

Saturated Throughput Analysis of IEEE 802.11e EDCA

Lixiang Xiong¹, Guoqiang Mao

*School of Electrical and Information Engineering
The University of Sydney,
Sydney, NW 2006, Australia*

Abstract

IEEE 802.11e standard has been published to introduce quality of service (QoS) support to the conventional IEEE 802.11 wireless local area network (WLAN). Enhanced Distributed Channel Access (EDCA) is used as the fundamental access mechanism for the medium access control (MAC) layer in IEEE 802.11e. In this paper, a novel Markov chain based model with a simple architecture for EDCA performance analysis under the saturated traffic load is proposed. Compared with the existing analytical models of EDCA, the proposed model incorporates more features of EDCA into the analysis. Firstly, we analyze the effect of using different arbitration interframe spaces (AIFSs) on the performance of EDCA. That is, the time interval from the end of the busy channel can be classified into different contention zones based on the different AIFSs used by different sets of stations, and these different sets of stations may have different transmission probabilities in the same contention zone. Secondly, we analyze the possibility that a station's backoff procedure may be suspended due to transmission from other stations. We consider that the contention zone specific transmission probability caused by the use of different AIFSs can affect the occurrence and the duration of the backoff suspension procedure. Based on the proposed model, saturated throughput of EDCA is analyzed. Simulation study is performed, which demonstrates that the proposed model has better accuracy than those in the literature.

Key words: IEEE 802.11e; EDCA; Markov chain; saturated throughput;

¹ Corresponding author

E-mail addresses: xlx@ee.usyd.edu.au (Lixiang Xiong), guoqiang@ee.usyd.edu.au (Guoqiang Mao)

1 INTRODUCTION

IEEE 802.11 wireless local area network (WLAN) [1] has been widely used for high speed wireless Internet access. Unfortunately the original IEEE 802.11 WLAN standard is based on the best-effort service model, and its fundamental access mechanism for the medium access control (MAC) layer, Distributed Coordination Function (DCF), can not satisfy the demand for better quality of service (QoS) support from multimedia applications. Thus IEEE 802.11e Enhanced Distributed Channel Access (EDCA) [2] is developed to introduce QoS support.

IEEE802.11e standard classifies traffic into four Access Categories (ACs), i.e., voice, video, best effort and background. AC based traffic prioritization is implemented by using a combination of AC specific parameters, which include arbitration interframe space (AIFS), the minimum contention window size (CW_{min}), the maximum contention window size (CW_{max}) and the transmission opportunity (TXOP) limit.

Consider a wireless channel which just returns idle from a busy state and a station is ready to transmit a frame. Before the station can start the transmission, it must sense the channel idle for a complete AIFS or EIFS (extended interframe space) from the end of the last busy channel. The selection of AIFS or EIFS depends on the type of the last channel busy event. If the last channel busy event is an unsuccessful transmission (e.g., a collision), the station must wait an EIFS, otherwise it must wait an AIFS. The duration of an AIFS is given by

$$AIFS = SIFS + AIFSN \times aTimeSlot, \quad (1)$$

where $aTimeSlot$ and SIFS (short interframe space) are determined by the physical layer characteristics, and AIFSN (AIFS number) is a non-negative integer depending on the AC. A higher priority AC has a smaller AIFSN. Fig. 1 depicts the relationship between different AIFSs, where the AIFS for a higher priority AC A, depicted as $AIFS[A]$, is smaller than the AIFS for a lower priority AC B, depicted as $AIFS[B]$. The duration of an EIFS is related to the duration of an AIFS by

$$EIFS = SIFS + ACK + AIFS, \quad (2)$$

where ACK is the time required to transmit an acknowledgment (ACK) frame. EIFS is also AC specific and we have $C = AIFS[B] - AIFS[A] = EIFS[A] - EIFS[B]$. After observing the channel idle for a complete AIFS or EIFS, the station needs to complete an additional backoff procedure before transmission. Should the station detect the channel busy before the completion of the idle AIFS or EIFS, the station will not start its backoff procedure until it observes the channel idle for another complete AIFS or EIFS depending on

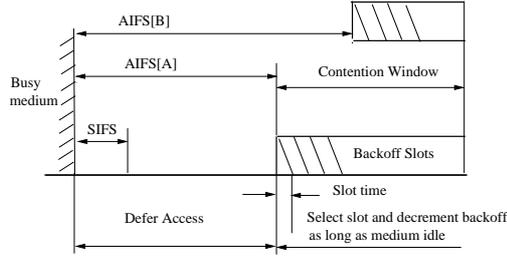


Fig. 1. The relations between different AIFSs in EDCA. [2]

the type of the channel busy event preceding the idle channel. As long as the channel is not idle for a complete AIFS or EIFS, the station keeps suspending the start of its backoff procedure. During the backoff procedure, the station decreases its backoff counter by one at the beginning of each time slot if the time slot immediately before is idle. When the backoff counter reaches zero at the beginning of an idle time slot, the station will start the transmission of the frame at the beginning of the next time slot. Should the station detect that the channel is busy in a time slot, it suspends its backoff procedure and freezes its backoff counter until it observes the channel idle for a complete AIFS or EIFS depending the type of the channel busy event preceding the idle channel. By then it may continue its backoff procedure: decreases its backoff counter by one or starts a transmission at the beginning of the time slot immediately following the completed AIFS/EIFS. If the channel becomes busy again before the completion of an AIFS/EIFS, the station will wait another complete AIFS or EIFS after the channel returns idle depending on the type of the channel busy event preceding the idle channel. As long as the channel is not idle for a complete AIFS/EIFS, the station keeps suspending its backoff procedure. To avoid any confusion, in this paper, we use the term “backoff suspension” to refer to the procedure that a station’s normal backoff procedure is interrupted by transmission from other stations, and the station must sense the channel idle for a complete AIFS/EIFS before it can start a new backoff procedure or resume the suspended backoff procedure.

The backoff counter is a uniformly distributed random integer, drawn from a predefined range $[0, CW]$. The value of CW is within an AC specific range, $[CW_{min}, CW_{max}]$. The higher priority AC has a smaller CW_{min}/CW_{max} . For the first transmission of a frame, CW is set to be CW_{min} . If the transmission is successful, the receiving station replies with an ACK frame after waiting a SIFS. If the transmission is unsuccessful, CW is doubled and the frame is retransmitted following the procedure described in the last paragraph. CW is doubled with each unsuccessful retransmission until it reaches CW_{max} . The frame will be discarded if the number of retransmissions reaches the maximum retransmission limit. Either the transmission is successful or the frame is finally discarded, the station resets its CW value to CW_{min} . Once the station gains access to the channel, it may continuously transmit a maximum amount of frames determined by the AC specific value, TXOP limit. A higher priority

AC has a larger TXOP limit.

Furthermore, some details of EDCA should be noted:

- In the case that a collision happens, colliding stations (i.e., stations involved in the collision) will wait an ACK timeout duration to detect the collision, and then they will wait an AIFS before starting another backoff procedure. According to [3], the sum of the ACK timeout duration and an AIFS is equal to an EIFS. Non-colliding stations (i.e., stations not involved in the collision) also wait an EIFS after a collision [4, clause 9.2.5.2, pp.77-79]. Fig. 2(a) depicts this situation. Therefore, all stations wait an AIFS from the end of the busy channel after a successful transmission, and wait an EIFS (or an equivalently AIFS+ACK timeout) from the end of the busy channel after a collision. For simplicity, we use the term “IFS” to represent both AIFS and EIFS in this paper, when there is no need to specify their difference.
- As mentioned earlier, a station decreases its backoff counter by one at the beginning of a time slot during its backoff procedure. This means whether the backoff counter is decreased or not depends on the channel status in the previous time slot. This backoff counter decrement is independent of whether the channel is busy or not in the current time slot. Furthermore, every time the station leaves a backoff suspension procedure after completing an IFS, its non-zero backoff counter will be decreased by one at the beginning of the immediately following time slot, and this decrement is independent of the channel status in that time slot [2, clause 9.9.1.3, pp.81-83], [5].
- When the backoff counter is decreased to zero at the beginning of a time slot, the station will start its transmission at the beginning of the next time slot provided that there is no transmission from other stations in the current time slot. Otherwise the station will enter into a backoff suspension state to wait a complete idle IFS and start its transmission at the beginning of the immediately following time slot [2, clause 9.9.1.3, pp.81-83], [5].

To investigate the performance of EDCA, an accurate analytical model is necessary. In addition to the effect of using different CW ranges that has been well investigated in the previous publications, we consider some important factors that should be carefully considered for an accurate analysis of EDCA performance:

Firstly, the effect of using different AIFSs should be carefully considered. In this paper, we assume that each station carries the traffic of an AC only for simplicity. Therefore stations can be classified into different sets based on their AIFS values. The stations in the same set have the same AIFS value. Different sets of stations will wait different IFSs in the backoff suspension procedure before they may access the channel in the normal backoff procedure, as shown in Fig. 2. Fig. 2 shows that priority A stations with a smaller $IFS[A]$ may

begin their backoff procedure and transmit after $IFS[A]$, while priority B stations with a larger $IFS[B]$ are still in the backoff suspension procedure and can not transmit. When $IFS[B]$ is completed, both sets of stations can begin their backoff procedure and transmit. Therefore the time period from the end of the busy channel can be classified into different intervals, referred to as contention zones in this paper, depending on the different transmission probabilities of different sets of stations in each zone caused by the use of different AIFSSs.

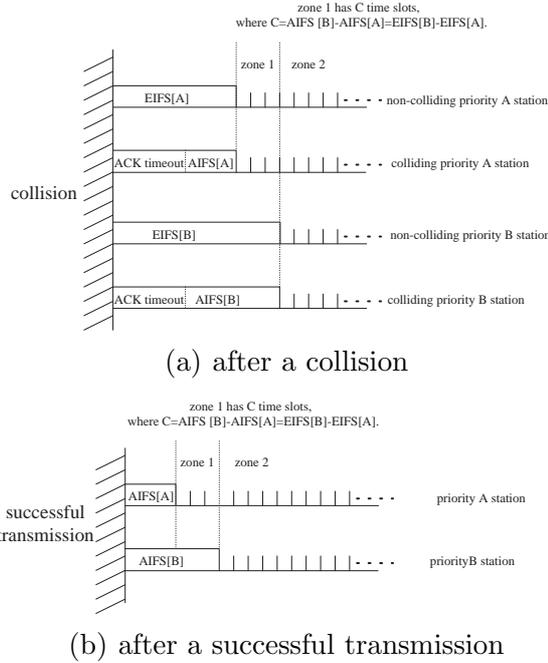


Fig. 2. The contention zone specific transmission probability.

Secondly, the possibility of backoff suspension should be analyzed. As mentioned earlier, before the start of a new backoff procedure, as well as every time the channel turns busy during the backoff procedure, the station may experience a backoff suspension procedure. The occurrence of backoff suspension depends on the channel status which is affected by the activities of other stations. Moreover, while a station is in the backoff suspension procedure, transmission from other stations may occur before the station completes an idle IFS. In this case the station must wait another complete idle IFS after the channel returns idle. Therefore the exact duration of each backoff suspension procedure is uncertain since it is affected by transmission from other stations. It is obvious that the occurrence and the duration of the backoff suspension procedure can affect the performance of EDCA.

A novel Markov chain based analytical model of EDCA with a simple architecture is proposed in this paper, where more features of EDCA are incorporated. Both the effect of using different AIFSSs and the effect of backoff suspension are considered, which results in a more accurate analysis.

The rest of this paper is organized as follows: Section 2 gives a brief introduction to previous work in this area; Section 3 introduces the proposed analytical model; saturated throughput of EDCA is analyzed in Section 4; simulation study is performed in Section 5; finally Section 6 concludes the paper.

2 Related Work

Some analytical models for EDCA have been proposed in the literature [6–19]. Most of them use the Markov chain approach [8–19]. However, there are some researchers trying to obtain a closed-form expression for the collision probability and the saturated throughput using elementary probability theory directly [6, 7]. In [6], a closed-form expression of the saturated throughput in EDCA is obtained based on an analytical model for DCF proposed in [20], where the average numbers of idle time slots and collisions between two consecutive successful transmissions in the system are first obtained using elementary probability theory, then the saturated throughput is calculated. A similar approach is used in [7], where Kuo *et al.* obtain a closed-form expression of the saturated throughput for each traffic class by obtaining the average values of some variables using elementary probability theory first, such as the initial backoff counter value and the number of collisions between two consecutive successful transmissions in the system, then the saturated throughput is calculated. We refer to the approach in [6, 7] as non-Markov approach. The major problem with the non-Markov chain approach is that in order to obtain a closed-form solution using elementary probability theory directly, significant simplification and approximation have to be made, thus they can not fully capture the complexity of EDCA, including the effects of using different AIFSs and backoff suspension. For example, the possibility that the backoff procedure of lower priority stations may be consecutively interrupted by transmission from higher priority stations is not considered in [7], and the effect of using different AIFSs is ignored in [6].

Compared with the non-Markov chain approach, the Markov-chain approach has a disadvantage that a closed-form solution is difficult to obtain. However, a well designed Markov chain model can capture the complexity of EDCA more easily than the non-Markov chain approach. Using Markov chain to analyze EDCA performance was originally started by a Markov chain model developed by Bianchi for analyzing legacy DCF [21]. In [21], two stochastic processes are used to construct a two-dimensional multiple-layer Markov chain model for modeling DCF. One stochastic process is used to represent the backoff counter, and the other is used to represent the number of consecutive retransmissions. Each layer in the Markov chain represents a backoff stage, and each state in a layer represents a specific backoff counter value at the corresponding backoff stage. A similar approach is used in most Markov chain models for

EDCA performance analysis with some modifications [8–19]. However some limitations exist in these approaches, which leaves room for us to develop a better model to achieve more accurate analytical result.

In [8–13], some Markov chain models are developed based on that in [21]. Different contributions are made to develop these Markov chain models so that they can be used for EDCA performance analysis, such as the post-collision analysis presented in [8], which considers the effect of using different AIFS, the delay analysis in [10], and the Z-transform approach in [11]. But a common problem exists among them: the possibility of backoff suspension is ignored or not clearly analyzed in their Markov chain model.

Compared with those in [8–13], models presented in [14–16] consider the possibility of backoff suspension. In [14, 15], the backoff suspension is considered by adding a transition for each backoff state, and this transition starts and ends at the same state. It represents that the backoff procedure may be suspended in the corresponding backoff state. In [16], the backoff suspension is considered by using some extra states to represent the possible backoff suspension that occurs in the corresponding backoff state. However, some potential flaws may exist. Firstly, the difference between the backoff suspension procedure and the normal backoff procedure is not considered in [14, 15]. In the backoff suspension procedure, a station must wait a complete idle IFS before it can decrease its backoff counter; while in the normal backoff procedure, a station only needs to wait an idle time slot in order to decrease its backoff counter. Secondly, the mandatory IFS before the start of a new backoff procedure and the possible backoff suspension procedure within this IFS interval are not considered in [14, 15]. Finally, the contention zone specific transmission probability caused by using different AIFSs is not considered in [16].

Finally, some Markov chain models consider both the effect of using different AIFSs and the effect of backoff suspension [17–19]. However, some flaws still exist among them.

In [17], two Markov chain models are created for two traffic classes separately. The model for the higher priority traffic class with a smaller AIFS is a two dimensional Markov chain model, which is similar to the popular Bianchi’s model for DCF [21], and two stochastic processes are used in the model to represent the backoff stage and backoff counter respectively. The model for the lower priority traffic class with a larger AIFS is a three dimensional Markov chain model, where the third dimension is a stochastic process representing backoff suspension. In the three dimensional Markov chain model, some extra “hold” states are added in each backoff state which represent some additional time slots in the backoff procedure of the AC with a larger AIFS. If the channel remains idle in a specific backoff state, a transition to the next backoff state will occur; otherwise a transition to the hold state will occur which represents

backoff suspension. Moreover, if the channel turns busy in the hold state because of transmission from the higher priority traffic class, a loop transition to the hold state itself will occur. Two problems exist in [17]. Firstly, both the Markov chain models in [17] consider that a station will keep retransmitting until the frame has been successfully transmitted. The possibility that the frame may be dropped after reaching the maximum retransmission limit is not considered. Secondly, the two dimensional Markov chain for the higher priority traffic class does not consider the possibility that the backoff procedure of a station with a higher priority traffic flow may also be suspended due to transmission from other stations. The model proposed in this paper uses a similar approach as that in [17], but we remove the above two problems.

In [18], two Markov chain models are used to model two traffic classes with different priority respectively, which are very similar to those in [17]. The authors of [18] also consider the possibility that the frame may be dropped after reaching the maximum retransmission limit, which has been ignored in [17]. However, they do not consider the possibility of backoff suspension for the higher priority traffic class.

In [19], a three-dimensional Markov model is created for each traffic class. The three state variables in the three-dimensional Markov chain represent the backoff stage, the backoff counter, and the number of idle time slots after a transmission respectively. In particular, the third state variable is a stochastic process representing the number of idle time slots since the end of the previous transmission. The effect of using different AIFSs has been properly considered through the third state variable in their model. In addition to its complexity, a potential problem in the model is that the model assumes a station will keep retransmitting until the frame has been successfully transmitted. The possibility that the frame may be dropped after reaching the maximum retransmission limit is not considered. This problem has been removed in our model.

Furthermore, some details of EDCA, including the backoff counter decrement rule following the end of an IFS and the exact timing of a transmission after the backoff counter reaches zero, are simply ignored or not correctly analyzed in the Markov chain models in [8–12, 14, 15, 17–19]. This may be caused by the fact that IEEE 802.11e standard was not finished yet when those papers were published. Although they may only have minor effects on EDCA performance analysis, including them in the Markov chain model certainly improves the accuracy.

3 A MARKOV CHAIN BASED MODEL FOR EDCA PERFORMANCE ANALYSIS

In this section, we present the proposed analytical model of EDCA using Markov chain. Firstly, the basic Markov chain models are proposed. Secondly, the transition probabilities for the proposed Markov chain models are analyzed, where the contention zone specific transmission probability caused by using different AIFSs is analyzed following the method in [8]. Finally, a solution for the Markov chain models is obtained. The following assumptions are made in our analysis.

- Traffic load is saturated. That is, traffic is always backlogged at each station.
- Only two ACs are considered: AC A and AC B. AC A has higher priority than AC B and $AIFS[A] < AIFS[B]$. However, our analysis can be easily extended to include more than two ACs.
- Each station carries traffic from one AC only. Thus a station may be referred to as a priority A station or a priority B station depending on the AC of the traffic it carries.
- Only one frame is transmitted in each TXOP.
- A WLAN with a fixed number of stations is considered in our analysis. The number of stations for AC A and AC B is denoted by n_A and n_B respectively. n_A and n_B are known numbers.
- The transmission probability of a station in a generic time slot is a constant, which is determined by its AC only. This is an assumption widely adopted in the area [8, 11–16, 18]. The transmission probabilities of a priority A station and a priority B station in a generic time slot are represented by τ_A and τ_B respectively. The values of τ_A and τ_B are unknown, which need to be solved. Here the term “generic time slot” refers to as the time slot following IFS because transmission within IFS is not possible.
- The wireless channel is ideal. That is, there are no noise, no external interference and hidden station problems. Moreover, the channel is perfectly synchronized. That is, all stations can immediately sense the channel busy or idle, and they can perform their backoff procedure synchronously.

3.1 Two Discrete Time Two-Dimensional Markov Chain Models

3.1.1 The Basic Markov Chain Models

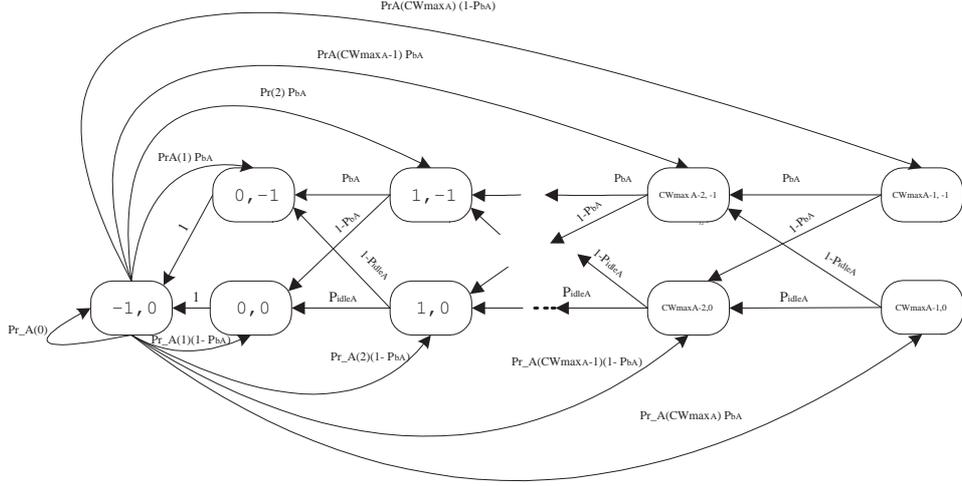
Fig. 3 illustrates two discrete time two-dimensional Markov chain models for an AC A station and an AC B station respectively. Each Markov chain model represents the channel contention procedure for a station of a specific AC. For the ease of drawing, we use the symbol “ C ” to represent $AIFS[B] - AIFS[A]$

as mentioned in Section 1. There are two stochastic processes within the Markov chain model. The first process, denoted by $w(t)$, is used to model the decrement of the backoff counter during the backoff procedure of the station. Here a special value of $w(t) = -1$ is used to represent the station's own transmission, which includes the idle $IFS[A]$ immediately following the end of the busy channel as no frame transmission is possible during this interval. The second process, denoted by $v(t)$, is used to model the backoff suspension procedure. $v(t) = 0$ indicates the station is in the normal backoff procedure or is transmitting its own frame. When the station is in the backoff suspension procedure, $v(t)$ is non-zero and its value represents the number of idle time slots after the idle $IFS[A]$ following the end of the busy channel. Here we use a special value of $v(t) = -1$ to represent a frame transmission from other stations, which also includes the idle $IFS[A]$ immediately following the end of the busy channel.

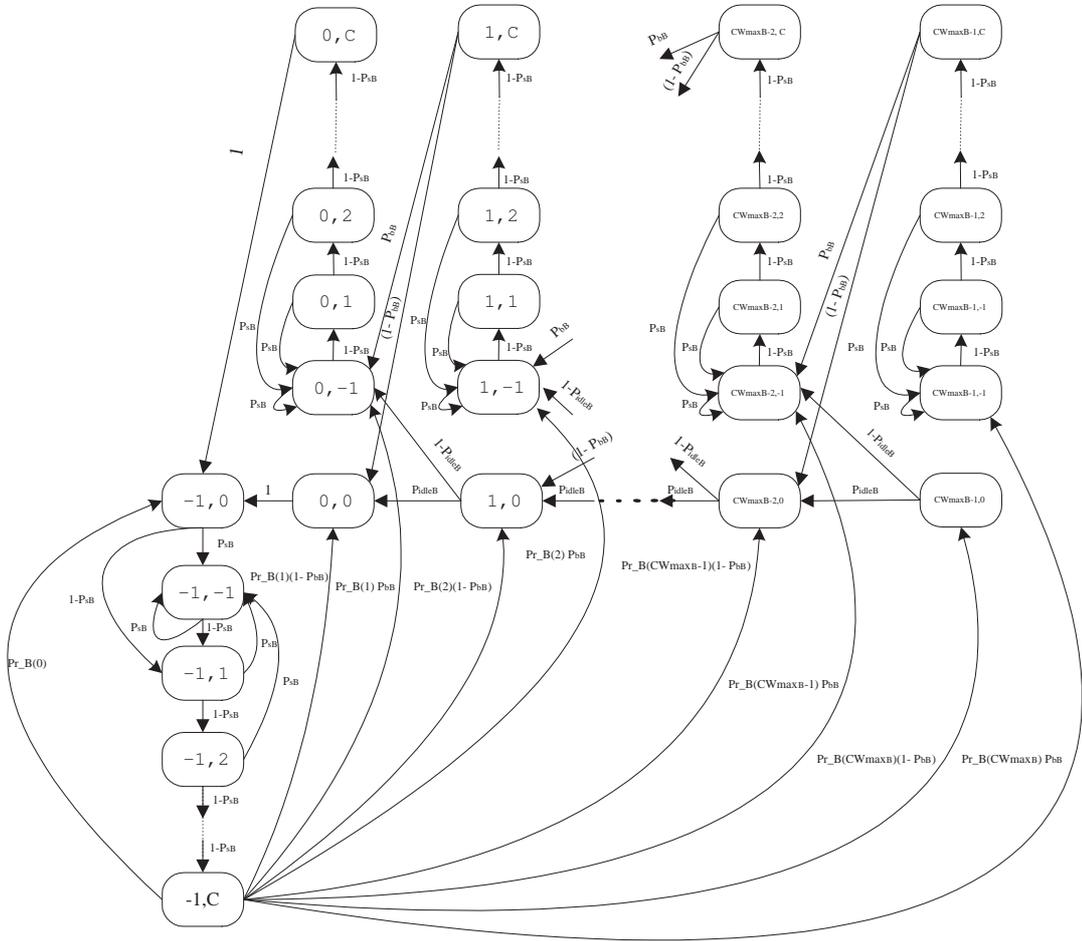
In both Markov chain models, states $(r, 0)$, $0 \leq r \leq CW_{max} - 1$ represent an idle time slot in the normal backoff procedure, where r represents the value of the backoff counter. States $(r, -1)$, $0 \leq r \leq CW_{max} - 1$ represents possible transmission from other stations which includes the idle $IFS[A]$ following the end of the busy channel, and r represents the corresponding value of the backoff counter. The special state $(-1, 0)$ is used to represent the station's own transmission, which includes the idle $IFS[A]$ following the end of the busy channel.

Another special state $(-1, -1)$ is used to represent possible transmission from other stations, which occurs before the completion of the IFS immediately following the end of the busy channel caused by the station's own transmission. As usual, this state also include the idle $IFS[A]$ following the end of the busy channel. This special state only exists for priority B stations because transmission from other priority A stations is possible before a priority B station completes the idle $IFS[B]$ immediately following its own transmission.

After leaving the state $(-1, -1)$, a priority B station may traverse each state $(-1, k)$, $1 \leq k \leq C$ if the channel remains idle. The state $(-1, k)$, $1 \leq k \leq C$ represents an idle time slot in the backoff suspension procedure, where k indicates the number of idle time slots after the idle $IFS[A]$ following the end of the last busy channel. If the channel turns busy due to transmission from other priority A stations before the state $(-1, C)$ is reached, the station will move back to the state $(-1, -1)$. After reaching the state $(-1, C)$, the station will start a backoff procedure with a random initial backoff counter. Similarly, states $(r, -1)$ and (r, k) , $0 \leq r \leq CW_{maxB} - 1$, $1 \leq k \leq C$ are used to model the backoff suspension procedure, which occurs when the normal backoff procedure has been started. A priority B station in state (r, C) , $1 \leq r \leq CW_{maxB} - 1$ may transit to either $(r-1, 0)$ or $(r-1, -1)$ depending on whether there is transmission from other stations.



(a) The Markov chain model for AC A



(b) The Markov chain model for AC B

Fig. 3. The Markov chain model for modeling the backoff procedure for a station of a specific AC.

For a priority A station, since no transmission is possible during the $IFS[A]$ following the end of the busy channel, the states $(-1, -1)$, $(-1, k)$, $(r, -1)$, and (r, k) , $0 \leq r \leq CW_{maxA} - 1$, $1 \leq k \leq C$ do not exist for priority A stations.

The embedding points of the Markov chain models can be readily determined from the earlier definition of the states. Fig. 4 depicts an example of the embedding points used in the Markov chain models. In this example, the channel turns busy because of a transmission at time point t . After the busy status ends, the channel will remain idle until $C+1$ time slots following the idle $IFS[A]$ have elapsed. The following time points, $t+k$, $1 \leq k \leq C + 2$ are located in the time slot boundary, as shown in Fig. 4. At time point t , transmitting priority A or priority B stations will enter the state $(-1, 0)$, and non-transmitting priority A or priority B stations will suspend their backoff procedure and enter the state $(r, -1)$, where the value of r is station specific. At time point $t+1$, following the completion of the idle $IFS[A]$ from the end of the busy channel, all priority A stations start or resume their backoff procedure. For transmitting priority A stations, they will start a new backoff procedure with a station specific random initial backoff counter r . As mentioned in Section 1, a station will decrease its backoff counter by one at the end of the idle IFS. Therefore all priority A stations will enter the state $(r-1, 0)$ at time point $t+1$, and their backoff counter will be decreased by one following each idle time slot. For priority B stations, they will start to traverse a series of states $(-1, k)$, (for transmitting priority B stations) or (r, k) (for non-transmitting priority B stations), $1 \leq k \leq C$ at time point $t+1$, and they will leave the state (r, C) at time point $t+C+1$ and enter the state $(r-1, 0)$ to begin or resume a normal backoff procedure. Here $r-1$ also represents that their backoff counter is decreased by one following the completion of the idle $IFS[B]$ following the end of the busy channel. Then all priority B stations can also decrease their backoff counter by one following each idle time slot by entering the corresponding state.

It should be noted that some special scenarios are not included in the aforementioned example for the ease of the drawing, such as non-transmitting priority B stations may enter the state $(-1, -1)$ at time point t , or at least one station has decreased its backoff counter to zero before the time point $t+C+1$ is reached. They can be explained more clearly in the following description for the one-step transition probabilities.

3.1.2 Transition probabilities

The one-step transition probabilities for the Markov chain model in Fig. 3(a) are explained in the following.

- (1) When a specific priority A station finishes a transmission and completes

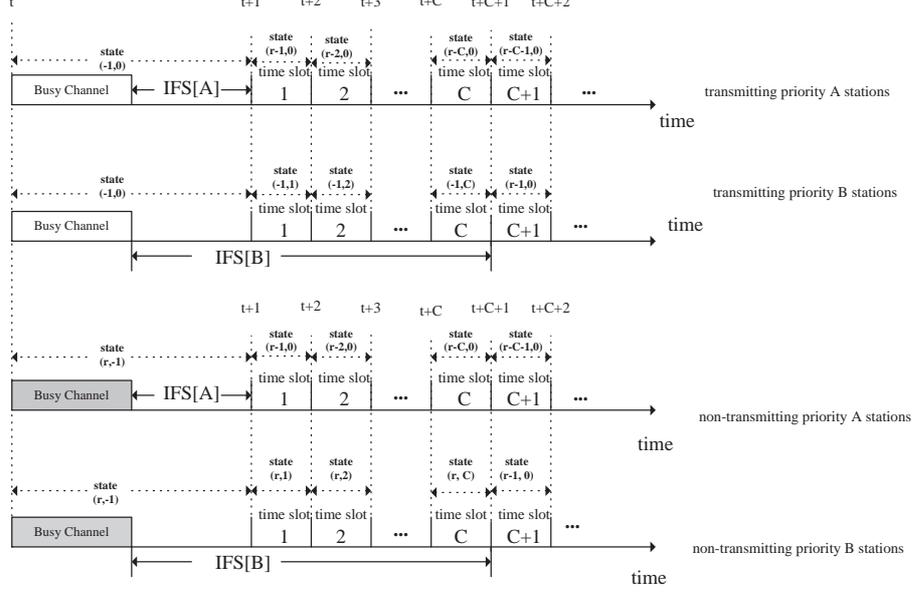


Fig. 4. An example of the embedding points used in the proposed model

the following idle $IFS[A]$, the station will leave the corresponding state $(-1,0)$ and move into the next state to start a new backoff procedure with an initial backoff counter r at the beginning of the immediately following time slot. As described in Section 1, the backoff counter will be decreased by one following the end of the $IFS[A]$. Therefore the backoff counter will be decreased to $r-1$ as the station reaches the next state. Moreover, the channel status at this moment decides the next state in the Markov chain model: the state $(r-1,-1)$ (if the channel turns busy with a probability P_{bA}), or the state $(r-1,0)$ (if the channel remains idle with a probability $1 - P_{bA}$).

$$\begin{cases} P\{(r-1,1)|(-1,0)\} = P_{bA}Pr_A(r), \\ P\{(r-1,0)|(-1,0)\} = (1 - P_{bA})Pr_A(r), \end{cases} \quad (3)$$

where $Pr_A(r)$ is the probability that the priority A station starts a new backoff procedure with a random initial backoff counter r . For the special case that the initial backoff counter is zero, the station may start a transmission at the beginning of the immediately following time slot independent of the channel status in this time slot:

$$P\{(-1,0)|(-1,0)\} = Pr_A(0). \quad (4)$$

- (2) If the station reaches the state $(r,0)$, it will reside in this state for an idle time slot. Then the station will decrease its backoff counter by one and move into the next state at the beginning of the immediately following time slot. The channel status at this moment decides the next state: the state $(r-1,-1)$ (if the channel turns busy with a probability $1 - P_{idleA}$) or

the state $(r-1,0)$ (if the channel remains idle with a probability P_{idleA}).

$$\begin{cases} P\{(r-1,-1)|(r,0)\} = 1 - P_{idleA}, \\ P\{(r-1,0)|(r,0)\} = P_{idleA}. \end{cases} \quad (5)$$

For the special case that r equals to zero, a station shall stay in the state $(0,0)$ for an idle time slot and start the frame transmission at the beginning of the immediately following time slot with a probability 1:

$$P\{(-1,0)|(0,0)\} = 1. \quad (6)$$

- (3) If the station reaches the state $(r,-1)$, it will stay in this state until the idle $IFS[A]$ following the end of the busy channel is completed. Then it will decrease its backoff counter by one and move into the next state at the beginning of the immediately following time slot. The channel status at this moment decides the next state: the state $(r-1,-1)$ (if the channel turns busy with a probability P_{bA}) or the state $(r-1,0)$ (if the channel remains idle for a probability $1 - P_{bA}$).

$$\begin{cases} P\{(r-1,-1)|(r,-1)\} = P_{bA}, \\ P\{(r-1,0)|(r,-1)\} = 1 - P_{bA}. \end{cases} \quad (7)$$

For the special case that r equals to zero, a station shall stay in the state $(0,-1)$ until the idle $IFS[A]$ following the end of the busy channel is completed, and the station will start a transmission at the beginning of the immediately following time slot with a probability 1:

$$P\{(-1,0)|(0,-1)\} = 1. \quad (8)$$

As for the Markov chain model in Fig. 3(b), its one-step transition probabilities are slightly different from those for the Markov chain model in Fig. 3(a), because extra states $(-1,-1)$, $(-1,k)$, $(r,-1)$ and (r,k) , $0 \leq r \leq CW_{maxB}-1$, $1 \leq k \leq C$ are used to represent the C (that is, $AIFS[B] - AIFS[A]$) idle time slots remaining in the $IFS[B]$ and the possible transmission from priority A stations during this time interval. The details of its one-step transition probabilities can be explained in the following.

- (1) When a specific priority B station finishes its frame transmission including the idle $IFS[A]$ following the end of the busy channel, it will leave the corresponding state $(-1,0)$. The station still needs to complete the C idle time slots remaining in its $IFS[B]$ before it can start a new backoff procedure. The station will move into the next state at the beginning of the immediately following time slot, and the channel status at this moment decides the next state: the state $(-1,-1)$ (if the channel turns busy with a probability P_{sB}), or the state $(-1,1)$ which represents that the

first idle time slot following the $IFS[A]$ can be elapsed (if the channel remains idle with a probability $1 - P_{sB}$).

$$\begin{cases} P\{(-1, -1)|(-1, 0)\} = P_{sB}, \\ P\{(-1, 1)|(-1, 0)\} = 1 - P_{sB}. \end{cases} \quad (9)$$

- (2) If the station enters the state $(-1, -1)$, it will stay in this state until the idle $IFS[A]$ following the end of the busy channel is completed. At the beginning of the immediately following time slot, the station will move to the next state. If the channel remains idle with a probability $1 - P_{sB}$, the station will move to the state $(-1, 1)$.

$$P\{(-1, 1)|(-1, -1)\} = 1 - P_{sB}. \quad (10)$$

If the channel turns busy with a probability P_{sB} , the station will remain in the state $(-1, -1)$ to wait transmission from other priority A stations.

$$P\{(-1, -1)|(-1, -1)\} = P_{sB}. \quad (11)$$

- (3) When the station moves into the state $(-1, k)$, $1 \leq k \leq C-1$ and completes an idle time slot, it will move into the next state at the beginning of the immediately following time slot. If the channel turns busy with a probability P_{sB} , the station will move back to the state $(-1, -1)$ to wait transmission from other priority A stations.

$$P\{(-1, -1)|(-1, k)\} = P_{sB}. \quad (12)$$

If the channel remains idle with a probability $1 - P_{sB}$, the station will move into the next state $(-1, k+1)$.

$$P\{(-1, k+1)|(-1, k)\} = 1 - P_{sB}. \quad (13)$$

- (4) When the priority B station moves into the state $(-1, C)$, it will wait the final idle time slot remaining in the $IFS[B]$ and start a new backoff procedure with an initial backoff counter r at the beginning of the immediately following time slot. Similar to the Markov chain model in Fig. 3(a), the station will decrease its backoff counter by one following the end of the $IFS[B]$ and move into the next state at the beginning of the immediately following time slot. If the channel turns busy with a probability P_{sB} , the station will move into the state $(r-1, -1)$.

$$P\{(r-1, -1)|(-1, C)\} = Pr_B(r)P_{sB}, \quad (14)$$

where $Pr_B(r)$ is the probability that the priority B station gets an initial backoff counter value r .

If the channel remains idle with a probability $1 - P_{sB}$, the station will move into the state $(r-1,0)$.

$$P\{(r-1,0)|(-1,C)\} = Pr_B(r)(1 - P_{sB}). \quad (15)$$

For the special case that r equals to zero, the station will start a transmission immediately independent of the channel status,

$$P\{(0,-1)|(-1,C)\} = Pr_B(0). \quad (16)$$

- (5) If the station enters the state $(r,0)$, it will stay in this state for an idle time slot, decrease its backoff counter by one and move into the next state at the beginning of the immediately following time slot. If the channel turns busy with a probability $1 - P_{idleB}$, it will move into the state $(r-1,-1)$.

$$P\{(r-1,-1)|(r,0)\} = 1 - P_{idleB}. \quad (17)$$

If the channel remains idle with a probability P_{idleB} , it will move into the state $(r-1,0)$.

$$P\{(r-1,0)|(r,0)\} = P_{idleB}. \quad (18)$$

For the special case that r equals to zero, the station will stay in the state $(0,0)$ for an idle time slot, and start a transmission at the beginning of the immediately following time slot with a probability 1.

$$P\{(-1,0)|(0,0)\} = 1. \quad (19)$$

- (6) If the station enters the state $(r,-1)$, $0 \leq r \leq CW_{maxB-1}$, the one-step transition probabilities between the state $(r,-1)$ and the states (r,k) , $1 \leq k \leq C-1$ are similar to those between the state $(-1,-1)$ and the states $(-1,k)$, $1 \leq k \leq C-1$:

$$\begin{cases} P\{(r,1)|(r,-1)\} = 1 - P_{sB}, \\ P\{(r,-1)|(r,-1)\} = P_{sB}, \\ P\{(r,k+1)|(r,k)\} = 1 - P_{sB}, \\ P\{(r,-1)|(r,k)\} = P_{sB}. \end{cases} \quad (20)$$

- (7) When the station reaches the state (r,C) , it will stay in this state for the final idle time slot in the $IFS[B]$, decrease its backoff counter by one, and move into the next state at the beginning of the immediately following backoff slot. If the channel turns busy at this moment with a probability P_{bB} , the station will move into the state $(r-1,-1)$.

$$P\{(r-1,-1)|(r,C)\} = P_{bB}. \quad (21)$$

If the channel remains idle with a probability $1 - P_{bB}$, the station will move into the state $(r-1,0)$.

$$P\{(r-1,0)|(r,C)\} = 1 - P_{bB}. \quad (22)$$

For the special case that r equals to zero, the station will wait an idle time slot in the state $(0,C)$ and start a transmission at the beginning of the immediately following backoff slot with a probability 1.

$$P\{(-1,0)|(0,C)\} = 1. \quad (23)$$

3.1.3 System Equations

Let $b_{A(r,k)}$ be the steady probability of state (r,k) in the Markov chain model in Fig. 3(a). The following system equations for this Markov chain model can be obtained due to the regularity of the Markov chain:

$$\left\{ \begin{array}{l} b_{A(CW_{maxA}-1,0)} = b_{A(-1,0)}Pr_{-A}(CW_{maxA})(1 - P_{bA}), \\ b_{A(CW_{maxA}-1,-1)} = b_{A(-1,0)}Pr_{-A}(CW_{maxA})P_{bA}, \\ \\ b_{A(r,0)} = b_{A(-1,0)}Pr_{-A}(r+1)(1 - P_{bA}) + b_{A(r+1,0)}P_{idleA} + b_{A(r+1,-1)}(1 - P_{bA}), \\ \text{for } 0 \leq r \leq CW_{max} - 1, \\ \\ b_{A(r,-1)} = b_{A(-1,0)}Pr_{-A}(r+1)P_{bA} + b_{A(r+1,0)}(1 - P_{idleA}) + b_{A(r+1,-1)}P_{bA}, \\ \text{for } 1 \leq r \leq CW_{max} - 2, \end{array} \right. \quad (24)$$

and

$$\sum b_{A(r,k)} = 1. \quad (25)$$

Since the state $(0,-1)$ represents the transmission procedure of the station, the corresponding steady probability $b_{A(-1,0)}$ should be equal to its transmission probability τ_A :

$$b_{A(-1,0)} = \tau_A, \quad (26)$$

where τ_A is the unknown probability to be solved.

Similarly, the system equations for the Markov chain model in Fig. 3(b) can

be obtained.

$$\left\{ \begin{array}{l}
b_{B(-1,-1)} = \frac{b_{B(-1,0)}[1-(1-P_{sB})^C]}{P_{sB}+(1-P_{sB})^C}, \\
b_{B(-1,1)} = (1 - P_{sB})(b_{B(-1,-1)} + b_{B(-1,0)}), \\
b_{B(-1,k)} = (1 - P_{sB})b_{B(-1,k-1)}, \\
\text{for } 2 \leq k \leq C, \\
\\
b_{B(CW_{maxB}-1,0)} = b_{B(-1,C)}Pr_B(CW_{maxB})(1 - P_{bB}), \\
b_{B(CW_{maxB}-1,-1)} = \frac{b_{B(-1,C)}Pr_B(CW_{maxB})P_{bB}}{P_{sB}+(1-P_{sB})^C}, \\
\\
b_{B(r,0)} = b_{B(-1,C)}Pr_B(r+1)(1 - P_{bB}) \\
+ b_{B(r+1,C)}(1 - P_{bB}) + b_{B(r+1,0)}P_{idleB}, \\
b_{B(r,-1)} = [b_{B(-1,C)}Pr_B(r+1)P_{bB} + b_{B(r+1,C)}P_{bB} \\
+ b_{B(r+1,0)}(1 - P_{idleB})]/[P_{sB} + (1 - P_{sB})^C], \\
\text{for } 0 \leq r \leq CW_{maxB} - 2, \\
\\
b_{B(r,1)} = (1 - P_{sB})b_{B(r,-1)}, \\
b_{B(r,k)} = (1 - P_{sB})b_{B(r,k-1)}, \\
\text{for } 0 \leq r \leq CW_{maxB} - 1 \text{ and } 2 \leq k \leq C.
\end{array} \right. \quad (27)$$

$$\sum b_{B(r,k)} = 1. \quad (28)$$

and

$$b_{B(-1,0)} = \tau_B, \quad (29)$$

where τ_B is the unknown probability to be solved.

3.2 Transition Probabilities

In this section, we analyze the unknown parameters in the transition probability equations shown in the last section, including P_{idleA} , P_{idleB} , P_{sB} , P_{bA} , P_{bB} , $Pr_A(r)$, and $Pr_B(r)$. It is organized as follows. Firstly, a new Markov chain model is used for analyzing the contention zone specific transmission probability, which results from the effect of using different AIFSs. Secondly, using the new Markov chain model, the AC specific average collision probabilities p_A and p_B are obtained. Also, the AC specific probabilities that the

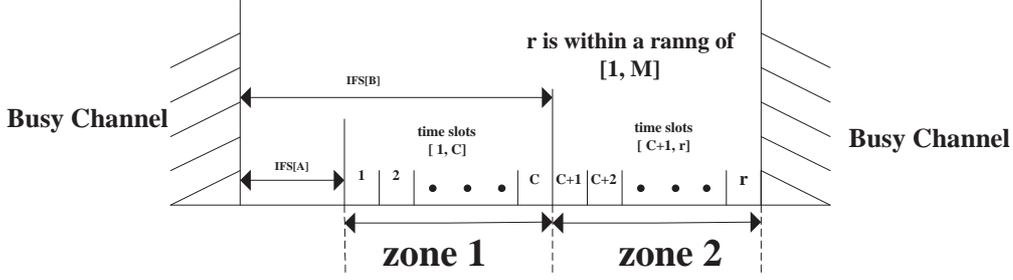


Fig. 5. Time slot distribution between two successive transmissions in the system channel remains idle in a time slot during the normal backoff procedure, P_{idleA} and P_{idleB} , are obtained. Thirdly, the transition probability that the channel turns busy in a time slot within the $IFS[B]$, P_{sB} , is obtained. Fourthly, the AC specific probabilities that the channel turns busy after IFS, P_{bA} and P_{bB} , are obtained. Finally, the AC specific transition probabilities $Pr_A(r)$ and $Pr_B(r)$ are analyzed by using another new Markov chain model.

3.2.1 A Markov Chain Model for Analyzing the Effect of the Contention Zone Specific Transmission Probability

Fig. 5 depicts the number of consecutive time slots between two successive transmissions in the WLAN. In Fig. 5, no station can transmit during the first $IFS[A]$ time interval from the end of the busy channel. During the time slots in the range of $[1, C]$ after the $IFS[A]$, referred to as zone 1, priority A stations which have completed their $IFS[A]$ may begin their backoff procedure and transmit, while priority B stations are still waiting for the completion of their $IFS[B]$ and can not transmit. During the time slots in the range of $[C+1, r]$, referred to as zone 2, priority B stations also begin their backoff procedure and may transmit by contending with priority A stations. Here r is bounded by M , which is the maximum number of possible consecutive time slots between two successive transmissions in the WLAN:

$$M = \min(CW_{maxA}, C + CW_{maxB}). \quad (30)$$

From Fig. 5, a new discrete time one-dimensional Markov chain model can be created, which is shown in Fig. 6. The stochastic process in this Markov chain model represents the number of consecutive idle time slots between two successive transmissions in the WLAN. The state (r) in the Markov chain model represents the r^{th} consecutive idle time slot starting from the end the last transmission in the WLAN, which includes the idle $IFS[A]$ following the end of the busy channel. The transition events following the states (r), $0 \leq r \leq C - 1$ represent the possible channel activity in zone 1, and the transition events following the states (r), $C \leq r \leq M$ represent the possible

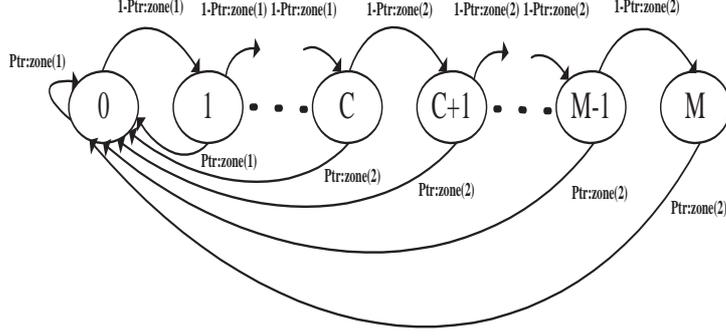


Fig. 6. The Markov chain model for modeling the number of consecutive idle time slots between two successive transmissions in the WLAN channel activity in zone 2.

The activity of this Markov chain is described by its one-step transition probabilities in the following.

- (1) In zone 1, if the channel status turns busy following the end of the r^{th} idle time slot, the system will move from state (r) to state (0):

$$P\{(0)|(r)\} = P_{tr:zone(1)}, \text{ for } 0 \leq r \leq C - 1, \quad (31)$$

where $P_{tr:zone(1)}$ is the probability that at least one priority A stations start the frame transmission at the beginning of a time slot in zone 1, given by

$$P_{tr:zone(1)} = 1 - (1 - \tau_A)^{n_A}. \quad (32)$$

- (2) If no transmission occurs the system will move from state (r) to state ($r+1$) with a probability ($1 - P_{tr:zone(1)}$):

$$P\{(r+1)|(r)\} = 1 - P_{tr:zone(1)}, \text{ for } 1 \leq r \leq C - 1. \quad (33)$$

- (3) In zone 2, both priority A stations and priority B stations begin their backoff procedure and may transmit. A transmission from either priority A or priority B stations can cause the system to return to state (0):

$$P\{(0)|(r)\} = P_{tr:zone(2)}, \text{ for } C \leq r \leq M - 1, \quad (34)$$

where $P_{tr:zone(2)}$ is the probability that there is at least one station starts the frame transmission in a time slot in zone 2, given by

$$P_{tr:zone(2)} = 1 - (1 - \tau_A)^{n_A} (1 - \tau_B)^{n_B}. \quad (35)$$

- (4) If no transmission occurs the system will move from state (r) to state

($r+1$) with a probability $(1 - P_{tr:zone(2)})$:

$$P\{(r+1)|(r)\} = 1 - P_{tr:zone(2)}, \text{ for } C \leq r \leq M-1. \quad (36)$$

- (5) When the system reaches the last state (M), a frame transmission will definitely occur after the corresponding time slot. Thus the system will return to state (0) with a probability 1:

$$P\{(0)|(M)\} = 1. \quad (37)$$

Using above transition probability equations and the regularity of the Markov chain, the relations between the steady probability $s_{(r)}$ for the Markov chain model can be obtained by

$$\begin{cases} s_{(r+1)} = (1 - P_{tr:zone(1)})s_{(r)}, \\ \text{for } 0 \leq r \leq C-1, \\ \\ s_{(r+1)} = (1 - P_{tr:zone(2)})s_{(r)}, \\ \text{for } C \leq r \leq M-1, \end{cases} \quad (38)$$

and

$$\sum_{r=0}^M s_{(r)} = 1. \quad (39)$$

Using equations (38) and (39), the steady probability $s_{(r)}$ can be solved:

$$s_{(0)} = \left[\frac{1 - (1 - P_{tr:zone(1)})^{C+1}}{P_{tr:zone(1)}} + (1 - P_{tr:zone(1)})^{C+1} (1 - P_{tr:zone(2)}) \frac{1 - (1 - P_{tr:zone(2)})^{M-C}}{P_{tr:zone(2)}} \right]^{-1}, \quad (40)$$

and

$$\begin{cases} s_{(r)} = (1 - P_{tr:zone(1)})^r s_{(0)}, \\ \text{for } 1 \leq r \leq C, \\ \\ s_{(r)} = (1 - P_{tr:zone(2)})^{r-C} s_{(0)} (1 - P_{tr:zone(1)})^C, \\ \text{for } C+1 \leq r \leq M. \end{cases} \quad (41)$$

3.2.2 p_A , p_B , P_{idleA} , and P_{idleB} ,

For a specific station transmitting its frame, collision may occur if one or more other stations start a transmission in the same time slot. The corresponding collision probability is determined by the composition of contending stations. In zone 1, only priority A stations can transmit and cause collisions. In zone 2, both priority A stations and priority B stations can transmit and collide with each other. Thus the collision probability for a priority A station should be contention zone specific, which can be obtained by

$$\begin{cases} p_{A:zone(1)} = 1 - (1 - \tau_A)^{n_A-1}, \\ p_{A:zone(2)} = 1 - (1 - \tau_A)^{n_A-1}(1 - \tau_B)^{n_B}, \end{cases} \quad (42)$$

For a priority A station in the backoff counter count-down procedure, it sees an “idle” time slot when no other stations start a transmission in the same time slot. Considering the contention zone specific transmission probability, the contention zone specific probability that a priority A station sees an idle time slot can be obtained by

$$\begin{cases} P_{idleA:zone(1)} = (1 - \tau_A)^{n_A-1}, \\ P_{idleA:zone(2)} = (1 - \tau_A)^{n_A-1}(1 - \tau_B)^{n_B}. \end{cases} \quad (43)$$

Thus, the average collision probability for a specific priority A station can be obtained as the sum of the weighted contention zone specific collision probability:

$$p_A = \sum_{r=1}^M s_{(r)} p_{A:zone_r}, \quad (44)$$

where $p_{A:zone_r}$ is the contention zone specific collision probability in the r^{th} time slot. Depending on whether the r^{th} time slot belongs to zone 1 or zone 2, $P_{idleA:zone(1)}$ or $P_{idleA:zone(2)}$ should be used for $p_{A:zone_r}$. $s_{(r)}$ is the steady probability of the state (r), which is obtained from (40) and (41).

Similarly, the average probability P_{idleA} that a specific priority A station in the backoff procedure sees an idle time slot can be obtained by

$$P_{idleA} = \sum_{r=1}^M s_{(r)} P_{idleA:zone_r}, \quad (45)$$

where $P_{idleA:zone_r}$ is the contention zone specific probability for a priority A station that the channel is idle in the r^{th} time slot. Depending on whether the r^{th} slot belongs to zone 1 or zone 2, $P_{idleA:zone(1)}$ or $P_{idleA:zone(2)}$ should be used for $P_{idleA:zone_r}$.

For a specific priority B station, all of its time slots are located in zone 2, where all stations may transmit. Thus its average collision probability can be simply obtained by

$$p_B = 1 - (1 - \tau_A)^{n_A}(1 - \tau_B)^{n_B-1}, \quad (46)$$

and so is the average probability that a specific priority B station has an idle time slot:

$$P_{idleB} = (1 - \tau_A)^{n_A}(1 - \tau_B)^{n_B-1}. \quad (47)$$

3.2.3 P_{sB}

As described in Section 1, a station suspending its backoff procedure may leave the backoff suspension procedure if the channel remains idle for an AC specific IFS interval from the end of the last busy channel. Any transmission from other stations during this time interval can stop the station from leaving the backoff suspension procedure.

For a priority A station, it needs to wait an idle $IFS[A]$ from the end of the last busy channel to leave the backoff suspension procedure. No transmission is possible during the $IFS[A]$ interval. Thus a priority A station can stay in the backoff suspension procedure for the duration of a single frame transmission only, and it will leave for the next state at the beginning of the immediately following time slot.

For a priority B station, it needs to wait an idle $IFS[B]$ from the end of the last busy channel to leave the backoff suspension procedure. According to Fig. 5, the C time slots in zone 1 are part of the $IFS[B]$, where transmission from priority A stations is possible. Thus, the probability P_{sB} that the channel turns busy in a time slot in zone1 for a specific priority B station can be obtained by

$$P_{sB} = 1 - (1 - \tau_A)^{n_A}. \quad (48)$$

3.2.4 P_{bA} and P_{bB}

According to Fig. 5, the time slot immediately following the $IFS[A]$ is located in zone 1, where only priority A station may transmit. Thus, the probability that the channel turns busy at the beginning of this time slot for a specific priority A station can be obtained by

$$P_{bA} = 1 - (1 - \tau_A)^{(n_A-1)}. \quad (49)$$

Also according to Fig. 5, the time slot immediately following the $IFS[B]$ is

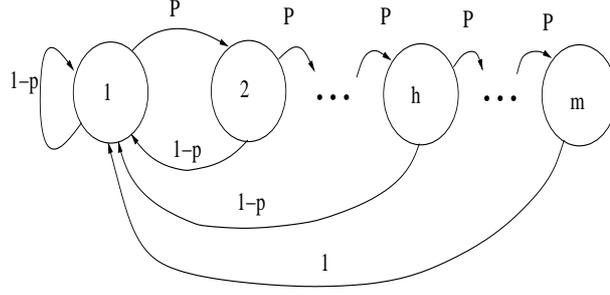


Fig. 7. The Markov chain model for modeling the number of the consecutive re-transmissions of a station

located in zone 2, where all other stations may transmit. Thus, the probability that the channel turns busy at the beginning of this time slot for a specific priority B station can be obtained by

$$P_{bB} = 1 - (1 - \tau_A)^{n_A} (1 - \tau_B)^{(n_B-1)}. \quad (50)$$

3.2.5 $Pr_A(r)$ and $Pr_B(r)$

As described in Section 1, the backoff counter is drawn randomly from the range $[0, CW]$ and the CW value is determined by the AC specific CW_{min} and CW_{max} values as well as the number of previous consecutive retransmissions. Therefore the probability of obtaining a specific backoff counter value r is related to the number of previous consecutive retransmissions. The Markov chain models shown in Fig. 3 do not explicitly consider the effect of consecutive retransmissions. Instead, its effect is considered in the probability $Pr_A(r)$ or $Pr_B(r)$ of obtaining a specific backoff counter r by weighting the probability of the number of consecutive retransmissions. For simplicity, we use the generic terms $Pr(r)$, p , CW_{min} , and CW_{max} in this section instead of the AC specific terms.

In order to obtain the probability that an AC specific station performs a specific number of consecutive retransmissions, a discrete time one-dimensional Markov chain model is created, as shown in Fig. 7. The stochastic process in this Markov chain model represents the number of consecutive retransmissions (including the first transmission of the frame) for a station at time t . Thus state (r) represents that the station is performing the r_{th} consecutive retransmission. In this Markov chain, state (h) represents the h_{th} consecutive retransmission in which the CW value reaches CW_{max} for the first time, and state (m) represents the m_{th} consecutive retransmission, which is the maximum retransmission limit. Both h and m are constants determined by the WLAN standard.

The activity of the Markov chain shown in Fig. 7 is governed by its one-step transition probabilities as follows:

- (1) If the r_{th} retransmission is unsuccessful, the system will move from state (r) to state (r+1) with a probability p :

$$P\{(r+1)|(r)\} = p, \text{ for } 1 \leq r \leq m-1, \quad (51)$$

where p is the AC specific average collision probability, which can be obtained from (44) or (46).

- (2) If the r_{th} consecutive retransmission is successful, the system will move from state (r) to state (1) with a probability $1-p$ and the station will start transmitting a new frame:

$$P\{(1)|(r)\} = 1-p, \text{ for } 1 \leq r \leq m. \quad (52)$$

- (3) when the maximum retransmission limit m is reached, the station will begin the first transmission of a new frame no matter whether the m_{th} consecutive retransmission is successful or not. Thus the system will return to state (1) with a probability 1:

$$P\{(1)|(m)\} = 1. \quad (53)$$

From (51), the relationship between two adjacent states can be obtained by

$$d_{(r+1)} = d_{(r)}p, \quad (54)$$

where $d_{(r)}$ is the corresponding steady probability for state (r).

Also due to the regularity of the Markov chain, the following relationship can be obtained:

$$\sum_{r=1}^m d_{(r)} = 1. \quad (55)$$

Thus the steady probability $d_{(r)}$ can be obtained:

$$d_{(r)} = p^{r-1}(1-p)/(1-p^m), \text{ for } 1 \leq r \leq m. \quad (56)$$

Since the backoff counter is a random integer uniformly distributed in the range $[0, CW]$, the probability of obtaining a specific backoff counter value from this range should be $\frac{1}{1+CW}$. Thus, the AC specific probability $Pr(r)$ of obtaining a specific backoff counter r can be obtained as the sum of the probability of obtaining a specific initial backoff counter r in the k_{th} consecutive retransmission weighted with the probability of the occurrence of the k_{th} consecutive retransmission:

$$Pr(r) = \sum_{k=1}^m \frac{d_{(k)}c_{(r)}}{CW(k)+1}, \quad (57)$$

where $d_{(k)}$ is the steady probability of performing the k_{th} consecutive retransmission, which obtained from (56); $CW(k)$ is the corresponding CW value in the k_{th} consecutive retransmission; and $c_{(r)}$ indicates whether the specific value r is included in the range $[0, CW(k)]$ or not (if yes, $c_{(r)}$ is 1, otherwise it is zero).

Based on the earlier analysis, an expression for the AC specific probability $Pr(r)$ can be obtained:

$$Pr(r) = \begin{cases} \sum_{k=1}^{h-1} \frac{d_{(k)}}{2^{k-1}CW_{min}+1} + \sum_{k=h}^m \frac{d_{(k)}}{CW_{max}+1}, \\ \text{for } 0 \leq r \leq CW_{min}, \\ \\ \sum_{k=j}^{h-1} \frac{d_{(k)}}{2^k CW_{min}+1} + \sum_{k=h}^m \frac{d_{(k)}}{CW_{max}+1}, \\ \text{for } 2^{j-1}CW_{min} + 1 \leq r \leq 2^j CW_{min}, \text{ and } 1 \leq j \leq h - 1, \\ \\ \sum_{k=h}^m \frac{d_{(k)}}{CW_{max}+1}, \\ \text{for } 2^{h-1}CW_{min} + 1 \leq r \leq CW_{max}, \end{cases} \quad (58)$$

where CW_{min} and CW_{max} are AC specific and known.

3.3 Summary

Finally, this section summarizes the relationship of earlier analysis.

- (1) In Section 3.1, two novel Markov chain models are created for each AC in the WLAN, which are shown in Fig. 3. The system equations for each Markov chain model are also obtained, as shown in equations (24)-(29). Those equations show the steady state $b_{(r,k)}$ in each Markov chain model can be expressed in the form of the AC specific transition probabilities, including P_{idleA} , P_{idleB} , P_{sB} , P_{bA} , P_{bB} , $Pr_A(r)$, and $Pr_B(r)$.
- (2) The above AC specific transition probabilities for each Markov chain model shown in Fig. 3 are analyzed in Section 3.2 and they can be expressed in terms of τ_A and τ_B .
- (3) By using the system equations in Section 3.1 and the transition probabilities expressed in terms of τ_A and τ_B Section 3.2, the steady state probability $b_{(r,k)}$ for both Markov chain models shown in Fig. 3 can be obtained in terms of τ_A and τ_B .
- (4) Finally two non-linear equations about τ_A and τ_B based on equations (25)

and (28) can be constructed for the AC specific Markov chain models shown in Fig. 3. The values of τ_A and τ_B can be obtained from the equations.

4 SATURATED THROUGHPUT ANALYSIS FOR EDCA

In this section, we shall analyze the saturated throughput of EDCA. We consider that the throughput is equal to the ratio of the effective payload to the time required for successfully transmitting the effective payload. The Markov chain model shown in Fig. 6 is used to obtain the throughput, and its state probabilities can be obtained after τ_A and τ_B are solved. This Markov chain model represents the time slot distribution between two successive transmissions in the WLAN. Two possible events may occur in a time slot:

- (1) At least one transmission occurs in the time slot. Depending on whether the time slot is in zone 1 or zone 2 a transmission may occur with a probability of $P_{tr:zone(1)}$ or $P_{tr:zone(2)}$. Furthermore, depending on whether the transmission is successful or not, two possibilities may occur:
 - (a) A successful transmission. That is, only one transmission from either a priority A station or a priority B station occurs in the time slot. The corresponding contention zone probability for a successful transmission can be obtained by

$$\begin{cases} P_{sucA:zone(1)} = n_A \tau_A (1 - \tau_A)^{n_A - 1}, \\ P_{sucA:zone(2)} = n_A \tau_A (1 - \tau_A)^{n_A - 1} (1 - \tau_B)^{n_B}, \\ P_{sucB:zone(1)} = 0, \\ P_{sucB:zone(2)} = n_B \tau_B (1 - \tau_B)^{n_B - 1} (1 - \tau_A)^{n_A}. \end{cases} \quad (59)$$

- (b) A collision. That is, two or more stations start transmitting in the same time slot. The corresponding contention zone specific collision probability can be obtained by

$$\begin{cases} P_{col:zone(1)} = P_{tr:zone(1)} - P_{sucA:zone(1)} - P_{sucB:zone(1)}, \\ P_{col:zone(2)} = P_{tr:zone(2)} - P_{sucA:zone(2)} - P_{sucB:zone(2)}. \end{cases} \quad (60)$$

- (2) No transmission occurs in the time slot. The corresponding contention zone specific probability for an idle time slot can be obtained by

$$\begin{cases} P_{idle:zone(1)} = 1 - P_{tr:zone(1)}, \\ P_{idle:zone(2)} = 1 - P_{tr:zone(2)}. \end{cases} \quad (61)$$

Therefore, the average effective payload for priority A stations can be obtained as:

$$E[A] = \sum_{r=1}^M P_{sucA:zone(r)} s_{(r)} E[P], \quad (62)$$

where $E[P]$ is the payload size of a frame, and $s_{(r)}$ can be obtained from (40) and (41). $E[P]$ is considered as a known constant. The effective payload for priority A station measures the effective amount of priority A traffic that is transmitted between two successive transmissions.

Similarly, the average effective payload for AC[B] stations can be obtained by

$$E[B] = \sum_{r=1}^M P_{sucB:zone(r)} s_{(r)} E[P]. \quad (63)$$

The average time duration between two successive transmission can be obtained as:

$$EL = \sum_{r=1}^M s_{(r)} [(P_{sucB:zone(r)} + P_{sucA:zone(r)})Ts + P_{col:zone(r)}Tc + P_{idle:zone(r)}aTimeSlot], \quad (64)$$

where Ts and Tc are time required for a successful transmission and a collision respectively. They are illustrated in Fig. 8 and can be obtained by

$$Ts = H + P + SIFS + ACK + AIFS_{min}, \quad (65)$$

and

$$Tc = H + P + EIFS_{min}, \quad (66)$$

where H is the time required for transmitting the physical layer header and the MAC layer header of a frame, P is the time required for transmitting the data payload of a frame, ACK is the duration for transmitting an ACK frame, $AIFS_{min}$ is the minimum AIFS used in the WLAN, and $EIFS_{min}$ equals to $SIFS + ACK + AIFS_{min}$. Here a basic access data rate determined by the WLAN physical layer is used for transmitting the physical layer header and ACK frame, while the payload data rate of sending MAC layer header and payload can be higher [22, p. 11].

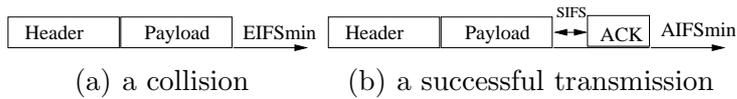


Fig. 8. Transmission duration

Finally, the throughput for each station of each AC can be obtained by

$$\begin{cases} \text{Throughput}_A = E[A]/EL/n_A, \\ \text{Throughput}_B = E[B]/EL/n_B. \end{cases} \quad (67)$$

5 Simulation Study

In this section, the theoretical analysis presented in the earlier sections is validated using simulation. Simulation is conducted using OPNET [23]. The impact of using different AIFSs and different CW sizes on network performance is analyzed. Finally, comparison is performed between theoretical results obtained using the proposed model and those in [16–19], which demonstrates that the proposed model has better accuracy.

The parameters used in the simulation are shown in Table 1. Four ACs are used in the simulation and their parameters are consistent with those defined in [2, Table 20df, p.49]. Two scenarios are simulated. In the first scenario, two ACs, i.e., voice and video, are used. This scenario is designed to investigate the effect of using different CW sizes since a common AIFS but different CW sizes are used by AC[voice] and AC[video] respectively. In the second scenario, two ACs, i.e., best effort and background, are used. The purpose of this scenario is to investigate the effect of using different AIFS, since a common CW size but different AIFS are used by AC[best effort] and AC[background] respectively. In both scenarios, there are equal number of stations in each AC.

Fig. 9 shows the simulation result as well as theoretical results obtained from the proposed model for the first scenario. The throughput of a station in a specific AC under different number of stations is shown. It is shown in the figure that theoretical results obtained from the proposed model generally agree very well with simulation results. As shown in the figure, by using different CW_{min} and CW_{max} , traffic is successfully classified into two different classes. Traffic with a smaller CW_{min} and CW_{max} can have better quality of service. When the number of stations in each AC is small, the difference in throughput for each AC is significant. When the number of stations in each AC increases, the difference in throughput decreases and throughput of both ACs decreases significantly due to more number of stations contending for bandwidth.

Fig. 10 shows the simulation result as well as theoretical results obtained from the proposed model for the second scenario. As shown in the figure, by using different AIFSs, traffic is successfully classified into two different classes, and this difference is more significant than that in the first scenario. Traffic with a smaller AIFS can have better quality of service. It should be noticed that

Table 1
WLAN parameter setting

PHY header	192 bits
MAC header	224 bits
Frame payload size	8000 bits
ACK frame size	PHY header+112 bits
MSDU frame size	PHY header+MAC header+ Frame payload
Physical layer	IEEE 802.11b DSSS [22]
Basic access data rate	1Mbps
Payload data rate	1Mbps
Time slot	20 μ s
SIFS	10 μ s
Maximum retransmission limit	7
AIFSN	AIFSN[voice]=2, AIFSN[video]=2, AIFSN[best effort]=3, AIFSN[background]=7
CW[voice]	$CW_{min} = 7, CW_{max} = 15$
CW[video]	$CW_{min} = 15, CW_{max} = 31$
CW[best effort]	$CW_{min} = 15, CW_{max} = 1023$
CW[background]	$CW_{min} = 15, CW_{max} = 1023$

when the number of stations in each AC increases, the lower priority traffic belonging to the background AC may be starved.

The effects of AIFS and CW size on traffic prioritization observed in the simulation results as well as theoretical results can be easily explained. Use of different AIFSNs introduces the contention zone specific transmission probability. Lower priority station may be excluded for transmission in some contention zone, which results in the possibility that some higher priority stations monopolize transmission opportunities and bandwidth. However, use of different CW sizes will only result in longer delay for lower priority stations and lower priority stations can still get the opportunity to transmit. Moreover, as shown in Fig. 9, when the number of voice and video stations increases, the throughput of both ACs drops severely. The reason is both AC[voice] and AC[video] have small AIFS and CW values. This enables stations to have a high transmission probability at a time slot, and accordingly their transmission will suffer a high collision probability when the number of stations is large. Therefore the majority of the available bandwidth is wasted on collision instead of successful transmission.

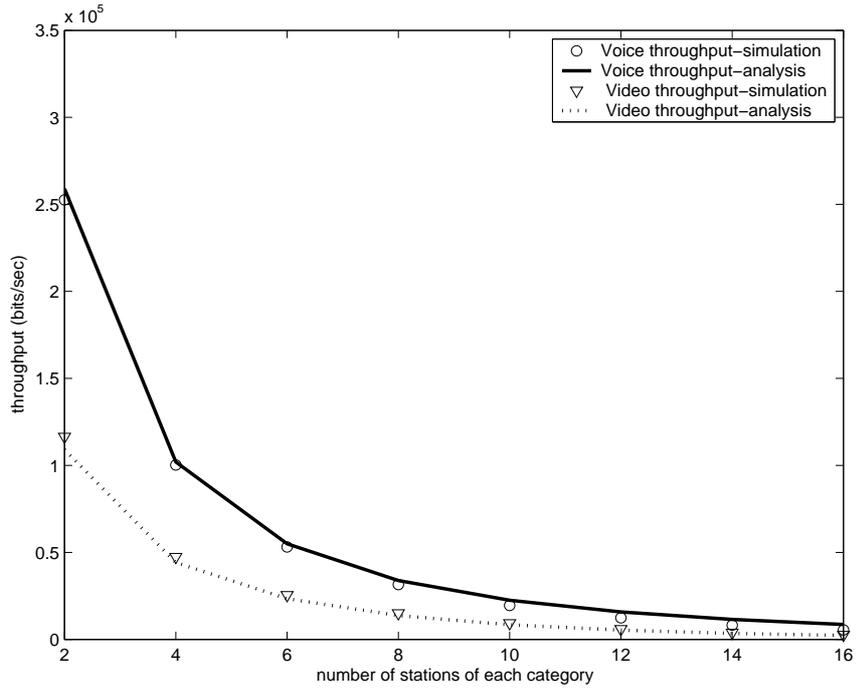


Fig. 9. Simulation and analysis results for AC[voice] and AC[video]

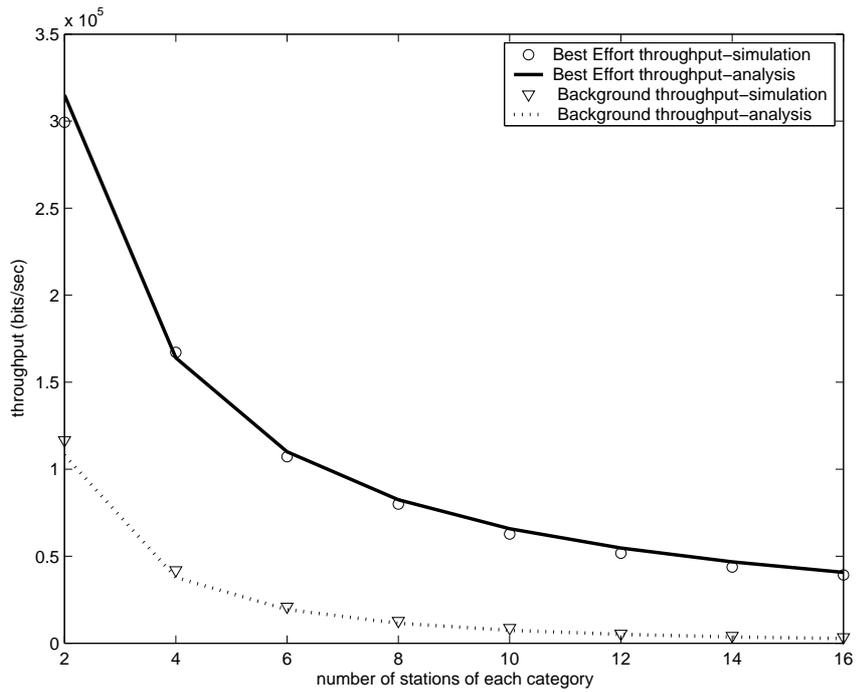


Fig. 10. Simulation and analysis results for AC[best effort] and AC[background]

Finally, a larger discrepancy between theoretical and simulation results at smaller number of stations is observed. It results from the assumption used in the model, that is, the transmission probability at a generic time slot is constant. As pointed out in [21], this assumption is more accurate when the

number of stations is larger.

5.1 Comparison

The results obtained in this paper is compared with those in [16–19]. Note that some typo errors existing in some models have been corrected in order to generate meaningful comparison. Firstly, equation (17) in [17] has been revised as $p1 = 1 - (1 - \tau_1)^{n_1-1} [P_{hold} + (1 - P_{hold})(1 - \tau_2)^{n_2-1}]$, because a wrong term P_{temp} instead of P_{hold} is used in [17]. This typo error has been confirmed by personal communication with the authors. Secondly, equation (2) in [19] considers that the probability of allocating a random initial backoff counter within a range of $[0 \text{ CW}]$ is 1, which is apparently incorrect and can lead to that a solution cannot be obtained. We revise the probability to $\frac{1}{CW+1}$, and the revised equation (2) is given by

$$\begin{cases} P^{(l)}(0, 0, k|i, j, 0) = \frac{1-P_{b,i}^{(l)}}{CW_0^{(l)}+1} & k \in [0, CW_0^{(l)}] \\ P^{(l)}(0, j+1, k|i, j, 0) = \frac{P_{b,i}^{(l)}}{CW_{j+1}^{(l)}+1} & k \in [0, CW_{j+1}^{(l)}] \\ P^{(l)}(0, m, k|i, m, 0) = \frac{P_{b,i}^{(l)}}{CW_m^{(l)}+1} & k \in [0, CW_m^{(l)}] \end{cases} \quad (68)$$

The results of the comparison are shown in Fig. 12-14. As shown in the results, the proposed model can achieve better accuracy than those in [16–19]. These results are expected as the proposed model captures the complexity of EDCA and removes some problems in [16–19]. These have been explained in detail in Section 2.

6 CONCLUSION

In this paper, a novel Markov chain model for EDCA performance analysis under the saturated traffic load was proposed. Compared with the existing analytical models of EDCA, the proposed model incorporated more features of EDCA into the analysis and eliminated their limitations. Both the effect of the contention zone specific transmission probability differentiation caused by using different AIFS and the effect of backoff suspension caused by transmission from other stations are considered. Based on the proposed model, the saturated throughput of EDCA was analyzed. Simulation study using OPNET was performed, which demonstrated that theoretical results obtained from the proposed model can closely match simulation results, and the proposed model has better accuracy than that in the literature.

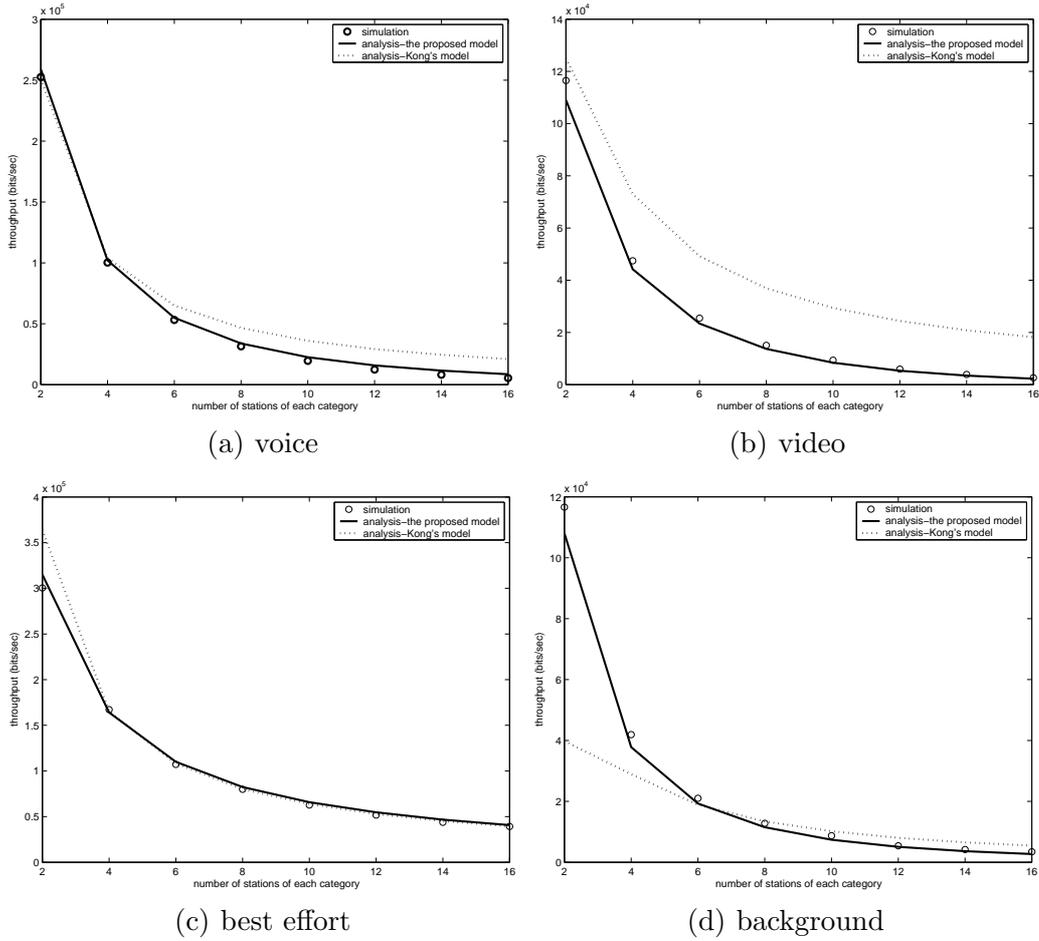


Fig. 11. Comparison results with the model in [16].

Despite the improvement, the analysis presented in this paper was based on the saturated throughput assumption. In a real network, traffic from a station is more likely to be non-saturated. Therefore a more interesting scenario will be throughput in non-saturated conditions. Moreover, wireless channel is characterized by the relatively higher bit error rate due to noise and interference. The effect of noise on EDCA performance should also be considered. These problems shall be addressed in our future research. These problems shall be addressed in our future research.

References

- [1] 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 (1999).
- [2] IEEE Standard for Information technology - Telecommunications and

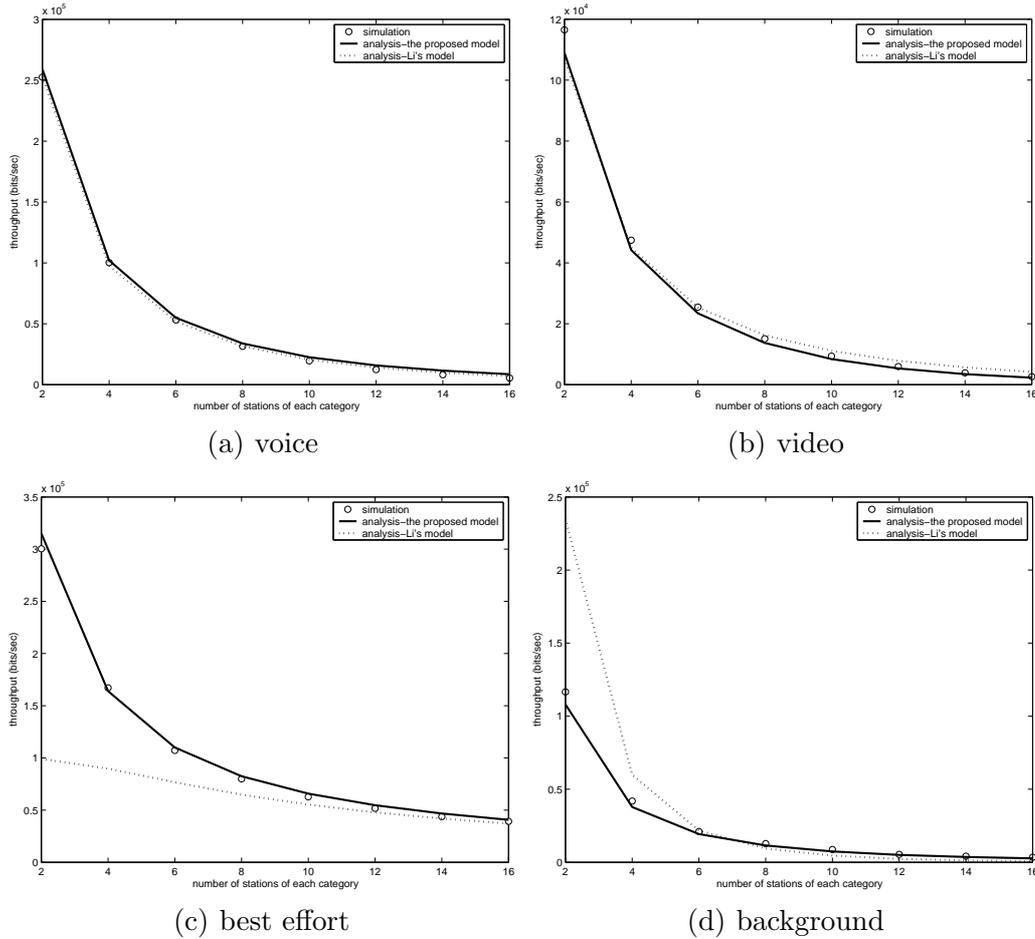


Fig. 12. Comparison results with the model in [17].

information exchange between systems - Local and metropolitan area networks
- Specific requirements Part 11: Wireless LAN Medium Access Control (MAC)
and Physical Layer (PHY) specifications Amendment 8: Medium Access Control
(MAC) Quality of Service Enhancements (2005).

- [3] G. Bianchi, I. Tinnirello, Remarks on IEEE 802.11 DCF performance analysis, *Communications Letters, IEEE* 9 (8) (2005) 765–767.
- [4] 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 (2003).
- [5] G. Bianchi, I. Tinnirello, L. Scalia, Understanding 802.11e contention-based prioritization mechanisms and their coexistence with legacy 802.11 stations, *Network, IEEE* 19 (4) (2005) 28–34.
- [6] Y. Chen, Q.-A. Zeng, D. Agrawal, Performance analysis of IEEE 802.11e enhanced distributed coordination function, in: *Networks, 2003. ICON2003. The 11th IEEE International Conference on*, 2003, pp. 573–578.

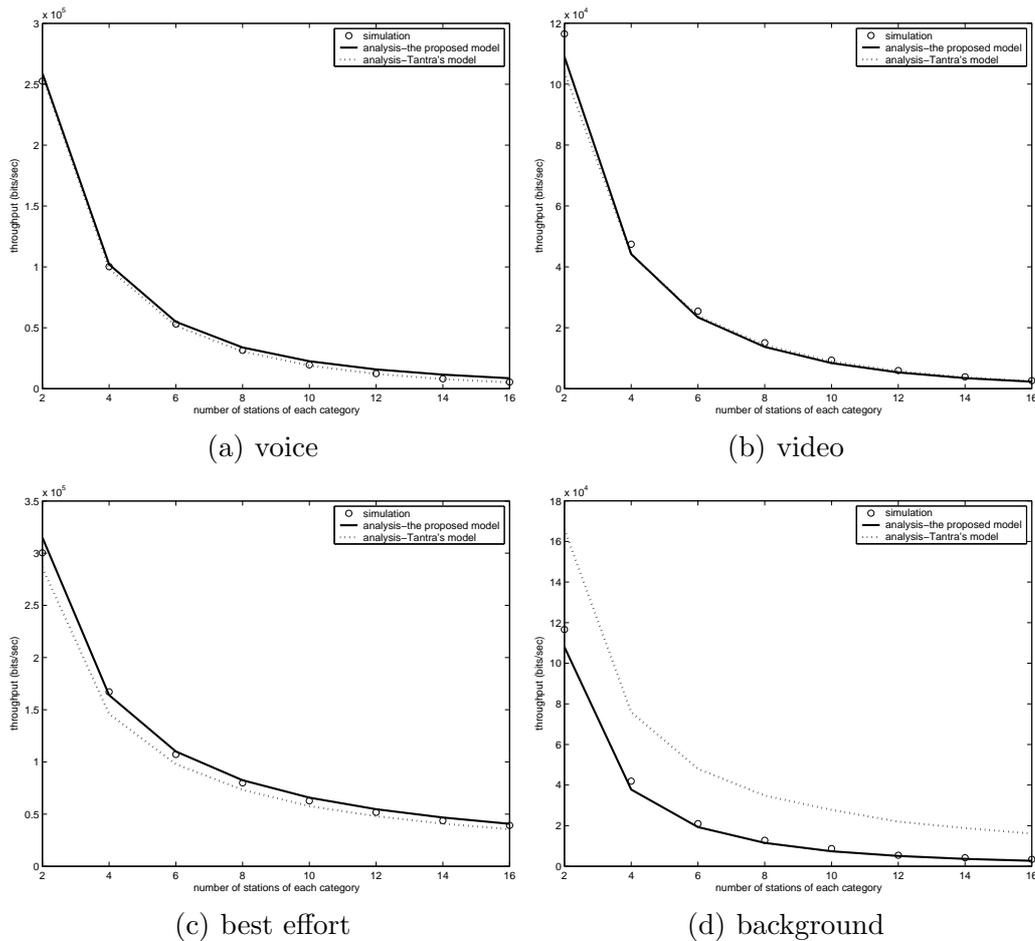


Fig. 13. Comparison results with the model in [18].

- [7] Y.-L. Kuo, C.-H. Lu, E. H.-K. Wu, G.-H. Chen, Y.-H. Tseng, Performance analysis of the enhanced distributed coordination function in the IEEE 802.11e, in: Vehicular Technology Conference, 2003. VTC 2003-Fall. 2003 IEEE 58th, Vol. 5, 2003, pp. 3488–3492 Vol.5.
- [8] J. Robinson, T. Randhawa, Saturation throughput analysis of IEEE 802.11e enhanced distributed coordination function, Selected Areas in Communications, IEEE Journal on 22 (5) (2004) 917–928.
- [9] K. Xu, Q. Wang, H. Hassanein, Performance analysis of differentiated QoS supported by IEEE 802.11e enhanced distributed coordination function (EDCF) in WLAN, in: Global Telecommunications Conference, 2003. GLOBECOM '03. IEEE, Vol. 2, 2003, pp. 1048–1053 Vol.2.
- [10] A. Banchs, L. Vullero, A delay model for IEEE 802.11e EDCA, Communications Letters, IEEE 9 (6) (2005) 508–510.
- [11] X. Chen, H. Zhai, Y. Fang, Enhancing the IEEE 802.11e in QoS Support: Analysis and Mechanisms, in: Quality of Service in Heterogeneous Wired/Wireless Networks, 2005. Second International Conference on, 2005, p. 23.

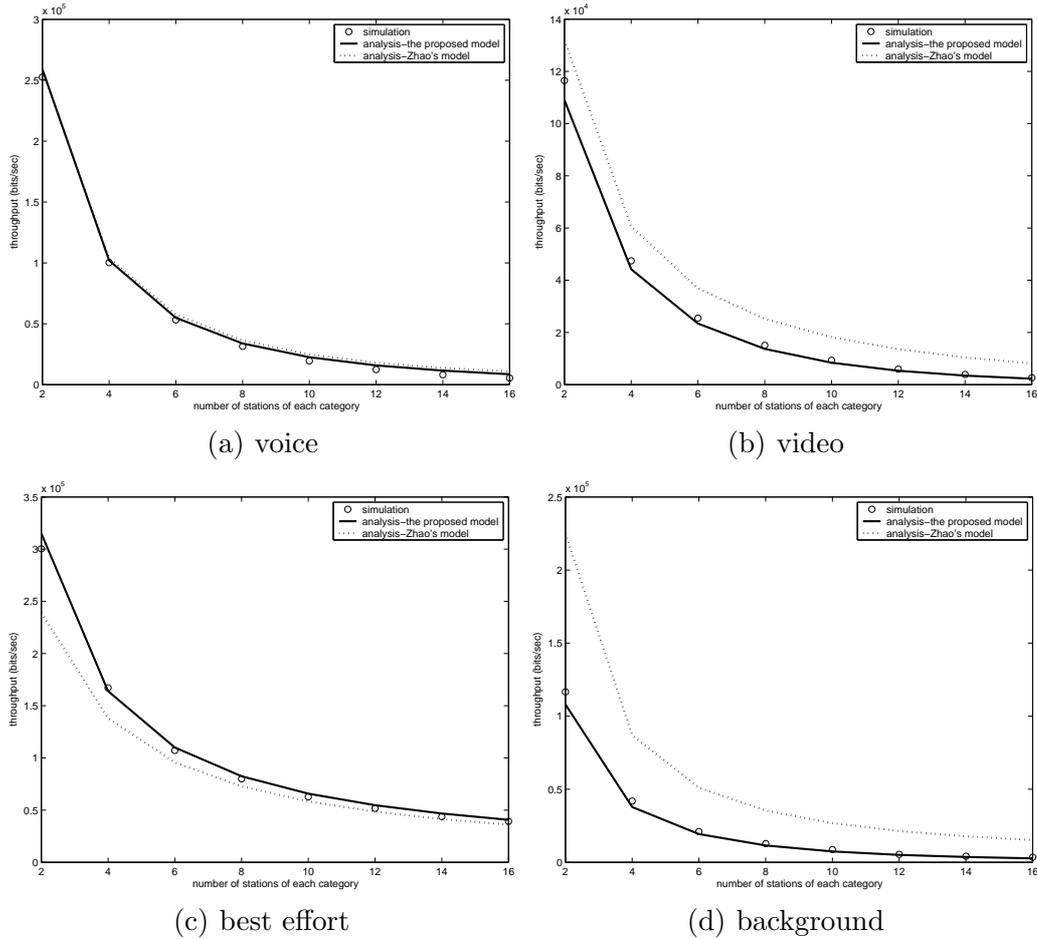


Fig. 14. Comparison results with the model in [19].

- [12] H. Zhu, I. Chlamtac, Performance analysis for IEEE 802.11e EDCF service differentiation, *Wireless Communications, IEEE Transactions on* 4 (4) (2005) 1779–1788.
- [13] J. Hui, M. Devetsikiotis, A Unified Model for the Performance Analysis of IEEE 802.11e EDCA, *Communications, IEEE Transactions on* 53 (9) (2005) 1498–1510.
- [14] Y. Xiao, Performance analysis of priority schemes for IEEE 802.11 and IEEE 802.11e wireless LANs, *Wireless Communications, IEEE Transactions on* 4 (4) (2005) 1506–1515.
- [15] P. Engelstad, O. Osterbo, Delay and Throughput Analysis of IEEE 802.11e EDCA with Starvation Prediction, in: *Local Computer Networks, 2005. 30th Anniversary. The IEEE Conference on*, 2005, pp. 647–655.
- [16] Z.-n. Kong, D. Tsang, B. Bensaou, D. Gao, Performance analysis of IEEE 802.11e contention-based channel access, *Selected Areas in Communications, IEEE Journal on* 22 (10) (2004) 2095–2106.
- [17] B. Li, R. Battiti, Supporting Service Differentiation with Enhancements of the IEEE 802.11 MAC Protocol: Models and Analysis, Technical Report of

Department of Computer Science and Telecommunications of University of Trento, no. DIT-03-024].

- [18] J. Tantra, C. H. Foh, A. Mnaouer, Throughput and delay analysis of the IEEE 802.11e EDCA saturation, in: Communications, 2005. ICC 2005. 2005 IEEE International Conference on, Vol. 5, 2005, pp. 3450–3454 Vol. 5.
- [19] J. Zhao, Z. Guo, Q. Zhang, W. Zhu, Performance study of MAC for service differentiation in IEEE 802.11, in: Global Telecommunications Conference, 2002. GLOBECOM '02. IEEE, Vol. 1, 2002, pp. 778–782 vol.1.
- [20] Y. C. Tay, K. C. Chua, A capacity analysis for the IEEE 802.11 MAC protocol, *Wirel. Netw.* 7 (2) (2001) 159–171.
- [21] G. Bianchi, Performance analysis of the IEEE 802.11 distributed coordination function, *Selected Areas in Communications, IEEE Journal on* 18 (3) (2000) 535–547.
- [22] Supplement to 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: Higher-speed physical layer extension in the 2.4 GHz band (2000).
- [23] OPNET University Program: <http://www.opnet.com/services/university/>.