Analysis of the 802.11e Enhanced Distributed Channel Access Function †‡

Inanc Inan *Member*, Feyza Keceli *Member*, and Ender Ayanoglu *Fellow*
Center for Pervasive Communications and Computing
Department of Electrical Engineering and Computer Science
The Henry Samueli School of Engineering
University of California, Irvine, 92697-2625
Email: {iinan, fkeceli, ayanoglu}@uci.edu

Abstract

We propose an analytical model for the performance analysis of the Enhanced Distributed Channel Access (EDCA) function of the IEEE 802.11e standard. The proposed Discrete-Time Markov Chain (DTMC) model incorporates the main QoS features of 802.11e, namely, CW, AIFS, and TXOP differentiation. Due to its specific design, the proposed DTMC model jointly considers the state of the Medium Access Control (MAC) layer buffer and the MAC differentiation mechanisms which facilitates a novel performance analysis framework at an arbitrary traffic load. Analytical and simulation results are compared to demonstrate the accuracy of the proposed approach for varying traffic load, EDCA parameters, and MAC layer buffer space.

I. Introduction

The IEEE 802.11 standard [1] defines the Distributed Coordination Function (DCF) which provides best-effort service at the Medium Access Control (MAC) layer of the Wireless Local Area Networks (WLANs). The recently ratified IEEE 802.11e standard [2] specifies the Hybrid Coordination Function (HCF) which enables prioritized and parameterized Quality-of-Service (QoS) services at the MAC layer, on top of DCF. The HCF combines a distributed contention-based channel access mechanism, referred to as Enhanced Distributed Channel Access (EDCA), and a centralized polling-based channel access mechanism, referred to as HCF Controlled Channel Access (HCCA).

In this paper, we confine our analysis to the EDCA scheme, which uses Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) and slotted Binary Exponential Backoff (BEB) mechanism as the

† This work is supported by the Center for Pervasive Communications and Computing, and by National Science Foundation under Grant No. 0434928. Any opinions, findings, and conclusions or recommendations expressed in this material are those of authors and do not necessarily reflect the view of the National Science Foundation.

‡ Portions of the work in this paper has been presented at IEEE Global Communications Conference, November 2007.

basic access method. The EDCA defines multiple Access Categories (AC) with AC-specific Contention Window (CW) sizes, Arbitration Interframe Space (AIFS) values, and Transmit Opportunity (TXOP) limits to support MAC-level QoS and prioritization [2].

In order to assess the performance of these functions, simulations or mathematical analysis can be used. Although simulation models may capture system dynamics very closely, they lack explicit mathematical relations between the network parameters and the performance measures. A number of networking functions would benefit from the insights provided by such mathematical relations. For example, analytical modeling is a convenient way to assist QoS-aware MAC scheduling and Call Admission Control (CAC) algorithms. On the other hand, analytical modeling can potentially be complex, where the effect of multiple layer network parameters makes the task of deriving a simple and accurate analytical model highly difficult. However, a set of appropriate assumptions may lead to simple yet accurate analytical models.

The assumption that every station has always data ready to transmit in its buffer anytime (in saturation) provides accurate asymptotic figures. As will be discussed further in Section III, the majority of analytical work on the performance of 802.11e EDCA (and of 802.11 DCF) assumes saturation. However, as also considered in a number of other analytical models, the saturation assumption is unlikely to be valid in practice given the fact that the demanded bandwidth for most of the Internet traffic is variable with significant idle periods. One of the key features of our EDCA analytical model is that it releases the saturation assumption. The model is shown to predict EDCA performance accurately for the whole traffic load range from a nonsaturated channel to a saturated medium for a range of traffic models.

The majority of analytical work on the performance of 802.11e EDCA (and of 802.11 DCF) assumes constant collision probability for any transmitted packet at an arbitrary backoff slot independent of the number of retransmissions it has experienced (slot homogeneity). A complementary assumption is the constant transmission probability for any AC at an arbitrary backoff slot independent of the number of retransmissions it has experienced. As will be discussed in Section III, these approximations lead to accurate analysis, especially in saturation. Our analysis also shows that the slot homogeneity assumption leads to accurate performance prediction even when the saturation assumption is released.

Following slot homogeneity assumption, we model the EDCA function via a three-dimensional Discrete-Time Markov Chain (DTMC). The main distinctions of the proposed DTMC from previously proposed models are that the proposed analytical model incorporates *i)* main EDCA QoS parameters, CW, AIFS, and TXOP, and *ii)* the finite size MAC layer queue (interface queue between Link Layer (LL) and MAC layer). The former feature of the model provides an accurate analysis for the main EDCA differentiation mechanisms, considering varying collision probabilities at different AIFS slots (which is a direct result of varying number of contending stations) and multiple packet transmissions within an EDCA TXOP. In the meantime, the latter feature enables consideration of the case of finite traffic load (rather than saturation) and also shows the significant effect of MAC layer buffer space on the EDCA performance. This is the first EDCA analytical model which incorporates main EDCA QoS parameters, CW, AIFS, and TXOP, for an arbitrary traffic load. Comparing with simulation results, we show that our model can provide accurate results for an arbitrary selection of AC-specific EDCA parameters at any load.

To enable analysis in the Markov framework, we assume constant probability of packet arrival per state (for the sake of simplicity, Poisson arrivals). On the other hand, we also show that the results hold for a range of traffic types.

II. EDCA Overview

The IEEE 802.11e EDCA is a QoS extension of IEEE 802.11 DCF. To support QoS, EDCA differentiates packets using different priorities and maps them to specific ACs. Each AC$_i$ within a station ($0 \leq i \leq 3$) having its own EDCA parameters contends for the channel independently. The convention of [2] is that the larger the index i is, the higher the priority of the AC is. Levels of services are provided through different assignments of the AC-specific EDCA parameters; AIFS, CW, and TXOP limits.

The EDCA function must sense the channel to be idle for a complete AIFS before it can start the transmission or the backoff countdown. $AIFSN$ is the AC-specific AIFS number, $SIFS$ is the length of the Short Interframe Space and T_{slot} is the duration of a time slot. Then, $AIFS = SIFS + AIFSN \cdot T_{slot}$ [2]. The higher priority ACs are assigned smaller AIFSN. Therefore, the higher priority ACs can either transmit or decrement their backoff counters while lower priority ACs are still waiting in AIFS. This results in higher priority ACs enjoying a lower average probability of collision and relatively faster progress through backoff slots.

The initial value of AC-specific CW is CW_{min}. At every retransmission the CW range is doubled, up to CW_{max}. The higher priority ACs may select backoff values from a comparably smaller CW range. This prioritizes the access since a smaller CW value means a smaller backoff delay before the transmission.

Upon gaining the access to the medium, each AC may carry out multiple frame exchange sequences as long as the total access duration does not go over the AC-specific TXOP limit. Within a TXOP, the transmissions are separated by SIFS. Multiple frame transmissions in a TXOP can reduce the contention overhead. A TXOP limit of zero corresponds to only one frame exchange per access.

An internal (virtual) collision within a station is handled by granting the access to the highest priority AC. Then, the lower priority ACs run the collision procedure as if an outside collision has occured [2].

III. Related Work

In this section, we provide a brief summary of the theoretical DCF and EDCA performance analysis in the literature.

The majority of previous work carries out the performance analysis for asymptotical conditions assuming saturation. Three major saturation performance models have been proposed for DCF; *i)* assuming constant collision probability for each station, Bianchi [3] developed a simple DTMC and the saturation throughput is obtained by applying regenerative analysis to a generic slot time, *ii)* Cali *et al.* [4] employed renewal theory to analyze a p-persistent variant of DCF with persistence factor p derived from the CW, and *iii)* Tay *et al.* [5] instead used an average value mathematical method to model the DCF backoff procedure and to calculate the average number of interruptions that the backoff timer experiences. Having the common assumption of slot homogeneity, these models define all different renewal cycles all of which lead to accurate saturation performance analysis. Similarly, Medepalli *et al.* [6] provided explicit expressions for average DCF cycle time. Pointing out another direction for future performance studies, Hui *et al.* [7] proposed using metamodeling techniques in order to find approximate closed-form mathematical models.

These major methods are modified by several researchers to include the extra features of the EDCA function in the saturation analysis. Xiao [8] and Kong *et al.* [9] extended [3] to analyze the CW and the AIFS differentiation, respectively. These models miss the correct treatment of varying collision probabilities at different AIFS slots due to varying number of contending stations. Many other models considered this issue employing different approaches [10]–[16]. Robinson *et al.* [10] proposed an averaging analysis on the collision probability for different contention zones during AIFS. Hui *et al.* [11] unified several major approaches into one approximate average model. Zhu *et al.* [12] proposed averaging the transition probabilities of the AC-specific Markov Chain based on the number and the parameters of high priority flows. Inan *et al.* [13], Tao *et al.* [14], and Zhao *et al.* [15] proposed 3-dimensional DTMCs which provide accurate treatment of AIFS and CW differentiation. Banchs *et al.* [16] used the notion of *k-slot time*. Extending [5], Chen *et al.* [17], Kuo *et al.* [18], and Lin *et al.* [19] studied mean value analysis to include AIFS and CW differentiation.

A few studies also investigated the effect of EDCA TXOPs on 802.11e performance for a saturated scenario [20]–[23]. The efficiency of burst transmissions with block acknowledgements is studied in [20], [21]. Tinnirello *et al.* [22] proposed different TXOP managing policies for temporal fairness provisioning. Peng *et al.* [23] proposed an analytical model to study the effect of burst transmissions and showed that improved service differentiation can be achieved using a novel scheme based on TXOP thresholds.

Markov analysis for nonsaturation also assumes slot homogeneity and the proposed models mainly extend [3]. For the nonsaturation analysis of DCF, Duffy *et al.* [24] and Alizadeh-Shabdiz *et al.* [25] proposed similar DTMCs, which assume a MAC layer buffer size of one packet. Our analysis shows that this assumption may lead to significant performance prediction errors. Cantieni *et al.* [26] assumed infinitely large station buffers and the MAC queue being empty with constant probability regardless of the backoff stage the previous transmission took place. Li *et al.* [27] proposed an approximate model for nonsaturation where only CW differentiation is considered.

A number of models employing queueing theory have also been developed for 802.11(e) performance analysis in nonsaturation. Tickoo *et al.* [28], [29] modeled each 802.11 node as a discrete time G/G/1 queue to derive the service time distribution based on an assumption that the saturated setting provides good approximation for certain quantities in nonsaturation. Chen *et al.* [30] employed both G/M/1 and G/G/1 queueing models on top of [8] which only considers CW differentiation. Lee *et al.* [31] and Engelstad *et al.* [32] analyzed the use of M/G/1 queueing model, while employing different nonsaturation Markov models to calculate necessary quantities. Medepalli *et al.* [33] built upon the average cycle time derivation [6] to obtain individual queue delays using both M/G/1 and G/G/1 queueing models. Foh *et al.* [34] proposed a Markov framework to analyze the performance of DCF under statistical traffic. This framework models the number of contending nodes as an M/E_j/1/k queue. Tantra *et al.* [35] extended [34] to include service differentiation in EDCA. However, such analysis is only valid for a restricted scenario where all nodes have a MAC queue size of one packet. Zhai *et al.* [36], [37] employed M/G/1/K and M/G/1 queueing models, respectively, to find the distribution of the DCF MAC layer service time.

A thorough literature survey shows that an EDCA analytical model which incorporates *all* of the main EDCA QoS parameters, CW, AIFS, and TXOP, for *any* traffic load has not been designed yet.

IV. EDCA Discrete-Time Markov Chain Model

Assuming slot homogeneity, we propose a novel DTMC to model the behavior of the EDCA function of any AC at any load. The main contribution of this work is that the proposed model considers the effect of all EDCA QoS parameters (CW, AIFS, and TXOP) on the performance for the whole traffic load range from a lightly loaded nonsaturated channel to a heavily congested saturated medium. Although we assume constant probability of packet arrival per state (for the sake of simplicity, Poisson arrivals), we show that the model provides accurate performance analysis for a range of traffic types.

The state of the EDCA function of any AC at an arbitrary time t depends on several MAC layer events that may have occured before t. We model the MAC layer state of an AC$_i$, $0 \leq i \leq 3$, with a 3-dimensional Markov process, $(s_i(t), b_i(t), q_i(t))$. The stochastic process $s_i(t)$ represents the value of the backoff stage at time t, i.e., the number of retransmissions that the packet to be transmitted currently has experienced until time t. The stochastic process $b_i(t)$ represents the state of the backoff counter at time t. Up to this point, the definition of the first two dimensions follows [3] which is introduced for DCF. In order to enable the accurate nonsaturated analysis considering EDCA TXOPs, we introduce another dimension which models the stochastic process $q_i(t)$ denoting the number of packets buffered for transmission at the MAC layer. Moreover, as the details will be described in the sequel, in our model, $b_i(t)$ does not only represent the value of the backoff counter, but also the number of transmissions carried out during the current EDCA TXOP (when the value of backoff counter is actually zero).

Using the assumption of independent and constant collision probability at an arbitrary backoff slot, the 3-dimensional process $(s_i(t), b_i(t), q_i(t))$ is represented as a Discrete-Time Markov Chain (DTMC) with states (j, k, l) and index i. We define the limits on state variables as $0 \leq j \leq r_i - 1$, $-N_i \leq k \leq W_{i,j}$ and $0 \leq l \leq QS_i$. In these inequalities, we let r_i be the retransmission limit of a packet of AC$_i$; N_i be the maximum number of successful packet exchange sequences of AC$_i$ that can fit into one TXOP$_i$; $W_{i,j} = 2^{min(j,m_i)}(CW_{i,min} + 1) - 1$ be the CW size of AC$_i$ at the backoff stage j where $CW_{i,max} = 2^{m_i}(CW_{i,min} + 1) - 1$, $0 \leq m_i < r_i$; and QS_i be the maximum number of packets that can buffered at the MAC layer, i.e., MAC queue size. Moreover, it is important to note that a couple of restrictions apply to the state indices.

- When there are not any buffered packets at the AC queue, the EDCA function of the corresponding AC cannot be in a retransmitting state. Therefore, if $l = 0$, then $j = 0$ should hold. Such backoff states represent the postbackoff process [1], [2], therefore called *postbackoff slots* in the sequel. The postbackoff procedure ensures that the transmitting station waits at least another backoff between successive TXOPs. Note that, when $l > 0$ and $k \geq 0$, these states are named *backoff slots*.
- The states with indices $-N_i \leq k \leq -1$ represent the negation of the number of packets that are

successfully transmitted at the current TXOP rather than the value of the backoff counter (which is zero during a TXOP). For simplicity, in the design of the Markov chain, we introduced such states in the second dimension. Therefore, if $-N_i \leq k \leq -1$, we set $j = 0$. As it will be clear in the sequel, the addition of these states enables EDCA TXOP analysis.

We make following definitions for the analysis. Let p_{c_i} denote the average conditional probability that a packet from AC$_i$ experiences either an external or an internal collision after the EDCA function decides on the transmission. Let $p_{nt}(l', T|l)$ be the probability that there are l' packets in the MAC buffer at time $t+T$ given that there were l packets at t and no transmissions have been made during the interval T. Similarly, let $p_{st}(l', T|l)$ be the probability that there are l' packets in the MAC buffer at time $t+T$ given that there were l packets at time t and a successful transmission has been made during the interval T. We assume the packets arrive at the AC queue with size QS_i according to a Poisson process with rate λ_i packets per second. The exponential interarrival distributions are independent, so p_{nt} and p_{st} only depend on the interval length T and are independent of time t. Using the probability distribution function of the Poisson process, the probability of k arrivals occuring in time interval t can be calculated as

$$\Pr(\eta_{t,i} = k) = \frac{\exp^{-\lambda_i t} (\lambda_i t)^k}{k!}. \tag{1}$$

Then, $p_{nt}(l', T|l)$ and $p_{st}(l', T|l)$ can be calculated as follows. Note that the finite buffer space is considered throughout calculations since the number of packets that may arrive during T can be more than the available queue space.

$$p_{nt}(l', T|l) = \begin{cases} \Pr(\eta_{T,i} = l' - l), & \text{if } l' < QS_i \\ 1 - \displaystyle\sum_{l'=l}^{QS_i - 1} \Pr(\eta_{T,i} = l' - l), & \text{if } l' = QS_i. \end{cases} \tag{2}$$

$$p_{st}(l', T|l) = \begin{cases} \Pr(\eta_{T,i} = l' - l + 1), & \text{if } l' < QS_i \\ 1 - \displaystyle\sum_{l'=l-1}^{QS_i - 1} \Pr(\eta_{T,i} = l' - l + 1), & \text{if } l' = QS_i. \end{cases} \tag{3}$$

We first provide the formulation for state transition probabilities used in the construction of the proposed DTMC model. As will be described in Section IV-A, the state transition probabilities are functions of collision probabilities and state durations. In Section IV-B, we describe the method to find the steady-state solution numerically. The transmission probability is formulated using the steady-state solution of the DTMC (which is a function of collision probabilities and state durations). Conversely, collision probabilities and state durations are functions of transmission probabilities. The solution of the nonlinear system is of interest for performance analysis. We carry out normalized throughput and average delay analysis in Sections IV-C and IV-D, respectively.

A. State Transition Probabilities

The nonzero state transition probabilities of the proposed Markov model for AC$_i$ are denoted as $P_i(j', k', l'|j, k, l)$ and are calculated as follows.

1) The backoff counter is decremented at the slot boundary. We define the postbackoff or the backoff slot as the slot time is defined in [3]. Then, for $0 \leq j \leq r_i - 1$, $1 \leq k \leq W_{i,j}$, and $0 \leq l \leq l' \leq QS_i$,

$$P_i(j, k-1, l'|j, k, l) = p_{nt}(l', T_{i,bs}|l). \tag{4}$$

It is important to note that the evolution of the proposed DTMC is not real-time and the state duration varies depending on the state. The average duration of a backoff slot $T_{i,bs}$ is calculated by (25) which will be derived. Note that, in (4), we consider the probability of packet arrivals during $T_{i,bs}$ (buffer size l' after the state transition depends on this probability).

2) We assume the transmitted packet experiences a collision with constant probability p_{c_i} (slot homogeneity). If the first packet exchange sequence is successful, the next state is -1 in the second dimension (showing that a packet has been successfully transmitted in the current TXOP). On the other hand, if a collision occurs, the next state in the first dimension depends on whether the retry limit is reached or not. In either case, a new backoff value is selected (the state in the second dimension has a value larger than or equal to 0). Note that the cases when the retry limit is reached and when the MAC buffer is full are treated separately, since the transition probabilities should follow different rules. Let $T_{i,s}$ and $T_{i,c}$ be the time spent in a successful transmission and a collision by AC$_i$, respectively, which will be derived. Then, for $0 \leq j \leq r_i - 1$, $1 \leq l \leq QS_i - 1$, and $l - 1 \leq l' \leq QS_i$,

$$P_i(0, -1, l'|j, 0, l) = (1 - p_{c_i}) \cdot p_{st}(l', T_{i,s}|l) \tag{5}$$

$$P_i(0, -1, QS_i - 1|j, 0, QS_i) = 1 - p_{c_i}. \tag{6}$$

For $0 \leq j \leq r_i - 2$, $0 \leq k \leq W_{i,j+1}$, and $0 \leq l \leq l' \leq QS_i$,

$$P_i(j+1, k, l'|j, 0, l) = \frac{p_{c_i} \cdot p_{nt}(l', T_{i,c}|l)}{W_{i,j+1} + 1}. \tag{7}$$

For $0 \leq k \leq W_{i,0}$, $0 \leq l \leq QS_i - 1$, and $\max(0, l-1) \leq l' \leq QS_i$,

$$P_i(0, k, l' | r_i - 1, 0, l) = \frac{p_{c_i}}{W_{i,0} + 1} \cdot p_{st}(l', T_{i,s} | l) \tag{8}$$

$$P_i(0, k, QS_i - 1 | r_i - 1, 0, QS_i) = \frac{p_{c_i}}{W_{i,0} + 1} \tag{9}$$

Note that we use p_{nt} in (7) although a transmission has been made. On the other hand, the packet has collided and is still at the MAC queue for retransmission as if no transmission has occured. This is not the case in (5) and (8), since in these transitions a successful transmission or a drop occurs. When the MAC buffer is full, any arriving packet is discarded as (6) and (9) imply.

3) Once the TXOP is started, the EDCA function may continue with as many packet SIFS-separated exchange sequences as it can fit into the TXOP duration. Let $T_{i,exc}$ be the average duration of a successful packet exchange sequence for AC$_i$ which will be derived in (23). Then, for $-N_i + 1 \leq k \leq -1$, $1 \leq l \leq QS_i$, and $\max(0, l-1) \leq l' \leq QS_i$,

$$P_i(0, k-1, l' | 0, k, l) = p_{st}(l', T_{i,exc} | l). \tag{10}$$

When the next transmission cannot fit into the remaining TXOP, the current TXOP is immediately concluded and the unused portion of the TXOP is returned. By design, this is implicitly included in the calculation of N_i in (24). Then, for $0 \leq k \leq W_{i,0}$ and $1 \leq l \leq QS_i$,

$$P_i(0, k, l | 0, -N_i, l) = \frac{1}{W_{i,0} + 1}. \tag{11}$$

The TXOP ends when the MAC queue is empty. Then, for $0 \leq k' \leq W_{i,0}$ and $-N_i \leq k \leq -1$,

$$P_i(0, k', 0 | 0, k, 0) = \frac{1}{W_{i,0} + 1}. \tag{12}$$

Note that no time passes in (11) and (12), so the definition of these states and transitions is actually not necessary for accuracy. On the other hand, they simplify the DTMC structure and symmetry.

4) If the queue is still empty when the postbackoff counter reaches zero, the EDCA function enters the idle state until another packet arrival. We make two assumptions; *i)* At most one packet may arrive during T_{slot} with constant probability ρ_i (considering the fact that T_{slot} is in the order of microseconds, the probability that multiple packets can arrive in this interval is very small), *ii)* if the channel is idle at the slot the packet arrives at an empty queue, the transmission will be successful at AIFS

completion without any backoff. The latter assumption is due to the following reason. While the probability of the channel becoming busy during AIFS or a collision occuring for the transmission at AIFS is very small at a lightly loaded scenario, the probability of a packet arrival to an empty queue is very small at a highly loaded scenario. As observed via simulations, these assumptions do not lead to any noticeable changes in the results while simplifying the Markov chain structure.

Note that p_{c_i} is equal to the probability of the channel being busy at an arbitrary slot when AC$_i$ is in idle state (0,0,0). If no packet arrives in the next slot, the EDCA function stays idle. This is considered in (13) by conditioning on the channel becoming busy or not in the next slot. The backoff counter is started if at least one packet arrives while the medium is busy. This case is considered in (14). If the channel is idle, but a packet arrives, an EDCA TXOP is started as denoted in (15).

$$P_i(0,0,0|0,0,0) = (1-p_{c_i}) \cdot (1-\rho_i) + p_{c_i} \cdot p_{nt}(0, T_{i,b}|0), \tag{13}$$

$$P_i(0,k,l|0,0,0) = \frac{p_{c_i}}{W_{i,0}+1} \cdot p_{nt}(l, T_{i,b}|0), \tag{14}$$

$$P_i(0,-1,l|0,0,0) = (1-p_{c_i}) \cdot \rho_i \cdot p_{nt}(l, T_{i,s}|0) \tag{15}$$

where $0 \leq k \leq W_{i,0}$ and $1 \leq l \leq QS_i$. Let $T_{i,b}$ in (13) and (14) be the length of a backoff slot given it is not idle. Note that actually a successful transmission occurs in the state transition in (15). On the other hand, the transmitted packet is not reflected in the initial queue size state which is 0. Therefore, p_{nt} is used instead of p_{st}. Note that $\rho_i = 1 - \Pr(\eta_{T_{slot},i} = 0)$.

Parts of the proposed DTMC model are illustrated in Fig. 1 for an arbitrary AC$_i$ with $N_i = 2$. Fig. 1(a) shows the state transitions for $l = 0$. Note that in Fig. 1(a) the states with $-N_i \leq k \leq -2$ can only be reached from the states with $l = 1$. Fig. 1(b) presents the state transitions for $0 < l < QS_i$ and $0 \leq j < r_i$. Note that only the transition probabilities and the states marked with rectangles differ when $j = r_i - 1$ (as in (8)). Due to space limitations, we do not include an extra figure for this case. Fig. 1(c) shows the state transitions when $l = QS_i$. Note also that the states marked with rectangles differ when $j = r_i - 1$ (as in (9)). The combination of these small chains for all j, k, l constitutes our DTMC model.

B. Steady-State Solution

Let $b_{i,j,k,l}$ be the steady-state probability of the state (j,k,l) of the proposed DTMC with index i which can be solved using (4)-(15) subject to $\sum_j \sum_k \sum_l b_{i,j,k,l} = 1$ (the proposed DTMC is ergodic and irreducible). Let τ_i be the probability that an AC$_i$ transmits at an arbitrary backoff or postbackoff slot

$$\tau_i = \frac{\left(\sum_{j=0}^{r_i-1}\sum_{l=1}^{QS_i} b_{i,j,0,l}\right) + b_{i,0,0,0}\cdot\rho_i\cdot(1-p_{c_i})}{\sum_{j=0}^{r_i-1}\sum_{k=0}^{W_{i,j}}\sum_{l=0}^{QS_i} b_{i,j,k,l}}. \tag{16}$$

Note that $-N_i \leq k \leq -1$ is not included in the normalization in (16), since these states represent a continuation in the EDCA TXOP rather than a contention for the access. The value of τ_i depends on the values of the average conditional collision probability p_{c_i} and the state durations $T_{i,bs}$, $T_{i,b}$, $T_{i,s}$ and $T_{i,c}$.

1) Average conditional collision probability p_{c_i}: The difference in AIFS of each AC in EDCA creates the so-called contention zones [10], [38]. In each contention zone, the number of contending stations may vary. In order to be consistent with the notation of [2], we assume $AIFS_0 \geq AIFS_1 \geq AIFS_2 \geq AIFS_3$. Let $d_i = AIFSN_i - AIFSN_3$. Also, let n^{th} backoff slot after the completion of $AIFS_3$ idle interval following a transmission period be in the contention zone x. Then, we define $x = \max\left(y \mid d_y = \max_z(d_z \mid d_z \leq n)\right)$. Therefore, x is assigned the largest index value within a set of ACs that have the largest AIFSN value which is smaller than or equal to $n + AIFSN_3$.

We can define $p_{c_{i,x}}$ as the conditional probability that AC$_i$ experiences either an external or an internal collision in contention zone x. Note $AIFS_x \geq AIFS_i$ should hold for AC$_i$ to transmit in x. Also, let the total number AC$_i$ flows be f_i. For the heterogeneous scenario in which each station has only one AC

$$p_{c_{i,x}} = 1 - \frac{\prod_{i':d_{i'}\leq d_x}(1-\tau_{i'})^{f_{i'}}}{(1-\tau_i)}. \tag{17}$$

In this paper, due to space limitations, we only investigate the situation when there is only one AC per station (no internal collisions can occur). We provide the extension of the proposed EDCA model for the case of larger number of ACs per station in [38].

We use the Markov chain shown in Fig. 2 to find the long term occupancy of contention zones. Each state represents the n^{th} backoff slot after completion of the AIFS$_3$ idle interval following a transmission period. The Markov chain model uses the fact that a backoff slot is reached if and only if no transmission occurs in the previous slot. Moreover, the number of states is limited by the maximum idle time between two successive transmissions which is $W_{min} = \min(CW_{i,max})$ for a saturated scenario. Although this is not the case for a nonsaturated scenario, we do not change this limit. As the comparison with simulation results show, this approximation does not result in significant prediction errors. The probability that at least one transmission occurs in a backoff slot in contention zone x is

$$p_x^{tr} = 1 - \prod_{i':d_{i'} \leq d_x} (1 - \tau_{i'})^{f_{i'}}. \tag{18}$$

Given the state transition probabilities as in Fig. 2, the long term occupancy of the backoff slots b'_n can be obtained from the steady-state solution of the Markov chain. Then, the AC-specific average collision probability p_{c_i} is found by weighing zone specific collision probabilities $p_{c_{i,x}}$ according to the long term occupancy of contention zones (thus backoff slots)

$$p_{c_i} = \frac{\sum_{n=d_i+1}^{W_{min}} p_{c_{i,x}} \cdot b'_n}{\sum_{n=d_i+1}^{W_{min}} b'_n} \tag{19}$$

where x is calculated depending on the value of n as stated previously.

Note that the average collision probability calculation in [10, Section IV-D] is a special case of our calculation for two ACs.

2) The state duration $T_{i,s}$ and $T_{i,c}$: Let $T_{i,p}$ be the average payload transmission time for AC$_i$ ($T_{i,p}$ includes the transmission time of MAC and PHY headers), δ be the propagation delay, T_{ack} be the time required for acknowledgment packet (ACK) transmission. Then, for the basic access scheme, we define the time spent in a successful transmission $T_{i,s}$ and a collision $T_{i,c}$ for any AC$_i$ as

$$T_{i,s} = T_{i,p} + \delta + SIFS + T_{ack} + \delta + AIFS_i \tag{20}$$

$$T_{i,c} = T_{i,p^*} + ACK\_Timeout + AIFS_i \tag{21}$$

where T_{i,p^*} is the average transmission time of the longest packet payload involved in a collision [3]. For simplicity, we assume the packet size to be equal for any AC, then $T_{i,p^*} = T_{i,p}$. Being not explicitly specified in the standards, we set $ACK\_Timeout$, using Extended Inter Frame Space (EIFS) as $EIFS_i - AIFS_i$. Note that the extensions of (20) and (21) for the RTS/CTS scheme are straightforward [38].

3) The state duration $T_{i,bs}$ and $T_{i,b}$: The average time between successive backoff counter decrements is denoted by $T_{i,bs}$. The backoff counter decrement may be at the slot boundary of an idle backoff slot or the last slot of AIFS following an EDCA TXOP or a collision period. We start by calculating the average duration of an EDCA TXOP for AC$_i$ $T_{i,txop}$ as

$$T_{i,txop} = \frac{\sum_{l=0}^{QS_i} b_{i,0,-N_i,l} \cdot ((N_i - 1) \cdot T_{i,exc} + T_{i,s}) + \sum_{k=-N_i+1}^{-1} b_{i,0,k,0} \cdot ((-k-1) \cdot T_{i,exc} + T_{i,s})}{\sum_{k=-N_i+1}^{-1} b_{i,0,k,0} + \sum_{l=0}^{QS_i} b_{i,0,-N_i,l}} \tag{22}$$

where $T_{i,exc}$ is defined as the duration of a successful packet exchange sequence within a TXOP. Since the packet exchanges within a TXOP are separated by SIFS rather than AIFS,

$$T_{i,exc} = T_{i,s} - AIFS_i + SIFS, \tag{23}$$

$$N_i = \max(1, \lfloor (TXOP_i + SIFS)/T_{i,exc} \rfloor). \tag{24}$$

Given τ_i and f_i, simple probability theory can be used to calculate the conditional probability of no transmission ($p_{x,i}^{idle}$), only one transmission from AC$_{i'}$ ($p_{x,i}^{suc_{i'}}$), or at least two transmissions ($p_{x,i}^{col}$) at the contention zone x given one AC$_i$ is in backoff [38]. Moreover, let x_i be the first contention zone in which AC$_i$ can transmit. Then,

$$T_{i,bs} = \frac{1}{1 - \sum\limits_{x_i < x' \leq 3} p_{z_{x'}}} \sum_{\forall x'} (p_{x',i}^{idle} \cdot T_{slot} + p_{x',i}^{col} \cdot T_{i,c} + \sum_{\forall i'} p_{x',i}^{suc_{i'}} \cdot T_{i',txop}) \cdot p_{z_{x'}} \tag{25}$$

where p_{z_x} denotes the stationary distribution for a random backoff slot being in zone x (define $d_{-1} = W_{min}$)

$$p_{z_x} = \sum_{n=d_x+1}^{\min\limits_{\forall x' \in [-1,3]} (d_{x'} | d_{x'} > d_x)} b'_n. \tag{26}$$

Note that, in (25), the fractional term before summation accounts for the busy periods experienced before AIFS$_i$ is completed.

The expected duration of a backoff slot given it is busy and one AC$_i$ is in idle state is calculated as

$$T_{i,b} = \sum_{\forall x'} \left(\frac{p_{x',i}^{col}}{1 - p_{x',i}^{idle}} \cdot T_{i,c} + \sum_{\forall i'} \frac{p_{x',i}^{suc_{i'}}}{1 - p_{x',i}^{idle}} \cdot T_{i',txop} \right) \cdot p_{z_{x'}}. \tag{27}$$

Together with the steady-state transition probabilities, (16)-(27) represent a nonlinear system which can be solved using numerical methods. We initialize the vector, τ, arbitrarily. Using (17)-(27), we calculate the parameters that are used in the definition of DTMC transition probabilities (in the first iteration, we skip (22), and use AC-specific TXOP limits instead). Once the DTMC transitions are defined numerically, we calculate the steady-state probabilities using simple probability theory. Next, τ is recalculated via (16) using the new steady-state solution. If the initial guess on τ is not equal to the result, a new estimate on τ is calculated by a weighted summation on the old and the new value, and the procedure is repeated. The iterations stop once the old and the new values converge to a solution. Note that when we compare the convergence time (such as in terms of the number of iterations) of the numerical solution technique we have used, the results for the proposed model are in a similar range with the previous models. The models we have implemented and compared are [3], [6], [9], [11], [24], [35], [39].

C. Normalized Throughput Analysis

The normalized throughput of a given AC$_i$, S_i, is defined as the fraction of the time occupied by the successfully transmitted information. Then,

$$S_i = \frac{p_{s_i}\zeta_{i,txop}T_{i,p}}{p_I T_{slot} + \sum_{i'} p_{s_{i'}} T_{i',txop} + (1 - p_I - \sum_{i'} p_{s_{i'}})T_c} \tag{28}$$

p_I is the probability of the channel being idle at a backoff slot, p_{s_i} is the conditional successful transmission probability of AC$_i$ at a backoff slot, and $\zeta_{i,txop} = (T_{i,txop} - AIFS_i + SIFS)/T_{i,exc}$. Note that, we consider $\zeta_{i,txop}$ and $T_{i,txop}$ in (28) to define the generic slot time and the time occupied by the successfully transmitted information in the case of EDCA TXOPs.

The probability of a slot being idle, p_I, depends on the state of previous slots. For example, conditioned on the previous slot to be busy ($p_B = 1 - p_I$), p_I only depends on the transmission probability of the ACs with the smallest AIFS, since others have to wait extra AIFS slots [13], [39], [40]. Generalizing this to all AIFS slots, p_I can be calculated as

$$p_I = \sum_{n=0}^{W_{min}} \gamma_n p_B (p_I)^n \cong \sum_{n=0}^{d_0-1} \gamma_n p_B p_I^n + \gamma_{d_0} p_I^{d_0} \tag{29}$$

where γ_n denotes the probability of no transmission occuring at the $(n+1)^{th}$ AIFS slot after $AIFS_3$. Substituting $\gamma_n = \gamma_{d_0}$ for $n \geq d_0$, and releasing the condition on the upper limit of summation, W_{min}, to ∞, p_I can be approximated as in (29). According to the simulation results, this approximation works well. Note that $\gamma_n = 1 - p_x^{tr}$ where $x = \max\Big(y \mid d_y = \max_z(d_z \mid d_z \leq n)\Big)$.

The probability of successful transmission p_{s_i} is conditioned on the states of the previous slots as well. This is again because the number of stations that can contend at an arbitrary backoff slot differs depending on the number of previous consecutive idle backoff slots. Therefore, for the heterogeneous case, in which each station only has one AC, p_{s_i} can be calculated as

$$p_{s_i} = \frac{N_i \tau_i}{(1-\tau_i)} \left(\sum_{n=d_i+1}^{d_0} \left(p_B p_I^{(n-1)} \prod_{i':0 \leq d_{i'} \leq (n-1)} (1-\tau_{i'})^{f_{i'}} \right) + (p_I)^{d_0} \prod_{\forall i'} (1-\tau_{i'})^{f_{i'}} \right). \tag{30}$$

D. Average Delay Analysis

Our goal is to find total average delay $E[D_i]$ which is defined as the average time from when a packet enters the MAC layer queue of AC$_i$ until it is successfully transmitted. D_i has two components; *i)* queueing time Q_i and *ii)* access time A_i. Q_i is the period that a packet waits in the queue for other packets in front to be transmitted. A_i is the period a packet waits at the head of the queue until it is transmitted successfully (backoff and transmission period). We carry out a recursive calculation to find $E[A_i]$. Then, another recursive calculation is carried out to find $E[D_i]$ by employing $E[A_i]$ in the recursion.

Let $A_i(j,k)$ denote the time delay from the current state (j,k,l) until the packet at the head of the AC$_i$ queue is transmitted successfully ($l \geq 1$). The initial condition on the recursive calculation is

$$A_i(r_i - 1, 0) = T_{i,s}. \tag{31}$$

Recursive delay calculations for $0 \leq j \leq r_i - 1$ are

$$A_i(j,k) = \begin{cases} A_i(j,k-1) + T_{i,bs}, & \text{if } 1 \leq k \leq W_{i,j} \\ (1-p_{c_i})T_{i,s} + p_{c_i}\left(\frac{\sum_{k'=0}^{W_{i,j+1}} A_i(j+1,k')}{W_{i,j+1}+1} + T_{i,c}\right), & \text{if } k = 0 \text{ and } j \neq r_i - 1. \end{cases} \tag{32}$$

Then,

$$E[A_i] = \frac{\sum_{k=0}^{W_{i,0}} A_i(0,k)}{W_{i,0}+1}. \tag{33}$$

Conversely, let $A_{i,d}(j,k)$ be the conditional access delay a packet experiences from the current state (j,k,l) until it is dropped due to the fact that the retry limit is reached. $A_{i,d}(j,k)$ can easily be calculated by modifying the recursive method of calculating $A_i(j,k)$ as we show in [38]. $E[A_{i,d}]$ is defined using (33) and replacing $A_i(0,k)$ with $A_{i,d}(0,k)$.

We perform another recursive calculation to calculate the total delay a packet experiences $D_i(j,k,l)$ (given that the packet arrives while the EDCA function is at state (j,k,l)). In the calculations, we account for the remaining access delay for the packet at the head of the MAC queue and the probability that this packet may be dropped due to the retry limit. We consider four different cases.

1) If the packet arrives during the idle state, the total delay is

$$D_i(0,0,0) = T_{i,s} \cdot (1-p_{c_i}) + (E[A_i] + T_{i,b}) \cdot p_{c_i} \cdot (1 - p_{i,0,drop}). \tag{34}$$

where $p_{i,j,drop} = (p_{c_i})^{r_i - j}$ denotes the packet loss probability conditioned on the fact that backoff has reached stage j. Note that (34) follows the assumptions and the reasoning in (13)-(15). The delay the dropped packets experience cannot be considered in a total delay calculation. Therefore, this case is excluded in the second term of (34) by multiplication with $1 - p_{i,0,drop}$ (as well as in (35) and (37)).

2) The total delay of an arrival during postbackoff is equal to the access delay. Then, for $0 \leq k \leq W_{i,j}$,

$$D_i(0,k,0) = A_i(0,k) \cdot (1 - p_{i,0,drop}). \tag{35}$$

3) If a packet arrives during the backoff of another packet, it is delayed at least for the remaining access time. Depending on the queue size, it may be transmitted at the current TXOP, or may be delayed till further accesses are gained. Then, for $0 \leq j \leq r_i - 1$, $0 \leq k \leq W_{i,j}$, and $1 \leq l \leq QS_i$,

$$\begin{aligned} D_i(j,k,l) =& (1 - p_{i,j,drop}) \cdot (A_i(j,k) + \min(N_i - 1, l - 1) \cdot T_{i,exc} \\ & + D_i(-1,-1,l - N_i)) + p_{i,j,drop} \cdot (A_{i,d}(j,k) + D_i(-1,-1,l-1)) \,. \end{aligned} \tag{36}$$

$D_i(-1,-1,l)$ shows the remaining total delay if the packet cannot be transmitted at the current TXOP the station gets. $D_i(-1,-1,l)$ is calculated recursively according to the value of l

$$D_i(-1,-1,l) = \begin{cases} 0, & \text{if } l \leq 0 \\ E[A_i] \cdot (1 - p_{i,0,drop}), & \text{if } l = 1 \\ \chi, & \text{if } l > 1 \end{cases} \tag{37}$$

$$\begin{aligned} \chi =& (1 - p_{i,0,drop}) \cdot (E[A_i] + \min(N_i - 1, l - 1) \cdot T_{i,exc} \\ & + D_i(-1,-1,l - N_i)) + p_{i,0,drop} \cdot (E[A_{i,d}] + D_i(-1,-1,l-1)) \,. \end{aligned} \tag{38}$$

4) If the packet arrives during a TXOP, it may be transmitted at the current TXOP, or it may wait for further accesses. Then, for $-N_i + 1 \leq k \leq -1$ and $1 \leq l \leq QS_i$,

$$D_i(j,k,l) = \min(k-1, l) \cdot T_{i,exc} + D_i(-1,-1,l-k+1). \tag{39}$$

Let the probability of any arriving packet seeing the EDCA function at state (j,k,l) be $\bar{b}_{i,j,k,l}$. Since we assume independent and exponentially distributed packet interarrivals, $\bar{b}_{i,j,k,l}$ can simply be calculated by normalizing $b_{i,j,k,l}$ excluding the states in which no time passes, i.e., $\forall (j,k,l)$ such that $(0, -N_i, 1 \leq l \leq QS_i)$ or $(0, -N_i \leq k \leq -1, 0)$. Note that $\bar{b}_{i,j,k,l} = 0$ for these states.

$$\bar{b}_{i,j,k,l} = \frac{b_{i,j,k,l}}{1 - \sum_{l=1}^{QS_i} b_{i,0,-N_i,l} - \sum_{k=-N_i}^{-1} b_{i,0,k,0}}. \tag{40}$$

Then, the total average delay a successful packet experiences $E[D_i]$ can be calculated by averaging

$D_i(j,k,l)$ over all possible states

$$E[D_i] = \sum_{\forall(j,k,l)} D_i(j,k,l) \cdot \bar{b}_{i,j,k,l}. \tag{41}$$

V. Numerical and Simulation Results

We validate the accuracy of the numerical results calculated via the proposed EDCA model by comparing them with the simulations results obtained from ns-2 For the simulations, we employ the IEEE 802.11e HCF MAC simulation model for ns-2.28 that we developed [41].

We consider ACs that transmit fixed-size User Datagram Protocol (UDP) packets. In simulations, we consider two ACs, one high priority and one low priority. Each station runs only one AC. Unless otherwise stated, the packets are generated according to a Poisson process with equal rate for both ACs. We set $AIFSN_1 = 3$, $AIFSN_3 = 2$, $CW_{1,min} = 15$, $CW_{3,min} = 7$, $m_1 = m_3 = 3$, $r_1 = r_3 = 7$. For both ACs, the payload size is 1034 bytes. The packet size is selected arbitrarily. The accuracy of the comparison of simulation and analytical results does not depend on the specific packet size. The simulation results are reported for the wireless channel which is assumed to be not prone to any errors during transmission. The errored channel case is left for future study. All the stations have 802.11g Physical Layer (PHY) using 54 Mbps and 6 Mbps as the data and basic rate respectively ($T_{slot} = 9\ \mu s$, $SIFS = 10\ \mu s$).

Fig. 3 shows the differentiation of throughput for two ACs when EDCA TXOP limits of both are set to 0 (1 packet exchange per EDCA TXOP). In this scenario, there are 5 stations for both ACs and they are transmitting to an AP. The normalized throughput per AC as well as the total system throughput are plotted for increasing offered load per AC. We have carried out the analysis for maximum MAC buffer sizes of 2 packets and 10 packets. The comparison between analytical and simulation results shows that our model can accurately capture the linear relationship between throughput and offered load under low loads, the complex transition in throughput between under-loaded and saturation regimes, and the saturation throughput. Although we do not present here, considerable inaccuracy is observed if the postbackoff procedure, varying collision probability among different AIFS zones, and varying service time among different backoff stages are not modeled correctly as proposed [38]. The results also present that the slot homogeneity assumption works accurately in a nonsaturated model for throughput estimation.

The proposed model can also capture the throughput variation with respect to the size of the MAC buffer. The results reveal how significantly the size of the MAC buffer affects the throughput in the transition period from underloaded to highly loaded channel. This also shows small interface buffer assumptions of [24], [25], and [35] can lead to considerable analytical inaccuracies. Although the total throughput for the small buffer size case has higher throughput in the transition region for the specific example, this cannot be generalized. The reason for this is that AC_1 suffers from low throughput when buffer sizes are larger (which affects the total throughput). When AC_3 has the opportunity to buffer more packets, it may access the medium more frequently depending on EDCA parameter settings (less idle time for AC_3), which decreases the chance for AC_1 transmissions. As will be presented, this behavior is not observed when TXOPs are set as in the scenario of Fig. 4 where AC_1 can send multiple packets in one access.

It is also important to note that the throughput performance does not differ significantly (around 1%-2%) for buffer sizes larger than 10 packets for the given scenarios. Therefore, we do not include such cases in order not to complicate the figures.

Fig. 4 depicts the differentiation of throughput for two ACs when EDCA TXOP limits are set to 1.504 ms and 3.008 ms for AC_3 and AC_1, respectively. For TXOP limits, we use the suggested values for voice and video ACs in [2]. It is important to note that the model works for an arbitrary selection of the TXOP limit. According to the selected TXOP limits, $N_1 = 11$ and $N_3 = 5$. The normalized throughput per AC as well as the total system throughput are plotted while increasing offered load per AC. We have carried out the analysis for maximum MAC buffer sizes of 2 packets and 10 packets. The model accurately captures the throughput for any traffic load. As expected, increasing maximum buffer size to 10 packets increases the throughput both in the transition and the saturation region. Note that when more than a packet fit into EDCA TXOPs, this decreases contention overhead which in turn increases channel utilization and throughput (comparison of Fig. 4 with Fig. 3). Although corresponding results are not presented here, the model works accurately for higher queue sizes in the case of EDCA TXOPs as well.

Fig. 5 displays the differentiation of throughput for two ACs when packet arrival rate is fixed to 2 Mbps and the station number per AC is increased. We have performed the analysis for the MAC buffer size of 10 packets with EDCA TXOPs enabled. The analytical and simulation results are well in accordance. As the traffic load increases, the differentiation in throughput between the ACs is observed.

Fig. 6 shows the normalized throughput for two ACs when offered load per AC is not equal. In this scenario, we set the packet arrival rate per AC_1 to 2 Mbps and the packet arrival rate per AC_3 to 0.5 Mbps. The analytical and simulation results are well in accordance. As the traffic load increases, AC_3 maintains linear increase with respect to offered load, while AC_1 experiences decrease in throughput due to larger settings of AIFS and CW if the total number of stations exceeds 22.

Note that Fig. 5 and Fig. 6 display the performance mainly in the vicinity of the transition region between under-loaded and saturation regimes. Analytical and simulation results may present marginal differences in this region. Since this region is presented in a smaller scale in Fig. 4 and Fig. 3 (in order to include under-loaded and saturation regions), the marginal differences between analytical and simulation results are not evident although they still exist.

In the design of the model, we assume constant packet arrival probability per state. The Poisson arrival process fits this definition because of the independent exponentially distributed interarrival times. We have also compared the throughput estimates obtained from the analytical model with the simulation results obtained using an On/Off traffic model in Fig. 7. A similar study has first been made for DCF in [24]. We modeled the high priority with On/Off traffic model with exponentially distributed idle and active intervals of mean length 1.5 s. In the active interval, packets are generated with Constant Bit Rate (CBR). The low priority traffic uses Poisson distributed arrivals. Note that we leave the packet size unchanged, but normalize the packet arrival rate according to the On/Off pattern so that total offered load remains constant to have a fair comparison. The analytical predictions closely follow the simulation results for the given scenario. We have observed that the predictions are more sensitive if the transition region is entered with a few number of stations (5 stations per AC). Due to space limitations, we can provide the results of a similar experiment for CBR traffic in [38].

Fig. 8 depicts the total average packet delay with respect to increasing traffic load per AC. We present the results for two different scenarios. In the first scenario, TXOP limits are set to 0 ms for both ACs. In the second scenario, TXOP limits are set to 1.504 ms and 3.008 ms for high and low priority ACs respectively. The analysis is carried out for a buffer size of 10 packets. As the results imply, the analytical results closely follow the simulation results for both scenarios. In the lightly loaded region, the delays are considerably small. The increase in the transition region is steeper when TXOP limits are 0. In the specific example, enabling TXOPs decreases the total delay where the decrease is more considerable for the low priority AC (due to selection of parameters). Since the buffer size is limited, the total average delay converges to a specific value as the load increases. Still this limit is not of interest, since the packet loss rate at this region is unpractically large. Note that this limit will be higher for larger buffers. The region of interest is the start of the transition region (between 2 Mbps and 3 Mbps for the example in Fig. 8). We also display other data points to show the performance of the model for the whole load span.

The model is also effective in predicting many other performance parameters such as packet loss rate, queue size distribution, etc. It captures the effect of varying AIFS and CW. Due to space limitations, we do not include these results in this paper. They are reported in a technical report [38].

VI. Conclusion

We have presented an accurate Markov model for analytically calculating the EDCA throughput and delay for the whole traffic load range. The presented model shows the homogeneous slot assumption (constant collision and transmission probability at an arbitrary backoff slot) holds in a variety of nonsaturation scenarios. The presented model accurately captures the linear relationship between throughput and offered load under low loads and the limiting behavior of throughput at saturation.

The key contribution of this paper is that the model accounts for all of the main EDCA differentiation mechanisms. The analytical model can incorporate any selection of AC-specific AIFS, CW, and TXOP values for any number of ACs. The model also considers varying collision probabilities at different contention zones which provides accurate AIFS differentiation analysis. Although not presented explicitly in this paper, the presented model can easily be extended for scenarios where the stations run multiple ACs (virtual collisions may take place) or RTS/CTS protection mechanism is used.

We also show that the MAC buffer size affects the EDCA performance significantly between underloaded and saturation regimes (including saturation) especially when EDCA TXOPs are enabled. The presented model captures this complex transition accurately. This analysis also points out the fact that including an accurate queue treatment is vital. Incorporating MAC queue states also enables EDCA TXOP analysis so that the EDCA TXOP continuation process is modeled in considerable detail. To the authors' knowledge this is the first demonstration of an analytic model including EDCA TXOP procedure for finite load.

Our model can easily be simplified to model DCF behavior. Moreover, after modifying our model accordingly, the analysis for the infrastructure WLAN can be performed (note that in a WLAN downlink traffic load may significantly differ from uplink traffic load).

Although the Markov analysis assumes the packets are generated according to Poisson process, the comparison with simulation results shows that the throughput analysis is valid for a range of traffic types such as CBR and On/Off traffic (On/Off traffic model is a widely used model for voice and telnet traffic).

VII. Acknowledgment

The authors would like to thank the anonymous reviewers whose comments improved the quality of the paper.

REFERENCES

[1] *IEEE Standard 802.11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications*, IEEE 802.11 Std., 1999.

[2] *IEEE Standard 802.11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications: Medium access control (MAC) Quality of Service (QoS) Enhancements*, IEEE 802.11e Std., 2005.

[3] G. Bianchi, "Performance Analysis of the IEEE 802.11 Distributed Coordination Function," *IEEE Trans. Commun.*, pp. 535–547, March 2000.

[4] F. Cali, M. Conti, and E. Gregori, "Dynamic Tuning of the IEEE 802.11 Protocol to Achieve a Theoretical Throughput Limit," *IEEE/ACM Trans. Netw.*, pp. 785–799, December 2000.

[5] J. C. Tay and K. C. Chua, "A Capacity Analysis for the IEEE 802.11 MAC Protocol," *Wireless Netw.*, pp. 159–171, July 2001.

[6] K. Medepalli and F. A. Tobagi, "Throughput Analysis of IEEE 802.11 Wireless LANs using an Average Cycle Time Approach," in *Proc. IEEE Globecom '05*, November 2005.

[7] J. Hui and M. Devetsikiotis, "Metamodeling of Wi-Fi Performance," in *Proc. IEEE ICC '06*, June 2006.

[8] Y. Xiao, "Performance Analysis of Priority Schemes for IEEE 802.11 and IEEE 802.11e Wireless LANs," *IEEE Trans. Wireless Commun.*, pp. 1506–1515, July 2005.

[9] Z. Kong, D. H. K. Tsang, B. Bensaou, and D. Gao, "Performance Analysis of the IEEE 802.11e Contention-Based Channel Access," *IEEE J. Select. Areas Commun.*, pp. 2095–2106, December 2004.

[10] J. W. Robinson and T. S. Randhawa, "Saturation Throughput Analysis of IEEE 802.11e Enhanced Distributed Coordination Function," *IEEE J. Select. Areas Commun.*, pp. 917–928, June 2004.

[11] J. Hui and M. Devetsikiotis, "A Unified Model for the Performance Analysis of IEEE 802.11e EDCA," *IEEE Trans. Commun.*, pp. 1498–1510, September 2005.

[12] H. Zhu and I. Chlamtac, "Performance Analysis for IEEE 802.11e EDCF Service Differentiation," *IEEE Trans. Wireless Commun.*, pp. 1779–1788, July 2005.

[13] I. Inan, F. Keceli, and E. Ayanoglu, "Saturation Throughput Analysis of the 802.11e Enhanced Distributed Channel Access Function," in *Proc. IEEE ICC '07*, June 2007.

[14] Z. Tao and S. Panwar, "Throughput and Delay Analysis for the IEEE 802.11e Enhanced Distributed Channel Access," *IEEE Trans. Commun.*, pp. 596–602, April 2006.

[15] J. Zhao, Z. Guo, Q. Zhang, and W. Zhu, "Performance Study of MAC for Service Differentiation in IEEE 802.11," in *Proc. IEEE Globecom '02*, November 2002.

[16] A. Banchs and L. Vollero, "Throughput Analysis and Optimal Configuration of IEEE 802.11e EDCA," *Comp. Netw.*, pp. 1749–1768, August 2006.

[17] Y. Chen, Q.-A. Zeng, and D. P. Agrawal, "Performance Analysis of IEEE 802.11e Enhanced Distributed Coordination Function," in *Proc. IEEE ICON '03*, September 2003.

[18] Y.-L. Kuo, C.-H. Lu, E. H.-K. Wu, G.-H. Chen, and Y.-H. Tseng, "Performance Analysis of the Enhanced Distributed Coordination Function in the IEEE 802.11e," in *Proc. IEEE VTC '03 - Fall*, October 2003.

[19] Y. Lin and V. W. Wong, "Saturation Throughput of IEEE 802.11e EDCA Based on Mean Value Analysis," in *Proc. IEEE WCNC '06*, April 2006.

[20] I. Tinnirello and S. Choi, "Efficiency Analysis of Burst Transmissions with Block ACK in Contention-Based 802.11e WLANs," in *Proc. IEEE ICC '05*, May 2005.

[21] T. Li, Q. Ni, T. Turletti, and Y. Xiao, “Performance Analysis of the IEEE 802.11e Block ACK Scheme in a Noisy Channel,” in *Proc. IEEE BroadNets ’05*, October 2005.

[22] I. Tinnirello and S. Choi, “Temporal Fairness Provisioning in Multi-Rate Contention-Based 802.11e WLANs,” in *Proc. IEEE WoWMoM ’05*, June 2005.

[23] F. Peng, H. M. Alnuweiri, and V. C. M. Leung, “Analysis of Burst Transmission in IEEE 802.11e Wireless LANs,” in *Proc. IEEE ICC ’06*, June 2006.

[24] K. Duffy, D. Malone, and D. J. Leith, “Modeling the 802.11 Distributed Coordination Function in Non-Saturated Conditions,” *IEEE Commun. Lett.*, pp. 715–717, August 2005.

[25] F. Alizadeh-Shabdiz and S. Subramaniam, “Analytical Models for Single-Hop and Multi-Hop Ad Hoc Networks,” *Mobile Networks and Applications*, pp. 75–90, February 2006.

[26] G. R. Cantieni, Q. Ni, C. Barakat, and T. Turletti, “Performance Analysis under Finite Load and Improvements for Multirate 802.11,” *Comp. Commun.*, pp. 1095–1109, June 2005.

[27] B. Li and R. Battiti, “Analysis of the IEEE 802.11 DCF with Service Differentiation Support in Non-Saturation Conditions,” in *QoFIS ’04*, September 2004.

[28] O. Tickoo and B. Sikdar, “Queueing Analysis and Delay Mitigation in IEEE 802.11 Random Access MAC based Wireless Networks,” in *Proc. IEEE Infocom ’04*, March 2004.

[29] ——, “A Queueing Model for Finite Load IEEE 802.11 Random Access MAC,” in *Proc. IEEE ICC ’04*, June 2004.

[30] X. Chen, H. Zhai, X. Tian, and Y. Fang, “Supporting QoS in IEEE 802.11e Wireless LANs,” *IEEE Trans. Wireless Commun.*, pp. 2217–2227, August 2006.

[31] W. Lee, C. Wang, and K. Sohraby, “On Use of Traditional M/G/1 Model for IEEE 802.11 DCF in Unsaturated Traffic Conditions,” in *Proc. IEEE WCNC ’06*, May 2006.

[32] P. E. Engelstad and O. N. Osterbo, “Analysis of the Total Delay of IEEE 802.11e EDCA and 802.11 DCF,” in *Proc. IEEE ICC ’06*, June 2006.

[33] K. Medepalli and F. A. Tobagi, “System Centric and User Centric Queueing Models for IEEE 802.11 based Wireless LANs,” in *Proc. IEEE Broadnets ’05*, October 2005.

[34] C. H. Foh and M. Zukerman, “A New Technique for Performance Evaluation of Random Access Protocols,” in *Proc. European Wireless ’02*, February 2002.

[35] J. W. Tantra, C. H. Foh, I. Tinnirello, and G. Bianchi, “Analysis of the IEEE 802.11e EDCA Under Statistical Traffic,” in *Proc. IEEE ICC ’06*, June 2006.

[36] H. Zhai, Y. Kwon, and Y. Fang, “Performance Analysis of IEEE 802.11 MAC Protocols in Wireless LANs,” *Wireless Communications and Mobile Computing*, pp. 917–931, December 2004.

[37] H. Zhai, X. Chen, and Y. Fang, “How Well Can the IEEE 802.11 Wireless LAN Support Quality of Service?” *IEEE Trans. Wireless Commun.*, pp. 3084–3094, November 2005.

[38] I. Inan, F. Keceli, and E. Ayanoglu, “Analysis of the IEEE 802.11e Enhanced Distributed Channel Access Function,” ArXiV cs.OH/ 0704.1833, April 2007. [Online]. Available: arxiv.org

[39] H. Wu, X. Wang, Q. Zhang, and X. S. Shen, “IEEE 802.11e Enhanced Distributed Channel Access (EDCA) Throughput Analysis,” in *Proc. IEEE ICC ’06*, June 2006.

[40] J. W. Tantra, C. H. Foh, and A. B. Mnaouer, “Throughput and Delay Analysis of the IEEE 802.11e EDCA Saturation,” in *Proc. IEEE ICC ’05*, June 2005.

[41] IEEE 802.11e HCF MAC model for ns-2.28. [Online]. Available: http://newport.eecs.uci.edu/ fkeceli/ns.htm

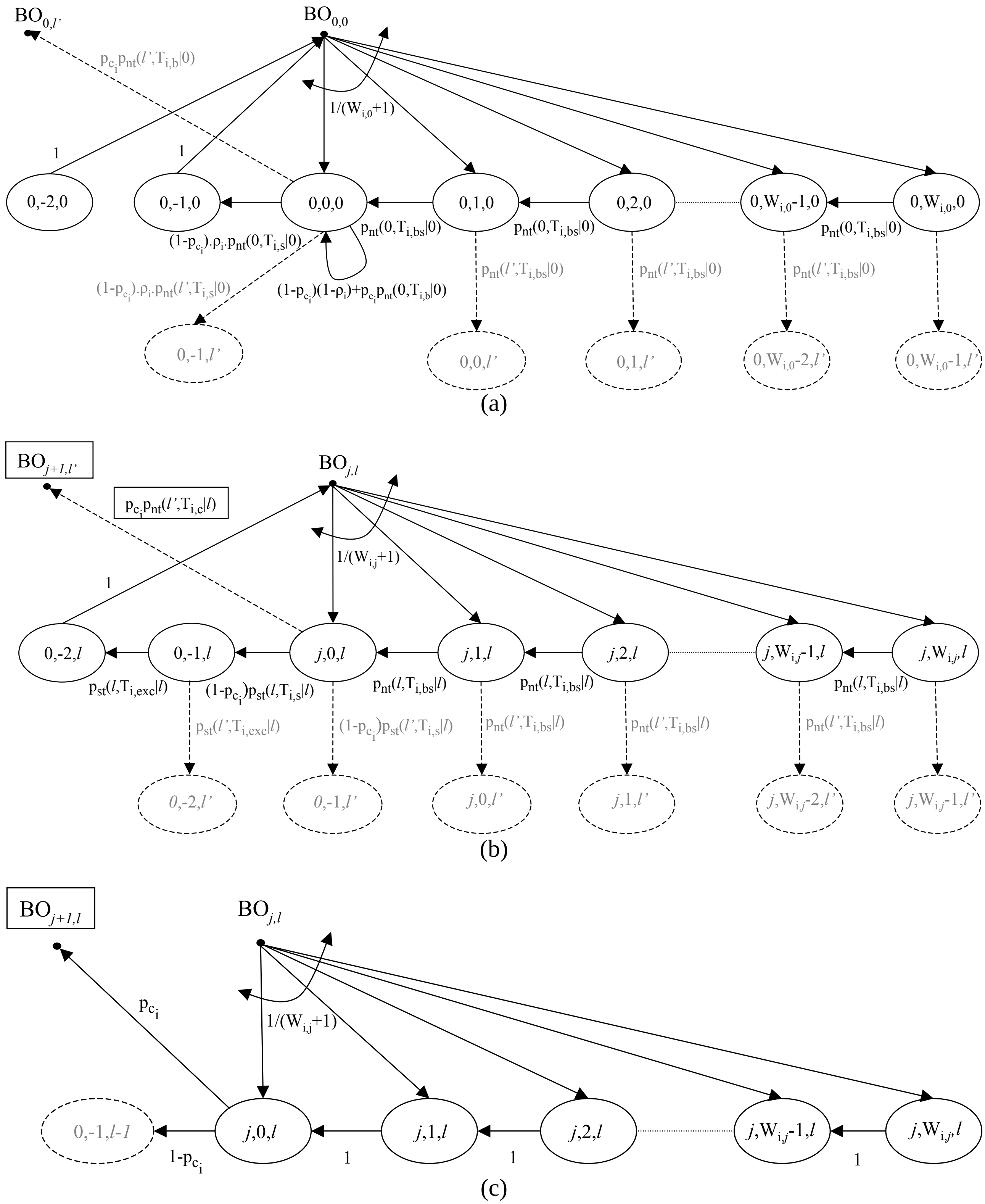

Fig. 1. Parts of the proposed DTMC model for N_i=2. The combination of these small chains for all j, k, l constitutes the proposed DTMC model. (a) $l = 0$. (b) $0 < l < QS_i$. (c) $l = QS_i$. Remarks: *i)* the transition probabilities and the states marked with rectangles differ when $j = r_i - 1$ (as in (8) and (9)), *ii)* the limits for l' follow the rules in (4)-(15), *iii)* each state has many incoming transitions (due to the change in the number of packets in the queue during the transition time) that are not shown in order not to complicate the figure.

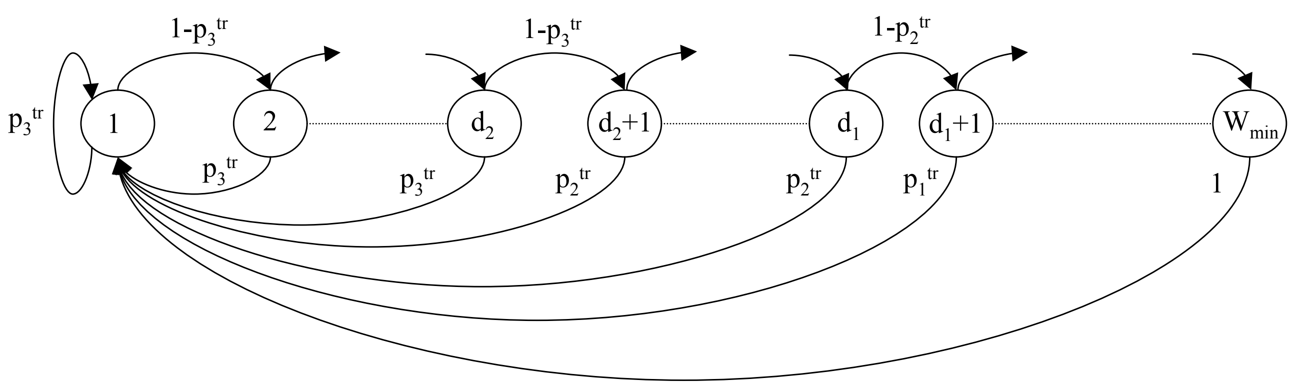

Fig. 2. Transition through backoff slots in different contention zones for the example given in Fig.**??**.

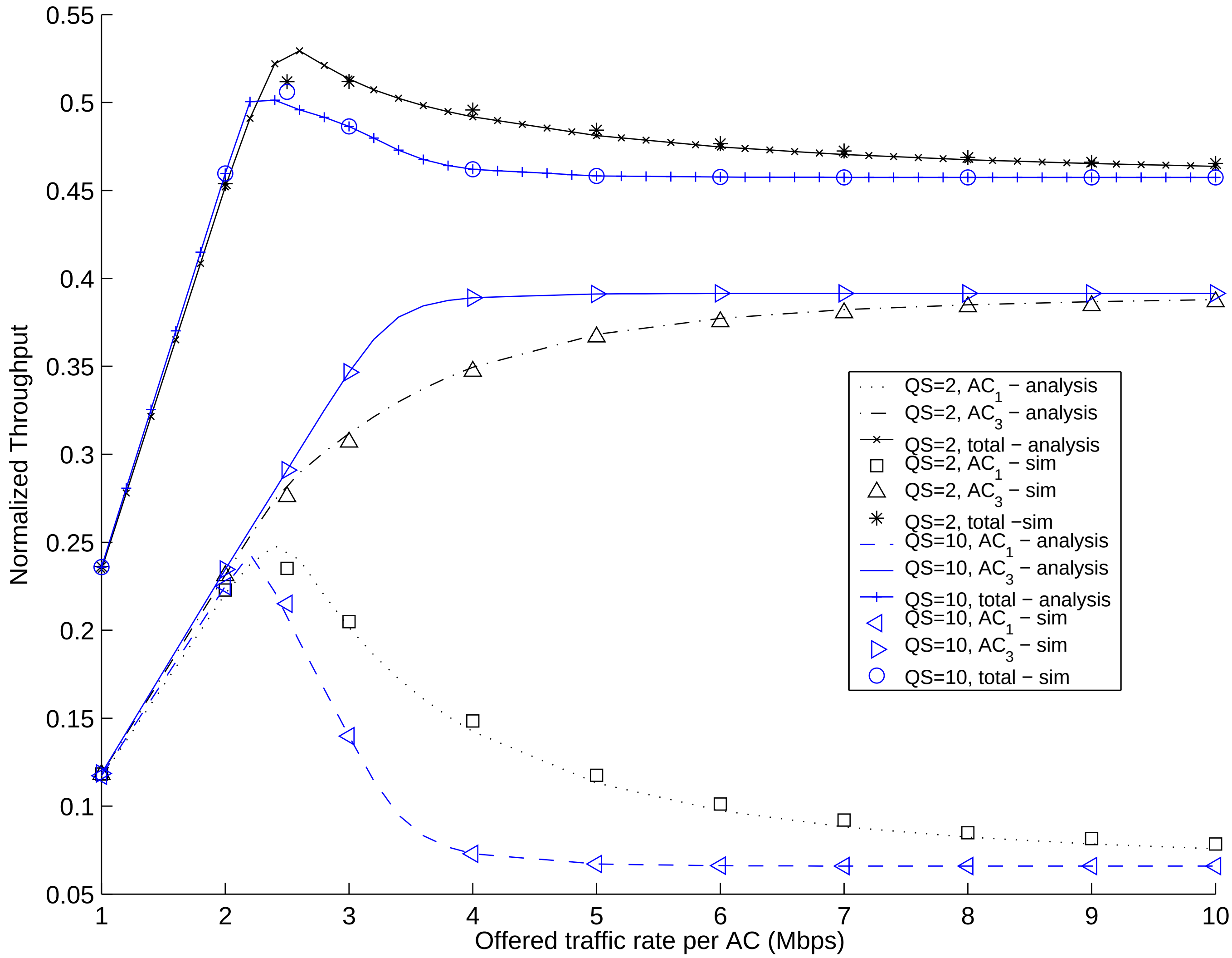

Fig. 3. Normalized throughput prediction of the proposed model for 2 AC heterogeneous scenario with respect to increasing load per AC at each station and varying MAC buffer size in basic access mode ($TXOP_3 = 0$, $TXOP_1 = 0$). Simulation results are also added for comparison.

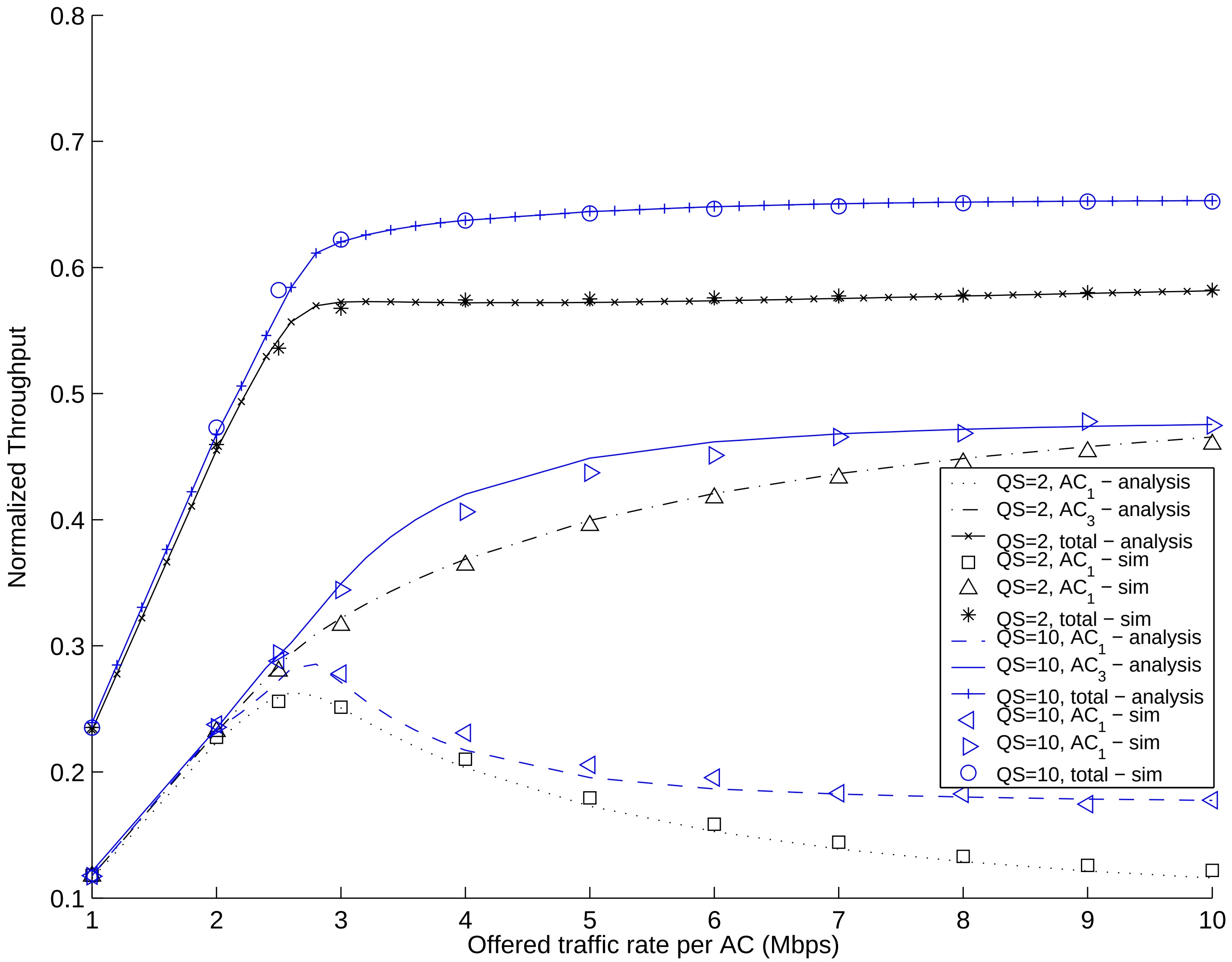

Fig. 4. Normalized throughput prediction of the proposed model for 2 AC heterogeneous scenario with respect to increasing load per AC at each station and varying MAC buffer size in basic access mode ($TXOP_3 = 1504ms$, $TXOP_1 = 3008ms$). Simulation results are also added for comparison.

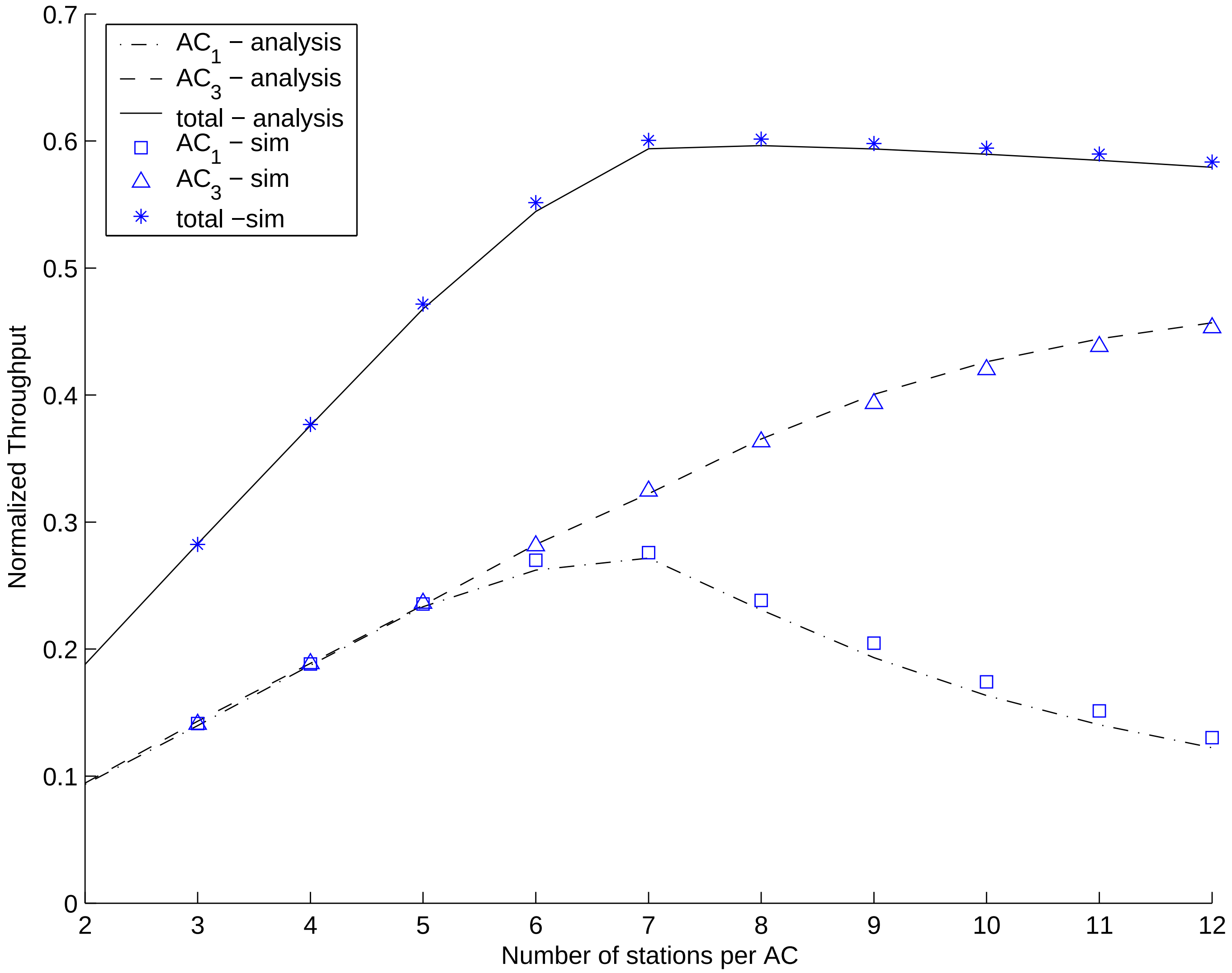

Fig. 5. Normalized throughput prediction of the proposed model for 2 AC heterogeneous scenario with respect to increasing number of stations when MAC buffer size is 10 packets and total offered load per AC is 2 Mbps ($TXOP_3 = 1504ms$, $TXOP_1 = 3008ms$). Simulation results are also added for comparison.

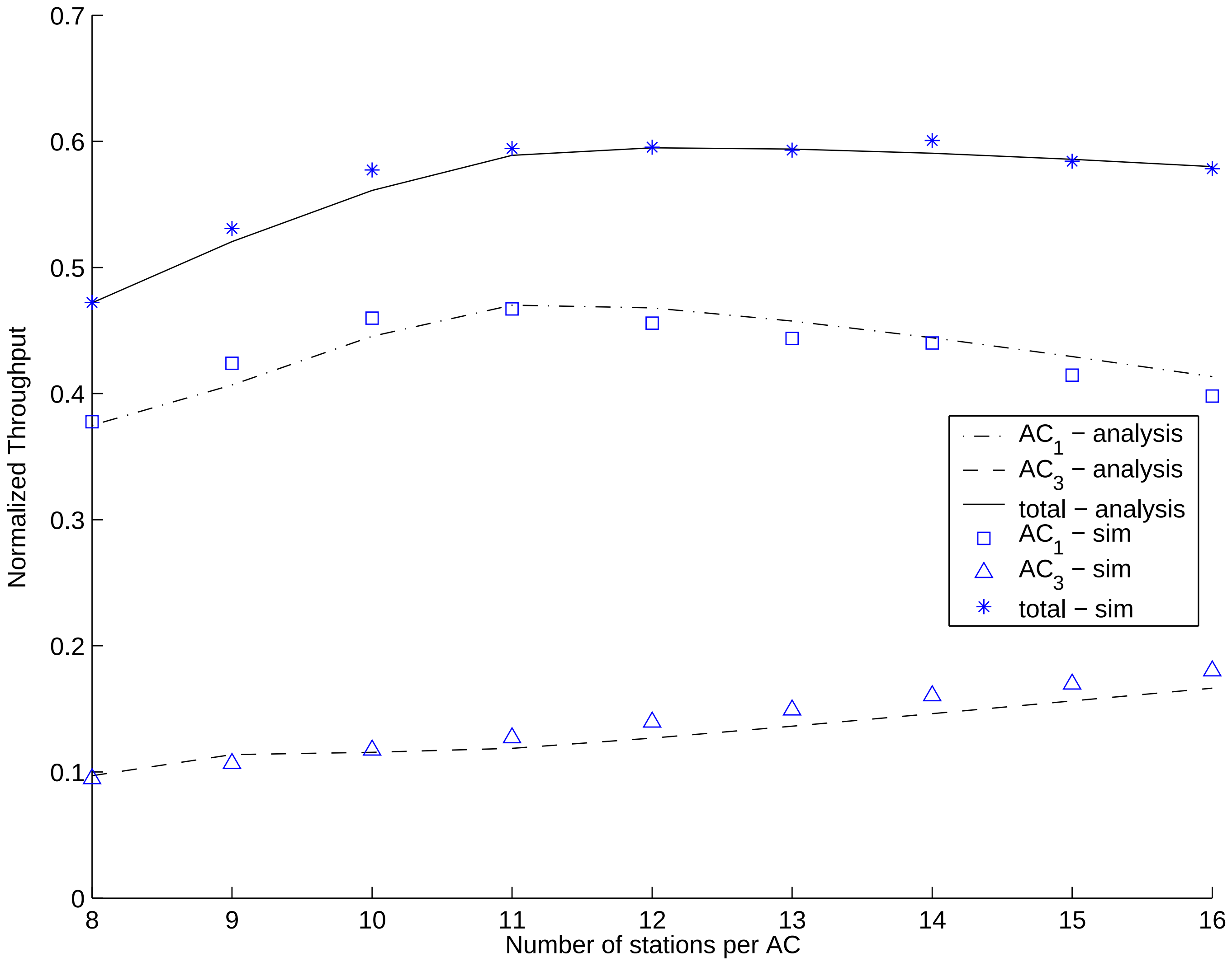

Fig. 6. Normalized throughput prediction of the proposed model for 2 AC heterogeneous scenario with respect to increasing number of stations when MAC buffer size is 10 packets ($TXOP_3 = 1504ms$, $TXOP_1 = 3008ms$). Total offered load per AC_3 is 0.5 Mbps while total offered load per AC_3 is 2 Mbps. Simulation results are also added for comparison.

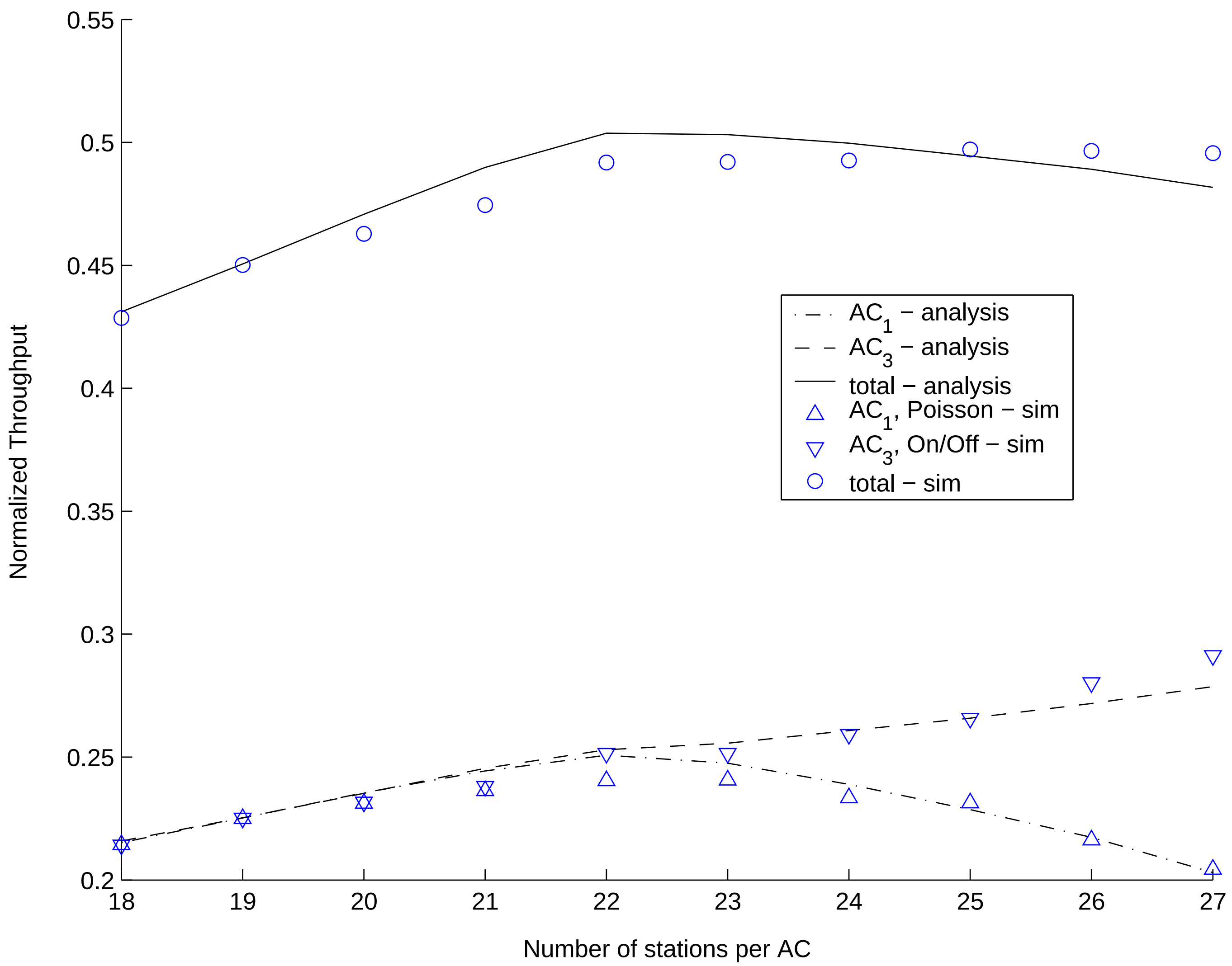

Fig. 7. Normalized throughput prediction of the proposed model for 2 AC heterogeneous scenario with respect to increasing number of stations when total offered load per AC is 0.5 Mbps ($TXOP_3 = 1504ms$, $TXOP_1 = 3008ms$). Simulation results are also added for the scenario when AC_3 uses On/Off traffic with exponentially distributed idle and active times both with mean 1.5s. AC_1 uses Poisson distribution for packet arrivals.

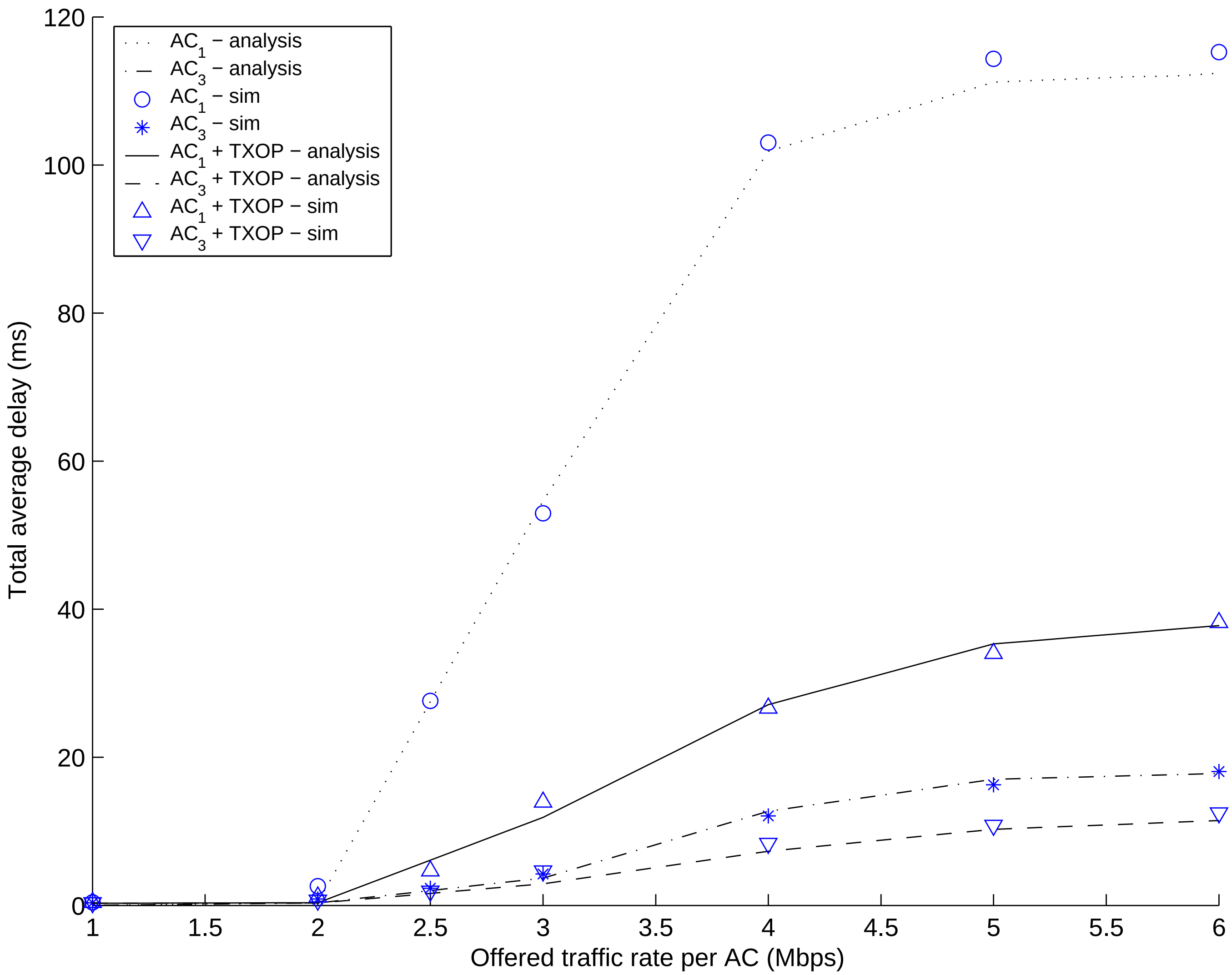

Fig. 8. Total average delay prediction of the proposed model for 2 AC heterogeneous scenario with respect to increasing load per AC at each station. In the first scenario, TXOP limits are set to 0 ms for both ACs. In the second scenario, TXOP limits are set to 1.504 ms and 3.008 ms for high and low priority ACs respectively. Simulation results are also added for comparison.