Throughput and Delay Analysis for the IEEE 802.11e Enhanced Distributed Channel Access

Zhifeng Tao and Shivendra Panwar

***Abstract*—In this letter, we propose a three-dimensional Markov chain model for the 802.11e enhanced distributed channel access (EDCA) mode. This model can be used to compute the maximum sustainable throughput and service delay distribution for each priority class when under saturation load. The new framework models the performance impact of major quality-of-service (QoS)-specific features (e.g., CWMin, CWMax, AIFS, internal collision resolution) of the 802.11e EDCA mode, and hence can provide an analytical approach to pick the parameter values associated with EDCA to meet the QoS requirements of each priority.**

***Index Terms*—IEEE 802.11e protocol, Markov model, performance evaluation, quality of service (QoS).**

I. Introduction

As IEEE 802.11e [1] approaches final ratification, the proposed basic quality-of-service (QoS) enhancements are deemed mature, and have drawn tremendous interest from both industry and academia. It has been shown, using simulation models [2]–[4], that the new protocol provides a significant improvement in QoS support over the 802.11 medium access control (MAC). [5] modifies the Markov chain in [7] to model resource sharing by different classes. [6] introduces a separate one-dimensional (1-D) Markov chain to be used along with Bianchi's model, so that the effect of varying interframe spaces can also be investigated. [8] takes a similar approach as [6], but revises Bianchi's model to study the effect of the postcollision period. Nevertheless, both [6] and [8] are limited by the various approximations (e.g., Markov chain decomposition) made therein.

In this letter, we propose a 3-D Markov chain model for the 802.11e protocol, which takes the new QoS mechanisms into consideration. We then develop a simple, yet efficient, numerical approach to solving this chain. Based upon this Markov model, we compute the throughput that different traffic classes can sustain, and the distribution of the channel-access delay that each head-of-line (HOL) packet experiences when the network is heavily loaded. This model is also an extension of the Markov chain proposed in [7]. The main contribution of this new model is that it captures all of the major QoS-specific features for the enhanced distributed channel access (EDCA) mode as described in [1], as well as providing a novel numerical approach to solving the Markov chain and calculating the *delay distribution*.

The rest of the letter is organized as follows. In Section II, we briefly introduce the new QoS scheme specified in EDCA. Interested readers are referred to [2] and [6] for more protocol details. In Section III-A, the 3-D model is defined, and an example is provided to demonstrate how this embedded Markov chain is constructed. Section III-B provides a set of equations that describes the Markov chain with general parameters. The numerical-solution approach to the equations is elaborated in Section III-C. We compute the saturation throughput and HOL packet service-delay distribution of each priority in Section III-D and III-E, respectively. The corresponding model validation and simulation results are presented in Section IV. Section V ends this letter with conclusions and related future work.

II. Protocol Description

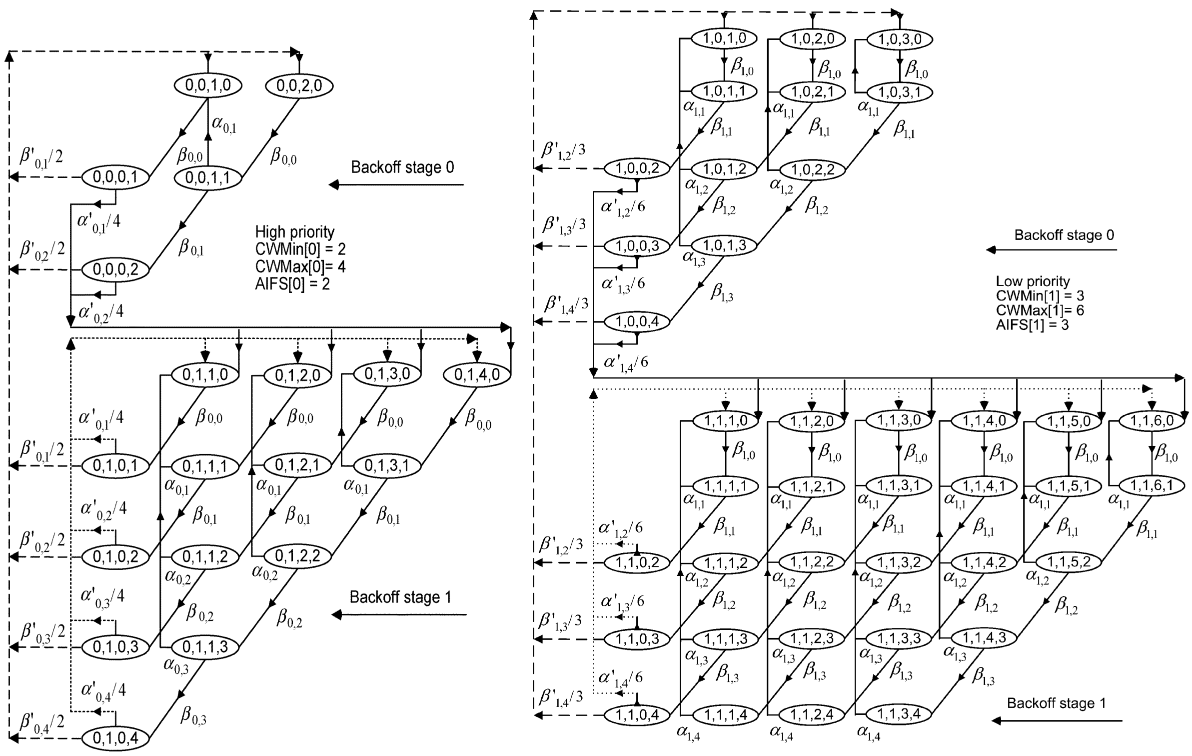
As an extension to the basic distributed coordination function (DCF) mechanism of the legacy 802.11, EDCA supports up to eight priorities in a station, which are mapped into four different access categories (ACs) at the MAC layer. A set of EDCA parameters, namely the arbitration interframe space (AIFS(AC)), minimum contention window size (CWMin(AC)), and maximum contention window size (CWMax(AC)), is associated with each AC to differentiate the channel access. AIFS(AC) is the time interval a packet of a given AC has to wait after the channel becomes idle before it can start the backoff process or transmit. After i collisions, the *backoff counter* for that particular AC is selected uniformly from range[1] $[1, 2^i \times \text{CWMin(AC)}]$, until i reaches the *backoff stage* m such that $2^m \times \text{CWMin(AC)} = \text{CWMax(AC)}$. At that point, the packet will still be retransmitted, if a collision occurs, with the backoff counters chosen from the range [1, CWMax(AC)]. When the total number of retransmissions equals the maximum number of allowable retransmissions [RetryLimit(AC)], no further retransmission is attempted, and the packet is discarded. Packets belonging to different ACs within a single station may collide with each other when their backoff counters decrement to zero simultaneously. This phenomenon is called an *internal collision* in 802.11e, and is prevented by letting the highest priority involved in the collision win the contention.

Paper approved by M. Zorzi, the Editor for Multiple Access of the IEEE Communications Society. Manuscript received October 23, 2004; revised April 23, 2005 and September 30, 2005. This paper was presented in part at the International Conference on Communications, Paris, France, June 2004.

The authors are with the Department of Electrical and Computer Engineering, Polytechnic University, Brooklyn, NY 11201 USA (e-mail: jefftao@photon.poly.edu; panwar@catt.poly.edu).

Digital Object Identifier 10.1109/TCOMM.2006.873066

[1]The Markov chain presented herein was initially developed based upon one of the first descriptions of 802.11e EDCF protocol [2]. As the 802.11e draft constantly evolves, changes have been introduced to the protocol. However, the pertinent revision, namely using $[0, 2^i \times \text{CW} - 1]$ instead of $[1, 2^i \times \text{CW}]$ in the backoff process, does not have a major impact on the analysis, and can also be readily handled by our model with only a minor modification.



(a) (b)

Fig. 1. Example Markov chain. (a) High priority: basic model. (b) Low priority: basic model.

III. Throughput and Delay Analysis

In our analysis, we assume that all nodes have chosen the same EDCA parameters. The network operates under the saturation condition (a.k.a. heavy load condition), where each priority queue within a station always has at least one packet awaiting transmission. Our Markov model does not consider frame errors, the hidden-node effect, or the capture effect. From now on, we will use the term *priority* instead of AC throughout our analysis. We also will adopt the less cumbersome notation $W_{\min}(c)$, $W_{\max}(c)$, and $D(c)$, to represent CWMin(AC), CWMax(AC), and AIFS(AC), respectively. Here, $0 \leq c \leq C-1$, and $c = 0$ corresponds to the highest priority. We further introduce three auxiliary variables in (1) to facilitate the mathematical derivation

$$\begin{cases} m_c \triangleq log_2\left(\frac{W_{\max}(c)}{W_{\min}(c)}\right), & c \in [0, C-1] \\ W_c^i \triangleq 2^i \times W_{\min}(c), & i \in [0, m_c], \quad c \in [0, C-1] \\ D_{j,i} \triangleq D(j) - D(i), & i, j \in [0, C-1], \quad j \geq i. \end{cases} \tag{1}$$

A. Markov Chain Definition

We use three state variables $[S^c(t), B^c(t), L(t)] (c \in [0, C-1])$, the first two of which are the values of the backoff stage and backoff time counter, respectively, to represent the state of the packet at the head of the priority c queue within a station. The third state variable, denoted $L(t)$, helps us locate the current time slot in an operation cycle, as explained next. $L(t) = k$ means that the observation time instance is $k (k \in [0, W\text{Max}])$ time slot(s) after the end of the last $D(0)$ period. $W\text{Max}$, defined as $\min_{c \in [0, C-1]}\{D_{c,0} + W_{\max}(c)\}$, is the largest number of backoff slots that any priority can continuously count down within an operation cycle, under the saturation condition. The observation points that we use to construct the embedded Markov chain are those at which state variables $[S^0(t), B^0(t), L(t)]$ change value. We will use $[c, i, j, k]$ and $p(c, i, j, k)$ to denote the state $[S^c(t) = i, B^c(t) = j, L(t) = k]$ and its corresponding steady-state probability, respectively. For example, Fig. 1(a) and 1(b) portray the Markov chain models for the high- and low-priority queues. The transition probabilities $\alpha_{c,k}$, $\alpha'_{c,k}$, $\beta_{c,k}$, and $\beta'_{c,k}$ used in these figures will be further defined in Section III-B.

B. Equation Formulation

Suppose there are N stations in an 802.11e network. Assume that $\forall i, j \in [0, C-1]$ and $j > i$, $W_{\max}(i) > D_{j,i}$. In this case, contention for channel access exists between any two priorities in the 802.11e network. For all $c \in [0, C-1]$, and $i \in [0, m_c]$, the Markov chain for priority c can be represented by (2). The variables $q_{c,i}$ in the equation will be specified later in (3).

Under the saturation condition, when the HOL packet of priority c is dropped or successfully transmitted, the next priority c packet becomes the HOL packet and selects an initial backoff counter value. This corresponds to the fact that the Markov chain is initially entered from state $[c, 0, j, 0]$ for each new packet and described by the first case in (2), shown at the bottom of the page. The second case of (2) represents the situation where the priority c queue is still in its AIFS period, while higher priority queues have started counting down. This corresponds to the transition from state $[1, i, j, 1]$ to $[1, i, j, 0]$ in Fig. 1(b). The third equation describes the transition from $[c, i, j, k]$ to $[c, i, j-1, k+1]$, where the channel is sensed idle and priority c decrements its backoff counter with probability $\beta_{c,k}$.

Equation (3), shown at the bottom of the page, relates the auxiliary variable $q_{c,i}$ with $p(c, i, j, k)$ and transition probabilities. The first case in (3) (i.e., $i = 0$) corresponds to the initial backoff stage. The second case, where $1 \le i \le m_c - 1$, represents any intermediate backoff stages. For the last stage (i.e., $i = m_c$), a collision always leads to a loopback to stage m_c, since we assume that an unsuccessful station will keep retransmitting with backoff window set to $W_{\max}(c)$.

$\alpha_{c,k}$ and $\beta_{c,k}$ are the probabilities that the wireless channel is observed to be busy or idle, respectively, at the kth slot during an operation cycle by a priority c HOL packet, whose backoff counter value j has not reached zero yet. Similarly, $\alpha'_{c,k}$ and $\beta'_{c,k}$ represent the probability that the channel is found to be busy or idle, which leads to a collision or success, respectively, during the kth slot in an operation cycle by a priority c packet, whose backoff counter value j equals zero. To relate $\alpha_{c,k}$, $\beta_{c,k}$, $\alpha'_{c,k}$, and $\beta'_{c,k}$ with the steady-state probability $p(c, i, j, k)$, we define $\tau_{c,k}$ as $P\{\text{A queue of priority } c \text{ counts down to } 0 | L(t) = k\}$. Note that a queue of priority c may not necessarily be able to initiate a transmission, when it counts down to zero, since it may be involved in an internal collision and lose the right to access the channel. We further make the assumption that $\tau_{c,k}$ is a constant with respect to state variables i and j, and only depends on the relative interval k in which a state transition occurs. Define $H(c)$ as the largest integer for each priority that satisfies $W_c^{H(c)} + D_{c,0} \le W\text{Max}$. Also, define $W_c^{-1} = 0$. Then, $\tau_{c,k}$ can be expressed as shown in (4) at the bottom of the page.

The first case in (4) corresponds to the scenario when the priority c queue is still in its AIFS period, and thus its conditional transmission probability is always zero. The second case represents the interval during which the priority c queue has completed its AIFS period, but has not counted down for more than $\min\{W\text{Max}, W_{\min}(c) + D_{c,0}\}$ time slots. The last case accounts for the limit imposed by $W\text{Max}$. Even though the model can handle it, for the sake of simplicity, we do not consider the case where $W_{\min}(c) + D_{c,0} \ge W\text{Max}$, which represents an unlikely choice of the QoS parameters.

For any $k \in [0, W\text{Max}]$, we express $\alpha_{c,k}$, $\beta_{c,k}$, $\alpha'_{c,k}$, $\alpha'_{c,k}$, and $\beta'_{c,k}$ in terms of $\tau_{c,k}$ in (5), where the impact of *internal collision resolution* is explicitly reflected

$$\begin{cases} 1 - \alpha_{c,k} = \beta_{c,k} = (1 - \tau_{c,k})^{N-1} \\ \quad \times \prod_{u=0}^{c-1} (1 - \tau_{u,k})^N \prod_{u=c+1}^{C-1} (1 - \tau_{u,k})^N \\ 1 - \alpha'_{c,k} = \beta'_{c,k} = \prod_{u=0}^{c-1} (1 - \tau_{u,k})^N \\ \quad \times \prod_{u=c}^{C-1} (1 - \tau_{u,k})^{N-1}. \end{cases}$$

$$\begin{cases} p(c,i,j,0) = q_{c,i} + \sum_{v=1}^{\min\left\{W_c^i + D_{c,0} - j, W\text{Max}\right\}} \alpha_{c,v} p(c,i,j,v), & j \in \left[1, W_c^i\right] \\ p(c,i,j,k) = \beta_{c,k-1} p(c,i,j,k-1), & j \in \left[1, W_c^i\right], \quad k \in [1, D_{c,0}] \\ p(c,i,j,k) = \beta_{c,k-1} p(c,i,j+1,k-1), & j \in \left[0, W_c^i - 1\right], \quad k + j \le W_c^i + D_{c,0} \\ & \quad k \in \left[D_{c,0} + 1, \min\left\{W_c^i + D_{c,0}, W\text{Max}\right\}\right] \end{cases} \tag{2}$$

$$\begin{cases} q_{c,0} = \sum_{v=0}^{m_c} \sum_{k=D_{c,0}+1}^{\min\{W\text{Max},\ W_c^v + D_{c,0}\}} \frac{\beta'_{c,k} p(c,v,0,k)}{W_{\min}(c)} \\ q_{c,i} = \sum_{k=D_{c,0}+1}^{\min\left\{W\text{Max}, W_c^{i-1} + D_{c,0}\right\}} \frac{\alpha'_{c,k} p(c,i-1,0,k)}{W_c^i}, & 1 \le i \le m_c - 1 \\ q_{c,m_c} = \sum_{k=D_{c,0}+1}^{\min\left\{W\text{Max}, \frac{W_{\max}(c)}{2} + D_{c,0}\right\}} \frac{\alpha'_{c,k} p(c,m_c-1,0,k)}{W_{\max}(c)} \\ \quad + \sum_{k=D_{c,0}+1}^{\min\{W\text{Max}, W_{\max}(c) + D_{c,0}\}} \frac{\alpha'_{c,k} p(c,m_c,0,k)}{W_{\max}(c)} \end{cases} \tag{3}$$

$$\tau_{c,k} = \begin{cases} 0, & k \in [0, D_{c,0}] \\ \dfrac{\sum_{i=v}^{m(c)} p(c,i,0,k)}{\sum_{i=v}^{m(c)} \left[\sum_{j=0}^{W_c^i + D_{c,0} - k} p(c,i,j,k)\right]}, & v \in [0, H(c)], \quad H(c) \ge 0 \\ & k \in \left[W_c^{v-1} + D_{c,0} + 1, \min\left\{W\text{Max}, W_c^v + D_{c,0}\right\}\right] \\ \dfrac{\sum_{i=H(c)+1}^{m(c)} p(c,i,0,k)}{\sum_{i=H(c)+1}^{m(c)} \left[\sum_{j=0}^{W_c^i + D_{c,0} - k} p(c,i,j,k)\right]}, & k \in \left[W_c^{H(c)} + D_{c,0} + 1, W\text{Max}\right], H(c) \ge 0 \end{cases} \tag{4}$$

C. Numerical Solution: An Iterative Approach

Since all the transition probabilities have been expressed as functions of $p(c,i,j,k)$, the Markov chain can be numerically solved, using an iterative approach outlined as follows.

For each Markov chain that corresponds to a specific priority c, we assume that it contains a total of M_c states, which are numbered 1 through m_c. Let $\vec{\pi_c}$ and $\vec{\pi}_c^0$ be the steady-state distribution vector and an arbitrary initial vector, respectively. $\boldsymbol{P}_c$ denotes the transition probability matrix of that Markov chain. For an ergodic irreducible Markov chain, $\lim_{L\to\infty} \vec{\pi}_c^0 \times \underbrace{\boldsymbol{P}_c \times \boldsymbol{P}_c \times \ldots \boldsymbol{P}_c}_{L} = \vec{\pi_c}$, independent of the value of initial vector $\vec{\pi}_c^0$. To solve the Markov chain, therefore, we first create $\vec{\pi}_c^0$ by assigning the entry state of the Markov chain $[c,0,j,0]$ with an equal probability of $1/W_{\min}(c)$ and setting all other initial-state probabilities to zero. Then, $\vec{\pi}_c^0$ can be plugged into (4) to calculate $\tau_{c,k}$, which, in turn, is used to compute the state-transition probability $\boldsymbol{P}_c$ by following (5). Multiplying $\vec{\pi}_c^0$ with $\boldsymbol{P}_c$ yields a new state-distribution vector $\vec{\pi}_c^1$, which is used to update the state-transition matrix $\boldsymbol{P}_c$. The new $\boldsymbol{P}_c$ is again multiplied with $\vec{\pi}_c^1$ to obtain yet another new state-distribution vector $\vec{\pi}_c^2$. Following this iterative approach, a convergence to the steady-state distribution $\vec{\pi_c}$ can be achieved. Our computation shows that the convergence occurs in approximately 30–100 iterations, depending on the size of the Markov chain.

D. Throughput

For a particular priority c, the wireless channel can witness four possible events occurring at any randomly chosen time slot. More specifically, the channel may be idle for a random backoff, or busy either due to a successful priority-c transmission or a collision involving that priority. It is also possible that priority-c queues in all the stations are in a *freeze period*, during which no priority-c queue within any station is transmitting, or decrementing its backoff counter, because at least one transmission of another priority is ongoing on the wireless medium. We use Pb_c, Ps_c, Pc_c, and Pf_c to denote the probabilities that the channel sees one of the four abovementioned events for priority-c, namely backoff, successful transmission, collision, and freeze, respectively. Assuming packets from different priority queues have the same size, Tb, Ts, Tc, and Tf can then represent the corresponding length of a backoff time slot, successful transmission, collision, or freeze period. Define the throughput (S_c) as the fraction of time the network transmits the packet payload bits of priority-c successfully. S_c then can be expressed as

$$\begin{aligned} S_c &= \frac{Ps_c \cdot \overline{\text{Payload}}}{Pb_c \cdot Tb + Ps_c \cdot Ts + Pc_c \cdot Tc + Pf_c \cdot Tf} \\ &\approx \frac{Ps_c \cdot \overline{\text{Payload}}}{Pb_c \cdot Tb + (Ps_c + Pc_c + Pf_c) \cdot Ts} \\ &= \frac{Ps_c \cdot \overline{\text{Payload}}}{Pb_c \cdot Tb + (1 - Pb_c) \cdot Ts}. \end{aligned} \tag{6}$$

The probability of having a success and a backoff can be written as shown in (7) at the bottom of the page.

In (7), $P\{k\}$ is the probability that a priority is in the state of $[c,i,j,k]$, which implies that k slots have elapsed since the end of the last $D(0)$ period. Since $P\{k\}$ only depends on k and is independent of priority c, it is sufficient to derive $P\{k\}$ for the high-priority case (i.e., $c = 0$), as shown in (8) at the bottom of the page.

Finally, by substituting (7) and (8) into (6), we can obtain the throughput for each priority.

E. Service-Delay Distribution

The service delay of a successfully delivered HOL packet is the time duration from its first backoff until it leaves the system. In order to facilitate the computation of service-delay distribution, the Markov model defined in Section III-A requires minor modifications. The procedure employed to compute the delay distribution, as well as the necessary changes to the original Markov chain, are summarized as follows.

1) We incorporate the possibility of a packet drop into our Markov model, which can be readily achieved by duplicating the last stage in the initial chain RetryLimit$(c) - m_c$ times, and letting any transmission attempt in the last

$$\begin{cases} Ps_c = \sum_{k=1}^{W\text{Max}-1} \left[P\{k\} \times P\left\{\text{A success for priority } c | L(t) = k\right\}\right] \\ \quad = \sum_{k=1}^{W\text{Max}-1} \left[P\{k\} \cdot N\tau_{c,k}(1-\tau_{c,k})^{N-1} \times \prod_{u=0}^{c-1}(1-\tau_{u,k})^N \times \prod_{u=c+1}^{C-1}(1-\tau_{u,k})^{N-1}\right] \\ Pb_c = \sum_{k=0}^{W\text{Max}-1} \left[P\{k\} \times P\left\{\text{A backoff slot for priority } c | L(t) = k\right\}\right] \\ \quad = \sum_{k=0}^{W\text{Max}-1} \left[P\{k\} \times \prod_{u=0}^{C-1}(1-\tau_{u,k})^N\right] \end{cases} \tag{7}$$

$$P\{k\} = \begin{cases} \sum_{i=0}^{m_0} \left[\sum_{j=1}^{W_0^i} p(0,i,j,k)\right], & k \in [0, D_{0,0}] \\ \sum_{i=v}^{m_0} \left[\sum_{j=0}^{W_0^i + D_{0,0} - k} p(0,i,j,k)\right], & v \in [0, H(0)], H(0) \geq 0 \\ & k \in \left[W_0^{v-1} + D_{0,0} + 1, \min\left\{W\text{Max}, W_0^v + D_{0,0}\right]\right\} \\ \sum_{i=H(0)+1}^{m_0} \left[\sum_{j=0}^{W_0^i + D_{0,0} - k} p(0,i,j,k)\right], & H(0) \geq 0, k \in \left[W_0^{H(0)} + D_{0,0} + 1, W\text{Max}\right] \end{cases} \tag{8}$$

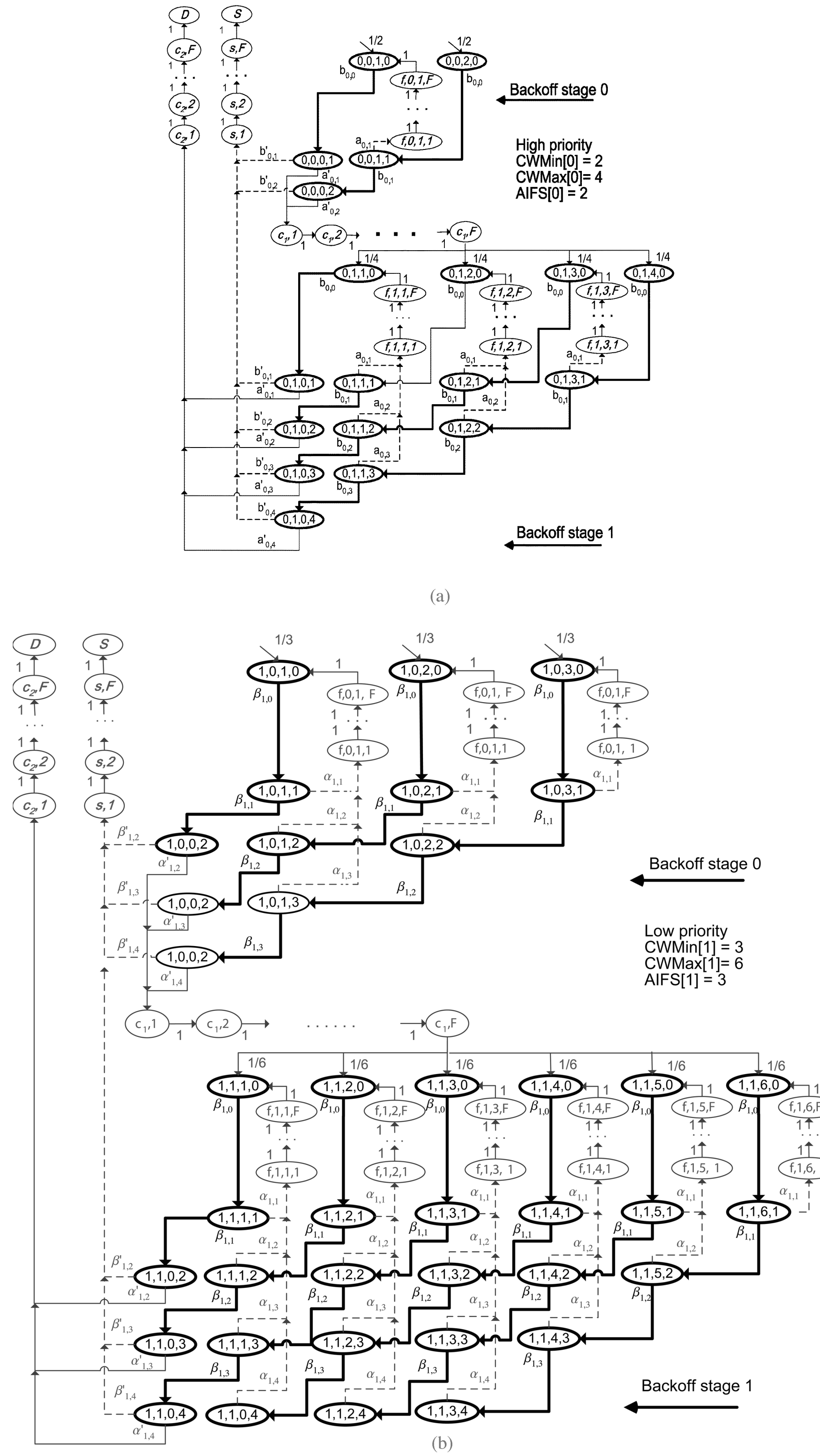


Fig. 2. Example Markov chain for delay-distribution analysis. (a) High priority. (b) Low priority.

backoff stage go back to the initial backoff stage, regardless of whether a successful delivery or collision occurs. The resultant Markov chain can be solved using the numerical approach outlined in Section III-C. A more detailed description of similar modifications can be found in [9].

2) In the embedded Markov model specified in Section III-A, the time interval between two consecutive observations may correspond to events of different durations. We further slot the channel busy period (i.e., a successful transmission, a collision, or a freeze period), where the

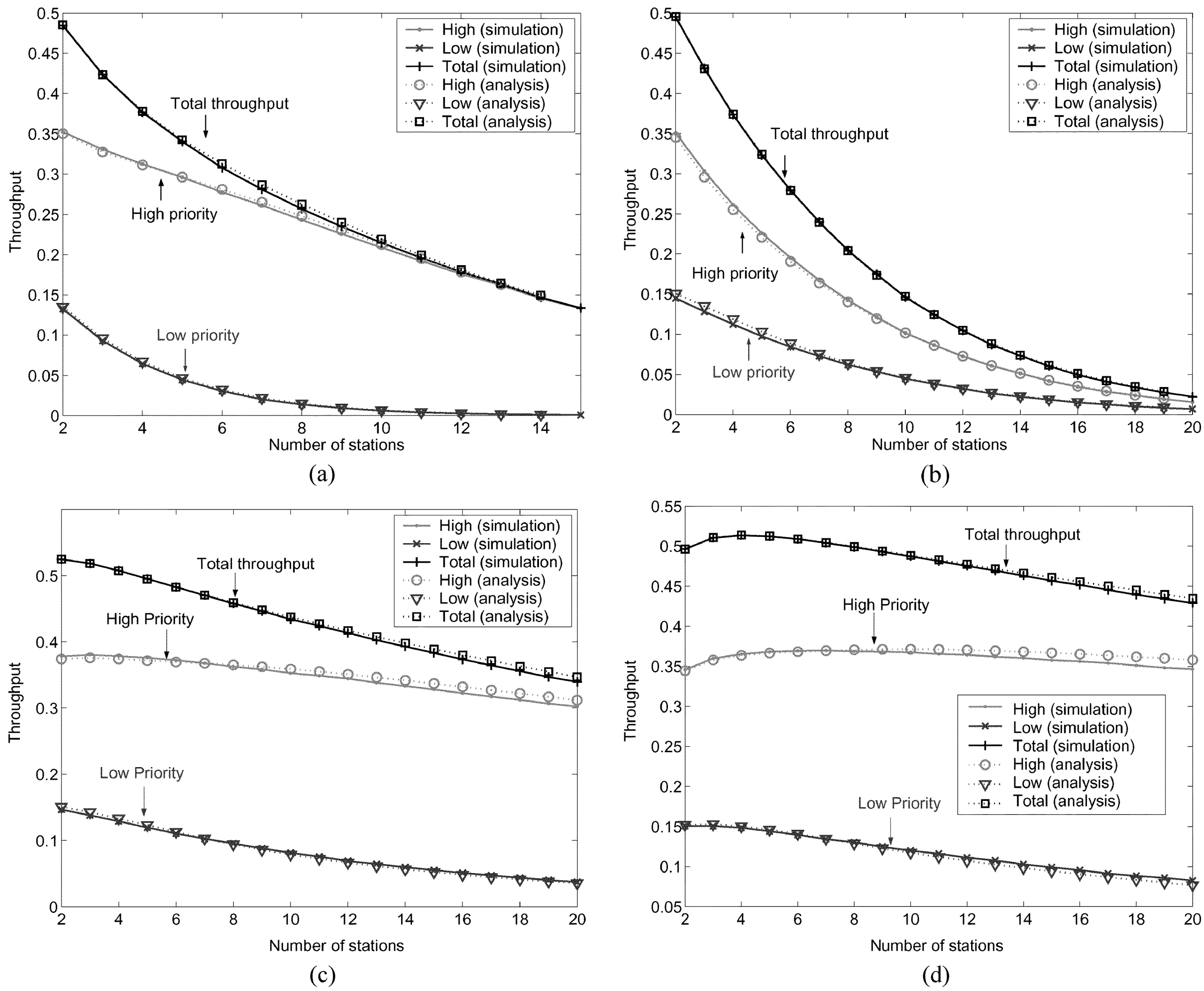


Fig. 3. Saturation throughput. (a) $D(0)/D(1) = 2/3$, $W_{\min}/W_{\mathrm{Max}}(0) = 8/16$, $W_{\min}/W_{\max}(1) = 8/16$. (b) $D(0)/D(1) = 2/2$, $W_{\min}/W_{\mathrm{Max}}(0) = 8/16$, $W_{\min}/W_{\max}(1) = 16/32$. (c) $D(0)/D(1) = 2/3$, $W_{\min}/W_{\mathrm{Max}}(0) = 32/64$, $W_{\min}/W_{\max}(1) = 64/128$. (d)$D(0)/D(1) = 2/3, W_{\min}/W_{\max}(0) = 64/128, W_{\min}/W_{\max}(1) = 128/256$.

unit slot time is equal to a single backoff time slot. As a result, for instance, each transition that represents a channel busy period in Fig. 1(a) and (b) is replaced by a set of states in Fig. 2(a) and (b). If we assume that all the data frames have the same size, each busy period then consists of F time slots, which can be approximated by $\lceil$Average channel busy time/Backoff time slot$\rceil$.

3) Two special sink states $\mathbf{S}$ and $\mathbf{D}$ are introduced to the Markov model in order to trap all the packets that are discarded or successfully transmitted, respectively.
For illustration purposes, we apply the necessary revisions described in steps 2 and 3 on the chains shown in Fig. 1(a) and (b), yielding two new chains depicted in Fig. 2(a) and (b) for high- and low-priority queues, respectively. All the new states and transitions are marked in thin lines. The states in Fig. 2(a) and (b) are obviously distinct. But for the sake of simplicity, we use the same labels for some of the states in these two chains.

4) Initialize the vector $\vec{\pi}_c^0$ for the new Markov chain obtained in step 3, and continue multiplying it with the corresponding $\boldsymbol{P}_c$. As illustrated in Fig. 2(a) and (b), the only nonconstant transition probabilities in the new chains are $\alpha_{i,j}$ and $\beta_{i,j}$, which assume the same values as their counterparts in the Markov chain specified in step 1. Keep track of the probability of staying in sink state $\mathbf{S}$ in vector $\vec{\pi}_c$. Once a difference in this probability between two consecutive $\vec{\pi}_c^{i-1}$ and $\vec{\pi}_c^i$ is detected, record the number of iterations i, and the associated probability p_i as well. Note that since the number of iterations is equal to the number of unit time slots, i essentially corresponds to one of the possible delay values that a frame experiences before it gets successfully delivered. Continue this procedure until the probabilities of $\mathbf{S}$ and $\mathbf{D}$ in $\vec{\pi}_c$ add up to 1. At this moment, all the collected values of i and the associated probabilities p_i, with proper normalization of course, constitute the service-delay distribution for successful frames.

IV. Model Validation and Simulation

To validate the analytical model, we have developed an event-driven custom simulator using the C programming language, which carefully models all the major features defined in 802.11e EDCA, including carrier sense multiple access with collision avoidance (CSMA/CA), multiple ACs, and internal collision resolution. The physical and MAC layer parameters used in the

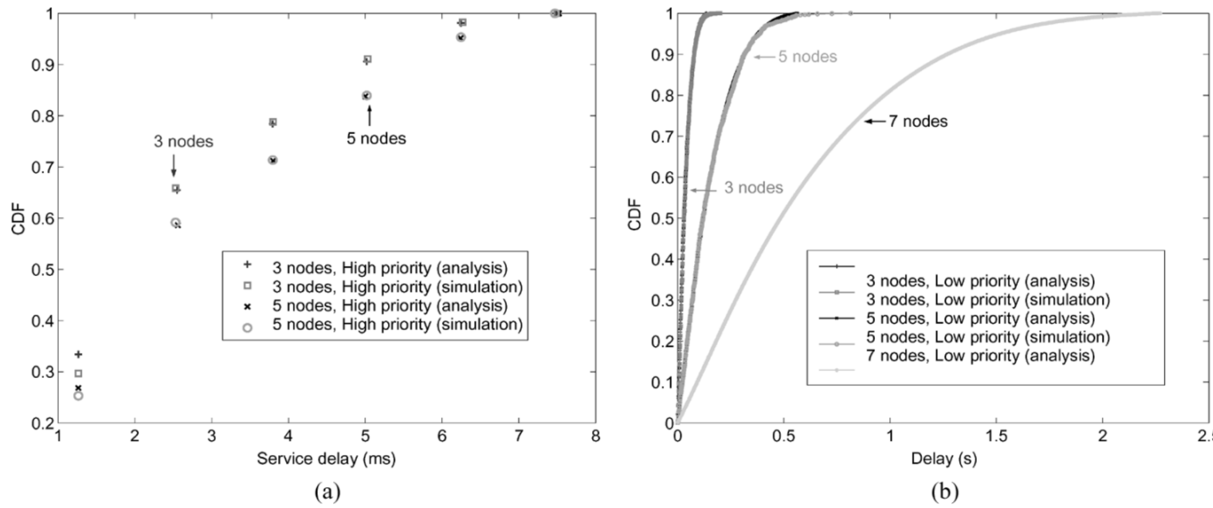


Fig. 4. Cumulative distribution of service delay. (a) High priority: $D(0)/D(1) = 2/3$, $W_{\min}/W_{\max}(0) = 2/4$, $W_{\min}/W_{\max}(1) = 3/6$. (b) Low priority: $D(0)/D(1) = 2/3$, $W_{\min}/W_{\max}(0) = 2/4$, $W_{\min}/W_{\max}(1) = 3/6$.

following discussion are specified in [10] and [11]. The data payload is 1024 B, and is transmitted at 11 Mb/s. Note that the simulation results for throughput presented hereafter are averaged over ten simulation runs. Since the boundary of 95% confidence interval is very close to the sample mean, it is not explicitly indicated in the figures for the sake of legibility.

Fig. 3 depicts the differentiation of throughput for two priorities in 802.11e EDCA. Fig. 3(a) and (b) illustrate the impact of AIFS versus that of $CW\mathrm{Min}/CW\mathrm{Max}$, respectively. The comparison reveals that AIFS has a more pronounced effect on service differentiation than $CW\mathrm{Min}/CW\mathrm{Max}$ alone, which has also been reported and developed further in [12]. Fig. 3(c) and (d) demonstrate that when we increase the $W_{\min}$ and $W_{\max}$ for both priorities to reasonably large values, the total throughput in Fig. 3(d) first increases and then decreases, which suggests that excessive collisions start to play a dominant role in determining the saturation throughput, as the network becomes more densely populated. Note that although this trend can be observed for both traffic types, the high-priority traffic still enjoys an appreciably slower throughput decay, thanks to the differentiation effect of $CW\mathrm{Min}/CW\mathrm{Max}$.

Fig. 4 shows the cumulative delay distribution for high- and low-priority packets, respectively. In Fig. 4(a), relative to the backoff time, the delays for a three-station and five-station network are mainly dominated by unsuccessful transmissions and the freeze period, which explains why the delays for high priority are centered around several discrete points corresponding to transmission events. The delay distribution for low-priority traffic in three different network scenarios are depicted in Fig. 4(b). It can also be seen that the service delay of low-priority packets in Fig. 4(b) is substantially higher than that of high priority in Fig. 4(a), again illustrating the prioritization effect of EDCA.

Meanwhile, due to this traffic prioritization, the throughput of low-priority traffic tends to diminish as the number of stations in the network grows. This renders the observation of successfully delivered low-priority packets to be a rare event in the simulation, which eventually causes difficulty in collecting adequate samples to derive a delay distribution for low-priority traffic. It is worthwhile to note that the analysis approach we propose does not have this problem, and hence, can generate a delay distribution more easily than simulation, at least with the parameter sets used herein. This also explains why only the analysis result of delay for a seven-node case is presented in Fig. 4(b). Since delay distribution plays a central role in determining the QoS for such critical applications as voice over IP, the ability to obtain this statistical metric is essential for any analytical model.

Also note that this analysis gives the worst-case maximum throughput and delay distribution for the low-priority traffic class, since the network operates under the saturation condition. When the applied load is less than saturated for high-priority traffic, we expect the throughput (delay) of low-priority traffic to be above (below) these values.

V. Conclusions and Future Work

In this letter, we develop a multidimensional Markov chain to analyze the saturation throughput and service-delay distribution of the 802.11e EDCA protocol. The analytical framework accurately models major QoS features introduced in EDCA, and hence, leads to results that match with simulations extremely well. The study also suggests that AIFS may have more appreciable influence on QoS differentiation than $CW\mathrm{Min}/CW\mathrm{Max}$.

As future work, incorporation of the $TXOP$ limit $[AC]$ [1] QoS feature into the Markov chain will be attempted. Moreover, the performance of networks with more than two traffic priorities will be evaluated within the proposed analytical framework. In addition, efforts to judiciously simplify the current model are also under consideration. Last, but not least, an extension of this Markov model will be applied to study the performance of 802.11e EDCA under nonsaturation conditions.

Acknowledgment

The authors would like to extend their gratitude to the anonymous reviewers for their insightful comments and constructive suggestions, which have helped improve the quality of this letter significantly.

REFERENCES

[1] *Draft Supplement to Part 11: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Medium Access Control (MAC) Enhancements for Quality of Service (QoS)*, IEEE 802.11e/D6.0, 2003.

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

[3] P. Garg, R. Doshi, R. Greene, M. Baker, M. Malek, and X. Cheng, "Using IEEE 802.11e MAC for QoS over Wireless," in *Proc. IEEE Int. Performance Comput. Commun. Conf.*, Phoenix, AZ, Apr. 2003, pp. 537–542.

[4] A. Lindgren, A. Almquist, and O. Schelen, "Quality of service schemes for IEEE 802.11 wireless LANs—An evaluation," *J. Special Issues Mobility of Syst., Users, Data, Comput.*, vol. 8, no. 3, pp. 223–235, Jun. 2003.

[5] Y. Xiao, "Enhanced DCF of IEEE 802.11e to support QoS," in *Proc. IEEE WCNC*, New Orleans, LA, Mar. 2003, vol. 2, pp. 1291–1296.

[6] S. Mangold, G. Hiertz, and B. Walke, "IEEE 802.11e wireless LAN—Resource sharing with contention based medium access," in *Proc. IEEE PIMRC*, Beijing, China, Sep. 2003, vol. 3, pp. 2019–2026.

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

[8] J. W. Robinson and T. S. Randhawa, "Saturation throughput analysis of IEEE 802.11e enhanced distributed coordination function," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 5, pp. 917–928, Jun. 2004.

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

[10] S. Mangold, S. Choi, G. Hiertz, O. Klein, and B. Walke, "Analysis of IEEE 802.11e for QoS support in wireless LANs," *IEEE Wireless Commun.*, vol. 10, no. 6, pp. 40–50, Dec. 2003.

[11] *Part 11: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Higher-Speed Physical Layer Extension in the 2.4 GHz Band*, IEEE 802.11b, 1999.

[12] C. Chou, K. Shin, and S. Shankar, "Interframe space (IFS) based service differentiation for IEEE 802.11 wireless LANs," in *Proc. IEEE Veh. Technol. Conf.*, Oct. 2003, vol. 3, pp. 1412–1416.