Performance Evaluation of Quality of Service in IEEE 802.11e Wireless LANs

Fei Peng and Victor C.M. Leung

Department of Electrical & Computer Engineering
The University of British Columbia
2332 Main Mall, Vancouver, BC, Canada V6T 1Z4
{feip,vleung}@ece.ubc.ca

Abstract. There have been many performance studies on the original IEEE 802.11 distributed coordinated function using both simulation and analytical methods. However, the recent IEEE 802.11e standard has not yet been investigated extensively. This paper proposes an accurate analytical model for the enhanced distributed channel access (EDCA), which is the core protocol of 802.11e. The model captures the operations of service differentiation using different contention window (CW) sizes, arbitrary interframe space (AIFS), and transmission opportunity (TXOP) limits in EDCA. Using this model, we derive throughput performance of EDCA access categories differentiated through the above mechanisms. The throughput and collision probability derived by our model are validated by simulation results, which show close agreements with the analytical results. Our model provides a useful tool for evaluating the impact of different parameters on the performance of EDCA service differentiation.

Keywords: IEEE 802.11e, MAC protocol, service differentiation, QoS, performance evaluation.

1 Introduction

The popularity of wireless local area networks (WLANs), especially those following the IEEE 802.11 standard, has generated much interest on improvements and modeling of the protocol. Many models have been proposed to analyze the protocol mechanisms of 802.11 distributed coordinated function (DCF), which employs carrier-sense multiple access with collision avoidance (CSMA/CA). Chhaya and Gupta [1] obtained the throughput of CSMA/CA using a simple model of the probabilities of capture in the presence of hidden stations. Bianchi [2] proposed a simple and accurate analytical model to compute the saturation throughput of 802.11 DCF. Ho, Huang, and Chen [3] presented approximate models that account for hidden terminals and capture effects. Cali, Conti and Gregori [4] improved 802.11 medium access control (MAC) performance by tuning persistent backoff strategies, and provided a thorough performance analysis. Tay and Chua [5] provided a good approximation of the saturation throughput of 802.11 DCF. However, these models [1-5] cannot be directly applied to analyze different access priorities in 802.11e.

J. Zheng et al. (Eds.): ADHOCNETS 2009, LNICST 28, pp. 407–421, 2010.

An area of major interest in recent research on IEEE 802.11 WLANs is focused on improvements of the standard to support quality of service (QoS). Enhancements for such purposes, commonly referred as 802.11e, have been incorporated in the 2007 release of the 802.11 standard [6]. 802.11e provides priority-based service differentiation through the enhanced distributed channel access (EDCA) mechanisms. Three traffic priority mechanisms are included in EDCA to provide quality of service differentiation: backoff contention window (CW) priority, arbitrary interframe space (AIFS) and transmission opportunity (TXOP) limit. A station supporting EDCA is called a (QoS) station (QSTA), which may use one of several access categories (ACs) to contend for channel access. The transmissions of MAC service data units (MSDUs) in each AC use a different set of values for the above traffic priority parameters, which result in a different level of access priority compared to other ACs.

Many models for 802.11e EDCA have been proposed by modifying or extending Bianchi's Markov chain model [2] to accommodate the differentiation of AIFS and/or CW; e.g., [7] extends Bianchi's model to study the backoff CW priority scheme, but it does not include the AIFS differentiation mechanism. To accommodate different AIFS values assigned to different ACs, [8] enlarges the original bi-dimensional Markov chain to tri-dimensional and [9] enlarges it even to multidimensional, which requires substantial computations due to its high complexity. Both [10] and [11] extend Bianchi's model to analyze the saturation throughput performance of EDCA, by considering AIFS differentiation without using a multi-dimensional Markov chain; however, these models are still very complex to analyze and furthermore, it is difficult to extend these models to more than two ACs. Also, the collision probabilities calculated by these models are not accurate for the analysis of throughput performance. Other than employing Markov chain methods, Bianchi in [12] has provided a rigorous analytical approach to model AIFS-based priority mechanisms, but the resulting model also suffers from a high complexity. In [13] the notion of k-slot time is presented, with which the probability of an empty slot in a randomly chosen k-slot time is derived. Rather rough approximate expressions for AIFS differentiation have been presented in [14][15]. Therefore, existing models for EDCA mostly considered only AIFS differentiation, and yet they suffer from either a high complexity [12-13] when the Markov chain approach is employed, or rough approximations [14-15] when the Markov chain approach is not employed. Finally, the performance impact of TXOP priorities between different ACs is rarely investigated in the current literature.

We note that the use of Markov chains to model EDCA dynamics is preferable as their accuracy has been proven. However, this approach has a scalability problem due to the rapid increase in the number of states as the number of ACs increases. Therefore, existing Markov chain models used to evaluate EDCA performance have considered no more than two ACs over a wide range of parameter settings. In this paper, we have refined Markov chain modeling of EDCA to enable accurate analysis of throughput and collision probability performance when different service differentiation schemes are employed in EDCA. Our model not only accounts for AIFS-based priority access, but also accounts for the use of different TXOP limits among different ACs. We show that numerical results obtained using our analytical model match the simulation results accurately when the ACK timeout during each collision period is taken into consideration. Our modeling method is relatively simple to apply to general parameter settings compared to other Markov chain models.

The outline of the paper is as follows. In Section 2, we briefly review the differences between EDCA and DCF. The analytical model is developed in Sections 3 and 4. Section 3 studies the average conditional probabilities of collision, successful transmission and packet transmission for different stations with multiple priorities. Based on these probabilities, we further present the throughput analysis in detail especially for TXOP in Section 4. Section 5 validates the analytical model by simulations and examines how the operation of EDCA differentiation mechanisms affects the provision of services. Final remarks are given in Section 6.

2 Comparison of DCF and EDCA

EDCA has been specified in [6] to add QoS support to DCF. It supports up to four queues in each QSTA associated with four specific ACs. A packet at the head of each queue contends for channel access independently from other queues in the QSTA and from other QSTAs. A different level of service is provided to each AC through a combination of three service differentiation mechanisms: CW, AIFS and TXOP.

2.1 Contention Window Based Priority

In DCF, the backoff counters are randomly selected from the interval $[0, CW-1]$, where the contention window CW is a function of the physical layer (PHY) specific *aCWmin* and *aCWmax* attributes. DCF adopts an exponential backoff scheme. At the first transmission attempt, CW is set equal to a value $CWmin$ called the minimum contention window. After each unsuccessful transmission, CW is doubled, up to a maximum value $CWmax=2^m \times CWmin$.

In contrast, the backoff counters in EDCA are selected at random from the interval $[1, CW]$, where the contention window CW is a function of the AC; specifically $CW[AC] \in [CWmin[AC], CWmax[AC]]$. With EDCA, after each successful transmission, the corresponding $CW[AC]$ will be set to $CWmin[AC]$. Once a transmission fails, CW will be calculated as follows: $CW[AC] = \min\{CWmax[AC], CW[AC] \times PF[AC]\}$ where $PF[AC]$ is the factor by which the current window size is increased when a frame transmission has failed; it equals 2 in the original IEEE 802.11 DCF and can be set to any real number larger than 1 in EDCA. QSTAs in different ACs can receive different service priority levels by choosing $CWmin[AC]$ and $CWmax[AC]$ appropriately. A higher $CWmin$ and/or $CWmax$ and/or PF chosen by an AC would result in less opportunity to access the channel for the AC and therefore a lower access priority for the AC relative to other ACs that have selected lower values for $CWmin$ and/or $CWmax$ and/or PF.

2.2 Arbitrary Interframe Space

In DCF, the backoff counter is decremented as long as the channel is sensed idle; it is "frozen" when a transmission is detected on the channel, and reactivated when the channel is sensed idle again for more than a DCF interframe space (DIFS). For each slot time interval, during which the channel stays idle, the backoff counter is decremented. The station transmits when the backoff counter reaches zero.

In EDCA, AIFS is used instead of DIFS, where *AIFS≥DIFS*. Each AC is assigned a different AIFS value to differentiate the QoS received by the AC. Stations that use a lower AIFS encounter fewer collisions and count down the backoff counter faster than the other stations; hence, they receive a better overall service than the other stations over a period of time. After waiting for *AIFS*[*AC*], each backoff counter is set to a random number from [0, *CW*[*AC*]] with the unit of time slot.

2.3 Transmission Opportunity Limit

EDCA places limits on channel occupancy using an AC-specific TXOP limit parameter, in contrast to a common limit for all stations in DCF. If multiple frame exchanges are allowed within the EDCA TXOP, it will reduce the network overhead since a station can send several MSDUs without contending for the channel between transmissions. Higher priority classes usually would be configured for longer TXOP limits than lower priority classes. It has been suggested [15] that this will increase the aggregate data throughput within a given service area for higher priority ACs. Our model will provide more insight into this suggestion in the following sections.

3 Average Conditional Collision Probability

3.1 Normal Contention Markov Backoff Process

We have done a number of simulation tests to find that Bianchi's model is accurate regarding a single station with a given priority. Therefore, the discrete time bidimensional Markov process we use to model backoff and transmission for stations within a certain AC is kept the same as his model except that the contention window ranges from 0 to *CW* instead of 0 to *CW*-1. Why the backoff counter ends at *CW* is explained in reference [6]. Here, we briefly illustrate the formulas in our model.

Denoting the stationary probability distribution of the backoff states as $b_{i,k}$, where k is the value of the backoff counter while the packet is retransmitted for the i-th time. We note the following relationships between backoff states:

$$b_{i,0} = p_c^i \cdot b_{0,0}, \quad i \in (0,m] \tag{1}$$

$$b_{i,k} = \frac{W_i - k}{W_i} \cdot p_c^i \cdot b_{0,1}, \quad i \in [0,m],\ k \in [0,W_i] \tag{2}$$

A solution for $b_{0,1}$ in terms of the average conditional collision probability p_c is found by imposing the normalization condition on the Markov process,

$$1 = \sum_{i=0}^{m} \sum_{k=0}^{W_i} b_{i,k} \tag{3}$$

Recalling our definition that transmissions occur whenever the backoff counter k reaches 0, we find the probability τ that a station transmits in a randomly chosen slot time

$$\tau = \sum_{i=0}^{m} b_{i,0} \,. \tag{4}$$

3.2 Average Conditional Probability

Having obtained the behavior of a single station with a given priority using Bianchi's Markov model, we develop a new model in this section to deal with systems containing stations with more than one ACs, with distinct numbers of stations in each AC and distinct AIFS values for each AC. In this case each AC corresponds to a different access priority level. Assuming that there are J ACs with $AIFS_j < AIFS_{j-1}$, $j \in (0, J-1]$. In such a system, we can define J contention zones such that stations with priority j, $j \in [J-i, J-1]$, are active in zone i, while stations with priority $j \in [J-i+1, J-1]$ are active in zone $i-1$. So, for the contention zone $i-1$, stations with priority $J-i+1$ or higher are active but stations with priority $J-i$ or lower are not. From the definition, it is clear that all the stations can be active in contention zone J while only stations having the highest priority, i.e., $J-1$, can be active in contention zone 1.

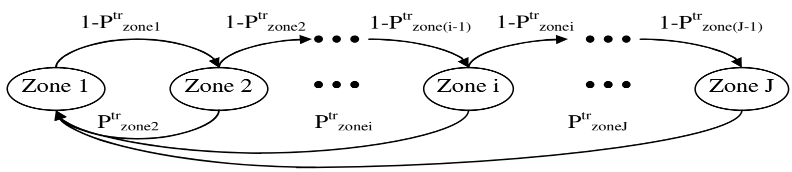

Fig. 1. Relationship between adjacent contention zones in a transmission period

The average probability of collision in each contention zone is found by weighing zone specific collision probabilities with the occupancy probabilities of the contention zones. The advantage of our model is that it can be readily extended to cope with systems with multiple AIFS priorities. Since a contention zone is reached only when no transmission has occurred in all the preceding zones in the current transmission period, and the probability of passing through each zone is assumed to be constant since we do not trace the evolution within each transmission period, we can use a Markov process to find the occupancy probabilities of the contention zones. Such a process is governed by the probability of at least one transmission occurring in contention zone i and is illustrated for convenience by Fig. 1.

Define $p^{tr}_{zone=i}$ as the probability that at least one transmission occurs in contention zone i

$$p^{tr}_{zone=i} = 1 - \prod_{j=1}^{i} (1 - \tau_{J-j})^{n_{(J-j)}} \tag{5}$$

The relationship between the occupancy probabilities of adjacent contention zones is

$$z_i = \left(1 - p^{tr}_{zone=i-1}\right) \cdot z_{i-1} \tag{6}$$

where z_i is the occupancy probability of contention zone i, and $p^{tr}{}_{zone(i\text{-}1)}$ is the probability of transmission in zone $i-1$. Different from [10] however, the relationship between the occupancy probabilities of last adjacent zones is,

$$z_J = \frac{\left(1 - p^{tr}_{zone=J-1}\right)}{p^{tr}_{zone=J}} \cdot z_{J-1} \tag{7}$$

Since we consider that each AC has a different AIFS value, the number of contention zones is equal to the number of priority levels (i.e., the number of ACs) so that the number of states in our model depicted in Fig. 1 is also fixed. This is different from Robinson's model [10] in which the number of states is not constant but related to the contention windows in each transmission period. With our model, the solution to the stationary distribution is given by,

$$z_1 = \frac{1}{1 + \sum_{i=1}^{J-2} \prod_{j=1}^{i} \left(1 - p^{tr}_{zone=j}\right) + \prod_{j=1}^{J-1} \left(1 - p^{tr}_{zone=j}\right) \Big/ p^{tr}_{zone=J}}$$

We note that the probability of a station successfully transmitting during a transmission period is by the definition the probability that this station transmits and no other active station transmits. Assuming that there are n_j (j=0, …, J–1) stations in priority j, these stations would be active only in zones [J–j, J]; thus we can express the probability that any station in priority j successfully transmitting in zone i as:

$$P^{s}_{j:zone=i} = \begin{cases} n_j \tau_j \dfrac{1 - p^{tr}_{zone=i}}{1 - \tau_j} & j \in [J-i, J-1],\ i \in [1, J] \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

A transmitted frame collides when one or more other stations also transmit during the slot time. The probability that a priority j station sends a frame in contention zone i but suffers a collision is

$$P^{c}_{j:zone=i} = 1 - \prod_{k=J-i}^{J-1} (1 - \tau_k)^{n_k} \Big/ (1 - \tau_j), \tag{9}$$

Having found p^{s}_{j} and p^{c}_{j} in contention zone i from (8) and (9), we obtain the average $\overline{p}_{cj}$ and $\overline{p}_{sj}$ by summing the zone specific conditional collision probabilities, weighted by respective contention zone occupancy probabilities:

$$\overline{p}_{cj} = \sum_{i=J-j}^{J} P^{c}_{j:zonei} \times z_i \ ,\ \overline{p}_{sj} = \sum_{i=J-j}^{J} P^{s}_{j:zonei} \times z_i \tag{10}$$

The expressions from (4) and (10) for τ and $\overline{p}_{cj}$, $j \in [0, J-1]$, respectively, are sufficient to form an exactly determined system of nonlinear equations, amenable to solution by numerical methods. Once found, τ and $\overline{p}_{cj}$ can be plugged into straightforward expressions for throughput that we will present in the following section.

4 Throughput Analysis

Let S be the normalized system throughput, defined as the fraction of time the channel is used to successfully transmit payload bits. S can be expressed as:

$$S = \frac{\text{E}[\textit{payload} \text{ transmitted in a transmission period}]}{\text{E}[\text{length of a transmission period}]} \tag{11}$$

We divide the channel occupancy time into three different components: (1) overhead of a successful transmission O_s, (2) collision time overhead O_c, (3) the data transmission burst. We specify the values of these components for two different access modes: basic access and RTS/CTS. Note that the physical header (H_{phy}) is transmitted using the PHY's basic rate R_b while the MAC frames, including MAC header (H_{mac}) and payload (P) and ACK are transmitted using the operational rate R. Letting δ be the propagation delay. For AC j (indicated by the subscript j), O_s and O_c in basic access mode (indicated by the subscript *bas*) and RTS/CTS mode (indicated by the subscript *rts*) are, respectively:

$$\begin{aligned} O_{sbas\,j} &= O_{cbas\,j} = AIFS_j\,, \\ O_{srts\,j} &= 2H_{phy}/R_b + RTS/R + 2\delta + 2SIFS + CTS/R + AIFS_j \\ O_{crts\,j} &= H_{phy}/R_b + RTS/R + AIFS_j + \delta \end{aligned} \tag{12}$$

With TXOP differentiation, the data transmission burst of priority j is composed of a given number k_j multiple of time units U_j, where U_j corresponds to the transmission time of a data frame. Letting E[P_j] be average payload of AC j, U_j can be expressed as,

$$U_j = H_{phy}/R_b + (H_{mac} + E[P_j])/R + \delta + SIFS + Ato \tag{13}$$

where $Ato = H_{phy}/R_b + ACK/R + \delta + SIFS$

In (12) and (13), *RTS*, *CTS* and *ACK* are the lengths of the RTS, CTS and ACK frames, respectively, including the MAC headers. Defining $TXOP_j$ for class j, k_j is then obtained by,

$$k_j = \frac{TXOP_j}{U_j}, \quad k_j \geq 1 \tag{14}$$

where $k_j = 1$ is the normal case in which each station is allowed to transmit one frame at a time, while $k_j > 1$ means that TXOP differentiation is adopted.

Because the use of AIFS differentiation may cause event probabilities to differ from one contention zone to the next, the expected length and payload of a transmission period must be summed over the contention zones of the transmission period weighted by the respective occupancy probabilities. For this we can use (6) and (7), and express the expected length and payload of a transmission period as

$$\begin{aligned} &\text{E}[\text{length of transmission period}] = E[\overline{T}_j] \\ &= \sigma + \overline{O}_{s\,j} + \overline{O}_{c\,j} + \overline{T}(k_j) \end{aligned} \tag{15}$$

where σ denotes the duration of a single timeslot and,

$$\overline{O}_{sj} = \sum_{i=1}^{J} z_i \cdot \{P^{s}_{j:zone=i} O_{sj}\}, \ \overline{O}_{cj} = \sum_{i=1}^{J} z_i \cdot \{P^{c}_{j:zone=i} O_{cj}\}. \tag{16}$$

Substituting O_{sj} and O_{cj} with $O_{sbas\,j}$, $O_{cbas\,j}$, $O_{srts\,j}$ and $O_{crts\,j}$ from (12) in (16), We can get $\overline{O}_{sbas\,j}$, $\overline{O}_{cbas\,j}$, $\overline{O}_{srts\,j}$ and $\overline{O}_{crts\,j}$, respectively. Additionally, $\overline{T}(k_j)$ for basic access and RTS/CTS, respectively, can be obtained as follows,

$$\overline{T}_{bas}(k_j) = \sum_{i=1}^{J} z_i \sum_{j=J-i}^{J-1} (P^{s}_{j:zone\,i} k_j U_j + P^{c}_{j:zone\,i} k^{*}_{j} U^{*}_{j}) \tag{17}$$

$$\overline{T}_{rts}(k_j) = \sum_{i=1}^{J} z_i \sum_{j=J-i}^{J-1} (P^{s}_{j:zonei} k_j U_j) \tag{18}$$

Assuming $\mathrm{E}[P^{*}_{j}]$ is the average length of the longest packet payload involved in a collision, from (13) and (14), we have,

$$U^{*}_{j} = H_{phy} / R_b + (H_{mac} + \mathrm{E}[P^{*}_{j}]) / R + \delta + SIFS + Ato \tag{19}$$

$$\text{and } k^{*}_{j} = \frac{TXOP_j}{U^{*}_{j}}, \ \ k^{*}_{j} \geq 1 \tag{20}$$

Note that the collision period in (15), (17) and (19) already includes the ACK timeout. This is as defined in [1] that during a collision, each station would have to wait for the ACK timeout before starting a new cycle of transmission. From (11), the expected payload information for a single timeslot can be obtained as,

$$\mathrm{E}[\overline{P}(k_j)] = \sum_{i=1}^{J} z_i \cdot \left\{P^{s}_{j:zone=i} k_j \mathrm{E}[P_j]\right\} \tag{21}$$

Combining (15)-(21), we can obtain throughput S for both the basic access and RTS/CTS mechanisms. The throughput differences between the normal case ($k_j = 1$) and TXOP enhancement ($k_j > 1$) for category j is determined by,

$$D^{txop}_{j} = \mathrm{E}[\overline{P}(k_j)] \cdot \overline{T}(1) + \mathrm{E}[\overline{P}(1)] \cdot \overline{T}(k_j), \ k_j > 1 \tag{22}$$

For the basic access mechanism, (17) is used to calculate (22) and for RTS/CTS, (18) is used. We can observe from (22) that the improvement due to TXOP compared to normal case has the following properties: i) better performance enhancements are obtained in RTS/CTS mode than in basic mode; ii) in the basic mode, TXOP does not necessarily improves the throughput as it depends on the collision payload distribution, TXOP size and probabilities of collisions and successful transmissions; iii) as long as $k > 1$, TXOP can improve the total throughput in RTS/CTS mode; however, the throughput of individual stations could be either increased or decreased, as ACs with longer TXOP limits would get a better chance to improve the performance at the expense of other ACs. The numerical results presented in the following section substantiate the above observations.

Table 1. 802.11e System Parameters and Access Category Parameters Using in Simulation and Analsysis

Frame payload	8000 bits
MAC header	224 bits
PHY header	192 bits
ACK	112 bits + PHY header
RTS	160 bits + PHY header
CTS	112 bits + PHY header
Channel bit rate	1 Mbit/s
Payload bit rate	11 Mbit/s
Propagation delay	1 μs
Slot time	20 μs
SIFS	10 μs
ShortRetryLimit	7
CWmin[0-3]	32
CWmax[[0-3]	1024
AIFS	SIFS + 2 × Slot time
TXOP	1 × Frame Payload

5 Model Validation

To validate the model, we compare the numerical results obtained from the analytical model with the results of simulations conducted in NS-2.26 [16]. Unless otherwise specified, the values of the parameters used to obtain numerical results for both the analytical model and the simulation runs are summarized in Table 1. The set of parameters is sufficiently general to cover most practical applications. All stations are configured according to the 802.11e system parameters and the specified AC parameters. Saturation conditions are created by using high rate constant bit rate traffic generators for all stations. In all the simulation results presented here, all transmitting stations contend to transmit fixed size user datagram protocol (UDP) packets to a single access point. With the exception of results for TXOP performance comparison, the results for AIFS and CW performance comparisons use basic access mode since the results for RTS/CTS mode are almost similar. The simulation program attempts to emulate as closely as possible the 802.11e standard as specified in [6].

In Figs. 2 and 3, stations differentiated only by CW size are subjected to increasing traffic loads. The contention window size is selected from the set {32, 40, 48, 56} for CWmin[3-0] and from the set {40, 48, 56, 1024} for CWmax[3-0]. In Figs. 4 and 5 stations are differentiated only by AIFS, in which AIFS[3-0] are valued as (5, 4, 3, 2), respectively. The performance differences of service differentiation only by the TXOP limits are illustrated in Figs. 6 and 7.

Figs. 2 and 3 show that under low load conditions, CW differentiation can be quite effective, but as the traffic load increases, its effectiveness suffers and stations may be starved of bandwidth. Decline in the effectiveness of CW differentiation under high

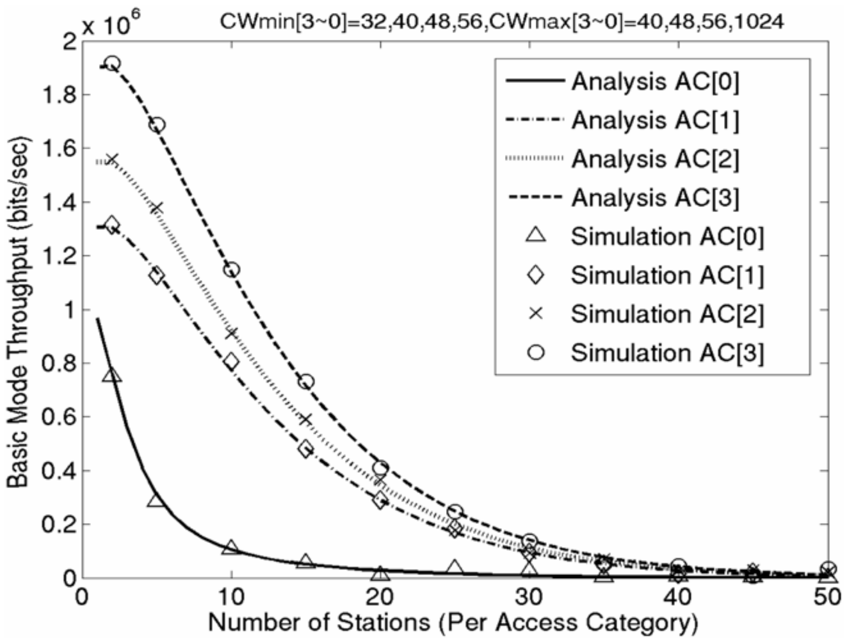

Fig. 2. Throughput of different ACs vs. traffic load (basic mode, CW Differentiation)

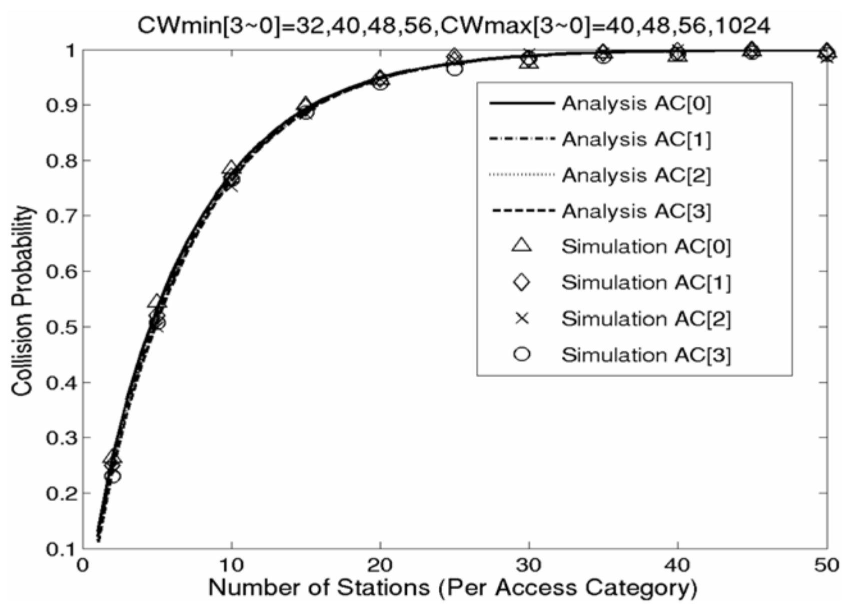

Fig. 3. Collision probabilities of different ACs vs. traffic load (basic mode, CW Differentiation)

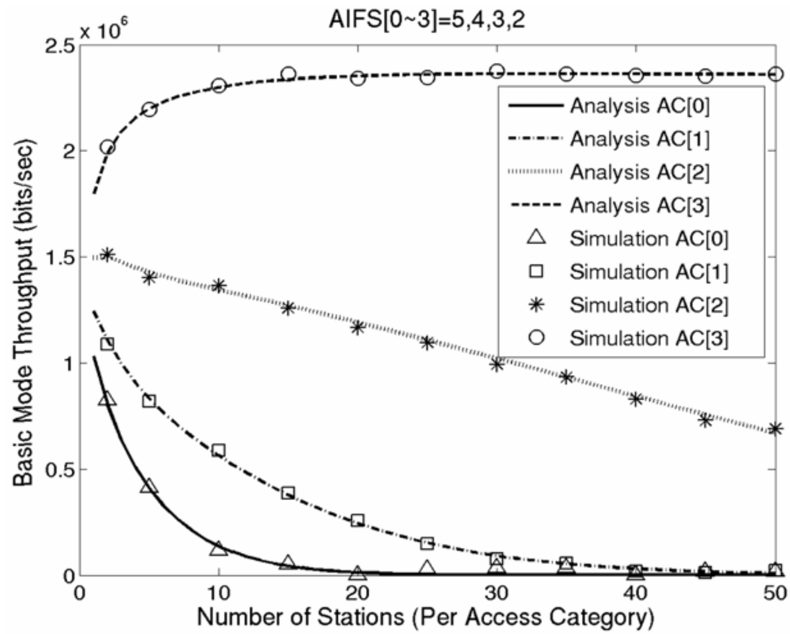

Fig. 4. Throughput of different ACs vs. traffic load (basic mode, AIFS differentiation)

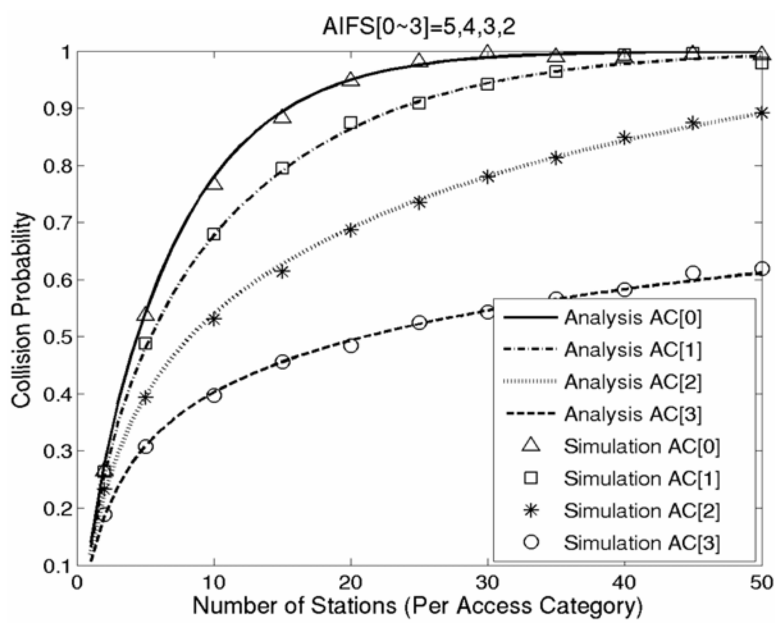

Fig. 5. Collision probabilities of different ACs vs. traffic load (basic mode, AIFS differentiation)

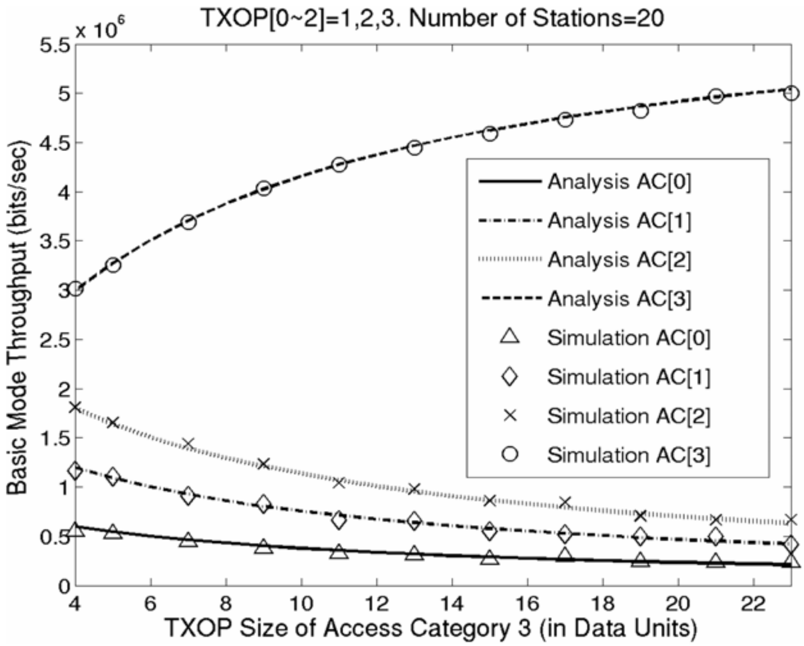

Fig. 6. Throughput of different ACs vs. TXOP[3] (basic mode, TXOP differentiation)

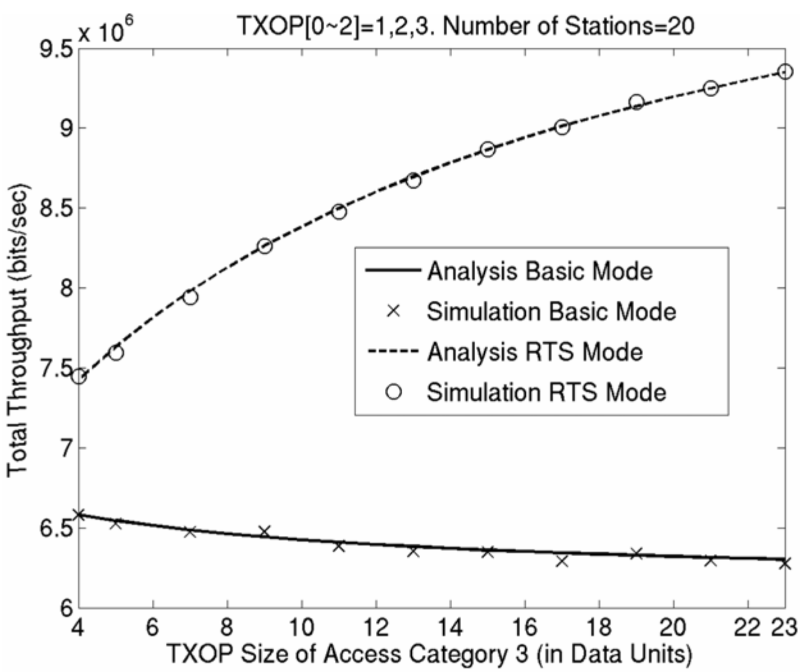

Fig. 7. Throughput of different ACs vs. TXOP[3] (RTS/CTS mode, TXOP differentiation)

load conditions can be explained by looking at how the probability of collision increases with load. As shown in Fig. 3, the collision probabilities for all ACs become close to 1 when the number of stations exceeds 35; at the same time, the tendency for CW differentiation to starve the high priority stations is apparent in Fig. 2. It is clear from (9) that when only CW differentiation is used, the collision probability of each AC would converge to 1 when load increases, which means that high priority stations can suffer performance degradation due to lower priority stations offering heavy loads. The case that low priority AC[0] stations starve earlier than higher priority AC[1]-AC[3] stations, as shown in Fig. 2, is only apparent when AC[0] stations have a maximum window size (1024 in this case) much greater than that of higher priority stations.

In contrast to CW differentiation, AIFS differentiation does not sacrifice service provided to high priority ACs when traffic loads of lower priority ACs are high. Fig. 4 shows that, with AIFS differentiation, the throughput of higher priority ACs would either increase, e.g., AIFS[3] or reduce at a slower rate (e.g., AIFS[2]) than that of lower priority ACs. The corresponding collision probability in Fig. 5 is also smaller for higher priority ACs compared to lower priority ACs. Because AIFS differentiation creates a contention zone where only high priority stations may transmit, it maintains a lower probability of collision longer for high priority stations. Since AIFS[i] < AIFS[i−1], stations in AC i would not suffer throughput degradation due to activities of stations belonging to lower ACs, but the opposite is true regarding activities of stations belonging to higher ACs. As shown in Fig. 4, since AC[3] has the highest priority, AIFS[3] could achieve much higher throughput than other ACs. However, with the higher throughput achieved by AIFS[3], the channel occupancy due to stations in AC[3] is very high, which further reduces the performance of the lower priority stations when traffic increases. As shown in Fig. 4, throughput reductions occur at a faster rate for lower priority stations (e.g., AIFS[0]) than higher priority stations (e.g., AIFS[1]) since AIFS[0] is affected by three higher priority ACs while AIFS[1] is only affected by two higher priority ACs. Furthermore, the increased load would cause reductions in the total throughput since the collision probability would increase for all ACs.

To evaluate the performance of TXOP, we set TXOP[0-2] as {1, 2, 3} number of unit packet size and increase TXOP[3] from 4 to 23. Fig. 6 shows that as the TXOP limit increases, the throughput of AC[3] increases at the expense of lower priority ACs, with a corresponding reduction in their throughput. The results agree with the observations from (22) that not all the ACs necessarily increase their throughput but those with larger TXOP limits can get a better chance to increase their throughput based on the lower probability due to collision. Fig. 7 shows that the RTS/CTS mechanism gives better improvements of total throughput than the basic access mode. These match our observations i) and iii) from (22). It also shows in Fig. 7 that the total throughput of basic mode can be reduced even with the increase in throughput for AC[3], which verifies observation ii) based on (22).

One of our objectives in presenting the above results is to show that the simulation results agree well with the analytical results over a wide range of system parameters. Other existing Markov chain models, which are more complex, have similarly been shown to be in close agreement with simulation results, albeit over much more limited range of system parameters (e.g., number of ACs) due to the difficulty in applying these methods to more general situations. Therefore we can conclude that our results are comparable to those using existing models when these models are applicable, but the simplicity of our model makes it suitable to be applied for system performance evaluations under much more general conditions.

6 Conclusions

In this paper, we have presented an improved analytical model to study the performance of EDCA service differentiation schemes, employing CW, AIFS and TXOP limits. The model achieves improved accuracy by taking into account of ACK timeouts in the collision periods. This model significantly reduces computation complexity and therefore can be effectively applied to general 802.11 WLAN operation environments that include several ACs. Close agreements between analytical results obtained using the model and simulation results give confidence to the validity of the model. Furthermore, we have presented analytical and simulation results which verify the general effectiveness of the EDCA service differentiation schemes. However, with CW differentiation, heavy traffic load from lower priority ACs can adversely affect the throughput of high priority ACs. AIFS provides effective service differentiation in that it preserves service to higher priority ACs at high loads, although it is prone to starving the lower priority ACs. We have also shown that the RTS/CTS mode can achieve better performance than the basic mode when TXOP differentiation is employed. The results presented this paper are beneficial to understanding how parameters should be chosen in EDCA to achieve the required QoS.

References

1. Chhaya, H.S., Gupta, S.: Performance modeling of asynchronous data transfer methods of IEEE 802.11 MAC protocol. Wireless Networks 3, 217–234 (1997)
2. Bianchi, G.: Performance analysis of the IEEE 802.11 DCF. IEEE J. Select. Areas Commun. 18, 535–547 (2000)
3. Ho, T.S., Chen, K.C.: Performance evaluation and enhancement of the CSMA/CA MAC protocol for 802.11 Wireless LAN's. In: Proc. IEEE PIMRC, Taipei, Taiwan, October 1996, pp. 392–396 (1996)
4. Cali, F., Conti, M., Gregori, E.: IEEE 802.11 Protocol: Design and Performance Evaluation of an Adaptive Backoff Mechanism. IEEE J. Select. Areas Commun. 18(19), 1774–1786 (2000)
5. Tay, Y., Chua, K.: A Capacity Analysis for the IEEE 802.11 MAC Protocol. Wireless Networks 7(2), 159–171 (2001)
6. IEEE 802.11-2007: IEEE Standard for Information technology – Telecommunications and information exchange between systems – Local and metropolitan area networks – Specific requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications (2007)
7. Xiao, Y.: Performance Analysis of Priority Schemes for IEEE 802.11 and IEEE 802.11e Wireless LANs. IEEE Transaction on Wireless Communications 4(4) (July 2005)
8. Tao, Z., Panwar, S.: Throughput and Delay Analysis for the IEEE 802.11e Enhanced Distributed Channel Access. IEEE Transactions on Communications 54(4) (April 2006)

9. Xu, K., Wang, Q., Hassanein, H.: Performance analysis of differentiated QoS supported by IEEE 802.11e enhanced distributed coordination function (EDCA) in WLAN. In: Proc. Globecom, December 2003, pp. 1048–1053 (2003)
10. Robinson, J.W., Randhawa, T.S.: Saturation Throughput Analysis of IEEE 802.11e Enhanced Distributed Coordination Function. IEEE J. Select. Areas Commun. 22(5) (June 2004)
11. Ramaiyan, V., Kumar, A., Altman, E.: Fixed Point Analysis of Single Cell IEEE 802.11e WLANs: Uniqueness, Multistability and Throughput Differentiation. In: ACM SIGMETRICS (2005)
12. Bianchi, G., Tinnirello, I.: Analysis of priority mechanisms based on differentiated inter frame spacing in CSMA-CA. In: Proc. VTC, pp. 1401–1405 (Fall 2003)
13. Banchs, A., Vollero, L.: Throughput Analysis and Optimal Configuration of 802.11e EDCA. Computer Networks 50, 1749–1768 (2006)
14. Zhu, H., Chlamtac, I.: Performance Analysis for IEEE 802.11e EDCA Service Differentiation. IEEE Transactions on Wireless Communications 4(4) (July 2005)
15. Romdhani, L., Bonnet, C.: Performance Analysis and Optimization of the 802.11e EDCA Transmission Opportunity (TXOP) mechanism. In: Third IEEE International Conference on Wireless and Mobile Computing, Networking and Communications, WiMob 2007 (2007)
16. http://www.tkn.tu-berlin.de/research/802.11e_ns2