UNIVERSITY
OF TRENTO

DEPARTMENT OF INFORMATION AND COMMUNICATION TECHNOLOGY

38050 Povo – Trento (Italy), Via Sommarive 14
http://www.dit.unitn.it

SUPPORTING SERVICE DIFFERENTIATION WITH
ENHANCEMENTS OF THE IEEE 802.11 MAC PROTOCOL:
MODELS AND ANALYSIS

Bo Li and Roberto Battiti

May 2003

Technical Report # DIT-03-024

[1] Supporting Service Differentiation with Enhancements of the IEEE 802.11 MAC Protocol: Models and Analysis

Bo Li and Roberto Battiti

Department of Computer Science and Telecommunications, University of Trento, 38050 POVO Trento, Italy
e-mail: {li, battiti}@dit.unitn.it

Abstract. As one of the fastest growing wireless access technologies, Wireless LANs must evolve to support adequate degrees of service differentiation. Unfortunately, current WLAN standards like IEEE 802.11 Distributed Coordination Function (DCF) lack this ability. Work is in progress to define an enhanced version capable of supporting QoS for multimedia traffic at the MAC layer. In this paper, we aim at gaining insight into three mechanisms to differentiate among traffic categories, i.e., differentiating the minimum contention window size, the Inter-Frame Spacing (IFS) and the length of the packet payload according to the priority of different traffic categories. We propose an analysis model to compute the throughput and packet transmission delays. In additions, we derive approximations to get simpler but more meaningful relationships among different parameters. Comparisons with discrete-event simulation results show that a very good accuracy of performance evaluation can be achieved by using the proposed analysis model.

Keyword: Wireless LAN, IEEE 802.11 MAC, Service Differentiation, Performance Evaluation

I. Introduction

The main objective of the next-generation broadband wireless networks is to provision seamless multimedia services to mobile users. In this context, one of the major challenges of the wireless mobile Internet is to provide suitable levels of Quality of Service over IP-based wireless access networks [1]. Current approaches to provide IP QoS include IntServ [2] and DiffServ [3] architectures. The IntServ architecture defines mechanisms for per-flow QoS management and provides tight performance guarantees for high priority flows, while the DiffServ architecture defines aggregated behavior for a limited number of performance

classes for which only statistical differentiation is provided, therefore achieving better scaling performance. Wireless access may be considered just another hop in the communication path for the whole Internet. Therefore, it is desirable that the architecture supporting quality assurances follows the same principles in the wireless networks as in the wireline Internet, assuring compatibility between the wireless and wireline parts. A good example for such a wireless technology is the IEEE 802.11 Distributed Coordination Function (DCF) standard [4], which is compatible with the current best-effort service model of the Internet.

In the literature, performance evaluation of 802.11 has been executed by using simulation [5] or by means of analytical models [6]-[11]. Constant or geometrically distributed backoff window sizes have been considered in [6], [7], and [8]. In [9], an exponential backoff with only two stages is modeled by using a two-dimensional Markov chain. In [10], a more general model that accounts for all the exponential backoff protocol details is proposed. In [11], instead of using stochastic analysis, the average value for a variable is used, which results in an approximate but effective analysis.

In order to support different QoS requirements for various types of service, a possibility is to support differentiation in the IEEE 802.11 MAC layer, as proposed in [12] - [15]. In [12], a simple priority scheme for IEEE 802.11 has been proposed, where a high priority station has a shorter waiting time when accessing the medium. In [13], a service differentiation scheme is proposed. The scheme uses two parameters of IEEE 802.11 MAC, the backoff interval and the IFS between each data transmission, to provide the differentiation. In [14], service differentiation is supported by setting different minimum contention windows for different types of services. The effectiveness is demonstrated by simulation. The work in [15] proposes three service differentiation schemes for IEEE 802.11 DCF. The first one is based on scaling the contention window according to the priority of each flow. The second one assigns different Inter-Frame Spacings to different traffic classes. The third one uses different maximum frame lengths. Moreover, an effective Contention Window (CW) resetting scheme to enhance the performance of IEEE 802.11 DCF is analyzed in [16], by extending the model proposed in [10]. In [17], both the Enhanced Distributed Coordination Function (EDCF) and the Hybrid Coordination Function (HCF), defined in the IEEE 802.11e draft are evaluated through simulation.

In order to gain a deeper insight into the modified IEEE 802.11 MAC with service differentiation support, system modeling and performance analysis are needed. By building on previous papers dealing with the analysis of the IEEE 802.11 MAC, we propose an

analysis model to compute the throughput and packet transmission delays in a WLAN with Enhanced IEEE 802.11 Distributed Coordination Function, with support for service differentiation. In our analytical model, service differentiation is supported by differentiating the contention window size, the Inter-Frame Spacing, and the packet length according to the priority of each traffic flow.

The paper is organized as follows: in Section II a brief description on the IEEE 802.11 DCF and enhancements is given. In Section III the model is defined and system performance measures, i.e., throughput and average packet delays, are obtained. In Section IV, some approximations are considered to obtain simpler formulas allowing a deeper insight into the system. Then, discrete-event simulation and numerical results obtained from the analysis are presented and discussed in Section V.

II. IEEE 802.11 Distributed Coordination and Enhancements

In the 802.11 MAC sub-layer, two services have been defined: the Distributed Coordination Function (DCF), which supports delay insensitive data transmissions, and the optional Point Coordination Function (PCF), based on polling and intended to support delay sensitive transmissions. The basic 802.11 MAC protocol, the Distributed Coordination Function (DCF), works as listen-before-talk scheme, based on Carrier Sense Multiple Access (CSMA) with a Collision Avoidance (CA) mechanism to avoid collisions that can be anticipated if terminals are aware of the duration of ongoing transmissions ("virtual carrier sense"). Following is a brief summary of the DCF protocol.

When the MAC receives a request to transmit a frame, a check is made of the physical and virtual carrier sense mechanisms. If the medium is not in use for an interval of DIFS, the MAC may begin transmission of the frame. If the medium is in use during the DIFS interval, the MAC selects a backoff time and increments the retry counter. The backoff time is randomly and uniformly chosen in the range $(0, W-1)$, W being the contention window. The MAC decrements the backoff value each time the medium is detected to be idle for an interval of one slot time. The terminal starts transmitting a packet when the backoff value reaches zero. When a station transmits a packet, it must receive an ACK frame from the receiver after SIFS (plus the propagation delay) or it will consider the transmission to have failed. If a failure happens, the station reschedules the packet transmission according to the given backoff rules. If there is a collision, the contention window is doubled, and a new backoff interval is

selected. At the first transmission attempt, W is set equal to a value CW_{min} called minimum contention window. After each unsuccessful transmission, W is doubled, up to a maximum value $CW_{max} = 2^m \cdot CW_{min}$.

The above-described two-way handshaking technique for the packet transmission is called *basic access* mechanism. Furthermore, in order to overcome the hidden station problem, 802.11 defines an optional Request-to-send/Clear-to-send (RTS/CTS) mechanism. In the RTS/CTS mechanism, before transmitting a data frame, the sender sends a short RTS frame. On receiving the RTS frame, the receiver transmits a CTS frame immediately. The RTS and CTS frames include information about the time required to transmit the next data frame, i.e., the first fragment, and the corresponding ACK response. Therefore, other stations close to the sender and hidden stations close to the receiver will not start any transmission during the advertised frame transmission period. In the paper, only the *basic access* mechanism is analyzed. The analysis method can be easily extended to the case of *RTS/CTS access* mechanism.

The basic DCF method is not appropriate for handling multimedia traffic requiring guarantees about throughput and delay. Because of this weakness, task group E of the IEEE 802.11 working group is currently working on an enhanced version of the standard called IEEE 802.11e. The goal of the extension is to provide a distributed access mechanism capable of service differentiation. A new access mechanism called Enhanced DCF (EDCF) has been selected [18]. It is shown by simulation that EDCF has better performance than PDF, and is more scalable [19]. In this paper, in the interest of conciseness, we are not interested in exploring all details of the new proposed standard but to gain insight into three of the building block used to achieve differentiation, i.e. differentiating the minimum contention window sizes, the Inter-Frame Spacing (IFS) and the lengths of packet payload according to the priority of each traffic category.

III. Performance Analysis

A. System Modeling
We assume that the channel conditions are ideal (i.e., no hidden terminals and capture) and that the system operates in saturation: a fixed number of traffic flows always have a packet available for transmission.

Because our analysis can be easily extended and for the sake of simplicity, only two

different types of traffic are considered with n_i traffic flows for traffic of type i $(i=1,2)$. We assume that type-1 traffic has a higher priority than type-2 traffic. Let $b_i(t)$ be the stochastic process representing the backoff time counter for a given traffic flow with type i $(i=1,2)$. Moreover, let us define for convenience $W_i = CW_{min,i}$ as the minimum contention window for traffic type i. Let m_i, "maximum backoff stage" be the value such that $CW_{max,i} = 2^{m_i} \cdot W_i$. Let $s_i(t)$ be the stochastic process representing the backoff stage $(0,1,...,m_i)$ for a given traffic flow with type i. Let DIF_i denote the length of the distributed inter-frame space (DIF) of traffic flows with type i. σ is the duration of an empty slot time. Moreover, we define $D \equiv \dfrac{DIF_2 - DIF_1}{\sigma}$ as the difference between DIF_2 and DIF_1 measured in the number of empty slots. In this paper, it is assumed that $DIF_2 \geq DIF_1$. Moreover, we only consider the case that D is an integer.

The key approximation in the model is that, at each transmission attempt for a traffic flow of type i, regardless of the number of retransmissions suffered, each packet collides with constant and independent probability p_i. This assumption has been shown by simulation to be very accurate if W_i and n_i are large [10]. The parameter p_i is referred to as conditional collision probability, the probability of a collision seen by a packet belonging to a traffic flow with type i at the time of its being transmitted on the channel. We *do not* assume that $p_1 = p_2$. Later, our analysis results justify this assumption.

In the following, we use a two-dimensional discrete-time Markov chain to model the behavior of a traffic flow with type 1. The states are defined as the combinations of two integers $\{s_1(t), b_1(t)\}$. The Markov chain can be presented as follows (see Fig.1a)

Case 1: Before packet transmissions,

$$P\{j,k \mid j,k+1\} = 1 \quad (k \in [0, 2^j \cdot W_1 - 2], j \in [0, m_1]) \tag{1.1}$$

Case 2: After packet transmissions,

$$\begin{cases} P\{j+1,k \mid j,0\} = \dfrac{p_1}{2^{j+1}W_1} & (j \in [0,m_1), k \in [0, 2^{j+1} \cdot W_1 - 1]) & (2.1) \\[3mm] P\{m_1,k \mid m_1,0\} = \dfrac{p_1}{2^{m_1}W_1} & (k \in [0, 2^{m_1} \cdot W_1 - 1]) & (2.2) \\[3mm] P\{0,k \mid j,0\} = \dfrac{(1-p_1)}{W_1} & (j \in [0,m_1], k \in [0, W_1 - 1]) & (2.3) \end{cases}$$

5

Equation (1.1) accounts for the fact that the backoff time is decremented at the beginning of each slot time. Equation (2.1) accounts for the fact that a collision occurs and the station doubles the backoff time. Equation (2.2) models the fact that once the backoff stage reaches the value m_1, it is not increased in subsequent packet transmissions. Equation (2.3) accounts for the fact that a new packet following a successful packet transmission starts with backoff stage 0.

Let's assume that DIF_2 is different from DIF_1, and therefore larger according to our definitions. When DIF_1 expires after the last transmission, type-1 traffic flows immediately step into their backoff stages, while type-2 traffic flows must still wait for the end of the longer inter-frame space DIF_2. Traffic of type 2 has a chance of being transmitted only if *all* traffic of higher priority (of type 1) select a backoff time sufficiently large that DIF_2 also expires before transmission is started. When this event occurs, competition to access the channel is determined by the minimum among backoff counters of all active traffic flows. Of course, the event never occurs in saturated conditions, and type-2 traffic is completely starved, if the difference D is larger than the maximum contention window size for type-1 traffic. We model this process by adding a sequence of *hold states* that type-2 traffic must enter before possibly starting to decrement the backoff counter.

Therefore, we model the behavior of a type-2 traffic flow with a three dimensional Markov chain. The states are defined as the combinations of three integers $\{s_2(t), b_2(t), d(t)\}$, where $0 \le d(t) \le D$. That is to say, for a *normal* state $\{s_2(t), b_2(t), 0\}$, the corresponding *hold state* is further sub-divided into D stages. Therefore, if $d(t) > 0$, the considered traffic flow is waiting at some stage of the *hold state*. For better understanding the state transitions of different stations carrying different types of traffic, we give an example for the state transitions of two different types of traffic flows in Fig. 2. In the example, for simplicity, we assume that there are only two traffic flows in the system. Flow 1 carries type-1 traffic, and Flow 2 carries type-2 traffic. Moreover, the DIF difference D between inter-frame separation is 1. In the figure, states for the two traffic flows are labeled in the corresponding time slot. In the time slot from $t1$ to $t2$, Flow 1 is at state $\{1,0\}$, and it begins sending a packet. Flow 2 is at state $\{0,2,0\}$. From $t2$ to $t3$, after the sending of the packet, Flow 1 begins another backoff process. At the same time, Flow 2 step into the *hold state* $\{0,2,1\}$ to wait for the end of DIF_2. From $t3$ to $t4$, Flow 1 further decreases its backoff counter from 2 to 1. Flow 2 resumes its backoff process and decreases its backoff counter from 2 to 1. From $t4$ to $t5$, both Flow 1

and Flow 2 try to send packets at the same time. Therefore, a collision occurs. From $t5$ to $t6$, Flow 1 begins a new backoff process in order to send the packet again. Flow 2 steps into hold state $\{0,0,1\}$. From $t6$ to $t7$, Flow 2 begins a new backoff process in order to send the collided packet again. Flow 1 continues its backoff process by decreasing its backoff time from 5 to 4.

In the following, we describe the state transitions mathematically:

Case 1: Before packet transmissions (see Fig. 1b)

$$
\begin{cases}
P\{j,k,0 \mid j,k+1,0\} = P_s & (3.1) \\
P\{j,k+1,1 \mid j,k+1,0\} = 1 - P_s & (3.2) \\
P\{j,k+1,d+1 \mid j,k+1,d\} = P_{s1} \quad (d \in [1, D-1]) & (3.3) \\
P\{j,k,0 \mid j,k+1,D\} = P_{s1} & (3.4) \\
P\{j,k+1,1 \mid j,k+1,d\} = 1 - P_{s1} \quad (d \in [1,D]) & (3.5)
\end{cases}
$$

where $k \in [0, 2^j \cdot W_2 - 2]$, $j \in [0, m_2]$. P_s ("probability of silence") is the probability that all the other traffic flows do not transmit under the condition that the considered traffic flow is at the state of $\{j,k+1,0\}$. Note that the considered traffic flow is not at the state of transmission. P_{s1} is the probability that all the type-1 traffic flows do not transmit. In order to make the model tractable, we assume that, regardless of the different stages for different *hold states*, P_{s1} is an independent constant probability. This assumption has been shown by our simulation to be very accurate if D is not so large that type-2 traffic flows are almost starved by type-1 traffic flows.

Equation (3.1) accounts for the fact that the backoff counter is decremented at the beginning of each slot time if all traffic flows are not in the state of transmission. Equation (3.2) means that for the considered type-2 traffic flows, if there are some other traffic flows in the state of transmission during the past time slot, its next state is a *hold state*. If the considered traffic flow is at a *hold state*, it transits into the next *hold state* with probability P_{s1} (see equation (3.3)). Or it goes back to the first *hold state* with probability $1 - P_{s1}$ because of the transmission of at least one type-1 traffic flow during the past time slot, see equation (3.5). Equation (3.4) shows that for the considered traffic flow, the traffic flow will step into normal state after the last *hold state* with probability P_{s1}.

Case 2: After packet transmission (see Fig. 1c)

7

$$
\begin{cases}
P\{j,0,1 \mid j,0,0\} = 1 & (4.1) \\[2mm]
P\{j,0,d+1 \mid j,0,d\} = P_{s1} & (d \in [1, D-1]) & (4.2) \\[2mm]
P\{j+1,k,0 \mid j,0,D\} = \dfrac{P_{s1} \cdot p_2}{2^{j+1} \cdot W_2} & (j < m_2, k \in [0, 2^{j+1} \cdot W_2 - 1]) & (4.3) \\[2mm]
P\{m_2,k,0 \mid m_2,0,D\} = \dfrac{P_{s1} \cdot p_2}{2^{m_2} \cdot W_2} & (k \in [0, 2^{m_2} \cdot W_2 - 1]) & (4.4) \\[2mm]
P\{0,k,0 \mid j,0,D\} = \dfrac{P_{s1} \cdot (1 - p_2)}{W_2} & (k \in [0, W_2 - 1]) & (4.5) \\[2mm]
P\{j,0,1 \mid j,0,d\} = 1 - P_{s1} & (d \in [1, D]) & (4.6)
\end{cases}
$$

where $j \in [0, m_2]$. Equation (4.1) accounts for the fact that after the transmission of a packet, the considered type-2 traffic flow steps into hold state with probability 1. If the considered traffic flow is at a hold state, it transits into the next hold state with probability P_{s1}, see equation (4.2), or it goes back to the first hold state with probability $1 - P_{s1}$ because of the transmission of at least one type-1 traffic flow during the past time slot, see equation (4.6). Equation (4.3) accounts for the fact that a collision occurs and the station carrying type-2 traffic doubles the backoff time after its sending of a packet. Equation (4.4) models the fact that once the backoff stage reaches the value m_2, it is not increased in subsequent packet transmissions. Equation (4.5) accounts for the fact that a new packet following a successful packet transmission starts with backoff stage 0.

From the above descriptions on the state transitions for traffic flows, we can see that with the introduction of the packet collision probabilities p_1 and p_2, hold states for type-2 traffic flows, and probabilities P_s and P_{s1}, it is possible to solve the Markov chain of a traffic flow independently.

B.   *Throughput Analysis*

First, we solve the Markov chain for type-1 traffic. Let $q_1(i, j)$, $i \in [0, m_1]$ and $j \in [0, 2^i W_1 - 1]$, be the stationary distribution of the chain. It is easy to find that

$$
\begin{cases}
q_1(i,0) = p_1^{\,i} \cdot q_1(0,0) & (0 < i < m_1) \\[2mm]
q_1(m_1,0) = \dfrac{p_1^{\,m_1}}{1 - p_1} \cdot q_1(0,0)
\end{cases}
\tag{5}
$$

and

$$
q_1(i,j) = \frac{2^i W_1 - j}{2^i W_1} \cdot q_1(i,0)
\tag{6}
$$

where $i \in [0, m_1]$ and $j \in [0, 2^i W_1 - 1]$. Because $\sum_{i=0}^{m_1} \sum_{j=0}^{2^i W_1 - 1} q_1(i,j) = 1$, we have

$$q_1(0,0) = \frac{2(1-2p_1)(1-p_1)}{(1-2p_1)(W_1+1) + p_1 W_1 [1-(2p_1)^{m_1}]} \tag{7}$$

τ_1 is defined as the probability that a station carrying type-1 traffic transmits in a randomly chosen slot time. We have

$$\tau_1 = \sum_{i=1}^{m_1} q_1(i,0) = \frac{2(1-2p_1)}{(1-2p_1)(W_1+1) + p_1 W_1 [1-(2p_1)^{m_1}]} \tag{8}$$

Please refer to [10] for detailed derivations on equations (5) to (8).

For type-2 traffic, let $q_2(i,j,k)$, $i \in [0, m_2]$, $j \in [0, 2^i W_2 - 1]$ and $k \in [1, D]$, be the stationary distribution of the chain. We derive

$$\begin{cases} q_2(i,0,0) = p_2^{\,i} \cdot q_2(0,0,0) & (0 < i < m_2) \\ q_2(m_2,0,0) = \dfrac{p_2^{\,m_2}}{1-p_2} \cdot q_2(0,0,0) \end{cases} \tag{9}$$

and

$$q_2(i,j,0) = \frac{2^i W_2 - j}{2^i W_2} \cdot q_2(i,0,0) \tag{10}$$

where $i \in m_2$ and $j \in [0, 2^i W_2 - 1]$. Furthermore, it can be easily found that

$$q_2(i,j,k) = \begin{cases} q_2(i,j,0) \cdot \dfrac{(1-P_s)}{P_{s1}^{\,D+1-k}} & (j>0, 1 \le k \le D) \\ q_2(i,j,0) \cdot \dfrac{1}{P_{s1}^{\,D+1-k}} & (j=0, 1 \le k \le D) \end{cases} \tag{11}$$

where $i \in [0, m_2]$. Please refer to the Appendix for the detailed derivations of equations (9) to (11). From above three equations, we can see that all the state probabilities can be expressed by $q_2(0,0,0)$. Because $\sum_{i=0}^{m_2} \sum_{j=0}^{2^i W_2 - 1} \sum_{k=0}^{D} q_2(i,j,k) = 1$, we can have:

$$q_2(0,0,0) = \frac{1}{\dfrac{1+\left(\sum\limits_{i=1}^{D} \dfrac{1}{P_{s1}^{\,i}}\right)}{1-p_2} + \dfrac{(1-2p_2)(W_2-1) + p_2 W_2 [1-(2p_2)^{m_2}]}{2(1-2p_2)(1-p_2)} \cdot \left[1 + (1-P_s) \cdot \left(\sum\limits_{i=1}^{D} \dfrac{1}{P_{s1}^{\,i}}\right)\right]} \tag{12}$$

Let us define P_{hold} as the probability that all stations carrying type-2 traffic are in hold states. We emphasize that, if one station carrying type-2 traffic is in *hold states*, all the other stations carrying type-2 traffic must be in hold states too. P_{hold} is expressed as

9

$$P_{hold} = \sum_{i=0}^{m_2} \sum_{j=0}^{2^i W_2 - 1} \sum_{k=1}^{D} q_2(i,j,k) \tag{13}$$

Moreover, we define τ_2 as the probability that a station carrying type-2 traffic transmits in a randomly chosen slot time under the condition that it is not in hold states. Therefore, we have

$$\tau_2 = \frac{\sum_{i=0}^{m_2} q_2(i,0,0)}{\sum_{i=0}^{m_2} \sum_{j=0}^{2^i W_2 - 1} q_2(i,j,0)} = \frac{2(1-2p_2)}{(1-2p_2)(W_2+1) + p_2 W_2 [1-(2p_2)^{m_2}]} \tag{14}$$

Now we can express P_s, the probability that all the other traffic flows, except for the considered type-2 traffic flow, are not in the state of transmission, as follows

$$P_s = (1-\tau_1)^{n_1} (1-\tau_2)^{n_2 - 1} \tag{15}$$

P_{s1}, the probability that all the traffic flows with type-1 are not transmitting, can be given as

$$P_{s1} = (1-\tau_1)^{n_1} \tag{16}$$

With the above probabilities defined, we can express the packet collision probabilities p_1 and p_2 as

$$p_1 = 1 - (1-\tau_1)^{n_1 - 1} [P_{temp} + (1-P_{temp})(1-\tau_2)^{n_2}] \tag{17}$$

$$p_2 = 1 - P_s = 1 - (1-\tau_1)^{n_1} (1-\tau_2)^{n_2 - 1} \tag{18}$$

Combining equations (8), (11) to (18) and by using SOR numerical method [20], we can get all the values for p_1, p_2, τ_1, τ_2, P_s, P_{s1} and P_{hold}.

In order to derive the system throughput, we define $Q(i,j)$ as the probability that there are a number of i type-1 stations and a number of j type-2 stations transmitting within a randomly selected slot. For $Q(0,0)$, with no transmitting station, we have

$$Q(0,0) = (1-\tau_1)^{n_1} [P_{hold} + (1-P_{hold})(1-\tau_2)^{n_2}] \tag{19}$$

For $Q(1,0)$, with only one type-1 station transmitting, we have

$$Q(1,0) = n_1 \tau_1 (1-\tau_1)^{n_1 - 1} [P_{hold} + (1-P_{hold})(1-\tau_2)^{n_2}] \tag{20}$$

For $Q(0,1)$, with only one type-2 station transmitting, we have

$$Q(0,1) = (1-\tau_1)^{n_1} (1-P_{hold}) n_2 \tau_2 (1-\tau_2)^{n_2 - 1} \tag{21}$$

For $Q(c_1, c_2)$ $(c_1 \geq 2, c_2 = 0)$, which means that there are some collisions occurring between type-1 traffic flows, we have

10

$$Q(c_1,c_2) = \binom{n_1}{c_1} \tau_1^{c_1} (1-\tau_1)^{n_1-c_1} \cdot [P_{hold} + (1-P_{hold})(1-\tau_2)^{n_2}] \tag{22}$$

For $Q(c_1,c_2)$ $(c_1 + c_2 \geq 2, c_2 \geq 1)$, we have

$$Q(c_1,c_2) = \binom{n_1}{c_1} \tau_1^{c_1} (1-\tau_1)^{n_1-c_1} \cdot (1-P_{hold}) \cdot \binom{n_2}{c_2} \tau_2^{c_2} (1-\tau_2)^{n_2-c_2} \tag{23}$$

The normalized system throughputs S can be defined and expressed as:

$$S \equiv \frac{\text{Average payload transmitted in a slot time}}{\text{Average length of a slot time}}$$

$$= S_1 + S_2 = \frac{Q(1,0) \cdot E[P_{Len,1}] + Q(0,1) \cdot E[P_{Len,2}]}{Q(0,0) \cdot \sigma + Q(1,0) \cdot T_{s,1} + Q(0,1) \cdot T_{s,2} + \sum_{\substack{0 \leq c_1 \leq n_1 \\ 0 \leq c_2 \leq n_2 \\ c_1+c_2 \geq 2}} Q(c_1,c_2) \cdot T_c(c_1,c_2)} \tag{24}$$

where S_1 and S_2 denote the throughputs contributed by type-1 and type-2 traffic flows, respectively. $E[P_{Len,i}]$ is the average duration to transmit the payload for type-i traffic (the payload size is measured with the time required to transmit it). For simplicity, with the assumption that all packets of type-i traffic have the same fixed size, we have $E[P_{Len,i}] = P_{Len,i}$. σ is the duration of an empty time slot. $T_{s,i}$ (see Fig. 3) is the average time of a slot because of a successful transmission of a packet of a type-i traffic flow. $T_{s,i}$ can be expressed as

$$T_{s,i} = PHY_{header} + MAC_{header} + E[P_{Len,i}] + SIF + \delta + ACK + DIF_1 + \delta \tag{25}$$

where δ is the propagation delay. $T_c(c_1,c_2)$, refer to Fig. 3, is the average time the channel is sensed busy by each station during a collision caused by simultaneous transmissions of c_1 type-1 stations and c_2 type-2 stations. It can be expressed as

$$T_c(c_1,c_2) = PHY_{header} + MAC_{header} + \max[\theta(c_1)P_{Len,1}, \theta(c_2)P_{Len,2}] + DIFS + \delta \tag{26}$$

and

$$\theta(x) \equiv \begin{cases} 1 & x > 0 \\ 0 & x = 0 \end{cases}$$

C. Packet Delay Analysis

For the quality of service of real-time multimedia it is important to know the time that a packet must wait for transmission over the IEEE 802.11 MAC. By analyzing the packet delay, we can derive an upper bound for the average delay.

For a type-i traffic flow, we define that the whole period of time within $T_{s,i}$ (see equation 25) is spent to send a packet. Therefore, for a type-i traffic flow, we define packet delay $T_{D,i}$ as the average time period between the instant of its finishing sending the former packets to the instant of beginning to send the next packet. Therefore, $T_{D,i}$ does not include the transmission time for a packet (see Fig. 4). In Fig. 4, detailed timing relationships between packet sending times and packet delays are illustrated.

Based on Fig. 4, it can be easily found that there is a simple relationship between $T_{D,i}$ and the throughput S_i:

$$s_i \equiv \frac{S_i}{n_i} = \frac{E[P_{Len,i}]}{T_{D,i} + T_{s,i}} \tag{27}$$

Therefore, $T_{D,i}$ for type-i traffic flows can be given as

$$T_{D,i} = \frac{E[P_{Len,i}]}{s_i} - T_{s,i} \tag{28}$$

IV. Approximate Analysis

In the above two sections, we have analyzed the system throughput and packet delay in saturation state. In order to gain a deeper insight into the whole system, we make some approximations to get simpler but more meaningful relationships among different parameters.

In the following discussion, it is assumed that $D = 0$, that is, we only consider the case that $DIF_1 = DIF_2$. We start from equations (17) and (18). Because $D = 0$, P_{hold} must also be 0. Therefore, we can easily derive

$$(1 - p_1)(1 - \tau_1) = (1 - p_2)(1 - \tau_2) = \prod_{j=1}^{2} (1 - \tau_j)^{n_j} \tag{29}$$

It can be seen that if $\tau_1 \neq \tau_2$, then one must have $p_1 \neq p_2$, which justifies our former assumption that conditional packet collision probabilities of different types of traffic flows may be different from each other. When the minimum contention window size $W_1 \gg 1$ and $W_2 \gg 1$, the transmission probabilities τ_1 and τ_2 are small, that is, $\tau_1 \ll 1$ and $\tau_2 \ll 1$. Therefore, from equation (29), we have the following approximation

$$p_1 \approx p_2 \tag{30}$$

under the condition of $W_1 \gg 1$ and $W_2 \gg 1$. Furthermore, when $W_1 \gg 1$, $W_2 \gg 1$ and

$m_1 \approx m_2$, we have the following approximation based on equations (8) and (14)

$$\frac{\tau_1}{\tau_2} \approx \frac{W_2}{W_1} \tag{31}$$

From equations (19), (20) and (24), we have

$$\frac{S_1}{S_2} = \frac{Q(1,0) \cdot E[P_{Len,1}]}{Q(0,1) \cdot E[P_{Len,2}]} = \frac{n_1 \tau_1 (1-\tau_2) E[P_{Len,1}]}{n_2 \tau_2 (1-\tau_1) E[P_{Len,2}]} \approx \frac{n_1 \tau_1 E[P_{Len,1}]}{n_2 \tau_2 E[P_{Len,2}]} \approx \frac{n_1 W_2 E[P_{Len,1}]}{n_2 W_1 E[P_{Len,2}]} \tag{32}$$

Then, from the above equation, we have

$$\frac{s_1}{s_2} \approx \frac{\dfrac{E[P_{Len,1}]}{W_1}}{\dfrac{E[P_{Len,2}]}{W_2}} \tag{33}$$

The above equation holds under the conditions of $D = 0$, $W_1 \gg 1$ and $W_2 \gg 1$ and $m_1 \approx m_2$. We can see that the throughput differentiation is mainly determined by the scaling of minimum contention window sizes and the length of packet payloads.

Moreover, from equation (28), we can see that under the condition $T_{s,i} \ll \dfrac{E[P_{Len,i}]}{s_i}$

$(i = 1,2)$, which always holds when the number of traffic flows $n_i \gg 1$ $(i = 1,2)$, we have

$$\frac{T_{D,1}}{T_{D,2}} \approx \frac{\dfrac{E[P_{Len,1}]}{s_1}}{\dfrac{E[P_{Len,2}]}{s_2}} \approx \frac{W_1}{W_2} \tag{34}$$

Equation (34) is another important approximation relationship obtained. We can see that packet delay differentiation among different types of traffic flows is mainly determined by the ratio of the corresponding minimum contention window sizes.

Assuming that all the W_i and $E[P_{Len,i}]$ are fixed, then from equations (33), (34), and (28) we can easily arrive at the following conclusions:

1. s_2 can be linearly expressed by s_1, that is, $s_2 \approx \dfrac{W_1 E[P_{Len,2}]}{W_2 E[P_{Len,1}]} \cdot s_1$. Therefore, s_2 reaches its

   maximum value when s_1 reaches its maximum value. That is to say, there is a case where all throughputs $s_i (i = 1,2)$ and the total throughput S of the system reach their maximum values at *almost* the same configuration.

2. $T_{D,2}$ can be linearly expressed by $T_{D,1}$, that is, $T_{D,2} \approx \dfrac{W_2}{W_1} \cdot T_{D,1}$. Therefore, $T_{D,2}$ reaches its

minimum value when $T_{D,1}$ reaches its minimum value. That is to say, there is a case where all the packet delays $T_{D,i}(i=1,2)$ reach their minimum values at *almost* the same configuration.

3. When all throughputs $s_i(i=1,2)$ reach their maximum value, all the $T_{D,i}$ will reach their minimum values at *almost* the same configuration.

The above properties show that by scaling W_i and $E[P_{Len,i}]$ one can obtain service differentiation and make the whole system easily controllable.

V. Results And Discussions

In this section, by using our analysis model we investigate the effect of the different parameters on the traffic differentiation and we validate the analysis assumptions by comparisons with a discrete-event simulation of the system. In addition, we assess the range of validity of the approximated formulas obtained in the previous analysis. We proceed by first considering the effect of varying a single crucial parameter on the performance differentiation and then by considering the validity of the approximated formulas for the case of equal inter-frame separations. In our examples, we assume that two types of traffic coexist in the system. The parameters for the system are summarized in Table 1 based on IEEE 802.11b.

Table 1.

System Parameters

MAC Header	272 bits
PHY Header	192 μs
ACK	112 bits +PHY header
Channel Bit Rate	11Mbps
Propagation Delay	1 μs
Slot Time	20 μs
SIFS	10 μs
DIFS	30 μs

The purpose of the first series of experiments is to verify the accuracy of our analysis model and assess the effect of varying the inter-frame separations on the performance. To this end, we keep all other parameters equal and only change the value of D. In Fig. 5 and Fig. 6, we validate our proposed analysis model by comparing simulation results and numerical results. For our simulator, which is implemented by using C++, we consider that there are 20

stations each carrying only one traffic flow. 5 of them carry type-1 traffic and the others carry type-2 traffic. In the simulation, ideal channel conditions (i.e., no hidden terminals and capture) are assumed. The other parameters are set as follows: $W_1 = W_2 = 64$, $m_1 = m_2 = 8$ and $E[P_{Len,1}] = E[P_{Len,2}] = 2000bytes$. Different simulation values are obtained by varying the IFS difference D. Each simulation values are obtained by running our discrete-event simulator to simulate the actual behavior of the system within the period of 5 hours. For each point, we run the simulator for 10 times with 10 different random seeds. Simulation results show that the standard deviation of the results caused by using different random seeds is less then one percent for each point, and therefore not visible in the figures.

In Fig. 5, throughput S, S_1 and S_2 versus IFS difference D are shown. In Fig. 6, average packet delays $T_{D,1}$, $T_{D,2}$ and the overall average packet delay per traffic flow $T_D \equiv \dfrac{n_1 T_{D,1} + n_2 T_{D,2}}{n_1 + n_2}$ are shown versus the IFS difference D. From these two figures, it can be seen that the simulation results agree very well with the theoretical ones. We can also see that the smaller D, the better the simulation results agree with the numerical ones, which suggests that the accuracy of our assumption of the probability P_{s1} being a constant is accurate and reasonable for this case. On the other hand, when D increases, the difference between simulation results and numerical ones becomes more evident. In order to understand the reason, let us consider an extreme case, where the maximum contention window size CW_{max} for type-1 traffic is smaller than D. It is evident that, in this case, after the period of transmission of a packet or collision between some packets, it is entirely impossible for type-2 traffic flows to get access to the channel. In this case, the backoff process for type-2 traffic flows will not proceed from one *normal state* to another. However, by referring to Fig. 1b and equations (15) and (16), it can be found that in most cases, one can always get positive and nonzero P_{s1} and P_s, which means that the backoff process of type-2 traffic flows will always go on, and type-2 traffic flows will eventually access the channel even in the above extreme case. Therefore, our model is only an approximation, needed to make our model tractable. Moreover, extensive simulations show that our model is very accurate as long as type-2 traffic flows are not heavily starved by type-1 traffic. From these two figures, we can quantify the differentiating effects caused by D. With the increase of D, more channel resources are allocated to type-1 traffic flows, and also the average packet transmission delay for type-1 traffic decreases. When D becomes larger, the rate for the performance improvement of type-

1 traffic decreases, and at the same time, performance for type-2 traffic becomes worse, which indicates that a very large D is not much helpful to improve the system performance.

The purpose of the second series of experiments for the case $D = 0$ is to verify the range of validity of the approximated formulas derived in the paper. In the experiments, two cases have been shown. In case 1, we set $E[P_{Len,1}] = E[P_{Len,2}] = 2000 bytes$. In case 2, different packet payloads for two types of traffic have been set, i.e., $E[P_{Len,1}] = 2000 bytes$ and $E[P_{Len,2}] = 200 bytes$. Other parameters are set as follows: $W_2 = 256$, $m_1 = m_2 = 2$, $n_1 = n_2 = 25$. Fig. 7 and Fig. 8 show a comparison between throughput and packet delays obtained by using accurate theoretical formulas and those obtained by using approximate formulas. In Fig. 7, the values of $\frac{s_1}{s_2}$ are compared with those of $\frac{W_2 \cdot E[P_{Len,1}]}{W_1 \cdot E[P_{Len,2}]}$. We can see that when $\frac{W_2}{W_1}$ is not large, that is, W_1 is not very small, the ratio $\frac{s_1}{s_2}$ can be approximated by $\frac{W_2 \cdot E[P_{Len,1}]}{W_1 \cdot E[P_{Len,2}]}$ fairly well, which verifies our approximation made in equation (33). Furthermore, we can verify that the length of the packet payload has an immediate effect on the throughput differentiation between different traffic classes. In Fig. 8, the values of $\frac{T_{D,2}}{T_{D,1}}$ are compared with those of $\frac{W_2}{W_1}$. Again, we can see that when $\frac{W_2}{W_1}$ is not large, the ratio $\frac{T_{D,2}}{T_{D,1}}$ can be approximated by $\frac{W_2}{W_1}$ fairly well, which verifies our approximation made in equation (34). Moreover, we notice that the corresponding values of $\frac{T_{D,2}}{T_{D,1}}$ for two different cases are overlapped (the maximum differences between these two cases are less then one percent), which indicates that packet payload almost has no influence on the values of $\frac{T_{D,2}}{T_{D,1}}$. That is to say, scaling the length of packet payload has little effect on the differentiation of packet delays between the different traffic classes.

Let us now consider the optimization of the overall performance of the system. In the former section, in the case of $D = 0$, and by using the approximated formulas, we point out that there is a case where all throughputs s_i and the total throughput S of the system reach

their maximum values at *almost* the same configuration. And we also point out that there is a case where all packet delays $T_{D,i}$ reach their minimum values at *almost* the same time. In Fig. 9 and Fig. 10, these conclusions are verified by using numerical results. In these figures, we keep the ratio $\dfrac{W_2}{W_1}$ constant and change the value W_1, in order to see the variations of throughput and packet delays of different types of traffic. In these two figures, two cases have been shown. Moreover, maximum throughput points and minimum packet delay points are shown. From Fig.9, we can see that for a particular case S, $s_i (i = 1,2)$ reaches their maximum values *almost* at the same W_1, which verifies our first conclusion made in the former section. And from Fig. 10, we can see that for a particular case $T_{D,i} (i = 1,2)$ reaches their maximum values *almost* at the same W_1, which verifies our second conclusion made in the former section. Moreover, by checking these two figures together, we can find that $s_i (i = 1,2)$ and corresponding $T_{D,i} (i = 1,2)$ reach their optimal points at exactly the same W_1, which in turn justifies our third conclusion made in the former section. From these two figures, by comparing case 1 and case 2, we find that, as expected, the system becomes more stable with the increase of maximum backoff stages m_1 and m_2. This is because that with larger m_1 and m_2 packet collision rates drop, which increases the utilization of the whole system.

Let us now consider the effect of the "traffic mixture" on the performance. To this end we fix the total number of flows and vary the relative portion of high- versus low-priority flows. In Fig. 11 and Fig. 12, we keep the total number of traffic flows constant, and change the number n_1 of type-1 traffic flows in the case of $D = 0$. In Fig. 11, throughputs s_1 and s_2 versus the number of type-1 traffic flows n_1 are shown. Different curves are for different values of the minimum contention window size W_1. In Fig. 12, the packet delays $T_{D,1}$ and $T_{D,2}$ for two types of traffic are shown in the same configuration. From Fig. 11 and Fig. 12, we can see that with the decrease of W_1, type-1 traffic flows gain priority over type-2 traffic flows: throughput s_1 becomes larger than s_2, and packet delay $T_{D,1}$ becomes smaller than $T_{D,2}$. Therefore, more bandwidth resources are allocated to type-1 traffic. However, we can see that in the case of large n_1 (such as $n_1 > 40$), both the performance on throughput and packet delays are worse than in the case of $W_1 = W_2 = 256$, which indicates that providing service differentiation with very large number of traffic flows belonging to higher priority

17

group makes the system performance worse than in the case of no service differentiation support. In this case, performance of all traffic flows, no matter if they belong to higher or lower priority classes, becomes worse.

The reason can be easily explained by referring to Fig. 13, that shows the packet collision rates p_1 and p_2 for two types of traffic. We can see that with the increase of n_1, collision rates p_1 and p_2 increase drastically, therefore reducing the bandwidth utilization. On the other hand, from Fig. 13, it can be seen that with the increase of W_1, the difference between p_1 and p_2 decreases, in agreement with the approximation obtained in equation (30). Moreover, referring to Fig. 11 and Fig. 12, if the number of traffic flows with higher priority is small, both throughput and packet delays for higher priority traffic flows are improved significantly with only small influence on the other traffic flows with lower priority. The above results indicate that the number of traffic flows with higher priority must be controlled to a suitable fraction of the total traffic by suitable access control schemes.

Finally, when one considers implementing some real system to support service differentiation, simplicity should also be considered. It is always the case that the simpler the scheme is, the lower the system costs. In our model, three service differentiation supporting mechanisms are analyzed. By adopting the scheme of differentiating minimum contention window sizes and packet payloads of different traffic types, simple relationships exist among throughput and packet delays, which is helpful to simplify the design of the whole system. On the other hand, in order to make the system as simple as possible, one should limit the number of parameters that can be adjusted. On the whole, by using the analysis model proposed in this paper, we can obtain deeper insight, which is important and helpful to the design of real systems. However, the implementation issues of real systems are beyond the scope of this paper, although they are very important research topics for our future work.

VI. Conclusions

We proposed an analysis model to compute the throughput and packet transmission delays in a WLAN with enhanced IEEE 802.11 Distributed Coordination Function, which supports service differentiation. In our analytical model, service differentiation is supported by scaling the contention window, the Inter-Frame Spacing (IFS) and the packet length according to the priority of each traffic flow. Comparisons with simulation results show that good accuracy of performance evaluations can be achieved by using the proposed model.

In particular, we evaluated the throughput and packet delay. It is shown that:

1. The settings of CW_{min} for different types of traffic flows have significant influence on the throughput and packet delay. One type of traffic gains priority over other types of traffic through a smaller minimum contention window size. More channel resources are occupied, with a smaller packet delay.

2. Traffic flows with shorter IFS obtain higher priority to access to the channel resources. However, excessive IFS values cause very long packet delays for traffic flows with lower priority, bordering on starvation.

3. The length of packet payload of different types of traffic directly influences the bandwidth allocation among different types of traffic flows. However, the differentiation of packet payload size has little influence on the differentiation of packet delays. Let us note that in noisy channel conditions, the typical situation in wireless LANs, longer payloads suffer a higher error probability and this fact may discourage applying payload length variability as a differentiation mechanism.

4. The number of traffic flows with higher priority must be limited to maintain the system working at a high performance regime by suitable access control or pricing schemes. See, for example, [21] for a possible adaptive scheme to regulate the appropriate number of transmitting stations.

5. By adopting the scheme of scaling minimum contention window sizes and packet payloads, approximate and simple relationships exist among throughput, packet delays and lengths of packet payload of different traffic types, which can be used for the optimal design of the whole system.

By using the proposed model, three different service differentiation schemes have been analyzed. The schemes are not mutually exclusive. The appropriate choice and setting of parameters for the control of a real-world system, including access control, is an interesting area of future research that can be benefit from the analysis presented in this paper.

References

[1]   Y. Cheng and W. H. Zhuang, "DiffServ resource allocation for fast handoff in wireless mobile Internet," IEEE Communications Magazine, vol. 40, no. 5, 2002, pp. 130-136.

[2]   R. Braden, D. Clark and S. Shenker, "Integrated services in the Internet architecture: an overview," IETF RFC 1633, Jun. 1994.

[3]   S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang and W. Weiss, "An architecture for differential

services," IETF RFC 2475, Dec. 1998.

[4]   Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications, IEEE Standard 802.11, Aug. 1999.

[5]   J. Weinmiller, M. Schlager, A. Festag, and A. Wolisz, "Performance study of access control in wireless LANs IEEE 802.11 DFWMAC and ETSI RES 10 HIPERLAN," Mobile Networks and Applications, vol. 2, pp. 55-67, 1997.

[6]   H. S. Chhaya and S. Gupta, "Performance modeling of asynchronous data transfer methods of IEEE 802.11 MAC protocol," Wireless Networks, vol. 3, pp. 217-234, 1997.

[7]   T. S. Ho and K. C. Chen, "Performance evaluation and enhancement of the CSMA/CA MAC protocol for 802.11 wireless LAN's," Proceedings of IEEE PIMRC, Taipei, Taiwan, Oct. 1996, pp.392-396.

[8]   F. Cali, M. Conti, and E. Gregori, "IEEE 802.11 wireless LAN: Capacity analysis and protocol enhancement," Proceedings of INFOCOM'98, San Francisco, CA, March 1998, vol. 1, pp. 142 -149.

[9]   G. Bianchi, L. Fratta, and M. Oliveri, "Performance analysis of IEEE 802.11 CSMA/CA medium access control protocol," Proceedings of IEEE PIMRC, Taipei, Taiwan, Oct. 1996, pp. 407-411.

[10]   G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," IEEE Journal on Selected Areas In Communications, vol. 18, no. 3, March 2000.

[11]   Y. C. Tay and K. C. Chua, "A Capacity Analysis for the IEEE 802.11 MAC Protocol," Wireless Networks, 7, 2001, pp. 159-171.

[12]   J. L. Sobrinho and A. S. Krishnakumar, "Distributed multiple access procedures to provide voice communications over IEEE 802.11 wireless networks," Proceedings GLOBECOM 1996, pp. 1689-1694.

[13]   J. Deng and R.S. Chang, "A priority scheme for IEEE 802.11 DCF access method," IEICE Transactions in Communications, vol. 82-B, no. 1, Jan 1999, pp. 96-102.

[14]   A. Veres, A. T. Campbell, M. Barry and L. H. Sun, "Supporting Service Differentiation in Wirelss Packet Networks Using Distributed Control," IEEE Journal on Selected Areas In Communications, vol. 19, no. 10, Oct 2001, pp. 2081-2093.

[15]   I. Aad and C. Castelluccia, "Differentiation Mechanisms for IEEE 802.11," Proceedings of IEEE Inforcom 2001, pp. 209-218.

[16]   W. Haitao, C. Shiduan, P. Yong, L. Keping, and M. Jian, "IEEE 802.11 distributed coordination function (DCF): analysis and enhancement," Proceedings of ICC 2002, pp. 605-609.

[17]   S. Mangold, S. Choi, P. May, O. Klein, G. Hietz and L. Stibor, "IEEE 802.11e wireless lan for quality of service," Proceedings of the European Wireless, Feb 2002.

[18]   M. Benveniste, G. Chesson, M. Hoehen, A. Singla, H. Teunissen, and M.Wentink, "EDCF proposed draft text," IEEE working document 802.11-01/131r1, March 2001.

[19]   A. Lindgren, A. Almquist, and O. Schelen, "Evaluation of quality of service schemes for IEEE 802.11 wireless LANs," Proceedings of IEEE Conference on Local Computer Networks (LCN 2001), November 15-16, 2001, pp. 348 –351.

[20]   G. Bolch, S. Greiner, H. de Meer, and K.S. Trivedi, Queueing Networks and Markov Chains: Modeling and Performance Evaluation with Computer Science Applications, Wiley-Interscience, 1998, pp. 140-144.

[21]   Roberto Battiti, Marco Conti, Enrico Gregori, Mikalai Sabel, "Price-based Congestion-Control in Wi-Fi Hot Spots," Proceedings of WiOpt'03 March 3-5, 2003, INRIA Sophia-Antipolis, France, pp. 91-100.

Appendix: Derivations on Equations (9), (10) and (11)

From Fig. 1b, for $i \in [0, m_2], 0 < j \le 2^i W_2 - 1$, we have

$$
\begin{cases}
q_2(i,j,k) = P_{s1} \cdot q_2(i,j,k-1) & (1 < k \le D) \\
q_2(i,j,1) = (1-P_s) \cdot q_2(i,j,0) + (1-P_{s1}) \cdot \displaystyle\sum_{k=1}^{D} q_2(i,j,k) \\
\qquad = (1-P_s) \cdot q_2(i,j,0) + (1-P_{s1}) \cdot q_2(i,j,1) \displaystyle\sum_{k=0}^{D-1} P_{s1}^{\,k} \\
\qquad = (1-P_s) \cdot q_2(i,j,0) + q_2(i,j,1) \cdot (1-P_{s1}^{\,D})
\end{cases}
\tag{A.1}
$$

From above equation, we have

$$
q_2(i,j,k) = \frac{1-P_s}{P_{s1}^{\,D+1-k}} \cdot q_2(i,j,0) \qquad (i \in [0,m_2], 0 < j \le 2^i W_2 - 1, 1 \le k \le D)
\tag{A.2}
$$

Moreover, for the case of $i \in [0, m_2], j = 0$, we have

$$
\begin{cases}
q_2(i,0,k) = P_{s1} \cdot q_2(i,0,k-1) & (1 < k \le D) \\
q_2(i,0,1) = q_2(i,0,0) + (1-P_{s1}) \cdot \displaystyle\sum_{k=1}^{D} q_2(i,0,k) \\
\qquad = q_2(i,0,0) + (1-P_{s1}) \cdot q_2(i,0,1) \displaystyle\sum_{k=0}^{D-1} P_{s1}^{\,k} \\
\qquad = q_2(i,0,0) + q_2(i,0,1) \cdot (1-P_{s1}^{\,D})
\end{cases}
\tag{A.3}
$$

then,

$$
q_2(i,0,k) = \frac{1}{P_{s1}^{\,D+1-k}} \cdot q_2(i,0,0) \qquad (i \in [0,m_2], 1 \le k \le D)
\tag{A.4}
$$

By combining equations (A.2) and (A.4), equation (11) can be obtained.

Next, we subdivided our discussions into three different cases:

Case 1: when $0 < i < m_2$

From the description on the Markov chain for type 2 traffic flows, we have

$$
\begin{cases}
q_2(i, 2^i W_2 - 1, 0) = q_2(i-1,0,D) \cdot \dfrac{P_{s1} \cdot p_2}{2^i W_2} \\
\qquad = q_2(i-1,0,0) \cdot \dfrac{p_2}{2^i W_2} \\
q_2(i,j,0) = P_s \cdot q_2(i,j+1,0) + P_{s1} \cdot q_2(i,j+1,D) \\
\qquad + q_2(i-1,0,D) \cdot \dfrac{P_{s1} \cdot p_2}{2^i W_2} & (0 \le j \le 2^i W_2 - 2) \\
\qquad = q_2(i,j+1,0) + q_2(i-1,0,0) \cdot \dfrac{p_2}{2^i W_2}
\end{cases}
\tag{A.5}
$$

From above equation, we have

$$q_2(i,j,0) = \frac{2^i W_2 - j}{2^i W_2} \cdot p_2 \cdot q_2(i-1,0,0) \quad (0 \le j \le 2^i W_2 - 1) \tag{A.6}$$

In equation (A.6), let $j = 0$, one derives

$$q_2(i,0,0) = p_2 \cdot q_2(i-1,0,0) \tag{A.7}$$

and then

$$q_2(i,j,0) = \frac{2^i W_2 - j}{2^i W_2} \cdot q_2(i,0,0) \quad (0 \le j \le 2^i W_2 - 1) \tag{A.8}$$

Case 2: when $i = m_2$

We have

$$\begin{cases} q_2(m_2, 2^{m_2} W_2 - 1, 0) = q_2(m_2-1,0,D) \cdot \dfrac{P_{s1} \cdot p_2}{2^{m_2} W_2} + q_2(m_2,0,D) \cdot \dfrac{P_{s1} \cdot p_2}{2^{m_2} W_2} \\[2mm] \qquad = [q_2(m_2-1,0,0) + q_2(m_2,0,0)] \cdot \dfrac{p_2}{2^{m_2} W_2} \\[2mm] q_2(m_2,j,0) = P_s \cdot q_2(m_2,j+1,0) + P_{s1} \cdot q_2(m_2,j+1,D) \\[2mm] \qquad + q_2(m_2-1,0,D) \cdot \dfrac{P_{s1} \cdot p_2}{2^{m_2} W_2} + q_2(m_2,0,D) \cdot \dfrac{P_{s1} \cdot p_2}{2^{m_2} W_2} \qquad (0 \le j \le 2^{m_2} W_2 - 2) \\[2mm] \qquad = q_2(m_2,j+1,0) + [q_2(m_2-1,0,0) + q_2(m_2,0,0)] \cdot \dfrac{p_2}{2^{m_2} W_2} \end{cases} \tag{A.9}$$

From equation (A.9), we have

$$q_2(m_2,j,0) = \frac{2^{m_2} W_2 - j}{2^{m_2} W_2} \cdot p_2 \cdot [q_2(m_2-1,0,0) + q_2(m_2,0,0)] \quad (0 \le j \le 2^{m_2} W_2 - 1) \tag{A.10}$$

By setting $j = 0$, from above equation and equation (A.7), we have

$$q_2(m_2,0,0) = \frac{p_2^{m_2}}{1 - p_2} \cdot q_2(0,0,0) \tag{A.11}$$

Combining equations (A.7) and (A.11), we can obtain equation (9). Moreover, we have

$$q_2(m_2,j,0) = \frac{2^{m_2} W_2 - j}{2^{m_2} W_2} \cdot q_2(m_2,0,0) \quad (0 \le j \le 2^{m_2} W_2 - 1) \tag{A.12}$$

Case 3: $i = 0$

We have the following equation

$$
\begin{cases}
q_2(0, W_2 - 1, 0) = \dfrac{P_{s1} \cdot (1 - p_2)}{W_2} \cdot \displaystyle\sum_{i=0}^{m_2} q_2(i, 0, D) \\[4mm]
\qquad\qquad\quad = \dfrac{(1 - p_2)}{W_2} \cdot \displaystyle\sum_{i=0}^{m_2} q_2(i, 0, 0) \\[4mm]
q_2(0, j, 0) = P_s \cdot q_2(0, j+1, 0) + P_{s1} \cdot q_2(0, j+1, D) \\[4mm]
\qquad\quad + \dfrac{P_{s1} \cdot (1 - p_2)}{W_2} \cdot \displaystyle\sum_{i=0}^{m_2} q_2(i, 0, D) \qquad (0 \le j \le W_2 - 2) \\[4mm]
\qquad = q_2(0, j+1, 0) + \dfrac{(1 - p_2)}{W_2} \cdot \displaystyle\sum_{i=0}^{m_2} q_2(i, 0, 0)
\end{cases}
\tag{A.13}
$$

then,

$$
q_2(0, j, 0) = \frac{W_2 - j}{W_2} \cdot (1 - p_2) \cdot \sum_{i=0}^{m_2} q_2(i, 0, 0) \quad (0 \le j \le W_2 - 1) \tag{A.14}
$$

From equation (9), the above equation can be reduced into the following equation

$$
q_2(0, j, 0) = \frac{W_2 - j}{W_2} \cdot q_2(0, 0, 0) \qquad (0 \le j \le W_2 - 1) \tag{A.15}
$$

From equations (A.8), (A.12) and (A.15), we can obtain equation (10).

Fig. 1a Markov model of backoff process for type 1 traffic

Fig. 1b Transitions out of state {j,k+1,0} for type 2 traffic

Fig. 1c Transitions out of state {j,0,0} for type 2 traffic

	t1		t2	t3	t4		t5	t6	t7
Flow 1	{1,0}		{0,2}	{0,1}	{0,0}		{1,5}	{1,4}	
Flow 2	{0,2,0}		{0,2,1}	{0,1,0}	{0,0,0}		{0,0,1}	{1,3,0}	

Note:
Flow 1: Carrying type 1 traffic with $W_1=8$, $m_1=2$
Flow 2: Carrying type 2 traffic with $W_2=8$, $m_2=2$
$D=1$

Fig. 2 An example for state transitions of two different type of traffic flows
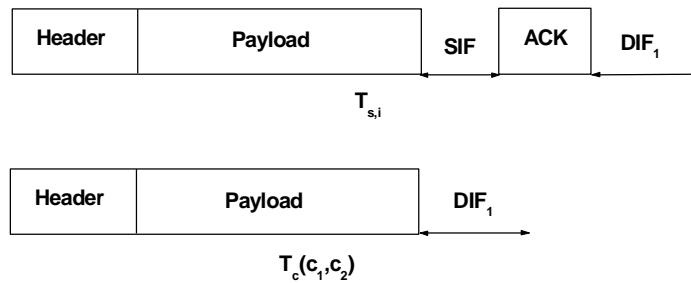
Fig. 3 Components for $T_{s,i}$ and $T_c(c_1,c_2)$

$T_{D,1}=Delay$
$T_{D,2}=Delay_1+Delay_2$
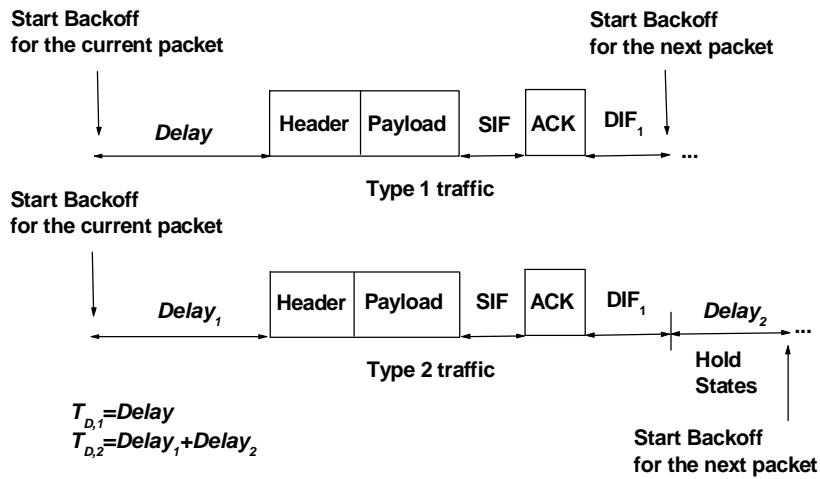
Fig. 4 Analysis of the packet delay

25
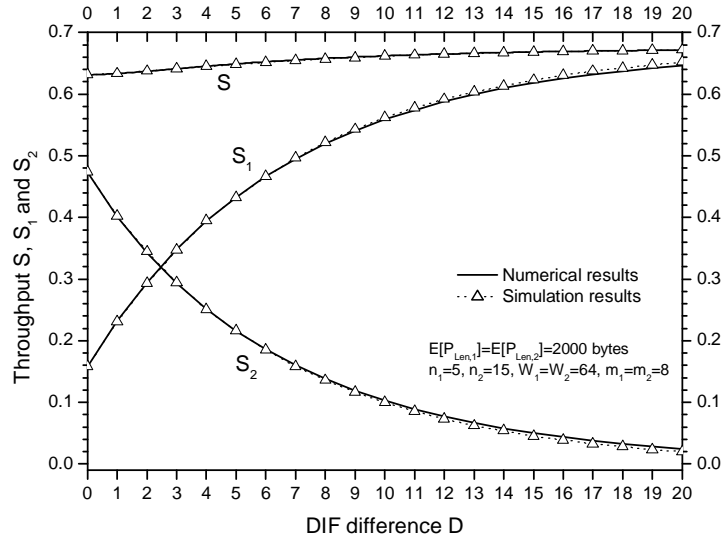
Fig. 5 Throughput S, S_1 and S_2 versus D
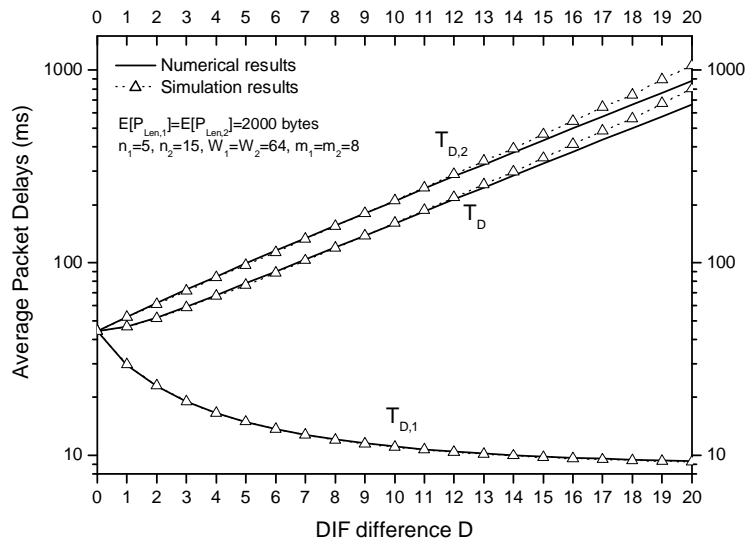
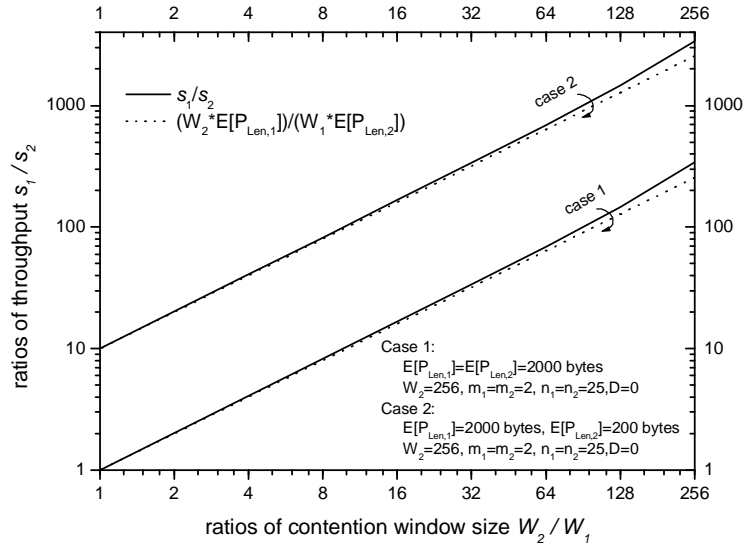Fig. 6 Average packet delays $T_{D,1}$, $T_{D,2}$ and T_D versus D

Fig. 7 Ratios of throughput s_1 / s_2 versus ratios of contention window size W_2 / W_1
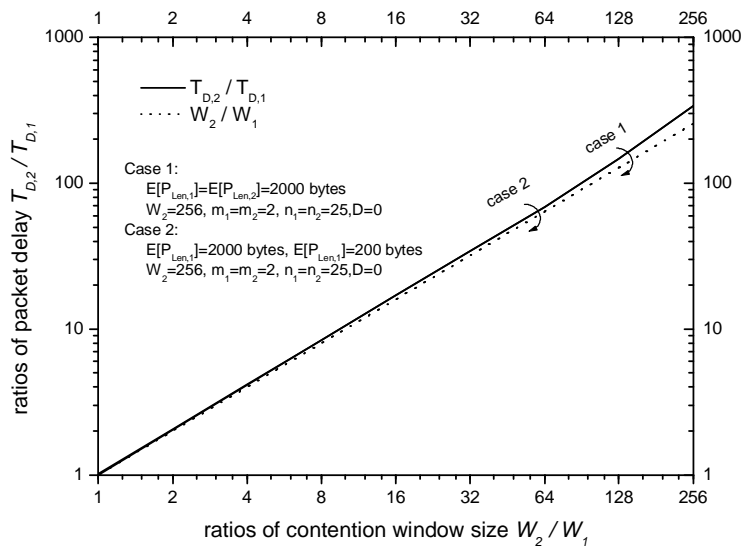

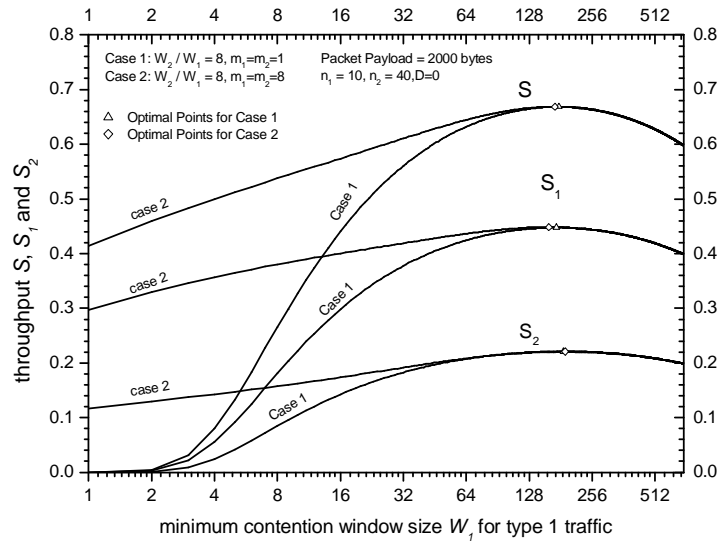Fig. 8 Ratios of packet delay $T_{D,2} / T_{D,1}$ versus ratios of contention window size W_2 / W_1

27

Case 1: $W_2 / W_1 = 8$, $m_1 = m_2 = 1$ Packet Payload = 2000 bytes
Case 2: $W_2 / W_1 = 8$, $m_1 = m_2 = 8$ $n_1 = 10$, $n_2 = 40$, D=0

△   Optimal Points for Case 1
◇   Optimal Points for Case 2

S

S_1

S_2

case 2

case 2

case 2

Case 1

Case 1

Case 1

throughput S, S_1 and S_2

minimum contention window size W_1 for type 1 traffic

Fig. 9 Throughput S, S_1 and S_2 versus contention window size W_1

Case 1: $W_2 / W_1 = 8$, $m_1 = m_2 = 1$
Case 2: $W_2 / W_1 = 8$, $m_1 = m_2 = 8$

△   Optimal Points for Case 1
◇   Optimal Points for Case 2

Packet Payload = 2000 bytes
$n_1 = 10$, $n_2 = 40$, D=0

Case 1

Case 1

Case 2

Case 2

$T_{D,2}$

$T_{D,1}$

packet delay $T_{D,1}$ and $T_{D,2}$ (ms)

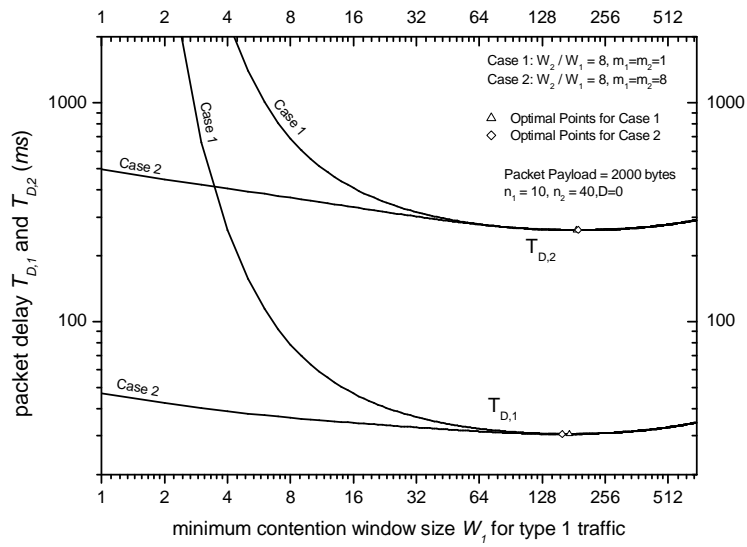minimum contention window size W_1 for type 1 traffic

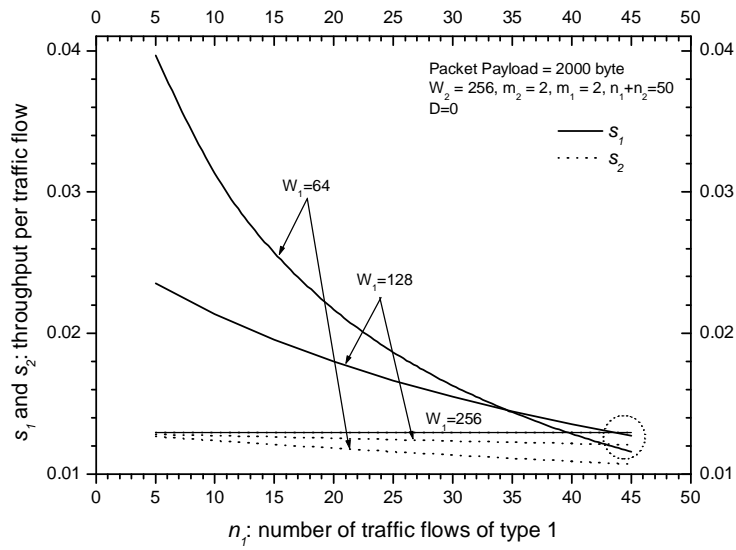Fig. 10 Packet delays $T_{D,1}$ and $T_{D,2}$ versus contention window size W_1

Fig. 11 Throughput s_1 and s_2 versus the number of traffic flows n_1
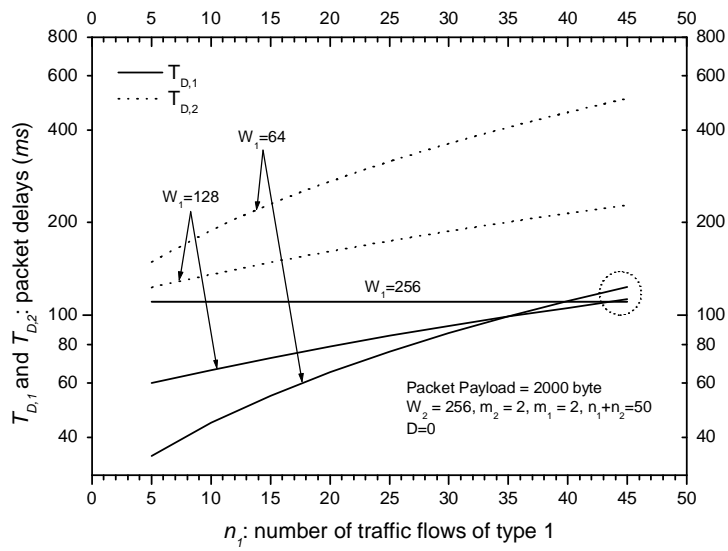

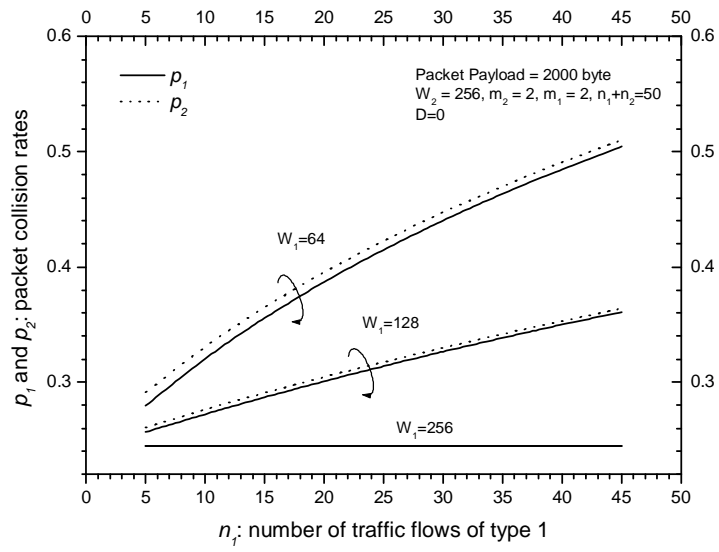Fig. 12 Packet delays $T_{D,1}$ and $T_{D,2}$ versus the number of traffic flows n_1

Fig. 13 Packet collision probabilities p_1 and p_2 versus the number of traffic flows n_1