Performance Analysis of IEEE 802.11 MAC Protocols in Wireless LANs

Hongqiang Zhai[1], Younggoo Kwon[2] and Yuguang Fang[1]

[1]Department of Electrical and Computer Engineering
University of Florida, Gainesville
Florida 32611-6130, USA
Tel: (352) 846-3043, Fax: (352) 392-0044
E-mail: zhai@ecel.ufl.edu, fang@ece.ufl.edu

[2]Department of Computer Engineering
Sejong University, 98, Gunja-dong, Kwangjin-gu,
Seoul, 143-747, Korea
Tel: 82-2-3408-3410, Fax: 82-2-3408-3667
E-mail: ygkwon@sejong.ac.kr

***Abstract*—IEEE 802.11 MAC protocol is the de facto standard for wireless LANs, and has also been implemented in many network simulation packages for wireless multi-hop ad hoc networks. However, it is well known that, as the number of active stations increases, the performance of IEEE 802.11 MAC in terms of delay and throughput degrades dramatically, especially when each station's load approaches to its saturation state. To explore the inherent problems in this protocol, it is important to characterize the probability distribution of the packet service time at the MAC layer. In this paper, by modeling the exponential backoff process as a Markov chain, we can use the signal transfer function of the generalized state transition diagram to derive an approximate probability distribution of the MAC layer service time. We then present the discrete probability distribution for MAC layer packet service time, which is shown to accurately match the simulation data from network simulations. Based on the probability model for the MAC layer service time, we can analyze a few performance metrics of the wireless LAN and give better explanation to the performance degradation in delay and throughput at various traffic loads. Furthermore, we demonstrate that the exponential distribution is a good approximation model for the MAC layer service time for the queueing analysis, and the presented queueing models can accurately match the simulation data obtained from ns-2 when the arrival process at MAC layer is Poissonian.**

Keywords—Performance Evaluation, IEEE 802.11 MAC, Wireless LANs, Queueing Analysis

I. INTRODUCTION

The Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) protocol used in the IEEE

This work was supported in part by the US Office of Naval Research under grant N000140210464 (ONR Young Investigator Award) and under grant N000140210554.

802.11 MAC protocol has been proposed as the standard protocol for wireless local area networks (LANs), which has also been widely implemented in many wireless testbeds and simulation packages for wireless multi-hop ad hoc networks.

However, there are many problems encountered in the higher protocol layers in IEEE 802.11 wireless networks. It has been observed that the packet delay increases dramatically when the number of active stations increases. Packets may be dropped either due to the buffer overflow or because of serious MAC layer contentions. Such packet losses may affect high layer networking schemes such as the TCP congestion control and networking routing maintenance. The routing simulations [1] [2] over mobile ad hoc networks indicate that network capacity is poorly utilized in terms of throughput and packet delay when the IEEE 802.11 MAC protocol is integrated with routing algorithms. TCP in the wireless ad hoc networks is unstable and has poor throughput due to TCP's inability to recognize the difference between the link failure and the congestion. Besides, one TCP connection from one-hop neighbors may capture the entire bandwidth, leading to the one-hop unfairness problem [3], [4], [5], [6].

Performance analysis for the IEEE 802.11 MAC protocol could help to discover the inherent cause of the above problems and may suggest possible solutions. Many papers on this topic have been published [7-11] [14] [17]. Cali [7], [8] derived the protocol capacity of the IEEE 802.11 MAC protocol and presented an adaptive backoff mechanism to replace the exponential backoff mechanism. Bianchi [9] proposed a Markov chain model for the binary exponential backoff procedure to analyze and compute the IEEE 802.11 DCF saturated throughput. All of these papers assume the saturated scenario where all stations always have data to transmit. Based on the saturated throughput in Bianchi's model, Foh and Zuckerman presented the analysis of the mean packet delay at different throughput for IEEE 802.11 MAC in [10]. Hadzi-Velkov also gave an analysis for the throughput and mean packet delay in the saturated case by incorporating frame-error rates [11]. Kim and Hou [17] analyzed the protocol capacity of IEEE 802.11 MAC with the assumption that the number of active stations having packets ready for transmission is large.

To the authors' best knowledge, there is no comprehensive study on the queue dynamics of the IEEE 802.11 wireless LANs. The delay analysis is limited to the derivation of mean value while the higher moments and the probability distribution function of the delay are untouched. And most of the current papers focused on the performance analysis in saturated traffic scenarios and the comprehensive performance study under non-saturated traffic situations is still open.

In this paper, to address the above issues, we first characterize the probability distribution of the *MAC layer packet service time* (i.e., the time interval between the time instant a packet starts to contend for transmission and the time instant that the packet either is acknowledged for correct reception by the intended receiver or is dropped). Based on the probability distribution model of the MAC layer packet service time, we then study the queueing performance of the wireless LANs at different traffic load based on the IEEE 802.11 MAC protocol. Then, we evaluate the accuracy of the exponential probability distribution model for the MAC layer service time in queueing analysis through both analytical approach and simulations.

II. PRELIMINARIES

A. *Distributed Coordination Function (DCF)*

Before we present our analysis for 802.11 MAC, we first briefly describe the main procedures in the DCF of 802.11 MAC protocol [12]. In the DCF protocol, a station shall ensure that the medium is idle before attempting to transmit. It selects a random backoff interval less than or equal to the current contention window (CW) size based on the uniform distribution, and then decreases the backoff timer by one at each time slot when the medium is idle (may wait for DIFS followed a successful transmission or EIFS followed a collision). If the medium is determined to be busy, the station will suspend its backoff timer until the end of the current transmission. Transmission shall commence whenever the backoff timer reaches zero. When there are collisions during the transmission or when the transmission fails, the station invokes the backoff procedure. To begin the backoff procedure, the contention window size CW, which takes an initial value of CWmin, doubles its value before it reaches a maximum upper limit CWmax, and remains the value CWmax when it is reached until it is reset. Then, the station sets its backoff timer to a random number uniformly distributed over the interval [0, CW) and attempts to retransmit when the backoff timer reaches zero again. If the maximum transmission failure limit is reached, the retransmission shall stop, CW shall be reset to CWmin, and the packet shall be discarded [12]. The RTS/CTS mechanisms and basic access mechanism of IEEE 802.11 are shown in Fig. 1.

B. System Modeling

Each mobile station is modeled as a queueing system, which can be characterized by the arrival process and the service time distribution. And the saturated status is reached if each station has heavy traffic and always has packets to transmit. The non-saturated status, i.e., under light or moderate traffic load, could be characterized by the non-zero probability that the queue length is zero.

The service time of the queueing system is the MAC layer packet service time defined in Section I. The IEEE 802.11 MAC adopts the binary exponential backoff mechanism for the transmission of each packet, which may collide with some other transmissions in the air at each transmission attempt. And the collision probability p_c is determined by the probability that there is at least one of other stations which will transmit at the same backoff time slot when the considered station attempts transmission. We assume that this probability does not change and is independent during the transmission of each packet regardless of the number of retransmission suffered. For the saturated case, this approximation has been used in [9] to derive the saturated throughput. And for the non-saturated case, the collision probability becomes more complex. It depends on the number of stations with packets ready for transmission and the backoff states of these stations. Between two transmission attempts at the considered station, other stations may complete several successful transmissions and/or encounter several collisions, and there may be new packet arrivals at stations no matter whether they are previously contending for transmission or not. Intuitively, this approximation becomes more accurate when the number of stations gets larger for both saturated and non-saturated case. For simplicity, we use the same approximation for both cases and argue that the collision probability does not change significantly as long as the input traffic rate from higher layer at each station are still the same during the service for each packet. Then we could model the binary exponential backoff mechanism as a Markov chain and make possible the derivation of the probability distribution of service time in the next section. Later in this paper, we will show that the analytical results from this approximation are consistent with the simulation results very well at the non-saturated case.

III. The Probability Distribution of the MAC Layer Service Time

A. *MAC Layer Service Time*

As described in section II, there are three basic processes when the MAC layer transmits a packet: the decrement process of the backoff timer, the successful packet transmission process that takes a time period of T_{suc} and the packet collision process that takes a time period of T_{col}. Here, T_{suc} is the random variable representing the period that the medium is sensed busy because of a successful transmission, and T_{col} is the random variable representing the period that the medium is sensed busy by each station due to collisions.

The MAC layer service time is the time interval from the time instant that a packet becomes the head of the queue and starts to contend for transmission to the time instant that either the packet is acknowledged for a successful transmission or the packet is dropped. This time is important when we examine the performance of higher protocol layers. Apparently, the distribution of the MAC layer service time is a discrete probability distribution because the smallest time unit of the backoff timer is a time slot. T_{suc} and T_{col} depend on the transmission rate, the length of the packet and the overhead (with a discrete unit, i.e., bit), and the specific transmission scheme (the basic access DATA/ACK scheme or the RTS/CTS scheme) [9] [12].

B. *Probability Generating Functions (PGF) of MAC Layer Service Time*

The MAC layer service time is a non-negative random variable denoted by random variable T_S, which has a discrete probability of p_i for T_S being t_{si} with the unit of one-bit transmission time or the smallest system clock unit, i=0,1,2,…. The PGF of T_S is given by

$$P_{T_S}(Z) = \sum_{i=0}^{\infty} p_i Z^{t_{si}} = p_0 Z^{t_{s0}} + p_1 Z^{t_{s1}} + p_2 Z^{t_{s2}} + \ldots \qquad (1)$$

and completely characterizes the discrete probability distribution of T_S , and has a few important properties as follows:

$$\begin{cases} P_{T_S}(1) = 1 \\ E[T_S] = \dfrac{\partial}{\partial Z} P_{T_S}(Z)\Big|_{Z=1} = P'_{T_S}(1) \\ VAR[X] = P''_{T_S}(1) + P'_{T_S}(1) - \{P'_{T_S}(1)\}^2 \end{cases} \qquad (2)$$

where the prime indicates the derivative.

To derive the PGF of the MAC layer service time, we will model the transmission process of each packet as a Markov chain in the following subsections. Here we first discuss how to drive the PGF of the service time from the Markov chain.

The state when the packet leaves the mobile station, i.e., being successfully transmitted or dropped, is the absorption state of the Markov chain for the backoff mechanism. To obtain the average transition time to the absorption state of the Markov chain, we can use the matrix geometric approach. However, in the case of Markov Chain for T_S with various transition times on different branches, it requires a new matrix formulation to accommodate different transition times, and its solution always accompanies extraneous complicated computations [13]. Here, we apply the generalized state transition diagram, from which we can easily derive the PGF of T_S and obtain arbitrary *n*th moment of T_S.

In the generalized state transition diagram, we mark the transition time on each branch along with the transition probability in the state transition diagram (the Markov chain). The transition time, which is the duration for the state transition to take place, is expressed as an exponent of Z variable in each branch. Thus, the probability generating function of total transition time can be obtained from the signal transfer function of the generalized state transition diagram using the well-known Mason formula [13][18].

To illustrate how the generalized Markov chain model works, we show one simple example for a MAC mechanism that allows infinite retransmissions for each packet without any backoff mechanisms. If the random variable F is defined as the duration of time taken for a state transition from the state "1" to "2" in Fig. 2, its PGF is simply the signal transfer function of the state transition. In Fig. 2, p is the collision probability, $1-p$ is the successfully transmitted probability, τ_1 is the collision time, and τ_2 is the successful transmission time. So the PGF of random variable F is

$$P_F(Z) = \frac{(1-p)Z^{\tau_2}}{1-pZ^{\tau_1}} \tag{3}$$

This satisfies (2), that is, $P_F(1)=1$ and its mean transition time is

$$P'_F(1) = \frac{p}{1-p}\tau_1 + \tau_2 \tag{4}$$

On the other hand, we can easily obtain the average collision/retransmission times N_C, i.e., *p/(1-p)*. Thus the average transition time can be directly obtained as $N_C \times \tau 1 + \tau 2$, which is the same as (4).

C. The processes of collision and successful transmission

We first study the RTS/CTS mechanisms. As shown in Fig. 1, the period of successful transmission T_{suc} equals to

$$T_{suc} = RTS + CTS + DATA + ACK + 3SIFS + DIFS \tag{5}$$

And the period of collision T_{col} equals to

$$T_{col} = RTS + SIFS + ACK + DIFS = RTS + EIFS \tag{6}$$

T_{col} is a fixed value and its PGF $C_t(Z)$ equals

$$C_t(Z) = Z^{RTS+EIFS} \tag{7}$$

T_{suc} is a random variable determined by the distribution of packet length. In the case that the length of DATA has a uniform distribution in [l_{min}, l_{max}], its PGF $S_t(Z)$ equals

$$S_t(Z) = Z^{RTS+CTS+ACK+3SIFS+DIFS}\frac{1}{l_{\max}-l_{\min}+1}\sum_{i=l_{\min}}^{l_{\max}} Z^i \tag{8}$$

In the case that the length of DATA is a fixed value l_D, its PGF $S_t(Z)$ equals

$$S_t(Z) = Z^{RTS+CTS+l_D+ACK+3SIFS+DIFS} \tag{9}$$

If the basic scheme is adopted, T_{col} is determined by the longest one of the collided packets. When the probability of three or more packets simultaneously colliding is neglected, its probability distribution can be approximated by the following equation,

$$\Pr\{T_{col}=i\} = \Pr\{l_1 = i, l_2 \le i\} + \Pr\{l_2 = i, l_1 \le i\} - \Pr\{l_1 = i, l_2 = i\},$$

where $l_i(i=1,2)$ is the packet length of the i_{th} collided packet. Thus we could obtain that

$$C_t(Z) \approx Z^{EIFS} \frac{1}{(l_{\max} - l_{\min} + 1)^2} \sum_{i=l_{\min}}^{l_{\max}} (2i - 2l_{\min} + 1) Z^i \quad (10)$$

$$S_t(Z) = Z^{SIFS+ACK+DIFS} \frac{1}{l_{\max} - l_{\min} + 1} \sum_{i=l_{\min}}^{l_{\max}} Z^i \quad (11)$$

for the case that the length of DATA has a uniform distribution in [l_{min}, l_{max}], or

$$C_t(Z) = Z^{l_D + EIFS} \quad (12)$$

$$S_t(Z) = Z^{l_D + SIFS + ACK + DIFS} \quad (13)$$

for the case that the length of DATA is a fixed value l_D.

D. *Decrement Process of Backoff Timer*

In the backoff process, if the medium is idle, the backoff timer will decrease by one for every idle slot detected. When detecting an ongoing successful transmission, the backoff timer will be suspended and deferred a time period of T_{suc}, while if there are collisions among the stations, the deferring time will be T_{col}.

As mentioned in section II, p_c is the probability of a collision seen by a packet being transmitted on the medium. Assuming that there are n stations in the wireless LAN we are considering and packet arrival processes at all the stations are independent and identically distributed, we observe that p_c is also the probability that there is at least one packet transmission in the medium among other (n-1) stations in the interference range of the station under consideration. This yields

$$p_c = 1 - [1 - (1 - p_0)\tau)]^{n-1} \quad (14)$$

where p_0 is the probability that there are no packets ready to transmit at the MAC layer in the wireless station under consideration, and τ is the packet transmission probability that the station transmits in a randomly chosen slot time given that the station has packets to transmit.

Let P_{suc} be the probability that there is one successful transmission among other (n-1) stations in the considered slot time given that the current station does not transmit. Then,

$$P_{suc} = \binom{n-1}{1}(1-p_0)\tau(1-(1-p_0)\tau)^{(n-2)} = (n-1)((1-p_c)^{(n-2)/(n-1)} + p_c - 1) \qquad (15)$$

Then $p_c - P_{suc}$ is the probability that there are collisions among other (n-1) stations (or neighbors).

Thus, the backoff timer has the probability of 1- p_c to decrement by 1 after an empty slot time σ, the probability P_{suc} to stay at the original state after T_{suc}, and the probability of $p_c - P_{suc}$ to stay at the original state after T_{col}. So the decrement process of backoff timer is a Markov process. The signal transfer function of its generalized state transition diagram is

$$H_d(Z) = (1-p_c)Z^{\sigma} / [1 - P_{suc}S_t(Z) - (p_c - P_{suc})C_t(Z)]\,. \qquad (16)$$

From above formula, we observe that $H_d(Z)$ is a function of p_c, the number of stations n and the dummy variable Z.

E. *Markov Chain Model for the Exponential Backoff Procedure*

Whenever the backoff timer reaches zero, transmission shall commence. According to the definition of p_c, the station has the probability 1- p_c to finish the transmission after T_{suc}, and the probability p_c to double contention window size and enter a new backoff procedure until the maximum retransmission limit is reached after T_{col}. Since the decrement process of backoff timer is a Markov process as discussed above, the whole exponential backoff procedure is also a Markov process.

Let W be the minimum value of contention window size CWmin plus 1. Following a similar procedure used in [9] and noticing that the transition probability at each branch of the Markov chain is different from [9] (which only denoted the value at the saturated status and did not consider that the contention window is reset after the maximum α times of retransmissions as defined in the protocols [12]), we can obtain (please refer to Appendix I)

$$\tau = \begin{cases} \dfrac{2(1-p_c^{\alpha+1})}{1-p_c^{\alpha+1}+(1-p_c)W(\sum_{i=0}^{\alpha}(2p_c)^i)} & ,\alpha \le m \\ \dfrac{2(1-p_c^{\alpha+1})}{1-p_c^{\alpha+1}+p_cW\sum_{i=0}^{m-1}(2p_c)^i+W(1-2^m p_c^{\alpha+1})} & ,\alpha > m \end{cases} \qquad (17)$$

where m is the maximum number of the stages allowed in the exponential backoff procedure (the definition is clarified below). We will use (14) and (17) in the queueing analysis to derive the collision probability at different input traffic in Section IV.

F. Generalized State Transition Diagram

Now, it is possible to draw the generalized state transition diagram for the packet transmission process as shown in Fig. 3. In Fig. 3, {s(t), b(t)} is the state of the bi-dimensional discrete-time Markov chain, where b(t) is the stochastic process representing the backoff timer count for a given station, and s(t) is the stochastic process representing the backoff stage with values (0, ..., α) for the station at time t. Let m be the "maximum backoff stage" at which the contention window size takes the maximum value, i.e., CWmax = 2^m(CWmin + 1) − 1. At different "backoff stage" i ∈ [0, α], the contention window size CW_i [1]= W_i − 1, where W_i = 2^i(CWmin + 1) if 0 ≤ i ≤ m, and W_i = CWmax + 1 if m ≤ i ≤ α.

As we defined before, the random variable T_S is the duration of time taken for a state transition from the start state (beginning to be served) to the end state (being transmitted successfully or discarded after maximum α times retransmission failures). Thus, its Probability Generating Function (PGF), denoted as *B(Z)* that is the function of p_c, *n* and *Z*, is simply the signal transfer function from the start state to the end state given by:

$$\begin{aligned} &HW_i(Z) = \begin{cases} \sum_{j=0}^{2^iW-1} H_d^{\ j}(Z)/(2^iW), & (0 \le i \le m) \\ HW_m(Z), & (m < i \le \alpha) \end{cases} \\ &H_i(Z) = \prod_{j=0}^{i} HW_j(Z), (0 \le i \le \alpha) \\ &B(Z) = (1-p_c)S_t(Z)\sum_{i=0}^{\alpha}(p_cC_t(Z))^i H_i(Z) + (p_cC_t(Z))^{\alpha+1}H_\alpha(Z) \end{aligned} \qquad . \qquad (18)$$

[1] The set of CW values shall be sequentially ascending integer power of 2, minus 1, beginning with CWmin, and continuing up to and including CWmax. [12]

Since $B(Z)$ can be expanded in power series, i.e.,

$$B(Z) = \sum_{i=0}^{\infty} \Pr(T_s = i) Z^i \,, \qquad (19)$$

we can obtain the arbitrary n^{th} moment of T_S by differentiation (hence the mean value and the variance), where the unit of T_S is slot. For example, the mean is given by

$$\mu^{-1} = E[T_S] = \frac{dB(Z)}{dZ}\Big|_{Z=1} \qquad (20)$$

where μ is the MAC layer service rate.

G. *Probability Distribution Modeling*

From the probability generation function (PGF) of the MAC layer service time, we can easily obtain the discrete probability distribution. Fig. 4 shows the probability distribution of the MAC service time at each discrete value. This example uses RTS/CTS mechanisms. The lengths of RTS/CTS/ACK conform to IEEE 802.11 MAC protocol. Data packet length is 1000 bytes and data transmission rate is 2 Mbps. The values of the parameters are summarized in Table I.

We notice that the envelope of the probability distribution is similar to an exponential distribution. If we use some continuous distribution to approximate the discrete one, it will give us great convenience to analyze the queueing characteristics. Fig.4 shows the approximate probability density distribution (PDF) of T_S and several well-known continuous PDFs including Gamma distribution, log-normal distribution, exponential distribution and Erlang-2 distribution. We observe that the log-normal distribution provides a good approximation for almost all cases (not only for cases at the high collision probability but also for cases at the low collision probability), and also has a very close tail distribution match with that of T_S. In addition, the exponential distribution seems to provide a reasonably good approximation except for cases at very low collision probability, where it is more like a deterministic distribution. Here, the PDF of T_S is obtained by assuming that the probability density function is uniform in a very small interval and is represented by a histogram while other continuous PDF is determined by the value of mean and/or variance of T_S. Here, we use 5 ms as the interval in the histogram because the distribution of the delay concentrates around the integer times of the successful transmission period for each packet which approximates 5 ms for packets with 1000 bytes long.

We also notice that p_c has different saturation values for different n. If the mobile station always has packets to transmit, i.e., in the saturation state, the idle probability p_0 takes the minimum value 0. So, according to formulae (14) and (17), we can obtain the saturation value of p_c by setting p_0 as 0 in Table II.

Fig.5 shows the distribution of T_S at different number of mobile stations, which mainly depends on p_c and hardly depends on n. Fig.6 shows the mean value of T_S at different collision probability. The maximum of T_S for different n, which is reached when p_c takes the saturation value, is marked. We observe that the distribution of T_S mainly depends on p_c and is determined by the number of the active stations at saturation status when p_c reaches the saturation value. We will discuss how to obtain the value of p_c at different traffic load in the following section.

IV. QUEUEING MODELING AND ANALYSIS

A. *Problem formulation:*

Many applications are sensitive to end-to-end delay and queue characteristics such as average queue length, waiting time, queue blocking probability, service time, and goodput. Thus, it is necessary to investigate the queueing modeling and analysis for wireless LANs to obtain such performance metrics.

A queue model can be characterized by the arrival process and the service time distribution with certain service discipline. We have characterized the MAC layer service time distribution in the previous section. In this paper, we assume that the packet arrivals at each mobile station follow the Poisson process or a deterministic distribution with average arrival rate λ. The packet transmission process at each station can be modeled as a general single "server". The buffer size at each station is K. Thus, the queueing model for each station can be modeled as an M/G/1/K when Poisson arrivals of packets are assumed.

B. The steady-state probability of the M/G/1/K queue

Let p_n represent the steady-state probability of n packets in the queueing system, and let π_n represent the probability of n packets in the queueing system upon a departure at the steady state, and let $\boldsymbol{P}=\{p_{ij}\}$ represent the queue transition probability matrix:

$$\mathrm{p_{ij}} = \Pr\{\mathrm{X_{n+1}} = \mathrm{j} \mid \mathrm{X_n} = \mathrm{i}\}, \qquad (21)$$

where X_n denotes the number of packets seen upon the nth departure.

To obtain p_{ij}, we define

$$\begin{aligned} \mathrm{k_n} &= \Pr\{\text{n arrivals during service time } Ts\} \\ &= \sum_{\mathrm{i}=0}^{\infty} \frac{\mathrm{e}^{-\lambda i}(\lambda i)^n}{n!} \Pr\{Ts = i\} \end{aligned}, \qquad (22)$$

where λ is the average arrival rate. We can easily obtain

$$\mathbf{P} = \{p_{ij}\} = \begin{bmatrix} k_0 & k_1 & k_2 & \cdots & k_{K-2} & 1-\sum_{n=0}^{K-2} k_n \\ k_0 & k_1 & k_2 & \cdots & k_{K-2} & 1-\sum_{n=0}^{K-2} k_n \\ 0 & k_0 & k_1 & \cdots & k_{K-3} & 1-\sum_{n=0}^{K-3} k_n \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & k_0 & 1-k_0 \end{bmatrix}. \qquad (23)$$

Moreover, we notice that

$$k_0 = B(e^{-\lambda}), k_n = \frac{\lambda^n}{(-1)^n n!} \frac{\partial^n B(e^{-\lambda})}{\partial \lambda^n}. \qquad (24)$$

where $B(e^{-\lambda})$ is obtained by replacing Z with $e^{-\lambda}$ in equation (18), i.e., the PGF of the MAC layer service time T_s.

According to the balance equation:

$$\pi P = \pi, \qquad (25)$$

where $\boldsymbol{\pi} = \{\pi_n\}$ and the normalization equation, we can compute the $\boldsymbol{\pi}$. For the finite system size K with Poisson input, we have [15]

$$p_0 = \frac{\pi_0}{\pi_0 + \rho}, p_n = \frac{\pi_n}{\pi_0 + \rho}(0 \le n \le K-1), p_K = 1 - \frac{1}{\pi_0 + \rho}, \quad (26)$$

where ρ is the traffic intensity and $\rho = \lambda E[T_S]$.

If we can approximate the distribution of MAC service time by an exponential distribution, the steady-state probability for the M/M/1/K model [15] is given by:

$$p_0 = [\Sigma_{i=0}^{K} \rho^i]^{-1}, \; p_i = (\rho)^i p_0, \; (0 \le i \le K). \quad (27)$$

C. *Conditional Collision Probability p_c and Distribution of MAC Layer Service Time*

From above derivation, we know that p_0 is a function of p_c, λ, n. So we can compute p_c under different values of λ and n with the help of (14) and (17) using some recursive algorithm. Thus, we can obtain the distribution of MAC service time at different offered load according to the results obtained in section III.

D. *Performance Metrics of the Queueing System*

The average queue length, blocking probability, and average waiting time including MAC service time are given by

$$L = \sum_{i=0}^{K} i \times p_i, \; p_B = p_k = 1 - \frac{1}{\pi_0 + \rho}, W = \frac{L}{\lambda(1 - P_B)} \quad (28)$$

E. *Throughput*

If we know the blocking probability p_B, then the throughput S at each station can be computed easily by

$$S = \lambda(1 - p_B)(1 - p_c^{\alpha+1}), \quad (29)$$

where $p_c^{\alpha+1}$ is the packet discard probability due to transmission failures.

F. Numerical Results

Fig.7 shows the results for the major performance metrics. All of them have a dramatic change around the traffic load of 1.1-1.5 Mbits/sec. This is because that the collisions increase significantly around this traffic load, resulting in much longer MAC service time for each packet.

From the results, we observe that all the metrics are dependent on the collision probability p_c. Fig.7 shows that p_c mainly depends on the total traffic in the non-saturated scenario. On the other hand, p_c is affected by the total number of packets attempting to transmit by all neighboring stations. In the non-saturated case, when all arriving packets are immediately served by the MAC layer, the queue length may reach zero and the corresponding station will not compete for the medium. However, in the saturated scenario, i.e., the stations always have packets to transmit, the total number of packets attempting to transmit equals to the total number of neighboring stations, hence p_c is mainly dependent on the total number of neighboring stations as we expect.

The MAC layer service time shows similar change at different offered load, because it is dependent on the p_c. All other performance metrics are dependent on the distribution of the MAC layer service time, so they show the similar change in the figures. The average queue length is almost zero at the non-saturated state and reaches almost maximum length at the saturated state. The average waiting time for each packet in the queue almost equals to zero at the non-saturated state and reaches several seconds at the saturated state. The queue blocking probability is zero at the non-saturated state when the traffic load is low, and linearly increases with the offered load at the saturated state. The throughput linearly increases with the offered load at the non-saturated state and maintains a constant value with different total number of transmitting stations at the saturated state. The throughput at saturated state decreases when the number of stations increases because collision probability climbs up with the number of stations. This is consistent with the results of saturation throughput in [9] where the author indicates that the saturated throughput decreases as n increases under a small initial size of the backoff window given a specific set of system parameters. In addition, the packet discarding probability at MAC layer is much smaller than the queue blocking probability.

In summary, all these results indicate that IEEE 802.11 MAC works well in the non-saturated state at low traffic load while its performance dramatically degrades at the saturated state, especially for the delay metric.

Besides, at the non-saturated state, the performance is dependent on the total traffic and indifferent to the number of transmitting stations. At the saturated state, the number of transmitting stations is much more important to the whole performance. The similar phenomena have been observed for the distribution of MAC service time shown in section III.

V. PERFORMANCE EVALUATION

A. *Simulation Enviroments*

In our simulation study, we use the ns-2 package [16]. The wireless channel capacity is set to 2Mbps. Data packet length is 1000 bytes, and the maximum queue length is 50. The radio propagation model is Two-Ray Ground model. We use different numbers of mobile stations in a rectangular grid with dimension 150m x 150m to simulate the Wireless LAN. All stations have the same rate of packet inputs. The MAC protocol uses the RTS/CTS based 802.11 MAC and other parameters are summarized in Table I.

B. *Probability Distribution of MAC Layer Service Time*

Fig.8 shows the simulation results of the MAC layer service time in the network with 17 mobile stations and total traffic of 0.2, 0.8 and 1.6 Mbps, respectively. It displays good match on the probability density functions between the analytical result and that from simulation. Notice that, similarly with Fig. 4, the PDFs shown in Fig. 8 are histogram approximations of the discrete probability distribution obtained from both analysis and simulations.

Our results indicate the distribution of MAC layer service time is independent of the packet input distribution whether it is deterministic or Poisson distributed. It mainly depends on the total traffic in the network before saturation and on the number of mobile stations after saturation, which is consistent with the analysis.

C. *Comparison of M/G/1/K and M/M/1/K approximations with simulation results*

Exponential distribution is a memoryless distribution. If we can model the MAC layer service time as this distribution, it will give us great convenience to predict the system performance, such as throughput, link delay, packet discarding ratio. The problem is how good this approximation is for our modeling.

As we said in section IV, the exponential distribution seems to be a good approximation for the MAC layer service time. In Fig. 9 and 10, we compare it with the derived discrete probability distribution in the queueing analysis to check its goodness to predict the MAC throughput, packet waiting time, queue blocking probability and average queue length. Here, we assume that the system has Poisson arrivals. We use two queueing models for these two distributions: M/M/1/K and M/G/1/K. Fig. 9 and 10 show the results for the WLAN with 9 mobile stations.

From Figs. 9 and 10, we observe that M/M/1/K model give a close approximation to the M/G/1/K model for some performance metrics. Both models have almost the same throughput and queue blocking probability. However, when the mobile stations are at the saturated state, M/M/1/K gives a large prediction error for the average queue length and average waiting time, and the difference is small except at the turning point between non-saturated state and the saturated state, where a dramatic change of the system performance is shown. The M/G/1/K model always provides better approximation for all performance metrics.

We also compare the results of queueing models with the simulation in Fig. 9 and 10. Two queueing models show very close approximations with the simulation results for all performance metrics when mobile stations are in the non-saturated state. However, there are distinct differences between them when the system is in the saturation state. This is because that the Markov chain model overestimates the average MAC layer service time about 10 % in the saturation state compared to the simulation results from ns-2, as showed in Fig.11. The reasons may be that the Markov chain model does not capture all the protocol details and/or the implementation considerations of IEEE 802.11 MAC protocols in ns-2. Thus, the simulation results have higher throughput, lower queue blocking probability, smaller average queue length and smaller average waiting time at saturated state.

With extensive simulations for different number of mobile stations in randomly generated wireless LANs, we have concluded that the Markov chain models seem to always give an upper bound of the average MAC layer service time. Thus, the queueing models using the distribution of the service time give a lower bound of the throughput, and upper bounds of queueing blocking probability, average queue length and average waiting time compared with simulations of ns-2. Therefore, our analytical models can always be useful to come up with the performance estimates for design purpose.

VI. CONCLUSIONS

In this paper, we have derived the probability distribution of the MAC layer service time. To obtain this distribution, we use the signal transfer function of generalized state transition diagram and expand the Markov chain model to the more general case for the exponential backoff procedure in IEEE 802.11 MAC protocols. Accurate discrete probability distribution and approximate continuous probability distributions are obtained in this paper. Based on the distribution of the MAC layer service time, we come up with a queueing model and evaluate the performance of the IEEE 802.11 MAC protocol in Wireless LANs in terms of throughput, delay, and other queue performance metrics. Our results show that at the non-saturated state, the performance is dependent on the total traffic and indifferent to the number of transmitting stations, and at saturated state, the number of transmitting stations affects the performance more significantly.

In addition, the analytical results indicate that exponential distribution may provide a good approximation for the MAC layer service time in the queueing analysis. The queueing models discussed in this paper can accurately estimate various performance metrics of WLAN in the non-saturated state which is the desired state for some application with a certain QoS requirement because there is no excessive queueing delay as that in saturated state. And for WLANs in the saturated state, the queueing models give a lower bound for the throughput, and upper bounds for queueing blocking probability, average queue length and average waiting time compared with simulation results obtained from ns-2.

REFERENCES

[1] J. Broch, D.A. Maltz, D.B. Johnson, Y. Hu, and J. Jetcheva, "A performance Comparison of Multihop Wireless Ad Hoc Network Routing Protocols," *Proc. IEEE/ACM MOBICOM'98*, Oct. 1998

[2] C. Perkins, E.M. Royer, S.R. Das, and M.K. Marina, "Performance Comparison of Two On-demand Routing Protocols for Ad Hoc Networks", *IEEE Personal Communications*, Feb. 2001

[3] G. Holland and N. Vaidya, "Analysis of TCP Performance over Mobile Ad Hoc Networks," *Proc. IEEE/ACM MOBICOM'99*, 1999

[4] S. Xu and T. Safadawi, "Does the IEEE 802.11 MAC Protocol Work Well in Multihop Wireless Ad Hoc Networks?" *IEEE Communications Magazine*, June 2001

[5] Z. Fu, P. Zerfos, H. Luo, S. Lu, L. Zhang, and M. Gerla, "The Impact of Multihop Wireless Channel on TCP Throuhput and Loss," *Proc. IEEE INFOCOM'2003*, March 2003.

[6] S. Pilosof, R. Ramjee, D. Raz, Y. Shavitt, and P. Sinha, "Understanding TCP fairness over Wireless LAN," *Proc. IEEE INFOCOM'2003, March 2003*

[7] F. Cali, M. Conti, and E. Gregori, "IEEE 802.11 Protocol: Design and Performance Evaluation of an Adaptive Backoff Mechanism," *IEEE journal on Selected Area in communications*, V18, N9, September 2000

[8] F. Cali, M. Conti, and E. Gregori, "IEEE 802.11 Wireless LAN: Capacity Analysis and Protocol Enhancement," *Proc. IEEE INFOCOM'1998*, March, 1998

[9] G. Bianchi, "Performance Analysis of the IEEE 802.11 Distributed Coordination Function," *IEEE Journal on Selected Area in Communications*, V18, N3, March 2000

[10] C. H. Foh and M. Zukerman, "Performance Analysis of the IEEE 802.11 MAC Protocol," *European Wireless 2002*, Feb. 2002, Florence, Italy

[11] Z. Hadzi-Velkov and B, Spasenovski, "Saturation Throughput – Delay Analysis of IEEE 802.11 DCF in Fading Channel," *Proc. IEEE ICC'2003*, May, 2003.

[12] *IEEE standard for Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications*, ISO/IEC 8802-11: 1999(E), Aug. 1999

[13] D. W. Choi, "Frame alignment in a digital carrier system-a tutorial", *IEEE Communications Magazine*, Volume: 28 2, Feb. 1990

[14] H. Wu, Y. Peng, K. Long, S. Cheng, and J. Ma, "Performance of reliable transport protocol over IEEE 802.11 wireless LAN: analysis and enhancement," *Proc. IEEE INFOCOM'2002*, June, 2002.

[15] D. Gross and C. M. Harris, "*Fundamentals of Queueing Theory*," 3rd ed., John Wiley & Sons, Inc, 1998

[16] K. Fall and K. Varadhan, editors, "*NS Notes and Documentation*," The VINT Project, UC Berkeley, LBL, USC/ISI, and Xerox PARC, April, 2002

[17] H. Kim and J. Hou, "Improving protocol capacity with model-based frame scheduling in IEEE 802.11-operated WLANs," *Proc. of ACM MobiCom'2003*, Sep. 2003.

[18] L.P.A. Robichaud, *Signal Flow Graphs and Applications*, Prentice-Hall, 1962

APPENDIX I. DERIVATION OF TRANSMISSION PROBABILITY

This section derives the transmission probability τ, i.e., the packet transmission probability that the station transmits in a randomly chosen slot time given that it has packets to transmit. We follow the similar notations in paper [9]. *{s(t), b(t)}* and W_i have been defined in section III. F. Let $\mathrm{P}\{i_1,k_1 \mid i_0,k_0\}$ be the short notation of one-step transition probability and $\mathrm{P}\{i_1,k_1 \mid i_0,k_0\} = \Pr\{s(t+1)=i_1, b(t+1)=k_1 \mid s(t)=i_0, b(t)=k_0\}$. Then the only non null one-step transition probabilities are

$$\begin{cases} \mathrm{P}\{i,k \mid i,k+1\} = 1 & k\in[0,W_i-2] \;\; i\in[0,\alpha] \\ \mathrm{P}\{0,k \mid i,0\} = (1-p_c)/W_0 & k\in[0,W_0-1] \;\; i\in[0,\alpha-1] \\ \mathrm{P}\{i,k \mid i-1,0\} = p_c/W_i & k\in[0,W_i-1] \;\; i\in[1,\alpha] \\ \mathrm{P}\{0,k \mid \alpha,0\} = 1/W_0 & k\in[0,W_0-1] \end{cases} \tag{30}$$

These equations account for the facts that: the backoff timer is decremented; the backoff timer starts from stage 0 after a successful transmission; the backoff timer starts from a new stage after an unsuccessful transmission; the contention window size is reset and the backoff timer starts from stage 0 when the maximum transmission failure limit is reached, respectively.

Let $b_{i,k} = \lim_{t\to\infty} \Pr\{s(t)=i, b(t)=k\}, 0\le i\le\alpha, 0\le k< W_i$ be the stationary distribution of the Markov chain. First, note that

$$b_{i-1,0} \cdot p_c = b_{i,0} \rightarrow b_{i,0} = p_c{}^i b_{0,0} \quad 0 < i \le \alpha \;, \tag{31}$$

and

$$b_{i,k} = \frac{W_i - k}{W_i} \cdot \begin{cases} b_{\alpha,0} + (1-p_c)\sum_0^{\alpha-1} b_{j,0} & i = 0 \\ p_c \cdot b_{i-1,0} & 0<\mathrm{i} \le \alpha \end{cases} \tag{32}$$

By means of equation (31), equation (32) can be simplified as

$$b_{i,k} = \frac{W_i - k}{W_i} b_{i,0} \quad (0 \le \mathrm{i} \le \alpha, \;\; 0 \le k \le W_i - 1) \tag{33}$$

Thus, $b_{0,0}$ can be finally determined by imposing the normalization condition, that simplifies as follows:

$$\begin{aligned} 1 &= \sum_{i=0}^{\alpha}\sum_{k=0}^{W_i-1} b_{i,k} = \sum_{i=0}^{\alpha} b_{i,0} \sum_{k=0}^{W_i-1} \frac{W_i - k}{W_i} = \sum_{i=0}^{\alpha} b_{i,0} \frac{W_i+1}{2} = \frac{b_{0,0}}{2}\sum_{i=0}^{\alpha} p_c{}^i \frac{W_i+1}{2} \\ &= \frac{b_{0,0}}{2} \begin{cases} \sum_{i=0}^{\alpha} p_c{}^i \frac{2^i W + 1}{2} & \alpha \le \mathrm{m} \\ \sum_{i=0}^{m-1} p_c{}^i \frac{2^i W + 1}{2} + \sum_{i=m}^{\alpha} p_c{}^i \frac{2^m W_i + 1}{2} & \alpha > \mathrm{m} \end{cases} \end{aligned} \tag{34}$$

As any transmission occurs when the backoff time counter equals zero, regardless of the backoff stage, the probability τ that a station, which has packets to transmit, transmits in a randomly chosen slot time is

$$\tau = \sum_{i=0}^{\alpha} b_{i,0} = \frac{1 - p_c{}^{\alpha+1}}{1 - p_c} b_{0,0} \,, \tag{35}$$

which can be simplified as

$$\tau = \begin{cases} \dfrac{2(1-p_c{}^{\alpha+1})}{1-p_c{}^{\alpha+1} + (1-p_c)W(\sum_{i=0}^{\alpha}(2p_c)^i)} & ,\alpha \le m \\ \dfrac{2(1-p_c{}^{\alpha+1})}{1-p_c{}^{\alpha+1} + p_c W \sum_{i=0}^{m-1}(2p_c)^i + W(1-2^m p_c{}^{\alpha+1})} & ,\alpha > m \end{cases} \tag{36}$$

TABLE I. IEEE 802.11 SYSTEM PARAMETERS

Channel Bit Rate	2 Mbit/s
PHY header	192 bits
MAC header	224 bits
Packet payload size	1000Bytes
Length of RTS	160bits + PHY header
Length of CTS	112bits + PHY header
Length of ACK	112bits + PHY header
Initial backoff window size (W)	31
Maximum backoff stages (m)	5
Short retry limit	7
Long retry limit	4

TABLE II. SATURATION VALUE OF COLLISION PROBABILITY

n	5	9	17	33	65
Max p_c	0.1781	0.2727	0.3739	0.4730	0.5692

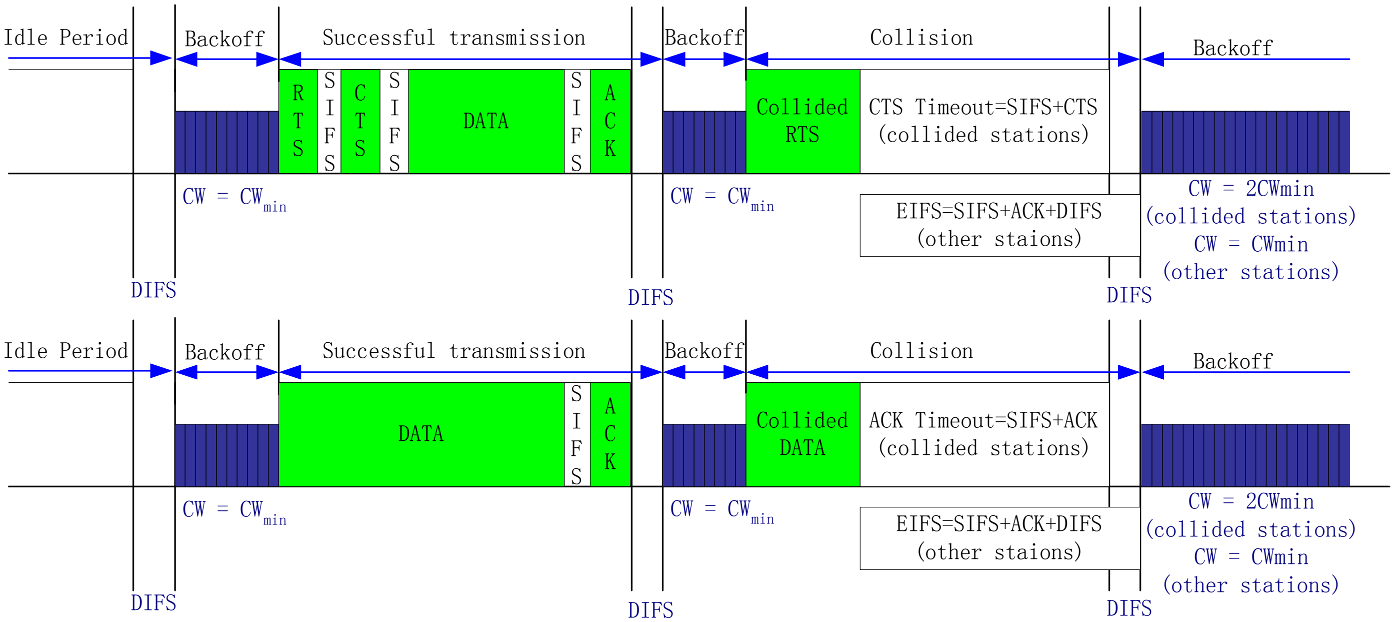

Figure 1. RTS/CTS mechanism and basic access mechanism of IEEE 802.11

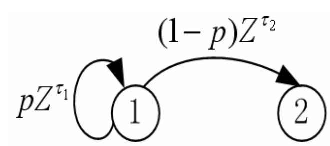

Figure 2. Generalized state transition diagram of one example

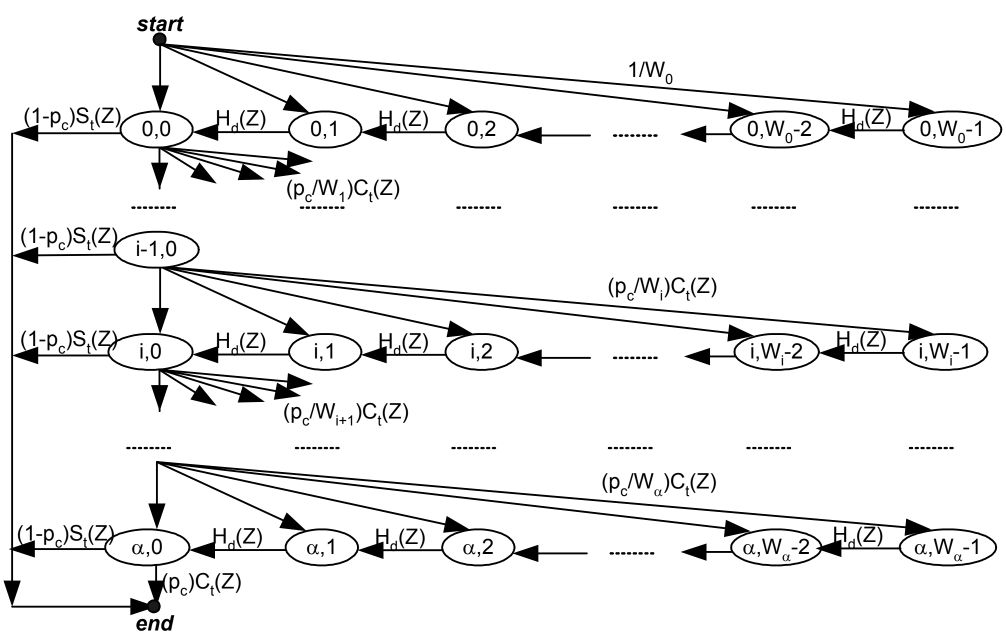

Figure 3. Generalized state transition diagram for transmission process

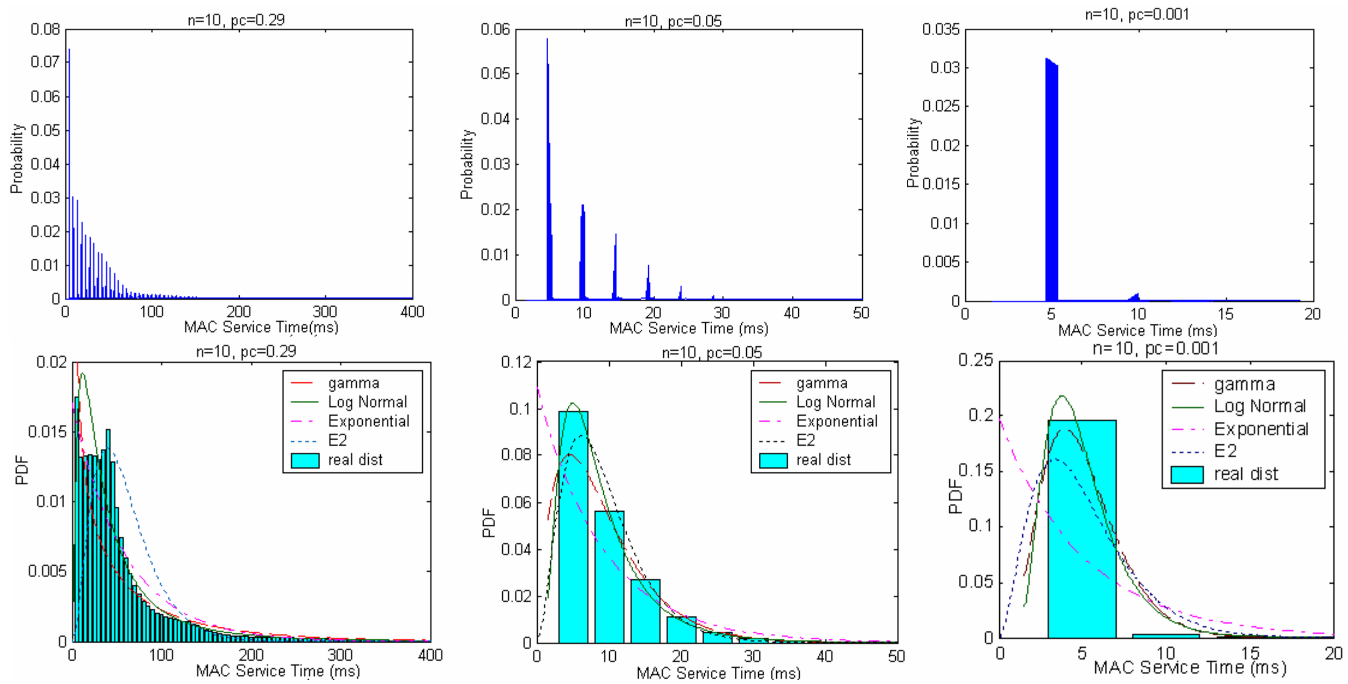

Figure 4. Probability Distribution of MAC Layer Service Time

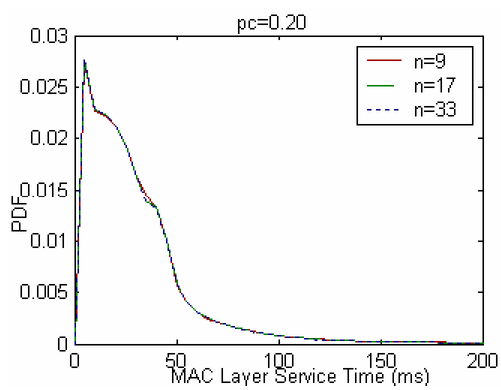

Figure 5. PDF of Service Time

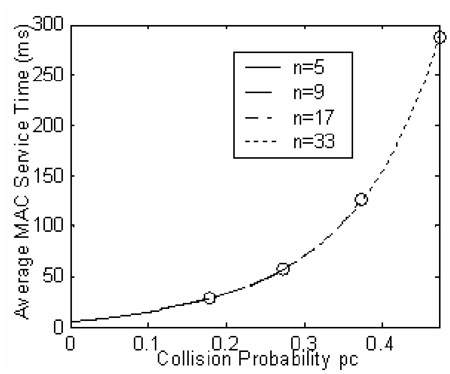

Figure 6. Mean of Service Time

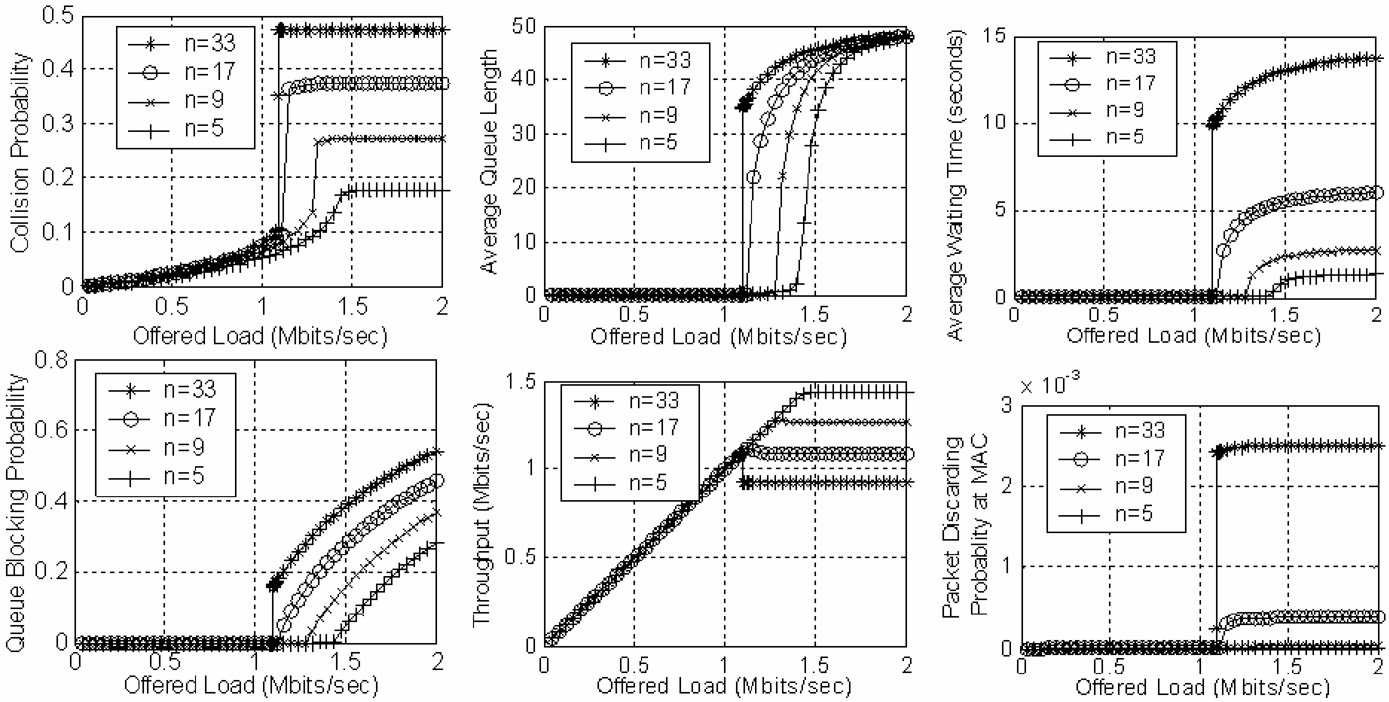

Figure 7. Queue Characteristics

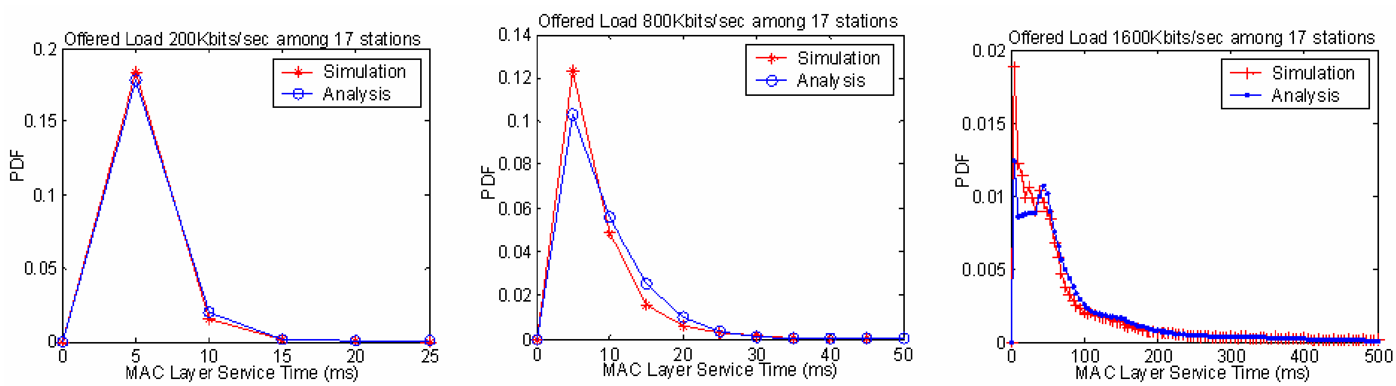

Figure 8. MAC Layer Packet Service Time

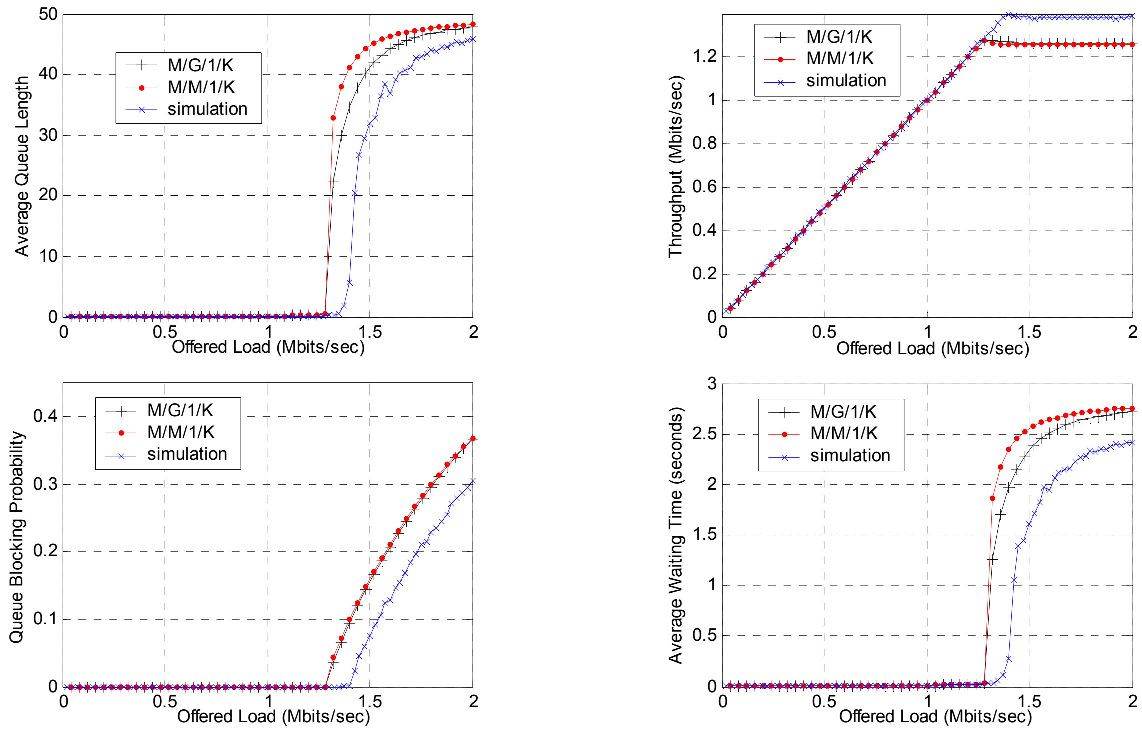

Figure 9. Comparisons between M/G/1/K, M/M/1/K models and simulation

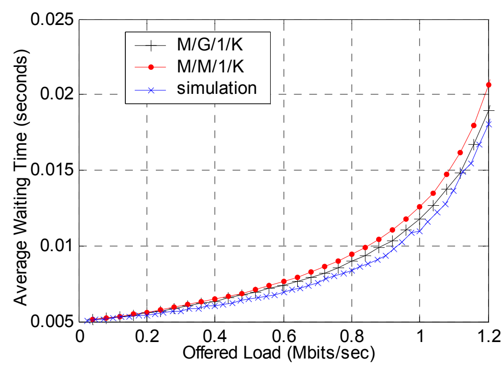

Figure 10. Average waiting time in non-saturated status

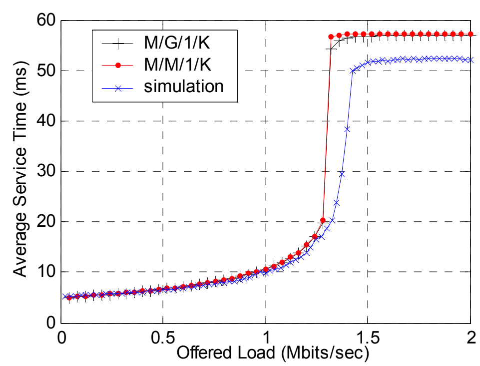

Figure 11. Average MAC layer service time