slot time. And $T_{n,s}$ denotes the corresponding average time that the channel is sensed busy because of a successful transmission. We have

$$P_{t,s} - \sum_{n\in t^{th}TC} P_{n,s} = \sum_{n=N_1+N_2+\ldots+N_{t-1}+1}^{N_1+N_2+\ldots+N_{t-1}+N_t} P_{n,s}. \tag{16}$$

Let $P_{t,c}$ denote the probability that a transmission with collision occurs in a slot time that the traffic flow with the largest payload belongs to the tth TC. And $T_{t,c}$ denotes the corresponding average time that the channel is sensed busy because of a collision in which a traffic flow, which belongs to the tth TC, has the largest payload length. The subscript t ranges [1, T]. Without loss of generality, let us assume that from the first to the Tth TC, the value of $T_{t,s}$ monotonously decreases. In that way, $P_{t,c}$ can be calculated, e.g.,

$$\begin{aligned} P_{t=1,c} &= 1 - \prod_{n=1}^{N_1}(1-\tau_n) - \sum_{n=1}^{N_1} P_{n,s} \\ &= 1 - \prod_{n=1}^{N_1}(1-\tau_n) - P_{t=1,s} \\ P_{t=2,c} &= \left(1 - \prod_{n=N_1+1}^{N_1 N_2}(1-\tau_n)\right) \times \prod_{n=1}^{N_1}(1-\tau_n) - \sum_{n=N_1+1}^{N_1+N_2} P_{n,s} \\ &= \left(1 - \prod_{n=N_1+1}^{N_1+N_2}(1-\tau_n)\right) \times \prod_{n=1}^{N_1}(1-\tau_n) - P_{t=2,s} \\ P_{t=3,c} &= \left(1 - \prod_{n=N_1+N_2+1}^{N_1+N_2+N_3}(1-\tau_n)\right) \\ &\quad \times \prod_{n=1}^{N_1+N_2}(1-\tau_n) - \sum_{n=N_1+N_2+1}^{N_1+N_2+N_3} P_{n,s} \\ &= \left(1 - \prod_{n=N_1+N_2+1}^{N_1+N_2+N_3}(1-\tau_n)\right) \times \prod_{n=1}^{N_1+N_2}(1-\tau_n) - P_{t=3,s} \end{aligned}$$

$$b_{n,0,0} = \begin{cases} \dfrac{2\lambda_n(1-f_np_n)(1-p_n)}{W_n(1-(f_np_n)^{m_n+1})(1-p_n)+(1-f_np_n)\left(1-p_n^{m_n+1}\right)(2\lambda_n-1)}, & \text{for} \quad m_n \le m_n' \\ \dfrac{2\lambda_n(1-f_np_n)(1-p_n)}{W_n\left(1-(f_np_n)^{m_n'+1}\right)(1-p_n)+(1-f_np_n)\left(1-p_n^{m_n+1}\right)(2\lambda_n-1)+W_nf_n^{m_n'}p_n^{m_n'+1}(1-f_np_n)\left(1-p_n^{m_n-m_n'}\right)}, & \text{for} \quad m_n > m_n' \end{cases} \tag{11}$$

$$\begin{aligned} &\ldots\ldots \\ P_{t=T,c} &= \left(1 - \prod_{n=N_1+\cdots+N_{T-1}+1}^{N_1+\cdots+N_{T-1}+N_T} (1-\tau_n)\right) \\ &\quad \times \prod_{n=1}^{N_1+\cdots+N_{T-1}} (1-\tau_n) - \sum_{n=N_1+\cdots+N_{T-1}+N_T}^{N_1+\cdots+N_{T-1}+N_T} P_{n,s} \\ &= \left(1 - \prod_{n=N_1+\cdots+N_{T-1}+1}^{N_1+\cdots+N_{T-1}+N_T} (1-\tau_n)\right) \\ &\quad \times \prod_{n=1}^{N_1+\cdots+N_{T-1}} (1-\tau_n) - P_{t=T,s}. \end{aligned} \tag{17}$$

To verify the correctness of (17), we have

$$(1-P_b) + \sum_{t=1}^{N_t} P_{t,s} + \sum_{t=1}^{N_t} P_{t,c} = 1.$$

Therefore, the normalized throughput of the nth traffic flow, S_n, is

$$\begin{aligned} S_n &= \frac{E[payload\ in\ a\ slot\ time]}{E[lenght\ in\ a\ slot\ time]} \\ &= \frac{P_{n,s}E[P_n]}{(1-P_b)\sigma + \sum_{t=1}^{N_t} P_{t,s}T_{t,s} + \sum_{t=1}^{N_t} P_{t,c}T_{t,c}}. \end{aligned} \tag{18}$$

In the numerator of (18), $\mathrm{E}[P_n]$ is the average packet length. In the denominator of (18), σ is the duration of an empty slot time.

Let packet header be $H = \mathrm{PHY_{hdr}} + \mathrm{MAC_{hdr}}$ and let the average propagation delay be δ. Based on the average payload length of the tth TC, we have the following expressions of $T_{t,s}$ and $T_{t,c}$. Note that the packet header H must include the additional MAC overhead due to the fragmentation if necessary

$$\begin{cases} T_{t,s}^{bas} = AIFS_t + H + E[P_t] + \delta + SIFS + ACK + \delta \\ T_{t,c}^{bas} = AIFS_t + H + E\left[P_t^*\right] + SIFS + ACK \end{cases} \tag{19}$$

$$\begin{cases} T_{t,s}^{rts} = AIFS_t + RTS + SIFS + \delta + CTS + SIFS \\ \qquad\quad +\delta + H + E[P_t] + \delta + SIFS + ACK + \delta \\ T_{t,c}^{rts} = AIFS_t + RTS + SIFS + CTS \end{cases} \tag{20}$$

where bas and rts means basic access method and RTS/CTS access method (mandatory for all data frames), respectively. $\mathrm{E}[P_t^*]$ is the average length of the longest packet payload in a collision. Based on the definition of $P_{t,c}$ used in (17), $\mathrm{E}[P_t^*] = \mathrm{E}[P_t]$. The ACK, RTS, and CTS are calculated as the transmission time of the corresponding frame length and the PHY overhead.

The aggregated throughput of N traffic flows is

$$S = \sum_{n=1}^{N} S_n. \tag{21}$$

C. TC Delay Analysis

In this paper, the frame delay time is defined as the time interval between two successive successful frame transmissions for a traffic category. It is possible that these two frames are not consecutive if a frame is dropped after exceeds the retry count. All traffic flows in the same traffic category have the same average frame delay. By this definition, the average frame delay of the nth traffic flow, $\mathrm{E}[D_n]$, can be calculated as

$$E[D_n] = \frac{E[P_n]}{S_n}. \tag{22}$$

V. Simulation Validation and Parameter Configuration

A. Traffic Models

The simulation uses a self-developed test-bed referred to the IEEE 802.11 MAC implementation in NS-2. In our simulation, we consider a heterogeneous traffic scenario with three types of traffic flows, i.e., voice, video, and data, in one basic service set (BSS). The traffic flow is characterized by its packet arrival pattern and payload statistics (the mean and variance of packet length). Different models are used to generate three types of traffic.

1) *Non-multimedia data traffic*: Data packets arrive from the upper layer as Poisson sequence, with exponentially distributed packet length. The mean packet length is 1024 octets. The average data throughput is 20 Kb/s.
2) *Voice traffic*: Voice traffic is characterized as a two state Markov ON/OFF [1], [6], [21]. The ITU-T G.711 speech codec has been selected to model good-quality voice calls, with 64-b/s bit-rate, 160-B-long packets plus 4-B-long compressed RTP/UDP/IP headers generated every 20 ms during a talking (On) period, and no packet generated in a listening (OFF) period. The mean value of ON/OFF period is 1.5 and 1.35 s, respectively. Voice traffic is time-critical, with the highest priority.
3) *Video traffic*: We model the video source rate by the first-order autoregressive Markov model [4], [8], [20]. Let $\lambda(n)$ represent the bit-rate of a video source during the nth video frame. We have $\lambda(\mathrm{n}) = \mathrm{a}\lambda(\mathrm{n}-1) + \mathrm{bw(n)}$ (bit/pixel), where $\mathrm{a} = 0.8781$, $\mathrm{b} = 0.1108$, and $w(n)$ is a sequence of independent Gaussian random variables with mean 0.572 and variance 1. We assume that the size of video source is H.263 QCIF (176 × 144) [17], and the rate is 10 f/s, which is appropriate for low power wireless terminals. Under this model, the video coding bit-rate is 132 Kb/s.

In order to better approximate the saturation condition, the interarrival time of the traffic model has been decreased when the number of traffic flows is low.

B. Validation of the Model

In this section, we validate the analytical model of EDCF by three steps: 1) in the default IEEE 802.11 DCF scheme without service differentiation, i.e., all traffic flows use the same contention parameters ($AIFS$, $CW_{\min}$ and $CW_{\max}$, and PF), and all traffic flows are data traffic flows with the same arrival pattern and payload statistics; 2) three traffic categories with different contention parameters, but all traffic flows are data traffic flows with the same arrival pattern and payload statistics; and 3) three traffic categories with different contention parameters, and each category is corresponding to one type of traffic flow, starting from the highest to the lowest priority as voice, video, and data.

In the first step, although there is no service differentiation, for convenience we still divide the traffic flows into three categories with exactly same set of contention parameters. In all three steps, the number of traffic flows in each category is integrated with a fixed ratio of 1:1:2.

The general DCF parameters used in both analytical model and simulation are shown in the first part of Table I. We choose a PHY with 2 Mb/s capacity DSSS DQPSK. The PHY overhead is always transmitted at 1 Mb/s (DBSK) [18]. The fragmentation threshold is chosen as the default value of 2312 B. Considering the average payload length defined in the traffic models, the basic access method is used by setting the RTS threshold at 2346 B, while the RTS/CTS access method is always used by setting the RTS threshold at 0 B.

The contention parameters used in three steps are also shown in Table I. As mentioned earlier, all traffic flows in Steps 1 and 2 have the homogeneous data flows, while heterogeneous traffic flows (data/video/voice) are used in Step 3. We retain the persistence factor of all traffic categories in all tests as the default value of 2 because of simple arithmetic, no division/mod operation needed, and simple random number generation.

Table II shows the values of $T_{n,s}$ and $T_{n,c}$, which are used in the analytical solution according to (19) and (20), for both the basic access method and RTS/CTS access method. Using the RTS/CTS method will effectively decrease the time wasted in collisions and backoff in the saturation condition. In the calculation, $\mathrm{PHY_{hdr}}$ is 192 μs because the PHY overhead is always transmitted in 10 Mb/s Db/sK, while $\mathrm{E}[P]$ is calculated with 2 Mb/s channel transmission rate.

Fig. 4 shows both the simulation and analytical (numerical) results of the normalized goodput in Steps 1, 2, and 3. And for each step, both basic scheme and RTS/CTS scheme have been tested. In all tests, the number of traffic flows in each category is in the ratio of 1:1:2. When the number of traffic flows is small, there is a small difference between the simulation and analytical result, which is because of the assumption of saturation condition we used in the analysis.

In Step 1, since all traffic categories have the same contention parameters, there is no goodput differentiation among three categories. In Fig. 4 (Step1), traffic category *data3* has a goodput twice as *data1* and *data2*, which is because *data3* has twice number of traffic flows as *data1* and *data2*. Therefore, without service differentiation and with homogeneous traffic, each single traffic flow acquires the same goodput as anyone else.

In Step 2, although we still have the homogeneous traffic, the service differentiation has been applied to three traffic categories. Fig. 4 (Step2) shows that the highest traffic category (*data1*) has acquired most of the channel goodput in the saturation condition, while the goodput of *data2* is slightly higher than that of *data3* even as the number of traffic flows is doubled in *data3*.

In Step 3, service differentiation has been applied to three traffic categories with heterogeneous traffic. Fig. 4 (Step3) shows the similar result as Fig. 4 (Step2). However, in this test, the average payload length of the highest priority traffic category (*voice*) is 164 B, which is much smaller than that of *data1* (8192 B) in Step2. Therefore, the channel experiences much more contention cycles, which cause more collision and backoff, in Step 3 than in Step 2. As a result, the normalized goodput decreases in Step 3, which is shown in both the simulation and analytical results in Fig. 4 (Step3).

In all three steps, we can see that the RTS/CTS scheme has a relative flat curve of normalized goodput, while the basic scheme has a decreasing curve with the increase of number of traffic flows. The consistent results are shown in both simulation and analytical figures.

Both numerical and simulation results of the average frame delay are shown in Fig. 5. In both Steps 2 and 3, the traffic flows with the highest priority have a very small average frame delay compared with traffic flows with lower priorities.

TABLE I
IEEE 802.11 PARAMETERS

General DCF Parameters			
Slot Time	20 µs	SIFS	10 µs
Max Propagation Delay	2 µs	Retry Limit (short/long)	7/4
RTS_threshold	0 Bytes for RTS/CTS, 2346 Bytes for basic		
Frag_threshold		2312 Bytes	
DSSS PreambleLength		144 bits	
DSSS PLCPHeaderLength		48 bits	
PLCP Trans Rate		1 Mbps (DBPSK)	
PPDU Trans Rate		2 Mbps (DQPSK)	
	Step1 Data1/Data2/Data3	**Step2** Data1/Data2/Data3	**Step3** Voice/Video/Data
AIFS (µs)	50/50/50	50/100/150	50/100/150
CW_{min} (slot)	15/15/15	15/31/63	15/31/63
CW_{max} (slot)	1023/1023/1023	255/511/1023	255/511/1023
PF	2/2/2	2/2/2	2/2/2
Ratio of # of flows	1:1:2	1:1:2	1:1:2

TABLE II
$T_{n,s}$ AND $T_{n,c}$ MEASURED IN μs

Step 1	**data1**	**data2**	**data3**
Avg. Packet Payload (bits)	8192	8192	8192
$T_{n,s}$ (basic) (µs)	4734	4734	4734
$T_{n,c}$ (basic) (µs)	4732	4732	4732
$T_{n,s}$ (rts) (µs)	5276	5276	5276
$T_{n,c}$ (rts) (µs)	580	580	580
Step 2	**data1**	**data2**	**data3**
Avg. Packet Payload (bits)	8192	8192	8192
$T_{n,s}$ (basic) (µs)	4734	4784	4834
$T_{n,c}$ (basic) (µs)	4732	4782	4832
$T_{n,s}$ (rts) (µs)	5276	5326	5376
$T_{n,c}$ (rts) (µs)	580	630	680
Step 3	**voice**	**video**	**Data**
Avg. Packet Payload (bits)	1312	13178.88	8192
$T_{n,s}$ (basic) (µs)	1294	7277.44	4834
$T_{n,c}$ (basic) (µs)	1292	7275.44	4832
$T_{n,s}$ (rts) (µs)	1836	7819.44	5376
$T_{n,s}$ (rts) (µs)	580	630	680

C. Configuration of Contention Parameters

Multiple contention parameters can be used to provide the service differentiation in the IEEE 802.11e. Every contention parameter has a different effect on the performance of the service differentiation. In this section, we investigate the effect on

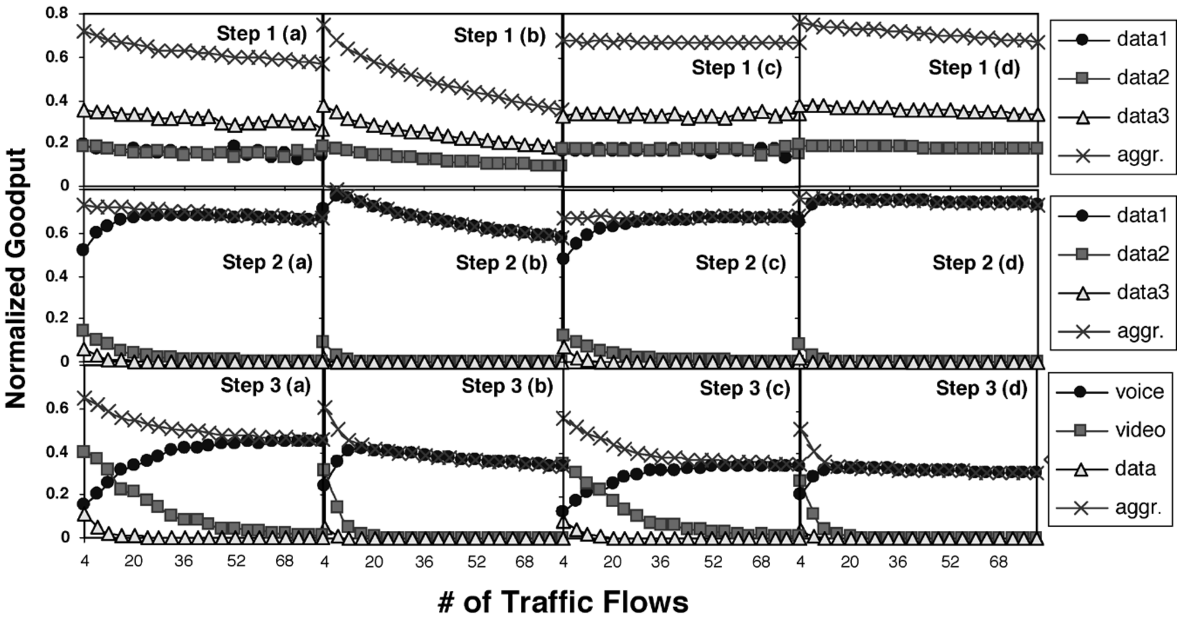


Fig. 4. Normalized saturation goodput. Step1: No service differentiation, with homogeneous traffic. Step2: Service differentiation, with homogeneous traffic. Step3: Service differentiation, with heterogeneous traffic. (a) Simulation-basic scheme. (b) Numerical-basic scheme. (c) Simulation-RTS/CTS scheme. (d) Numerical-RTS/CTS scheme.

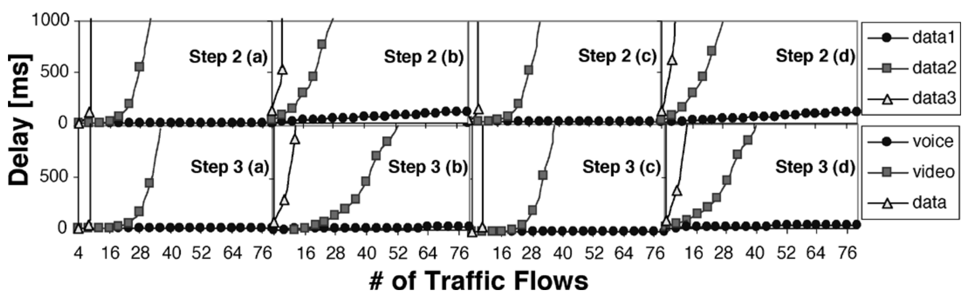


Fig. 5. MAC transmission delay. Step 2: Service differentiation, with homogeneous traffic. Step 3: Service differentiation, with heterogeneous traffic. (a) Numerical-basic scheme. (b) Simulation-basic scheme. (c) Numerical-RTS/CTS scheme. (d) Simulation-RTS/CTS scheme.

service differentiation based on each contention parameter (i.e., $AIFS, CW_{\max}, CW_{\min}$, and PF). The results are very helpful for choosing an appropriate configuration of contention parameters of TCs in each station. The configurations of contention parameters of all four tests are shown in Table III.

First, we want to observe the effect of changing parameter $AIFS$ in Test 1. We investigate a BSA with two traffic categories (1-high and 2-low), each with ten traffic flows. We only change $AIFS_1$ of the first category, and lock all other contention parameters in the system. The results are shown in Fig. 6, where Test 1(a) depicts the traffic priority τ_n (i.e., the stationary probability that a station transmits a packet in a randomly selected slot time). Test 1(b) depicts the conditional collision probability p_n (i.e., the probability of collision seen by a packet being transmitted on channel). Tests 1(c) and (d) depict the normalized goodput for the basic and RTS/CTS access methods, respectively. As proposed in IEEE802.11e, the difference between $AIFS_i$ and basic $DIFS$ should be an even number multiplies the slot time. The default value of $AIFS$ is $DIFS$. In our study, we have chosen some values of $AIFS_2$ beyond this rule just in order to obtain a better result with smooth curve.

TABLE III
CONFIGURING CONTENTION PARAMETERS

	Test 1	Test 2	Test 3	Test 4
AIFS (μs)	*/50	50/50	50/50	50/50
$CW_{\min}$ (slot)	15/15	15/15	*/15	15
$CW_{\max}$ (slot)	1023/1023	1023/*	1023/1023	1023/1023
PF	2/2	2/2	2/2	2/*
# of flows	10/10	10/10	10/10	10/10

(1: high priority flow / 2: low priority flow), * is the changing parameter

Based on the suggestion in IEEE 802.11e, $CW_{1,\max}$ in Test 2, $CW_{2,\min}$ in Test 3 and PF_2 in Test 4 are chosen from [31, 1023], [15, 511], and [1.0, 3.0], respectively. The results are also shown in Fig. 6.

By comparing all four tests, we can see that $AIFS$ and $CW_{\min}$ have much larger influence on service differentiation than $CW_{\max}$. This is because the effect of a different $CW_{\max}$ will not show up unless any specific frame experiences multiple transmission failures, while $AIFS$ and $CW_{\min}$ are able to affect all frames in transmission. Fig. 6 Test 3 shows that a steep increase/decrease of goodput of the high/low priority

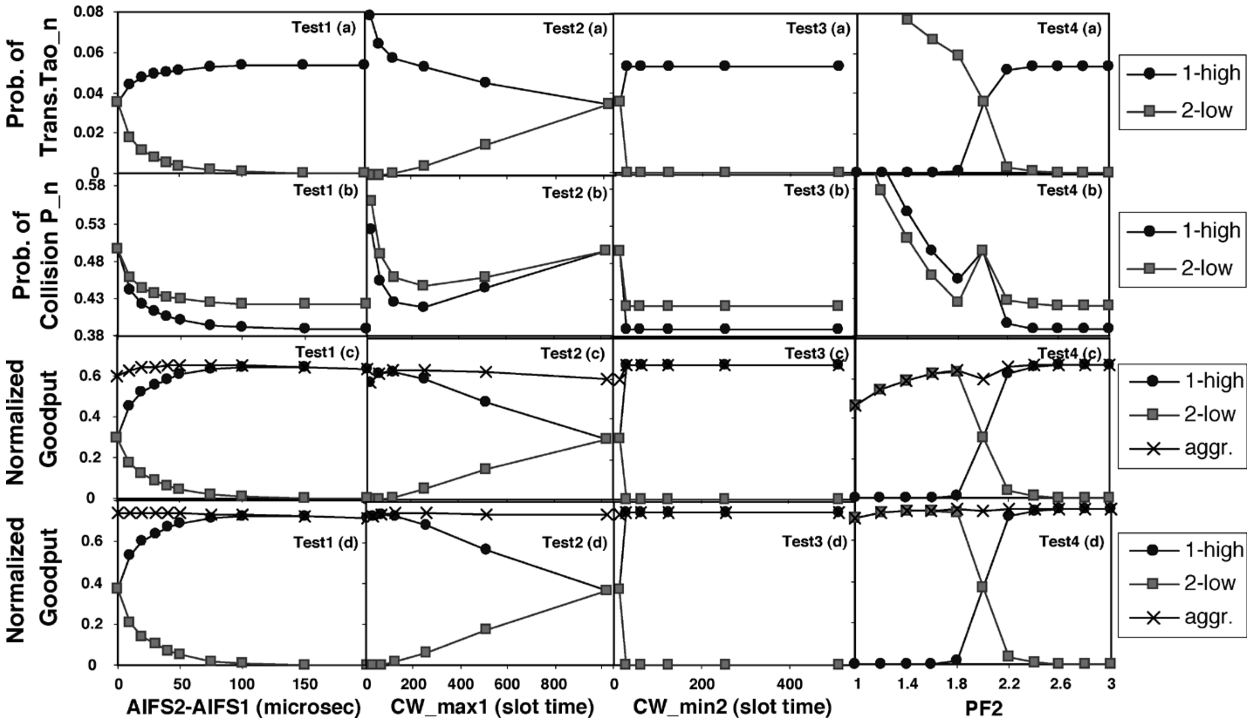


Fig. 6. Test 1: Service differentiation by changing $AIFS_1$ (50–250 μs). Test 2: Service differentiation by changing $CWmax_1$ (31–1023 slot time). Test 3: Service differentiation by changing $CWmin_2$ (15–511 slot time). Test 4: Service differentiation by changing PF_2 (1.0–3.0). (a) Probability of transmission τ_n. (b) Conditional probability of collision p_n. (c) Normalized goodput-basic scheme. (d) Normalized goodput-RTS/CTS scheme.

traffic occurs if $CW_{2,\min}$ is changed from 15 to 31, while Fig. 6 Test 2 shows a steady increase/decrease by changing $AIFS_2$ from 50 to 70 to 90 μs. Therefore, changing $AIFS$ can achieve better fine-granularity-scalability (FGS) than changing $CW_{\min}$. However, as we can see in Test 2, changing $CW_{\max}$ is able to achieve even better FGS than changing $AIFS$. Since changing PF directly affects the size of all the contention windows of retransmissions, Test 4 shows similar result as Test 3. Slight increment/decrement of PF will result in significant service differentiation between different priorities, while further changes of PF do not make much more difference.

The following is a summary.

1) $AIFS$ and $CW_{\max}$ are more appropriate for moderate service differentiation, with less starvation of low priority traffic flows. Moreover, by carefully choosing the value of $AIFS$ and $CW_{\max}$, a smooth service gradient from the low to the high priority may be achieved.
2) In the situation that high priority traffic flows have hard QoS requirement, i.e., the throughput must be guaranteed, $CW_{\min}$ is more appropriate for distinct service separation. Low-priority traffic flows will be starved in presence of high priority flows.
3) The implementation of nontwo PF will introduce more costly computation, while tuning PF has the similar result with tuning $CW_{\min}$. Therefore, we believe that it is not a good option for service differentiation.

VI. Conclusion

In this paper, we introduced an analytical model of EDCF access method of the IEEE 802.11e. This model can be used to calculate the traffic priority and throughput under the saturation condition. Simulation validation of the proposed analytical model was provided. Based on the proposed model, we analyzed the effect on service differentiation for each contention parameter. The contention parameters can be configured appropriately at each station to achieve better performance of service differentiation. The accuracy of the analytical model is based on the assumption of saturation condition and good channel condition. Further research will be done to study the transmission delay and to improve the accuracy of the model by investigate the system when both assumption are not fully valid.

References

[1] J. Deng and R. S. Chang, "A priority scheme for IEEE 802.11 DCF access method," *IEICE Trans. Commun.*, vol. E82-B, no. 1, pp. 96–102, Jan. 1999.

[2] 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. IEEE INFOCOM*, New York, Jun. 23–27, 2002, pp. 599–607.

[3] H. L. Truong and G. Vannuccini, "The IEEE 802.11e MAC for quality of service in wireless LANs," in *Proc. Int. Conf. Advances in Infrastructure and Mobile Technologies on the Internet* , L'Aquila, Italy, Jan. 6–12, 2003.

[4] T. Suzuki and S. Tasaka, "Performance evaluation of priority-based multimedia transmission with the PCF in an IEEE 802.11 standard wireless LAN," in *Proc. 12th IEEE Int. Symp. Personal, Indoor and Mobile Radio Communications*, vol. 2, Sep. 2001, pp. G70–G77.

[5] F. Calí, M. Conti, and E. Gregori, "IEEE 802.11 wireless LAN: Capacity analysis and protocol enhancement," in *Proc. INFOCOM*, San Francisco, CA, March 29–April 2 1998, pp. 142–149.

[6] G. Anastasi and L. Lenzini, "QoS provided by the IEEE 802.11 wireless LAN to advanced data applications: A simulation analysis," *ACM/Baltzer Wireless Netw.*, vol. 6, no. 2, pp. 99–108, 2000.

[7] I. Aad and C. Castelluccia, "Differentiation mechanisms for IEEE 802.11," in *Proc. IEEE INFOCOM*, vol. 1, Anchorage, AK, Apr. 2001, pp. 209–218.

[8] A. Banchs and X. P'erez, "Providing throughput guarantees in IEEE 802.11 wireless LAN," in *Proc. Wireless Communications Networking Conf.*, vol. 1, 2002, pp. 130–138.

[9] J.-Y. Yeh and C. Chen, "Support of multimedia services with the IEEE 802.11 MAC protocol," in *Proc. IEEE Int. Conf. Communications*, vol. 1, 2002, pp. 600–604.

[10] Y. C. Tay and K. C. Chua, "A capacity analysis for the IEEE 802.11 MAC protocol," *Wireless Netw.*, vol. 7, no. 2, pp. 159–171, Mar./Apr. 2001.

[11] W. M. Moh, D. Yao, and K. Makki, "Wireless LAN: Study of hidden-terminal effect and multimedia support," in *Proc. 7th Int. Conf. Computer Communications Networks*, Lafayette, LA, Oct. 1998, pp. 422–431.

[12] A. Lindgren, A. Almquist, and O. Schelén, "Quality of service schemes for IEEE 802.11: A simulation study," in *Proc. 9th Int, Workshop Quality of Service* , 2001, pp. 281–287.

[13] C. Andren, "IEEE 802.11 wireless LAN: Can we use it for multimedia?," *IEEE Multimedia*, vol. 5, no. 2, pp. 84–89, Apr.–Jun. 1998.

[14] S. Sharma, *Analysis of 802.11b MAC: A QoS, Fairness, and Performance Perspective*. Stony Brook, NY: TR, Dept. Comput. Sci, Stony Brook Univ., Jan. 2003.

[15] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 3, pp. 535–537, Mar. 2000.

[16] K. K. Leung, B. McNair, L. J. Cimini, and J. H. Winters, "Outdoor IEEE 802.11 cellular networks: MAC protocol design and performance," in *Proc. IEEE Int. Conf. Communications*, New York, Apr. 2002, pp. 595–599.

[17] M. Freytes, C. E. Rodriguez, and C. A. Marques, "Real-time H.263+ video transmission on 802.11 wireless LAN," in *Proc. Int. Conf. Information Technology: Coding and Computing*, Las Vegas, NV, Apr. 2–4, 2001, pp. 125–129.

[18] M. Gast, *802.11 Wireless Networks: The Definitive Guide*. Sebastopol, CA: O'Reilly, 2002.

[19] N. Reid and R. Seide, *802.11 (Wi-Fi) Networking Handbook*. New York: McGraw-Hill/Osborne, 2003.

[20] B. Maglaris, D. Anastassiou, P. Sen, G. Karlsson, and J. D. Robbins, "Performance models of statistical multiplexing in packet video communications," *IEEE Trans. Commun.*, vol. 36, no. 7, pp. 834–844, Jul. 1988.

[21] C. Coutras, S. Gupta, and N. B. Shroff, "Scheduling of real-time traffic in IEEE 802.11 wireless LANs," *Wireless Netw.*, vol. 6, pp. 457–466, 2000.

[22] P. Chatimisios, V. Vitsas, and A. C. Boucouvalas, "Throughput and delay analysis of IEEE 802.11 protocol," in *Proc. 5th IEEE Int. Workshop Network Appliances*, Liverpool, U.K., Oct. 2002, pp. 168–174.

[23] Y. Ge and J. Hou, "An analytical model for service differentiation in IEEE 802.11," in *IEEE Int. Conf. Communications*, vol. 2, 2003, pp. 1157–1162.

[24] Z. Hadzi-Velkov and B. Spasenovski, "Saturation throughput—delay analysis of IEEE 802.11 DCF in fading channels," in *Proc. IEEE Int. Conf. Commun.*, vol. 1, 2003, pp. 121–126.

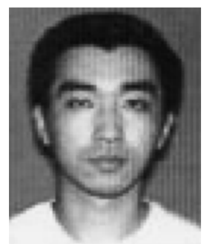
[25] S. Choi, J. D. Prado, S. Shankar, and S. Mangold, "IEEE 802.11e contention-based channel access (EDCF) performance evaluation," in *Proc. IEEE Int. Conf. Commun.*, vol. 2, 2003, pp. 1151–1156.

[26] Y. Xiao, "Backoff-based priority schemes for IEEE 802.11," in *Proc. IEEE Int. Conf. Commun.*, vol. 3, Anchorage, AK, May 2003, pp. 1568–1572.

[27] F. Eshghi and A. K. Elhakeem, "Performance analysis of ad hoc wireless LANs for real-time traffic," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 2, pp. 204–215, Feb. 2003.

[28] B. E. Mullins, N. J. Davis IV, and S. F. Midkiff, "A wireless local area network protocol that improves throughput via adaptive control," in *IEEE Int. Conf. Commun.*, vol. 3, Montreal, QC, Canada, June 8–12, 1997, pp. 1427–1431.

[29] IEEE 802.11 WG, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification, Aug. 1999.



Hua Zhu received the Ph.D. degree in Electrical Engineering from the University of Texas at Dallas, Texas.

Since 2005, he has been working for San Diego Research Center, Inc., San Diego, CA, as a Research Engineer. His research interests include all layers of wireless communication systems. His particular interest is in L2/3 air interface design, performance analysis, and optimization for ad hoc and sensor networks.

Imrich Chlamtac (M'86–SM'86–F'93) received the Ph.D. degree in computer science from the University of Minnesota, Minneapolis.

Since 1997, he has been the Distinguished Chair in Telecommunications, University of Texas at Dallas. He also holds the titles of: the Sackler Professor at Tel Aviv University, Tel Aviv, Israel; The Bruno Kessler Honorary Professor at the University of Trento, Trento, Italy, where he is currently on sabbatical; and University Professor at the Technical University of Budapest, Budapest, Hungary. He has published close to 300 papers in referred journals and conferences and is the coauthor of *Local Area Networks* (New York: Lexington Books, 1981, 1982, 1984) and of *Mobile and Wireless Networks Protocols and Services* (New York: Wiley, 2000).s He serves as the founding Editor-in-Chief of the *ACM/URSI/Kluwer Wireless Networks* (WINET), the *ACM/Kluwer Mobile Networks and Applications* (MONET) journals, and the *SPIE/Kluwer Optical Networks Magazine*.

Dr. Chlamtac was IEEE *Network Magazine's* 2000 Editor's Choice and was the winner of the 2001 ACM Sigmobile annual award and the IEEE ComSoc TCPC 2002 Award for contributions to wireless and mobile networks and of multiple best paper awards in wireless and optical networks. He is a Fellow of the ACM, a Fulbright Scholar, and an IEEE Distinguished Lecturer.