

# A Simple Model for Non-Homogeneous and Non-Saturated IEEE 802.11 DCF Infrastructure BSS

Eng Hwee Ong, Jamil Y. Khan and Kaushik Mahata

School of Electrical Engineering and Computer Science

The University of Newcastle, Australia, NSW 2308

Email: {enghwee.ong, jamil.khan, kaushik.mahata}@newcastle.edu.au

**Abstract**—We propose a simple unified analytical model to analyze the IEEE 802.11 DCF infrastructure BSS performance in terms of MAC delay, PLR, and throughput efficiency under non-homogeneous and non-saturation conditions. Our model accounts for: (i) asymmetric traffic load between an AP and its associated STAs of an infrastructure BSS; (ii) transition from the non-saturation to saturation mode (and vice-versa) from an AP perspective; (iii) heterogeneous traffic flows between STAs; and (iv) heterogeneous wireless channel conditions between BSSs of a multi-AP hotspot scenario, all in a single unifying framework. More specifically, we integrate a Markov chain model in conjunction with a finite queueing model to analyze the QoS performance of DCF infrastructure BSS, which will be useful for capacity analysis and the design of network control mechanisms. Extensive analyses and simulations have unveiled that the improper modeling or ignorant of backoff freezing for an infrastructure BSS will result in overly conservative bounds which will lead to low network utilization when deployed as admission control, particularly, in heavy load scenarios.

## I. INTRODUCTION

The throughput analysis of the IEEE 802.11 networks under saturation conditions is first introduced by Bianchi [1] and has since been extensively studied in the literature for an IBSS, also known as ad hoc networks, to support data communications with relaxed delay constraints. With the increasing popularity of real-time services, e.g., VoWLAN, the QoS performance analysis of metrics such as MAC delay and PLR is becoming more important. The key reason is that these QoS metrics are particularly useful in the capacity analysis of delay-sensitive VoWLAN as they can be utilized as upper bounds when designing network control mechanism, e.g., admission or load control. In reality, STAs have to operate under non-saturation conditions in order to support delay-sensitive voice traffic and achieve maximum throughput [2]. Consequently, Bianchi's model, which assumes saturation conditions, cannot be directly applied to derive these QoS metrics for VoWLAN. To be more specific, the proper capacity analysis of VoWLAN lies in the ability to model the transition from non-saturation to saturation mode (and vice-versa) which is critical for admission control as the quality of all voice connections will be compromised when the network capacity is exceeded. Furthermore, bulk of the existing IEEE 802.11 deployments in hotspots, enterprises, and campuses are configured as an infrastructure BSS with the basic access scheme of the DCF where STAs are associated with an AP. Although there is a

plethora of analytical models developed for an IBSS, not many are devoted to an infrastructure BSS.

In this paper, a simple unified analytical model is proposed to analyze the performance of the IEEE 802.11 DCF infrastructure BSS VoWLAN. This modeling approach follows closely to the works reported in [3] which incorporates retransmission limit to Bianchi's model, and [4] which integrates Bianchi's model with standard queueing models to derive the QoS metrics of MAC delay and PLR, as well as throughput efficiency. Collectively, these are known as the performance metrics of an AP. The performance metrics of an AP are of particular interest as the AP relays all traffic to and from WLAN, and consequently will be the capacity bottleneck of an infrastructure BSS. In addition, the work presented in [5] is incorporated to augment the analytical model for operating under both ideal and error-prone channel conditions. Furthermore, the importance of modeling backoff freezing, i.e., freezing of backoff counter when medium is busy, as in [6] for an infrastructure BSS, and the consequences if ignored or improperly modeled are exhaustively discussed.

The remainder of this paper is organized as follows. Section II presents the mathematical analysis which composes of the Markov model analysis, average MAC service time analysis, and queueing model analysis to achieve the key performance metrics of MAC delay, PLR, and throughput efficiency for the AP of an infrastructure BSS. Section III validates the proposed analytical model with OPNET simulations. Section IV investigates the effect of backoff freezing for an infrastructure BSS, and the consequences if ignored or improperly modeled are revealed. Finally, Section V concludes this paper.

## II. UNIFIED ANALYTICAL MODEL

In this section, a unified analytical model is proposed as illustrated in Fig. 1. The random backoff process is first modeled using a discrete time Markov chain which is solved numerically to obtain the transmission and its failure probabilities. Subsequently, closed-form expressions for the average MAC service time are derived from the random backoff process based on the transmission failure probability. This average MAC service time, i.e.,  $1/\mu$  is then used in conjunction with the  $M/M/1/K$  model to obtain the average MAC delay, PLR, and throughput efficiency for each STA and the AP. To the best of the author's knowledge, there is no prior

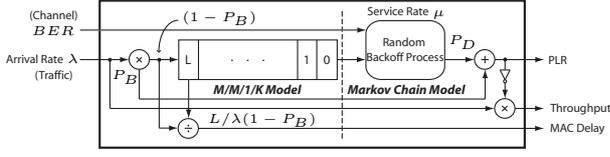


Fig. 1. Unified analytical model: Markov chain model in conjunction with finite queueing model.

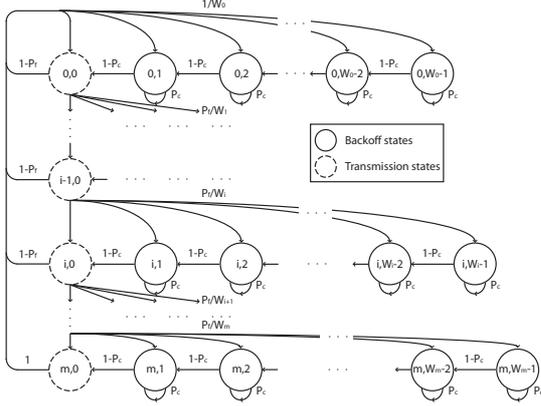


Fig. 2. Discrete time Markov chain transition diagram.

analytical model that offers the performance analysis from a unified perspective in order to pragmatically capture non-homogeneous operating conditions which span across both non-saturation and saturation modes.

### A. Markov Chain Model Analysis

A discrete time Markov chain in Fig. 2 is used to study the random backoff behavior of any STAs by modeling it as a two-dimensional process  $\{s(t), b(t)\}$  where  $s(t)$  and  $b(t)$  are stochastic processes representing the backoff stage and backoff time counter, respectively of the tagged STA at time  $t$ . The key assumptions in this analysis are summarized as follows:

- Collision probability  $P_c$  and transmission failure probability  $P_f$  of a packet transmission remain constant, and they are independent of the number of previous retransmissions.
- An AWGN wireless channel is considered. Thus, each bit has the same bit error probability, and bit errors are i.i.d. over the entire frame.
- Link adaptation and the effects of distance are ignored. Hence, STAs have fixed PHY data rate and the same BER, respectively.
- No hidden terminals are considered. Therefore, collisions will occur only in MAC data frames but not ACK frames.

The probability of transmission  $\tau$  that a STA transmits in a randomly chosen slot time on the condition that the STA has packets to transmit can be derived as

$$\tau = \begin{cases} \frac{2(1-P_c)(1-2P_f)(1-P_f^{m+1})}{\Theta}, & m \leq m' \\ \frac{2(1-P_c)(1-2P_f)(1-P_f^{m+1})}{\Phi}, & m > m' \end{cases} \quad (1)$$

where

$$\begin{aligned} \Theta &= W_0 (1 - (2P_f)^{m+1}) (1 - P_f) \\ &\quad + (1 - 2P_c) (1 - P_f^{m+1}) (1 - 2P_f), \\ \Phi &= W_0 (1 - (2P_f)^{m'+1}) (1 - P_f) \\ &\quad + (1 - 2P_c) (1 - P_f^{m'+1}) (1 - 2P_f) \\ &\quad + 2^{m'} W_0 (P_f^{m'+1}) (1 - P_f^{m-m'}) (1 - 2P_f). \end{aligned} \quad (2)$$

The derivations of (1) by following [5] and [7] are omitted due to space. Note that for  $P_c = 0$ , (1) reduces to the model of [5] which does not consider backoff freezing. From (1), the probability of transmission  $\tau$  depends on the collision probability  $P_c$  and transmission failure probability  $P_f$  which are still unknown.

Now, consider the case of  $n$  STAs where the per-STA quantities are subscripted with the STA label  $a = 1, \dots, n$ . To compute  $P_{c_a}$ , each packet transmitted by the tagged STA is assumed to have a constant and independent collision probability. Accordingly, the probability that medium is idle as seen by the tagged STA is

$$1 - P_{c_a} = \prod_{b \neq a} [1 - (1 - P_{0_b}) \tau_b] \quad (3)$$

where  $P_{c_a}$  is the collision probability as seen by the tagged STA.  $\tau_b$  is the packet transmission probability that other STAs transmit in a randomly chosen slot time given that they have packets to transmit.  $1 - P_{0_b}$  is the probability that other STAs have a non-empty queue by assuming that they can be modeled as a finite queue as in [4]. Essentially,  $1 - P_{0_b}$  functions as a scaling factor of  $\tau_b$  in the saturation mode by assuming that  $\tau_b$  in the non-saturation mode is proportional to  $1 - P_{0_b}$ . The subscripts  $a$  and  $b$  reflect the non-homogeneous network model [8] where the traffic generated by each STA and wireless channel conditions between BSSs may be different, and the fact that the AP of an infrastructure BSS has much higher traffic load than its associated STAs.

Similarly, to compute  $P_{f_a}$ , each packet transmitted by the tagged STA is assumed to have a constant and independent failure probability. A transmission failure is deemed to occur when either a collision or frame error happens by assuming collision and frame error as two independent events. It then follows that the transmission failure probability as seen by the tagged STA is

$$P_{f_a} = 1 - (1 - P_{c_a}) (1 - FER) \quad (4)$$

where  $FER$  is the frame error rate which can be computed as in [5]. For  $n$  STAs, (1) gives an expression for the per-STA transmission probability  $\tau_a$  where  $a = 1, \dots, n$  is the STA label. Hence, (1) and (4) form  $2n$  coupled non-linear equations which can be solved numerically by fixed point iteration technique for  $P_{f_1}, \dots, P_{f_n}$  and  $\tau_1, \dots, \tau_n$ .

### B. Average MAC Service Time Analysis

First, it is observed that the duration of each backoff state in the Markov chain is a random variable. More specifically, each backoff state could be occupied by one of the five virtual events with the corresponding time slot duration of: (i) successful transmission  $T_{s_a}$ ; (ii) unsuccessful transmission with ACK frame error  $T_{e_a}^{ack}$ ; (iii) unsuccessful transmission

with collision  $T_{c_a}$ ; (iv) unsuccessful transmission with data frame error  $T_{e_a}^{data}$ ; and (v) idle slot  $T_{idle}$ , according to a discrete and non-uniform slotted time scale. Although this analysis considers the basic access scheme of the DCF, it can be easily extended to incorporate the four-way handshake procedure of the RTS/CTS mechanism. It is important to note that voice frames are typically transmitted using the basic access scheme for reducing overheads due to their small payload size. Furthermore, one voice packet corresponds to one MAC frame without link layer fragmentation. According to [5], the five different time slot durations for basic access scheme are

$$\begin{cases} T_{s_a} = 2T_{PHY} + T_{DATA_a} + 2\delta + T_{SIFS} + T_{ACK} + T_{DIFS} \\ T_{e_a}^{ack} = T_{s_a} \\ T_{c_a} = T_{PHY} + T_{DATA_a} + \delta + T_{EIFS} \\ T_{e_a}^{data} = T_{c_a} \\ T_{idle} = \sigma \end{cases} \quad (5)$$

where

$$T_{EIFS} = T_{SIFS} + T_{PHY} + T_{ACK} + \delta + T_{DIFS}. \quad (6)$$

$T_{PHY}$  is the duration of PLCP overheads,  $T_{DATA_a}$  is the expected time taken by the tagged STA to transmit a data frame including MAC overheads,  $\delta$  is the propagation delay, and  $\sigma$  is a PHY-dependent slot time. Note that  $\sigma$ ,  $T_{SIFS}$ ,  $T_{DIFS}$ , and  $T_{EIFS}$  are defined in the IEEE 802.11 standard. From (5),  $T_{e_a}^{data} = T_{c_a}$  for the DCF as transmitter cannot differentiate between collisions and data frame errors with positive ACK. On the other hand,  $T_{e_a}^{ack} = T_{s_a}$  as other STAs can still correctly decode the duration field from the successfully received data frame even though the ACK frame is in error.

Now, the expected length of a backoff slot time can be expressed as

$$E[slot_a] = (1 - P_{c_a})\sigma + T_{s_a}(P_{s_a} + P_{e_a}^{ack}) + T_{c_a}(P_{col_a} + P_{e_a}^{data}) \quad (7)$$

where

$$P_{col_a} = P_{c_a} - P_{s_a} - P_{e_a}^{data} - P_{e_a}^{ack}. \quad (8)$$

In the first approximation, it is noted that  $E[slot_a]$  can be rewritten as

$$E[slot_a] = (1 - P_{c_a})\sigma + T_c P_{c_a} \quad (9)$$

if  $T_{s_a}$  and  $T_{c_a}$  of the tagged STA in (7) are equal. Specifically, VoWLAN is typically configured as an infrastructure BSS in a wireline-to-wireless topology where BSS consists of one AP,  $N - 1$  WLAN STAs, and  $N - 1$  ethernet STAs which are connected through a wireline backbone. In such a scenario, the traffic load flowing through the AP is  $N - 1$  times that of a WLAN STA when considering 2-way voice conversations between WLAN and ethernet STAs. As a matter of fact, the AP transmits half of the voice traffic to WLAN STAs. Therefore,  $L_{DATA}$  of the tagged STA can be reasonably approximated as the weighted mean of  $l$  different packet sizes in an infrastructure BSS in order to consider STAs with heterogeneous traffic flows. By symmetry, the STA label subscript of  $T_{DATA_a}$  is dropped such that

$$T_{DATA} = \frac{8L_{DATA}}{R_{DATA}}, \quad L_{DATA} = \left( \frac{\sum_l \lambda_l PLEN_l}{\sum_l \lambda_l} + L_{MAChdr} \right) \quad (10)$$

where  $R_{DATA}$  is the PHY data rate and  $L_{MAChdr}$  is the size of the MAC header.  $\lambda_l$  and  $PLEN_l$  are the arrival rate and packet length of  $l$  different packet sizes in an infrastructure BSS, respectively. Note that (10) also implies that the STA label subscript of the time slot durations in (5) can be omitted.

Once the expected length of backoff slot time is known, the average MAC service time is computed in two parts, viz. the expected time spent in the backoff states and expected time spent in the transmission states, according to the discrete time Markov chain transition diagram illustrated in Fig. 2. First, the expected number of backoff states  $b_{i,k}$ , where  $i \in [0, m]$ ,  $k \in [1, W_i - 1]$ , encountered by the tagged STA before its packet arrives at stage  $i$  can be expressed as

$$E[BO_a] = \sum_{i=0}^m p_{f_a}^i \cdot \frac{W_i - 1}{2} = \begin{cases} \frac{W_0}{2} \left( \frac{1 - (2P_{f_a})^{m+1}}{1 - 2P_{f_a}} \right) - \frac{1}{2} \left( \frac{1 - P_{f_a}^{m+1}}{1 - P_{f_a}} \right), & m \leq m' \\ \frac{W_0}{2} \left( \frac{1 - (2P_{f_a})^{m'+1}}{1 - 2P_{f_a}} + \frac{2^{m'} P_{f_a}^{m'+1} (1 - P_{f_a}^{m-m'})}{1 - P_{f_a}} \right) - \frac{1}{2} \left( \frac{1 - P_{f_a}^{m+1}}{1 - P_{f_a}} \right), & m > m' \end{cases} \quad (11)$$

Owing to the fact that a packet is dropped when it experiences another collision after reaching the last backoff stage  $m$ , i.e., after  $m + 1$  collisions, the expected number of backoff states  $b_{i,k}$ , where  $i \in [0, m]$ ,  $k \in [1, W_i - 1]$ , encountered by the tagged STA before its packet is dropped can be written as

$$E[BO_{drop}] = \sum_{i=0}^m \frac{W_i - 1}{2} = \begin{cases} \frac{W_0(2^{m+1} - 1) - (m+1)}{2}, & m \leq m' \\ \frac{W_0(2^{m'+1} - 1) + W_0 2^{m'}(m-m') - (m+1)}{2}, & m > m' \end{cases} \quad (12)$$

It then follows that the expected time spent by the tagged STA in the backoff states  $b_{i,k}$ , where  $i \in [0, m]$ ,  $k \in [1, W_i - 1]$ , conditioned on successful packet delivery is

$$E[T_{BO_a}] = \left( \frac{E[BO_a] - P_{f_a}^{m+1} \cdot E[BO_{drop}]}{1 - P_{f_a}^{m+1}} \right) E[slot_a] \quad (13)$$

where  $p_{f_a}^{m+1}$  is the probability that the tagged STA's packet is dropped after exceeding its retry limit, and  $1 - p_{f_a}^{m+1}$  is the probability that the tagged STA's packet is not dropped. In other words, expression (13) gives the expected time spent in the backoff states only for packets that are successfully received at the destination, whereas packets dropped due to retry limit do not contribute to the average MAC service time computation as in [3]. Similarly, the expected time spent in the transmission states  $b_{i,0}$ , where  $i \in [0, m]$ , conditioned on successful packet delivery by modeling the number of transmissions per packet of the tagged STA as geometrically distributed with the probability of success  $1 - P_{f_a}$  can be

expressed as

$$\begin{aligned}
E[T_{TX_a}] &= [(1 - P_{f_a}) T_s + P_{f_a} (1 - P_{f_a}) (T_c + T_s) + \dots \\
&+ P_{f_a}^{m-1} (1 - P_{f_a}) (mT_c - T_c + T_s) + \dots \\
&+ P_{f_a}^m (1 - P_{f_a}) (mT_c + T_s)] \frac{1}{1 - P_{f_a}^{m+1}} \\
&= T_s + T_c \left[ \frac{P_{f_a}}{(1 - P_{f_a}) (1 - P_{f_a}^{m+1})} \right. \\
&\left. (1 - (m + 1) P_{f_a}^m + m P_{f_a}^{m+1}) \right]. \quad (14)
\end{aligned}$$

Without loss of generality,  $E[T_{TX_a}]$  can be rewritten as

$$\begin{aligned}
E[T_{TX_a}] &= T_c \left[ 1 + \left( \frac{P_{f_a}}{(1 - P_{f_a}) (1 - P_{f_a}^{m+1})} \right) \right. \\
&\left. (1 - (m + 1) P_{f_a}^m + m P_{f_a}^{m+1}) \right] \quad (15)
\end{aligned}$$

which is immediate from (5) and (10). Finally, the closed-form of the average MAC service time can be expressed as the total amount of time spent by the tagged STA in both the backoff and transmission states given by

$$E[T_{S_a}] = E[T_{Bo_a}] + E[T_{TX_a}]. \quad (16)$$

Note that expression (16) is consistent with the one found in [9].

### C. Queueing Model Analysis

In a realistic networking scenario, most of the MAC frames will carry higher layer packets such as TCP/IP or RTP/UDP/IP in their payload for non-real-time or real-time applications, respectively. These applications are typically sensitive to the end-to-end delay and queue characteristics such as average queue length, MAC delay, queue blocking probability, and throughput. Thus, it will be imperative to analyze the queueing model in order to obtain such performance metrics for admission control and capacity analysis in VoWLAN. Under the assumptions of Poisson arrivals and exponential service time, the queue of each STA and the AP can be analyzed by using the  $M/M/1/K$  model where the steady state probabilities are readily obtained from [10] as

$$\begin{aligned}
P_0 &= \begin{cases} \frac{1-\rho}{1-\rho^{K+1}}, & \rho \neq 1 \\ \frac{1}{K+1}, & \rho = 1 \end{cases}, \\
P_n &= \rho^n P_0, \quad n \in [0, K], \quad (17)
\end{aligned}$$

which are stable even for  $\rho > 1$ . The average queue length is given by

$$L_q = \begin{cases} \frac{\rho}{1-\rho} - \frac{\rho(K\rho^K+1)}{1-\rho^{K+1}}, & \rho \neq 1 \\ \frac{K(K-1)}{2(K+1)}, & \rho = 1 \end{cases}. \quad (18)$$

Accordingly, the average number of packets in the system and MAC delay by relations from the Little's formula are given by

$$\begin{cases} L = L_q + \frac{\lambda(1-P_B)}{\mu} \\ W = \frac{L}{\lambda(1-P_B)} \end{cases}, \quad P_B = P_K. \quad (19)$$

In order to consider heterogeneous traffic flows between STAs, it is assumed that the queue of the AP can store  $K$  packets, independent of their sizes. Such a logical buffer can be achieved easily by using virtual memory mapping which has

TABLE I  
SYSTEM PARAMETERS OF THE IEEE 802.11B PHY.

System Parameters	Notations	802.11b (HR/DSSS)
Slot time	$\sigma$	20 $\mu$ s
SIFS duration	$T_{SIFS}$	10 $\mu$ s
DIFS duration	$T_{DIFS}$	50 $\mu$ s
Propagation delay	$\delta$	1 $\mu$ s
PLCP preamble duration	$T_{PLCPpre}$	144 $\mu$ s (long) 72 $\mu$ s (short)
PLCP header duration	$T_{PLCPhdr}$	48 $\mu$ s (long) 24 $\mu$ s (short)
Total PLCP overheads duration	$T_{PHY}$	192 $\mu$ s
PHY data rate	$R_{DATA}$	11 Mbps
PHY control rate	$R_{CON}$	1 Mbps
MAC header size including 32 bit FCS	$L_{MAChdr}$	224 bits
MAC payload size	$L_{PLD}$	1000 bytes
MAC data frame size	$L_{DATA}$	$L_{MAChdr} + L_{PLD}$
MAC ACK frame size	$L_{ACK}$	112 bits
Minimum CW size	$CW_{min}$	31
Maximum CW size	$CW_{max}$	1023
Maximum CW increasing factor	$m'$	5
Retry limit (Maximum backoff stage)	$m$	6
Bit error rate	$BER$	$0, 10^{-5}$

become a reality with the recent advances of high performance network processors [11].

Again, consider the case of  $n$  STAs where the per-STA quantities are subscripted with the STA label  $a = 1, \dots, n$ . The  $PLR_a$  of the tagged STA is then computed by assuming that the probability of blocking  $P_{B_a}$  and the probability of packet drop due to retry limit  $P_{D_a}$  are two independent events as

$$PLR_a = 1 - (1 - P_{B_a})(1 - P_{D_a}), \quad P_{D_a} = P_{f_a}^{m+1} \quad (20)$$

where  $P_{f_a}$  is the transmission failure probability from (4) that occurs according to a Bernoulli process. It is now trivial to compute the throughput efficiency (or normalized throughput) of each STA and the AP by

$$\bar{S}_a = \frac{8L_{DATA}}{R_{DATA}} \cdot \begin{cases} \lambda_a (1 - PLR_a), & a \in [1, N-1] \\ \sum_{b=1}^{N-1} \lambda_b (1 - PLR_a), & a = N \end{cases}. \quad (21)$$

The expressions (17) – (21) are of key importance since they relate traffic intensity  $\rho$  (function of arrival rate and service rate) and wireless channel conditions (function of service rate) to the key performance metrics of MAC delay, PLR, and throughput efficiency. These performance metrics are crucial for proper admission control and provide insights into the capacity analysis of VoWLAN so that its saturation point can be accurately predicted.

### III. MODEL VALIDATION

The unified analytical model presented in the previous sections is validated by comparing numerical and simulation results obtained based on the system parameters in Table I. The simulation models are developed using OPNET<sup>TM</sup> Modeler<sup>®</sup> 14.5 with Wireless Module.

There is a good agreement between the analysis and simulations in Fig. 3, which confirms the accuracy of modeling assumptions, particularly, during the transition from the non-saturation to saturation mode. The overestimation of the collision probability and MAC service time in the non-saturation region is due to the fact that both post-backoff and the possibility of immediate transmission after medium has been idle for a DIFS duration are not modeled in the Markov chain, which is originally designed by Bianchi under saturation assumption.

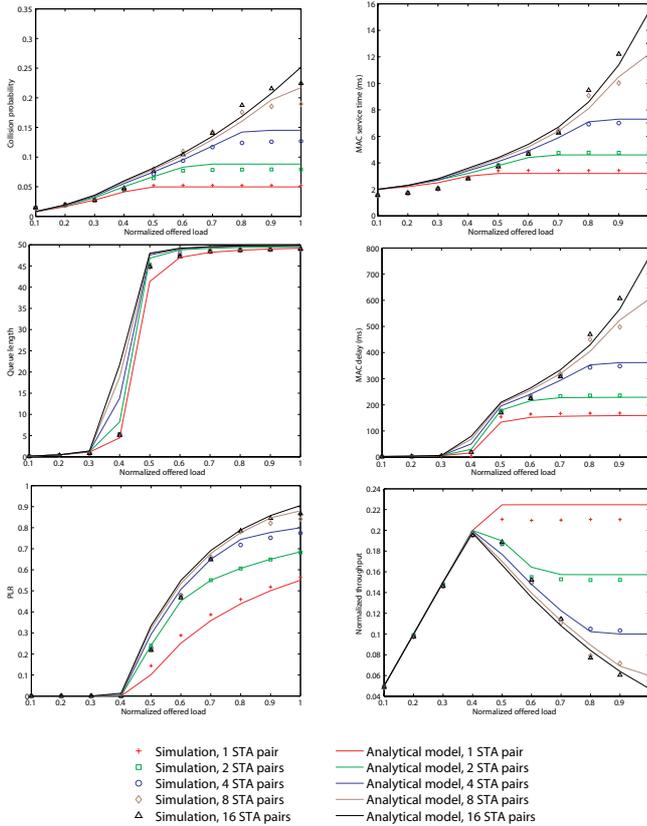


Fig. 3. Model validation: Analysis vs. OPNET simulation for homogeneous CBR traffic source, HR/DSSS PHY @ 11 Mbps,  $L_{PLD} = 1000$  bytes, and  $BER = 10^{-5}$  with different number of STAs and varying traffic arrival rates.

Consequently, these result in a higher collision probability and longer MAC service time. However, the MAC delay and PLR in this non-saturation region will be insignificant, and all offered load will be successfully transmitted.

Apart from the above observations, the analysis has also captured a number of important characteristics. First, a linear relationship between the offered load and throughput exists under the non-saturation mode in which the throughput increases with the offered load. Second, the maximum throughput is reached before saturation in both analysis and simulation when the number of STA pair is more than one. Furthermore, the point where the maximum throughput occurs is relatively insensitive to the number of STA pairs, but rather it is dependent on the offered load. Third, the saturation throughput during high offered load has similar behavior to that of Bianchi's model. Specifically, the saturation throughput decreases as the number of STA pairs increases for small initial CW sizes of 8, 16, 32, and 64. Fourth, the transition from the non-saturation to saturation mode where a marked increase in the collision probability, MAC service time, queue length, MAC delay, and PLR, as well as the corresponding decrease in the throughput are successfully captured.

It is also clear that the MAC service time is dependent on the collision probability. Therefore, it exhibits similar trends that correlate very well to the collision probability over the different range of offered load. On the other hand, all the other

performance metrics derived from the queueing analysis, viz. queue length, MAC delay, PLR, and throughput efficiency are dependent on the MAC service time. Considering all facts, the Markov chain analysis provides an upper bound of the average MAC service time. Consequently, the queueing analysis that is based on the average MAC service time gives the upper bounds of the average queue length, MAC delay, and PLR, whereas it gives the lower bound of the throughput efficiency. Collectively, these are desirable for reliable admission control and capacity analysis.

#### IV. EFFECT OF BACKOFF FREEZING

The effect of backoff freezing for an infrastructure BSS, and the consequences if backoff freezing is not considered or improperly modeled in an infrastructure BSS will be examined in this section. In order to have a uniform comparison, the models of Chatzimisios and Xiao are augmented with the  $M/M/1/K$  queueing model. From Fig. 4(a), inaccuracies in the collision probability are observed from the models of Chatzimisios [3], Zhai [4], and Xiao [7]. The fundamental reason of this phenomenon is due to whether and how backoff freezing is modeled. Note that Chatzimisios's model does not account for backoff freezing and is used as a benchmark for comparison. It is clear that Zhai's model has the greatest inaccuracy. In fact, the authors notice the overestimation of the MAC service time in the saturation mode, but they suggest that the Markov chain may not have captured all the protocol implementations. They further conclude that the overestimation is a reasonable upper bound of their simulation results. However, the reason for this discrepancy is because backoff freezing is accounted in the signal transfer function of their generalized state transition diagram, which models the backoff decrement process, instead of in the Markov chain used to derive the transmission probability of a packet.

It is evident from Fig. 4(a) that modeling backoff freezing in the backoff decrement process explicitly results in huge overestimation of the collision probability, especially, in the saturation region. This can be easily justified by observing expression (3) and the fact that the probability of a non-empty queue increases with the increasing MAC service time. Accordingly, modeling backoff freezing in the backoff decrement process actually increases the MAC service time and consequently the collision probability of AP instead of reducing the transmission probability of STAs and collision probability of AP as expected. The reason for this counterintuitive behavior is attributed to the fact that backoff freezing in the backoff decrement process induces a *positive feedback* phenomenon. Specifically, a slight increase in the MAC service time results in an increase of the collision probability, which in turn causes the MAC service time to increase further. This catastrophic effect will manifest in the AP and STA to cause serious implications for the queueing model analysis, which directly depends on the average MAC service time, as the AP transits into the saturation mode. In particular, the performance metrics of MAC delay and PLR will be overly overestimated while throughput efficiency will be grossly underestimated.

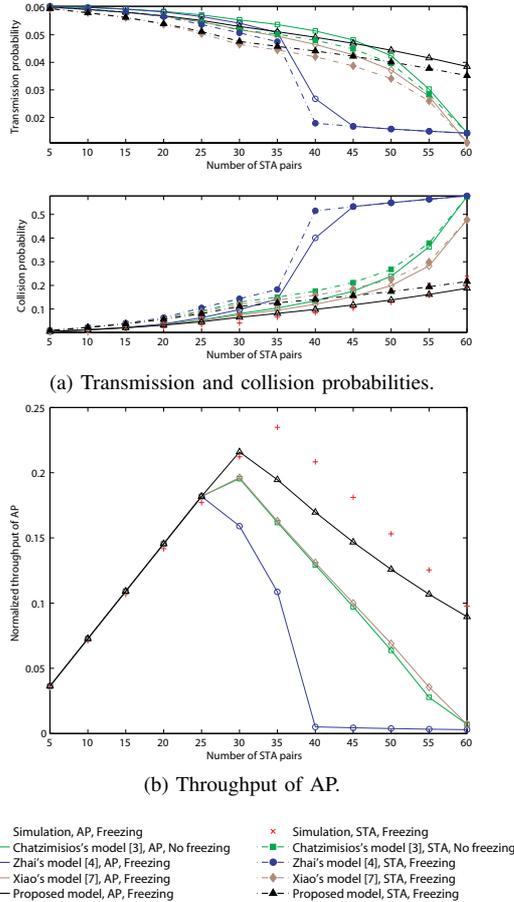


Fig. 4. Effect of freezing for an infrastructure BSS under non-saturation condition with homogeneous CBR traffic source, HR/DSSS PHY @ 11 Mbps,  $\lambda = 10$  frames/sec,  $L_{PLD} = 1000$  bytes, and BER = 0.

Hence, backoff freezing should not be independently modeled in the backoff decrement process.

On the other hand, Xiao's model considers backoff freezing in both the backoff decrement process and Markov chain. The results in Fig. 4(a) show that the collision probability is overestimated. In fact, it can be observed from Fig. 4(b) that the throughput efficiency obtained using Xiao's model is similar to that of Chatzimisios's model which does not consider backoff freezing. The only exception is the collision probability which yields a slightly lower value as compared to Chatzimisios's model when the number of STA pairs increases. This outcome is intuitive as the transmission probability of STAs is expected to reduce with backoff freezing, owing to the fact that STAs now spend a longer time in each backoff state. Again, by the relation of expression (3), the decrease of transmission probability of STAs will then cause the collision probability of AP to decrease correspondingly. When the collision probability of AP decreases, the MAC service time, MAC delay, and PLR of AP will also decrease whilst the throughput efficiency will increase. This is the desired effect of backoff freezing. However, when backoff freezing is also accounted in the backoff decrement process, the average time spent in each backoff state, i.e., the MAC service time is

immediately increased. This inadvertently negates the effect of backoff freezing in the Markov chain which has decreased the MAC service time initially. As a result, the performance metrics derived by Xiao's model is by large the same as Chatzimisios's model, except for the collision probability of AP which has decreased slightly. Hence, backoff freezing should be modeled only in the Markov chain as in the proposed model where the numerical results are found to be in high agreement with the simulation results.

## V. CONCLUSION

A simple unified analytical model is proposed to analyze the IEEE 802.11 DCF infrastructure BSS VoWLAN. Through this analysis, the key performance metrics of MAC delay, PLR, and throughput efficiency have been accurately obtained to enable efficient and effective admission control. Furthermore, they have provided insights into the capacity analysis of VoWLAN where its saturation point can be predicted by capturing the transition from the non-saturation to saturation mode of operation. The impact of backoff freezing on infrastructure BSS has also been investigated which revealed that backoff freezing should be properly modeled in order to derive accurate performance metrics and consequently tight bounds for admission control or capacity analysis. In fact, both the improper modeling and ignorant of backoff freezing will manifest as overly conservative bounds, particularly, in heavy load situations. This implies that residual capacity will be created, which leads to low network utilization when such models are deployed as admission or load control.

## REFERENCES

- [1] G. Bianchi. Performance analysis of the IEEE 802.11 distributed coordination function. *IEEE Journal on Selected Areas in Communications*, 18(3):535–547, March 2000.
- [2] H. Zhai, X. Chen, and Y. Fang. How well can the IEEE 802.11 wireless LAN support quality of service? *IEEE Transactions on Wireless Communications*, 4(6):3084–3094, November 2005.
- [3] P. Chatzimisios. *Performance Modelling and Enhancement of Wireless Communication Protocols*. PhD thesis, Bournemouth University, Poole, UK, December 2004.
- [4] H. Zhai, Y. Kwon, and Y. Fang. Performance analysis of IEEE 802.11 MAC protocols in wireless LANs. *Wirel. Commun. Mob. Comput.*, 4(8):917–931, November 2004.
- [5] Q. Ni, T. Li, T. Turletti, and Y. Xiao. Saturation throughput analysis of error-prone 802.11 wireless networks. *Wirel. Commun. Mob. Comput.*, 5(8):945–956, November 2005.
- [6] E. Ziouva and T. Antonakopoulos. CSMA/CA performance under high traffic conditions: Throughput and delay analysis. *Comput. Commun.*, 25(3):313–321, February 2002.
- [7] Y. Xiao. Performance analysis of IEEE 802.11e EDCF under saturation condition. In *Proc. IEEE International Conference on Communications, 2004. ICC '04*, volume 1, pages 170–174, June 2004.
- [8] D. Malone, K. Duffy, and D. Leith. Modeling the 802.11 distributed coordination function in nonsaturated heterogeneous conditions. *IEEE/ACM Transactions on Networking*, 15(1):159–172, February 2007.
- [9] P. Raptis, V. Vitsas, P. Chatzimisios, and K. Paparrizos. Voice and data traffic analysis in IEEE 802.11 DCF infrastructure WLANs. In *Proc. Second International Conference on Advances in Mesh Networks, 2009. MESH 2009*, pages 37–42, June 2009.
- [10] D. Gross and C. M. Harris. *Fundamentals of Queueing Theory*. John Wiley & Sons, Inc., New York, NY, USA, 3rd edition, 1998.
- [11] R. Sharmila, M. V. Lakshmi Priya, and R. Parthasarathi. An active framework for a WLAN access point using Intel's IXP1200 network processor. In *Proc. High Performance Computing, HiPC 2004*, pages 71–80, December 2004.