How Well Can the IEEE 802.11 Wireless LAN Support Quality of Service?

Hongqiang Zhai, *Student Member, IEEE*, Xiang Chen, *Member, IEEE*, and Yuguang Fang, *Senior Member, IEEE*

***Abstract*—This paper studies an important problem in the IEEE 802.11 distributed coordination function (DCF)-based wireless local area network (WLAN): how well can the network support quality of service (QoS). Specifically, this paper analyzes the network's performance in terms of maximum protocol capacity or throughput, delay, and packet loss rate. Although the performance of the 802.11 protocol, such as throughput or delay, has been extensively studied in the saturated case, it is demonstrated that maximum protocol capacity can only be achieved in the nonsaturated case and is almost independent of the number of active nodes. By analyzing packet delay, consisting of medium access control (MAC) service time and waiting time, accurate estimates were derived for delay and delay variation when the throughput increases from zero to the maximum value. Packet loss rate is also given for the nonsaturated case. Furthermore, it is shown that the channel busyness ratio provides precise and robust information about the current network status, which can be utilized to facilitate QoS provisioning. The authors have conducted a comprehensive simulation study to verify their analytical results and to tune the 802.11 to work at the optimal point with maximum throughput and low delay and packet loss rate. The simulation results show that by controlling the total traffic rate, the original 802.11 protocol can support strict QoS requirements, such as those required by voice over Internet protocol (VoIP) or streaming video, and at the same time achieve high channel utilization.**

***Index Terms*—IEEE 802.11, protocol capacity, quality of service (QoS), wireless local area networks (WLANs).**

I. Introduction

BECAUSE of its simple deployment and low cost, the IEEE 802.11 wireless local area network (WLAN) [12] has been widely used in recent years. It contains two access methods, i.e., distributed coordination function (DCF) and point coordination function (PCF), with the former being specified as the fundamental access method. Despite its popular use, currently only best effort traffic is supported in DCF. Section II describes the 802.11 protocol in more detail.

For the IEEE 802.11 WLAN to continue to thrive and evolve as a viable wireless access to the Internet, quality of service (QoS) provisioning for multimedia services is crucial. As shown in Table I, for real-time, streaming, and nonreal-time (or best effort) traffic, the major QoS metrics include bandwidth, delay, delay jitter, and packet loss rate [14], [15]. Guaranteeing QoS for multimedia traffic, however, is not an easy task given that the 802.11 DCF is in nature contention based and distributed, and hence renders effective and efficient control very difficult. In addition, other problems such as hidden terminals or channel fading make things worse. To address these challenges, current research works ([1], [23], [30] and references therein) and the enhanced DCF (EDCF) defined in the IEEE 802.11e draft [7], [13] tend to provide differentiated service rather than stringent QoS assurance.

However, we have not yet well understood the question of how well the IEEE 802.11 WLAN can support QoS when many researchers start to believe that service differentiation is the best that the 802.11 can achieve. In this paper, we endeavor to address this problem through both theoretical analysis (Section III) and simulations (Section IV).

We develop an analytical model to assess the capability of 802.11 for supporting major QoS metrics, i.e., throughput, delay and delay variation, and packet loss rate. While current literature on performance analysis is focused on the derivation of throughput or delay in the saturated case, we find that the optimal operating point for the 802.11 to work at lies in the nonsaturated case.[1] At this point, we analytically show that maximum throughput is achieved almost independent of the number of active nodes and that delay and delay variation are low enough to satisfy stringent QoS requirements of real-time traffic. Thus, 802.11 WLAN can perform very well in supporting QoS, as long as it is tuned to the optimal point. Since an accurate indicator of the network status is essential to effective tuning, we also demonstrate that the channel busyness ratio, which is easy to obtain and accurately and timely represents network utilization, can be used to design schemes such as call admission control or rate control in WLAN. Due to page limit, we will present such schemes in a subsequent paper.

In Section V, we show that our analytical results are still valid even when the effect of channel fading is taken into account. Also, we discuss the possible implications arising due to the employment of a prioritized 802.11 DCF. Finally, Section VI concludes this paper.

Manuscript received May 24, 2004; revised October 12, 2004: accepted November 15, 2004. The editor coordinating the review of this paper and approving it for publication is W. Liao. This work was supported in part by the U.S. Office of Naval Research under the Young Investigator Award N000140210464 and under Grant N000140210554, and by the National Science Foundation under Faculty Early Career Development Award ANI-0093241 and under Grant ANI-0220287.

The authors are with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611-6130 USA (e-mail: zhai@ecel.ufl.edu; xchen@ecel.ufl.edu; fang@ece.ufl.edu).

Digital Object Identifier 10.1109/TWC.2005.857994

[1]Note that a similar fact has been observed for Aloha or Slotted Aloha, where maximum throughput is achieved only when traffic arrives at a certain rate [2].

TABLE I
QoS REQUIREMENTS FOR MULTIMEDIA SERVICES

Class	Application	One-way transmission delay	Delay variation	Packet loss rate
Real-time	VoIP, videoconferencing	< 150ms(preferred), < 400ms(limit)	$1ms^*$	1%(video), 3%(audio)
Streaming	Streaming audio and video	up to $10s$	$1ms^*$	1%
Best effort	E-mail, file transfer, web browsing	minutes or hours	N/A	Zero
* Playout buffer (or jitter buffer) can be used to compensate for delay variation				

II. PRELIMINARIES

A. Operations of the IEEE 802.11

The basic access method in the IEEE 802.11 MAC protocol is DCF, which is based on carrier sense multiple access with collision avoidance (CSMA/CA). Before starting a transmission, each node performs a backoff procedure, with the backoff timer uniformly chosen from [0, CW − 1] in terms of time slots, where CW is the current contention window. When the backoff timer reaches zero, the node transmits a DATA packet. If the receiver successfully receives the packet, it acknowledges the packet by sending an acknowledgment (ACK). If no ACK is received within a specified period, the packet is considered lost; so the transmitter will double the size of CW, choose a new backoff timer, and start the above process again. When the transmission of a packet fails for a maximum number of times, the packet is dropped. To avoid collisions of long packets, the short request to send/clear to send (RTS/CTS) frames can be employed.

Note that the IEEE 802.11 MAC also incorporates an optional access method called PCF, which is only usable in infrastructure network configurations and is not supported in most current wireless cards. In addition, it may result in poor performance as shown in [24] and [27]. In this paper, we thus focus on the 802.11 DCF.

B. Related Work

To date, two threads of research have examined the property or performance of the IEEE 802.11: performance analysis, and performance and/or QoS enhancements.

Performance analysis: The first thread was devoted to building analytical models to characterize the behavior of 802.11 and deriving protocol capacity or delay performance [4], [6], [9]–[11], [25], [28], [29]. In [4], Bianchi proposed a Markov chain model for the binary exponential backoff procedure. By assuming that the collision probability of each node's transmission is constant and independent of the number of retransmissions, he derived the saturated throughput for the IEEE 802.11 DCF. Based on the saturated throughput derived in Bianchi's model, Foh and Zuckerman [9] used a Markovian state dependent single server queue to analyze the throughput and mean packet delay. Cali *et al.* [6] studied the 802.11 protocol capacity by using a p-persistent backoff strategy to approximate the original backoff in the protocol. Again, the focus is on the saturated throughput. In addition to collisions, Hadzi-Velkov and Spasenovski took the effect of frame error rate into account in their analysis of saturated throughput and delay [10]. We derived an approximate probability distribution of the service time and, based on the distribution, analyzed the throughput and average delay [28], [29]. As noticed, most works were focused on the analysis of throughput and delay in the saturated case. Moreover, none of these systematically considered the delay and delay variation in the nonsaturated case, let alone obtained accurate estimates for them.

Performance and/or QoS enhancements: The second thread of the research on the 802.11 DCF explored various ways to improve throughput [3], [5], [17], [20] or provide prioritized service, namely, service differentiation [1], [16], [21]–[23], [26].

Based on the work in [6], Cali *et al.* attempted to approach the protocol capacity by replacing the exponential backoff mechanism with an adaptive one [5]. Kim and Hou developed a model-based frame-scheduling algorithm to improve the protocol capacity of 802.11 [17]. Two fast collision resolution schemes were proposed by Bharghavan [3] and Kwon *et al.* [20]. The idea is to use two channels or to adjust backoff algorithms to avoid collisions, thereby providing higher channel utilization.

To provide service differentiation, Ada and Castelluccia [1] proposed to scale the contention window and use a different interframe spacing or maximum frame length for services of different priorities. In [23], two mechanisms, i.e., virtual medium access control (MAC) and virtual source, were proposed to enable each node to provide differentiated services for voice, video, and data. By splitting the transmission period into a real-time one and a nonreal-time one, the real-time traffic is supported with QoS guarantee [21]. However, the DCF mode was dramatically changed. The Blackbust in [22] provided high priority for real-time traffic. Unfortunately, it imposes special requirements on high-priority traffic and is not fully compatible with the existing 802.11 standard. In summary, if the semantics of the 802.11 DCF is maintained, all the works mentioned above can only support service differentiation.

Our paper can be considered to be a convergence between these two threads of research; however, it improves on both sides. We thoroughly study the QoS performance of 802.11 in terms of throughput, delay and delay variation, and packet loss rate. Moreover, we discover the optimal operating point at which, in addition to achieving the theoretical maximum throughput, 802.11 WLAN is capable of supporting strict QoS requirements for real-time traffic rather than only providing prioritized service.

III. ANALYTICAL STUDY OF THE IEEE 802.11

This section focuses on the analysis of the performance of the IEEE 802.11 DCF. Note that in the following analysis, the hidden terminal problem is ignored. This is because in a typical wireless LAN environment, every node can sense all the others' transmissions, although it may not necessarily be able to correctly receive the packets from all other nodes.

A. Maximum Throughput and Available Bandwidth

To simplify the analysis and yet reveal the characteristics of the IEEE 802.11 MAC protocol, we assume that the traffic is uniformly distributed among the nodes. The total number of nodes is n. The transmission probability for each node in any time slot is p_t. Note that here a time slot at the MAC layer could be an empty backoff time slot, a period associated with successful transmission, or a period associated with collision [4], [12]. Obviously, we obtain

$$\begin{cases} p_i = (1-p_t)^n \\ p_s = np_t(1-p_t)^{n-1} \\ p_c = 1 - p_i - p_s \end{cases} \tag{1}$$

where p_i is the probability that the observed backoff time slot is idle, p_s is the probability that there is one successful transmission, and p_c is the collision probability that there are at least two concurrent transmissions at the same backoff time slot. If we define T_{suc} as the average time period associated with one successful transmission and T_{col} as the average time period associated with collisions, we know [12]

$$\begin{aligned} T_{\text{suc}} &= \text{rts} + \text{cts} + \overline{\text{data}} + \text{ack} + 3\text{sifs} + \text{difs} \\ T_{\text{col}} &= \text{rts} + \text{sifs} + \text{cts} + \text{difs} = \text{rts} + \text{eifs} \end{aligned} \tag{2}$$

for the case where the RTS/CTS mechanism is used and

$$\begin{aligned} T_{\text{suc}} &= \overline{\text{data}} + \text{ack} + \text{sifs} + \text{difs} \\ T_{\text{col}} &= \overline{\text{data}^*} + \text{ack\_timeout} + \text{difs} \end{aligned} \tag{3}$$

for the case where there is no RTS/CTS mechanism, where $\overline{\text{data}}$ and $\overline{\text{data}^*}$ (please refer to [4] for the derivation of $\overline{\text{data}^*}$) are the average length, in seconds, for the successful transmission and collision of the data packets, respectively. Thus, it can easily be obtained that

$$\begin{cases} R_i = \frac{p_i\sigma}{p_i\sigma + p_sT_{\text{suc}} + p_cT_{\text{col}}} \\ R_b = 1 - R_i \\ R_s = \frac{p_sT_{\text{suc}}}{p_i\sigma + p_sT_{\text{suc}} + p_cT_{\text{col}}} \end{cases} \tag{4}$$

where σ is the length of an empty backoff time slot, R_i is the channel idleness ratio, R_b is the channel busyness ratio, and R_s is the channel utilization. Once we obtain R_s, the normalized throughput s is expressed as

$$s = \frac{R_s \times \overline{\text{data}}}{T_{\text{suc}}} \tag{5}$$

and the absolute throughput is s times the bit rate for data packets.

In most cases, we are more interested in the packet collision probability p observed at each individual node since it can be used to calculate QoS metrics for the traffic traversing the node.

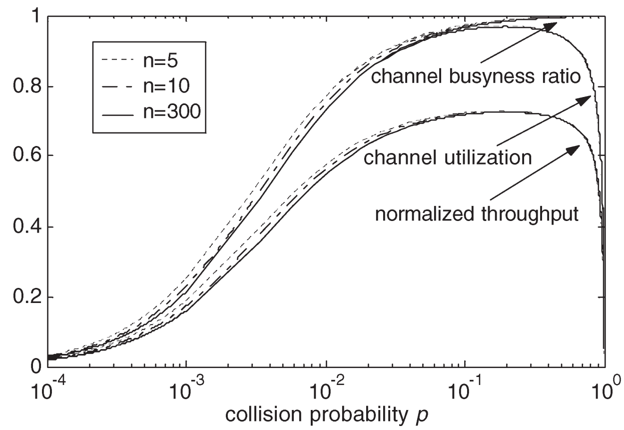


Fig. 1. Channel busyness ratio and utilization.

TABLE II
IEEE 802.11 System Parameters

Bit rate for DATA packets	2 Mbps
Bit rate for RTS/CTS/ACK	1 Mbps
PLCP Data rate	1 Mbps
Backoff Slot Time	20 μs
SIFS	10 μs
DIFS	50 μs
Phy header	192 bits
MAC header	224 bits
DATA packet	8000 bits + Phy header + MAC header
RTS	160 bits + Phy header
CTS, ACK	112 bits + Phy header

In other words, p is the probability that one node encounters collisions when it transmits. Also, p is the probability that there is at least one transmission among the neighbors in the observed backoff time slot. We thus link p to p_t as

$$p = 1 - (1-p_t)^{n-1}. \tag{6}$$

It can be seen that the collision probability increases with the increase in the number of neighboring nodes or in the traffic at each of these nodes. In this sense, p reflects the information about both the number of neighboring nodes and the traffic distribution at these nodes.

According to the above equations, we can express R_b, R_s, and s as a function of p, which are shown in Fig. 1. All the parameters involved are indicated in Table II and most are the default values in the IEEE 802.11. In Fig. 1, three cases, i.e., $n = 5$, 10, and 300, are considered. It is important to note that for each specific n, there exists a maximum value of p, denoted by MAX(p), at which the network operates in the saturated status, i.e., each of the n nodes always has packets in the queue and thus keeps contending for the channel. Based on the works in [4], [28], and [29], we know that in saturated status, the larger the number of nodes, the greater the collision probability. More precisely, MAX(p) $= 0.105$, 0.178, 0.290, 0.546, 0.701, 0.848 for $n = 3$, 5, 10, 50, 128, 300, respectively. Next, we present some important observations from Fig. 1.

1) Channel Busyness Ratio: An Accurate Sign of Network Utilization: First, we find that the channel busyness ratio is an injective function of the collision probability. In fact, this can easily be proved. Second, when $p \leqslant 0.1$, R_b is almost the same as R_s, namely

$$R_s \approx R_b. \tag{7}$$

This is not hard to understand. When the collision probability p is very small, the channel resource wasted in collisions is so minor that it can be ignored. Third, the normalized throughput almost stays unchanged when p increases from 0.1 to 0.2, although it reaches the maximum value around $p = 0.2$. Finally, maximum throughput is almost insensitive to the number of active nodes. Given these observations and the fact that throughput is proportional to R_s, we could therefore use the measured channel busyness ratio R_b to accurately estimate the throughput from zero to the maximum value. Note that this is very simple and useful to each node: it can monitor the throughput of the whole WLAN by simply measuring the channel busyness ratio, which can be easily done since IEEE 802.11 is a CSMA-based MAC protocol working on physical and virtual carrier sensing mechanisms. On the other hand, when R_b exceeds a certain threshold th_b, severe collisions can be observed in the WLAN.

2) Maximum Throughput: Fig. 1 also shows that throughput begins to decrease when p is greater than a certain value and could decrease to zero when p becomes very large. To ensure that the network is always working with a high throughput, it is important for us to find the critical turning point, i.e., when IEEE 802.11 will achieve the maximum throughput and how the maximum throughput depends on network characteristics such as the number of node n and traffic.

Combining (1), (4), and (6), we can write R_s as a function of p. To obtain maximum throughput, we take the derivative of R_s with respect to p and let it equal 0, i.e.,

$$\frac{d}{dp} R_s = 0. \tag{8}$$

Meanwhile, we know that p is upper bounded by $\mathrm{MAX}(p)$. Therefore, if p_{root} is the root of (8), we obtain the value of p, denoted by p^*, with which the maximum throughput is achieved, i.e.,

$$p^* = \mathrm{MIN}\left(p_{\mathrm{root}}, \mathrm{MAX}(p)\right). \tag{9}$$

By applying p^* to (5), we get the maximum normalized throughput of IEEE 802.11 at different n, as shown in Fig. 2(a) and (b). Here, two important points are noted.

Maximum throughput is achieved in the nonsaturated case rather than in the saturated case when $n > 5$: This is the very reason that we argue the network should work in nonsaturated case. When $n > 5$, the normalized throughput arrives at the maximum value around $p = 0.196$, much smaller than the collision probability in the saturated status, i.e., $\mathrm{MAX}(p)$, as clearly seen in Fig. 2(a). $p = 0.196$ means that there are five or six nodes simultaneously contending for the channel, which can be derived from the inverse function of $\mathrm{MAX}(p)$ as shown earlier. In addition, the maximum throughput achieved is not sensitive to the number of nodes n. It is rather stable as n increases.

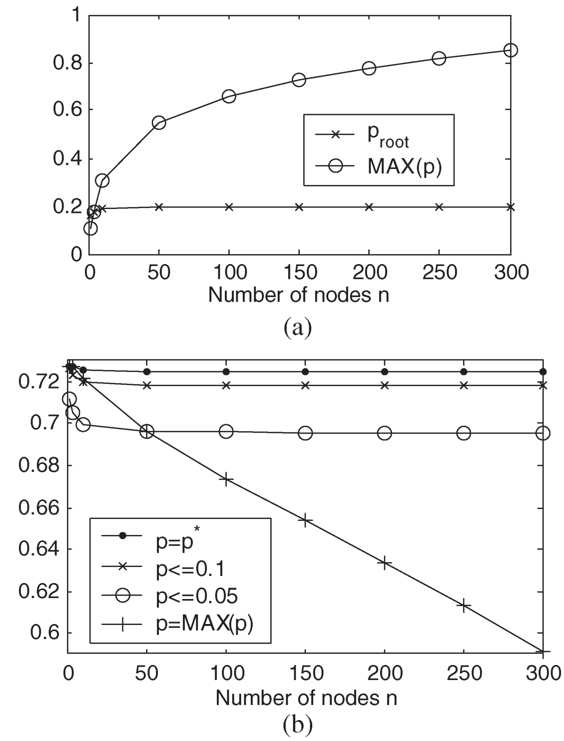


Fig. 2. (a) p_{root} and the collision probability in the saturated status with RTS/CTS and payload size of 8000 bits. (b) Maximum normalized throughput with different constraints on p with RTS/CTS and payload size of 8000 bits.

Maximum throughput can be achieved by controlling the total input traffic rate if no modification to the MAC protocol is allowed: As revealed in Fig. 2(b), rather than letting $p = p^*$ for each n, if we simply let $p \leqslant 0.1$ or $p \leqslant 0.05$, the achieved normalized throughput only drops by 0.96% and 4.2%, respectively, compared to the maximum normalized throughput. This is a very nice and important feature in the sense that as long as each node in the network can keep the collision probability p below a certain value, say 0.1, instead of p^*, which is dependent on n, maximum throughput is well approached. Thus, by maintaining a small collision probability in WLAN, which can be done through controlling the total input traffic rate, we can achieve high throughput. This in fact is consistent with our observation in Fig. 1, where $R_b \approx R_s$ when $p \leqslant 0.1$.

Note that in addition to achieving high throughput, keeping a small collision probability helps reduce delay. Since the time wasted due to collision could be neglected, the contention delay is very small, which is crucial in providing low delay for real-time traffic and will be discussed in detail in Section III-B.

3) Available Bandwidth: The total available bandwidth BW_a of the WLAN or the available traffic rate the network could further accommodate can be easily obtained by subtracting the current throughput from the maximum throughput.

Although it is not easy for each individual node to know the current total throughput if it does not decode everything received, the node can be aware of the available bandwidth by virtue of the channel busyness ratio, which could be easily

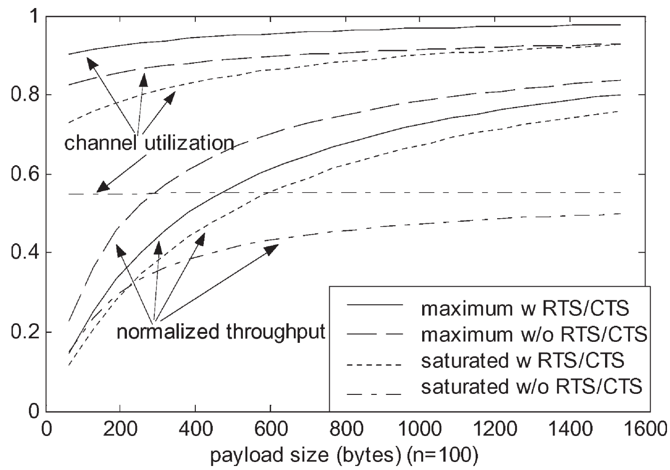


Fig. 3. Impact of payload size and the RTS/CTS mechanism.

acquired as described earlier. Especially, when $p \leqslant 0.1$, $R_b \approx R_s$. Thus, BW_a can be calculated as

$$\mathrm{BW}_a = \begin{cases} \dfrac{\mathrm{BW}(th_b - R_b)\overline{\mathrm{data}}}{T_{\mathrm{suc}}} & (th_b > R_b) \\ 0 & (th_b \leqslant R_b) \end{cases} \tag{10}$$

where BW is the transmission rate in bits per second for the DATA packets and th_b is a threshold of R_b and proportional to the maximum throughput.

4) Impact of Payload Size and the RTS/CTS Mechanism: Thus far, we have conducted our analysis without considering the impact of payload size and the RTS/CTS scheme on the throughput. In this subsection, we study this impact. Fig. 3 presents the analytical results, where RTS/CTS is used or not used, and various payload sizes are considered.

We find that no matter whether RTS/CTS is used, throughput increases along with payload size. But this is not necessarily true for channel utilization, such as the case that RTS/CTS is not used in the saturated case. The reason is the following. In the saturated case, given n, p is fixed. According to (1), (3), (4), and (6), R_s is almost unchanged and $R_s \approx p_s/(p_s + p_c)$.

It can also be observed that the maximum throughput is higher in the case that RTS/CTS is not used than in the case that RTS/CTS is used, no matter how large the payload size is. This is because maximum throughput is obtained when p is relatively small and thus the impact of collisions due to long data packets could be ignored. As a result, if RTS/CTS is not used, the MAC overhead is reduced, which results in higher throughput. On the contrary, in the saturated case where the collision probability is much higher, the use of RTS/CTS does improve the throughput, especially when the payload size is large. This is because the impact of collisions due to long data packets becomes significant in the saturated case and cannot be ignored; the exchange of RTS/CTS avoids long packet collisions and thus reduces MAC overhead. Note that for the payload size that is shorter than about 220 B in this parameter setting, the use of RTS/CTS is counterproductive because of its relatively high overhead compared with the short payload size.

To sum up, to maximize system throughput, the basic access without the RTS/CTS mechanism is desired as long as we can keep the collision probability at a relatively small value.

B. Delay and Delay Variation

In this subsection, we study delay and delay variation performance, which is an integral part of QoS provisioning in 802.11 WLAN. As we know, the delay in the network comprises three components, i.e., propagation delay, transmission delay, and queuing delay. Note that in WLAN, transmission delay contains a variable amount of delay caused by MAC layer collisions and thus is not fixed. Henceforth, we define the sum of the propagation delay and transmission delay as the service time at the MAC layer, which is the time period from the instant that a packet begins to be serviced by the MAC layer to the instant that it is either successfully transmitted or dropped after several failed retransmissions.

In the following, we will give an analysis of the service time and the queuing delay. Then, the estimates of delay and delay variation are derived.

1) Service Time Distribution:

Markov chain model for the service time: After examining the transmission procedure introduced in Section II-A, we can conclude that the only outside factor is the collision probability p when the node attempts the transmission. As discussed in the previous section, p is determined by the number of neighboring nodes and the traffic distribution at those nodes. Thus, we could assume that p is independent of the backoff state of the node under consideration, although it is still dependent on the backoff states of other nodes. We can therefore model the stochastic process of the service time as a Markov chain since the future state only depends on the current state. Clearly, the transition probabilities are dependent on the collision probability p, thus the service time distribution is a function of p.

Probability generating function of the service time: The service time for each packet consists of multiple backoff time slots that could be empty slots, collision slots, or successful transmission slots. As mentioned earlier, since the length of an empty backoff slot is a fixed value and T_{suc} or T_{col} depends on the length of the header and data packet, which are discrete in bits, it is suitable to model the service time distribution as a discrete probability distribution. To facilitate analysis, this distribution is completely described with the probability generating function (PGF).

By applying the signal transfer function to the generalized state transfer diagram of Markov chain, we have derived the PGF of the service time $G_{Ts}(Z)$, which is quite accurate as verified by ns-2 simulations ([28], [29]). On the other hand

$$G_{Ts}(Z) = \sum_{i=0}^{\infty} p_i Z^{ts_i} \tag{11}$$

where $ts_i (i \geqslant 0)$ are all the possible discrete values of service time Ts and $p_i = \Pr\{Ts = ts_i\}$. We also found that given p, the service time distribution is almost insensitive to n, while n only influences the maximum value of p as shown in Fig. 2(a).

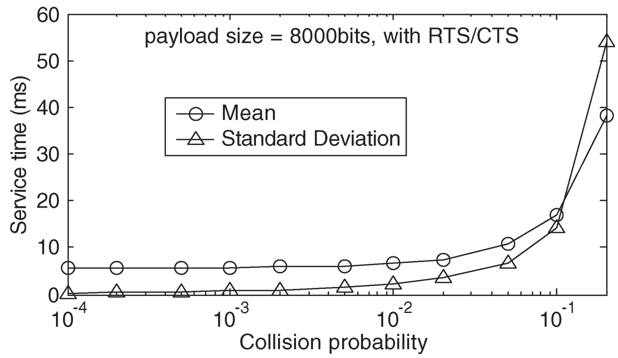


Fig. 4. Mean and standard deviation of service time.

Thus, the following delay analysis is valid for different n and we need not specify the value of n.

Mean and variance of the service time: Given (11), it is easy to obtain any moment of the service time Ts by taking the derivative of $G_{Ts}(Z)$ with respect to Z. Specifically, the mean and variance are

$$E[Ts] = \frac{\partial}{\partial Z} G_{Ts}(Z)|_{Z=1} = G'_{Ts}(1)$$
$$\text{Var}[Ts] = G''_{Ts}(1) + G'_{Ts}(1) - [G'_{Ts}(1)]^2. \qquad (12)$$

Fig. 4 demonstrates the mean and variance of service time as a function of collision probability p. It can be seen that when $p \geqslant 0.1$, both the mean and the variance increase exponentially with p. On the other hand, we have found that when $p \leqslant 0.1$, the achieved throughput is almost the same as the maximum achievable throughput. To provide a delay guarantee for some delay-sensitive applications such as voice over Internet protocol (VoIP) and achieve approximately maximum throughput, WLAN should keep the collision probability less than 0.1.

2) Packet Delay Bound and Delay Variation Estimate: Because there is typically one shared outgoing queue for all packets from different applications in each mobile node, we can model each node as a queuing system. In the queuing system, the packet arrival process is determined by the aggregate traffic behavior of all applications that emit packets to the MAC layer; the service time follows the distribution described in the previous subsection. After building such a queuing model, we can derive accurate estimates of delay and delay variation in the nonsaturated case. Notice that the number of packets waiting in the queue N_q almost equals zero in the nonsaturated case especially for $p \leqslant 0.1$ as shown in [28] and [29] and verified in our simulation later. Otherwise, each node will contend for the channel in most times and result in a much higher p.

Delay bound with known packet arrival rate: We start the analysis with a simple case, i.e., the packet arrival follows some process with a known (or estimated) arrival rate. If the arrival process is Poissonian, the system can be modeled as an M/G/1 system [18]. Accordingly, the mean of the packet delay T, which consists of the waiting time in the queue and the service time, is

$$E[T] = \overline{Ts} + \frac{\lambda \overline{Ts^2}}{2(1-\rho)} \qquad (13)$$

where λ is the average arrival rate of the input traffic and $\rho = \lambda \times \overline{Ts} < 1$. If the arrival process follows a general distribution, then we get a G/G/1 system, for which we have an upper bound T_U for T [19], i.e.,

$$E[T] \leqslant \overline{Ts} + \frac{\lambda\left(\sigma_a^2 + \sigma_{Ts}^2\right)}{2(1-\rho)} \equiv T_U \qquad (14)$$

where σ_{Ts} and σ_a are the standard deviations of the service time and packet arrival process, respectively.

So far, these results hold when $\rho < 1$ for the system with infinite buffer. The actual delay upper bound should be less than T_U because we do not count the packets dropped due to limited buffer, which will have a long delay in the system with infinite buffer. In fact, because we are only interested in the nonsaturated case with an almost empty queue, the above results are thus accurate.

Delay bound and delay variation with unknown packet arrival rate: In the previous paragraph, we only give the mean of the packet delay T with the estimation dependent on the specific packet arrival process and on the accurate estimate of λ. In reality, however, this approach could be infeasible if λ is hard to estimate when the instantaneous packet arrival rate at each individual node changes dramatically. We thus embark on deriving accurate estimates for delay and delay variation in a more general case, i.e., without any knowledge about λ.

Let Ts_i denote the service time for the ith packet at a node under consideration. Since the backoff timer is reset for every packet to be transmitted [12], $\{Ts_1, Ts_2, \ldots\}$ are independent and identically distributed (i.i.d.) random variables. Let T_i be the system time (or delay) of the ith packet including the service time and the waiting time in the queue, R_i be the residual service time seen by the ith packet, and N_i be the number of packets found waiting in the queue by the ith packet at the instant of arrival.

Based on such notations, we obtain

$$T_i = Ts_i + R_i + \sum_{j=i-N_i}^{i-1} Ts_j. \qquad (15)$$

As previously discussed, N_i almost equals 0 in the nonsaturated case, so we can approximate T_i as

$$T_i \cong Ts_i + R_i. \qquad (16)$$

Notice that R_i is the residual service time of the $(i - N_i - 1)$th packet, thus we have

$$Ts_i \leqslant T_i \leqslant Ts_i + Ts_{i-N_i-1}. \qquad (17)$$

By taking expectations on both sides of (17), we have

$$E[Ts] \leqslant E[T] \leqslant 2E[Ts]. \qquad (18)$$

Since it is difficult to derive the variation of R_i in general cases, we use the standard deviation of the service time σ_{Ts} to approximate that of T_i, i.e., σ_T as

$$\sigma_{Ts} \leqslant \sigma_T \approx k\sigma_{Ts} \qquad (19)$$

where k is a constant value. From ns-2 simulation results as presented later, $k = 1$, or 2 gives a good approximation.

In fact, by applying the Residual Life Theorem [18], we could obtain more accurate approximations of $E[T]$ and σ_T. Let r be the residual service time observed at any time instant during the service. If the service time distribution is $F_{Ts}(x)$, then the probability density function (pdf) of r, denoted by $f_r(x)$, can be expressed as $\mu(1 - F_{Ts}(x))$, where $\mu = (1/E[Ts])$. We thus have

$$E[r] = \int_0^\infty r f_r(x)dx = \frac{\mu}{2}E[Ts^2]$$

$$E[r^2] = \int_0^\infty r^2 f_r(x)dx = \frac{\mu}{3}E[Ts^3]$$

$$E[R] = 0P(\text{idle}) + E[r]P(\text{busy}) = \frac{r_0\mu}{2}E[Ts^2]$$

$$E[R^2] = 0P(\text{idle}) + E[r^2]P(\text{busy}) = \frac{r_0\mu}{3}E[Ts^3]$$

$$\text{Var}[R] = \frac{r_0\mu}{3}E[Ts^3] - \left(\frac{r_0\mu}{2}E[Ts^2]\right)^2 \quad (20)$$

where $r_0 = P(\text{busy})$ is the probability that the server is busy, i.e., there is one packet contending for the channel or being transmitted. Because $r_0 \leqslant 1$, we obtain

$$\begin{aligned} E[T] &\approx E[Ts] + E[R] \leqslant E[Ts] + \frac{E[Ts^2]}{2E[Ts]} \\ &\equiv T_{UR} \end{aligned} \quad (21)$$

$$\begin{aligned} \text{Var}[R] &= \frac{(r_0 - r_0^2)\,\mu}{3}E[Ts^3] \\ &\quad + r_0^2\left(\frac{\mu}{3}E[Ts^3] - \left(\frac{\mu}{2}E[Ts^2]\right)^2\right) \\ &= \frac{(r_0 - r_0^2)\mu}{3}E[Ts^3] + r_0^2\text{Var}[r] \\ &\leqslant \frac{\mu}{12}E[Ts^3] + \text{Var}[r] \\ &= \frac{5\mu}{12}E[Ts^3] - \left(\frac{\mu}{2}E[Ts^2]\right)^2 \end{aligned} \quad (22)$$

$$\begin{aligned} \text{Var}[T] &\approx \text{Var}[Ts] + \text{Var}[R] \\ &\leqslant \text{Var}[Ts] + \frac{5E[Ts^3]}{12E[Ts]} - \left(\frac{E[Ts^2]}{2E[Ts]}\right)^2 \\ &\equiv \sigma_{T_{UR}}^2. \end{aligned} \quad (23)$$

Fig. 5(a) illustrates both the lower bound and the upper bound for the packet delay T. We can see that the upper bound and lower bound are very close, thus we can characterize the delay with high accuracy, although the exact value is not available. As expected, when $p < 0.1$, T_{UR} is tighter than $2 \times E[Ts]$. This is desirable since we focus on the nonsaturated case where p is small. As revealed by the bounds, the mean of the system delay T is small: 5 ms $< E[T] <$ 10 ms when $p \leqslant 0.01$, and $E[T] <$ 30 ms when $p \leqslant 0.1$. This is sufficient for real-time applications such as VoIP.

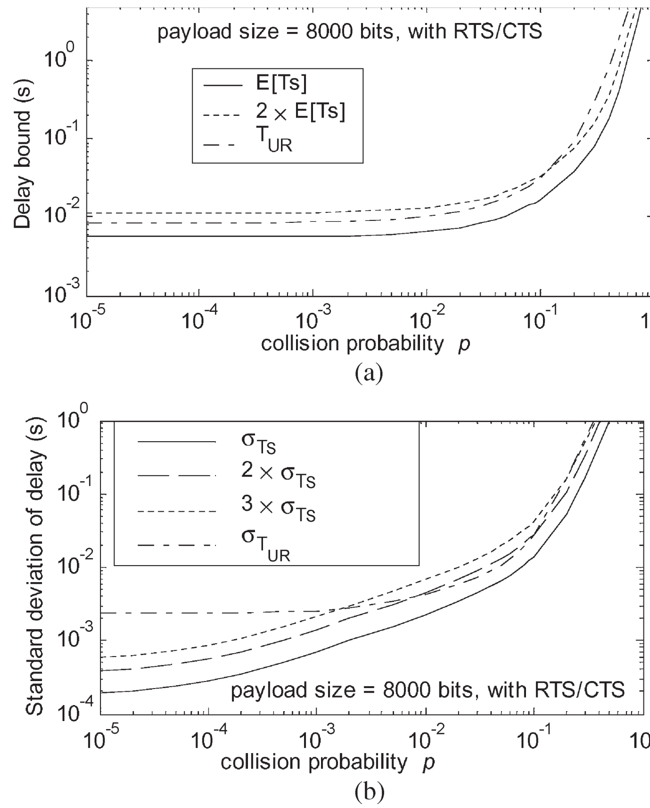


Fig. 5. (a) Delay bound. (b) Standard deviation of delay.

The standard deviation of the system delay is illustrated in Fig. 5(b). As shown in the figure, it is also small: $\sigma_{Ts} <$ 30 ms when $p \leqslant 0.1$. When $p \leqslant 0.02$, the standard deviation is much smaller than $E[Ts]$ (and than $E[T]$ since $E[Ts]$ is the lower bound). Note that $\sigma_{T_{UR}}$ is relatively large when $p \leqslant 0.002$. This is because the approximation in (22) uses $r_0 \geqslant 0.5$; however, r_0 should have been smaller than 0.5 when $p \leqslant 0.002$.

As a special case, if the packet arrival process is Poissonian, then $r_0 = \rho = \lambda E[Ts] < 1$. Thus

$$\begin{aligned} E[T] &\approx E[Ts] + E[R] = E[Ts] + \frac{1}{2}\lambda E[Ts^2] \\ &\equiv T_{URM} \end{aligned} \quad (24)$$

$$\begin{aligned} \text{Var}[T] &\approx \text{Var}[Ts] + \text{Var}[R] \\ &= \text{Var}[Ts] + \frac{1}{3}\lambda E[Ts^3] - \left(\frac{1}{2}\lambda E[Ts^2]\right)^2 \\ &\equiv \sigma_{T_{URM}}^2. \end{aligned} \quad (25)$$

Finally, we comment on the results of delay and delay variation. First, all the above results are derived for the nonsaturated case, which means the traffic intensity $\rho < 1$ and the collision probability $p \leqslant 0.1$. Second, the approximation in (16) relies on the assumption that there is no bulk arrival. Although this assumption is common in the analysis of queuing systems and is true for both the Poisson arrival process and the deterministic arrival process, in practice, bursty traffic such as transmission control protocol (TCP) traffic violates it. Consequently, the bursty traffic induces not only longer waiting time in the queue but also higher collision probability in the burst period leading to longer service time. For the above results to remain valid, it is necessary to regulate arriving traffic at the MAC layer.

C. Packet Loss Rate

At the MAC layer, a packet may be lost due to queue overflow or MAC collisions. Once a packet is queued, a node attempts to transmit it for a certain number of times, denoted by α. If all the attempts fail due to collisions, the packet gets dropped. Given the collision probability p, the packet dropping probability P_d due to MAC collisions is

$$P_d = p^{\alpha}. \tag{26}$$

When the packet blocking probability P_{block}, i.e., the probability that the queue is full when a packet arrives, is very small, as for the nonsaturated case where $N_q \cong 0$ as mentioned earlier, the total packet loss rate P_l of the queuing system can be approximated as P_d, i.e.,

$$P_l \approx p^{\alpha}. \tag{27}$$

We see when $p \leqslant 0.1$ and $\alpha = 7$ [12], $P_l \leqslant 10^{-7}$. Obviously, this satisfies the packet loss requirements of most applications such as VoIP.

On the contrary, a much higher packet loss rate is expected if the network is in the saturated case for the following reason. On one hand, the collision probability p gets significantly large, resulting in considerable packet losses due to collisions. On the other hand, each packet experiences a much longer system delay in the saturated case compared to that in the nonsaturated case, which leads to a full queue at most times and hence blocks newly arriving packets.

Before ending this section, we make a few remarks about the analytical model. Note that all the performance metrics are expressed as a function of the collision probability. However, obtaining the collision probability is not easy. There are two possible approaches. One is to analytically derive the collision probability, which requires full knowledge of the traffic arrival models at the node of interest and at all the other nodes in the network as well. The other is to measure it through experiments. Unfortunately, it is not amenable to practical measurement due to the lack of measured values or the inability of each node to distinguish collisions from channel fading. Therefore, we propose the channel busyness ratio as a good substitute for the collision probability for the following reasons. First, as mentioned earlier, the channel busyness ratio is an injective function of the collision probability. This indicates that the channel busyness ratio can also serve as the input of the analytical model. Unlike the collision probability, the channel busyness ratio is easy to measure in practice because IEEE 802.11 is essentially based on carrier sensing. Second, as shown for the nonsaturated case, the channel busyness ratio can accurately represent the channel utilization or the normalized throughput, and hence can be used to facilitate network control mechanisms such as call admission control over the real-time traffic and rate control over the best effort traffic. Accordingly, all the performance metrics are presented as a function of the channel busyness ratio in the following simulation results.

IV. Simulation Study of the IEEE 802.11

The simulation study in this section serves two purposes. First, it is aimed at verifying our analytical study in Section III. Second, while our analytical results have shown that IEEE 802.11 can operate at an optimal point that leads to maximum throughput, low delay, and almost zero packet loss rate, they do not reveal a specific way to achieve this optimal operating point. Thus, we demonstrate how to reach and retain the optimal point through simulations.

A. Simulation Configuration

The simulation study is conducted using the ns-2 simulator. The IEEE 802.11 system parameters are summarized in Table II. The RTS/CTS mechanism is used. We simulate different numbers of mobile stations in WLAN. Every node initiates an identical user datagram protocol (UDP)/constant bit rate (CBR) traffic flow to a randomly selected neighbor. The queue length at each node is ten packets.

As revealed, whether the network operates in the nonsaturated or saturated case can be determined by controlling the collision probability p. Also, the optimal operating point lies where $p \approx 0.1$. Without changing the 802.11 protocol, we use two techniques to control p in order to locate the optimal point. One is to schedule the start time of the UDP flows, which will be described below; the other is to gradually increase the sending rate of each flow from 0. In contrast, the saturated case can be easily simulated by boosting the traffic load to a much higher level than what the network can support.

1) Deterministic Minimum-Collision-Probability Scheduling (DPS): To minimize collision probability, we schedule UDP flows in such a way that the start time of one flow is separated from another by a constant period t_{int}/n, where t_{int} is the packet interarrival time for each flow. So if the aggregate traffic rate is less than the network capacity, i.e., the network can handle all the arriving packets from each flow, the collision probability could be reduced to zero. In this case, there is no queuing delay and the system delay is the random backoff time plus one packet transmission time. We call this scheduling deterministic minimum-collision-probability scheduling.

2) Distributed Randomized Scheduling (DRS): However, in a distributed WLAN environment, it is very difficult for each node to exactly know the start time of all the flows and schedule its own flows accordingly to avoid collisions. Therefore, to simulate a more realistic scenario, we cannot adopt the deterministic scheduling described above. We thus employ a simple yet effective scheduling algorithm that starts each flow at randomized times. Specifically, the start time of each flow is uniformly chosen in $[0, t_{\text{int}}]$, which keeps all the nodes from contending for the channel at the same time. As a result, the collision probability is reduced and no node needs to care about other nodes' transmission schedule.

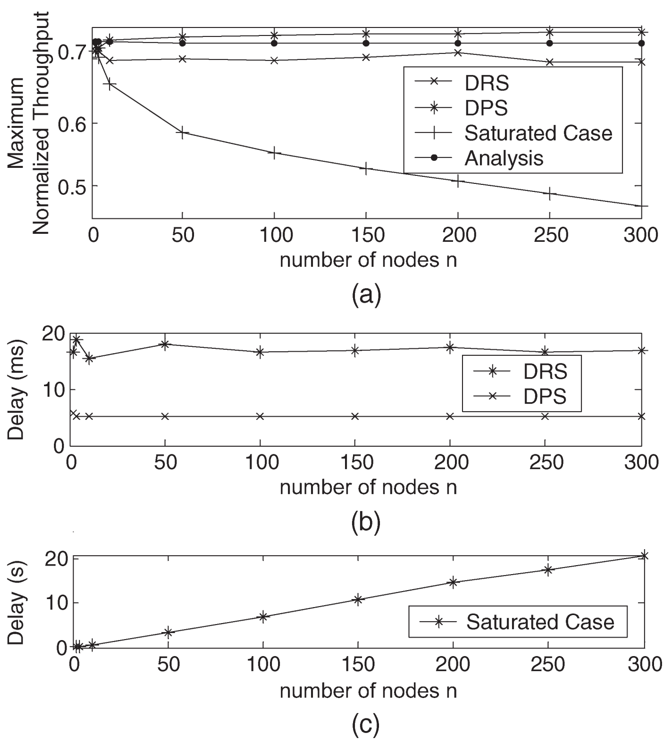


Fig. 6. Simulation results when payload size $= 8000$ bits.

B. Simulation Results

1) Saturated Case Versus Nonsaturated Case: In Fig. 6, for the nonsaturated case, we see that the normalized throughput that DPS achieves is slightly higher than the theoretical maximum throughput since it uses perfect scheduling and hence reduces the collision probability to zero. Likewise, the normalized throughput that DRS achieves is close to the theoretical maximum throughput since it greatly reduces the collision probability. On the contrary, the throughput in the saturated case is much lower. As is consistent with the analytical results, the nonsaturated throughput is almost independent of the number of nodes, whereas the saturated throughput declines significantly with the increase in node number. For delay, we see that there is difference in orders of magnitude for these two cases. Also, while the delay stays almost unchanged in the nonsaturated case as the number of node increases, it increases in the saturated case. This is due to the fact that in the latter case each node always has packets to transmit and keeps contending for the channel, which greatly increases the collision probability. As a result, each packet suffers from both long queuing delay and service time. Note that DPS enjoys a shorter delay than DRS since it reduces the collision probability more effectively.

2) Optimal Operating Point: As Fig. 6 shows, DRS yields a comparable performance with that of DPS; we thus use DRS as our scheduling algorithm henceforth. By gradually increasing the sending rate of each flow, we are able to locate the optimal operating point as shown in Figs. 7 and 8. While Fig. 7 presents the performance of throughput, delay, and delay variation as a function of the channel busyness ratio, Fig. 8 shows the behavior of average queue length and packet loss rate when input traffic increases.

Two important observations are made. First, we observe that there is a turning point in all the curves where the channel busyness ratio is about 0.95. Before that point, as the input traffic increases, the throughput keeps increasing, the delay and delay variation are small and almost unchanged, the queue at each node is empty, and the packet loss rate is zero. Note that with the small delay and delay variation, the delay requirements of real-time traffic can be adequately supported. After that point, the queue and the collision probability form a positive feedback loop. A slightly larger collision probability causes the queue to build up. The queue, even with one packet always in it, will force the MAC to keep contending for the channel, thereby exponentially increasing the collision probability, which in turn forces more packets to accumulate in the queue. Then, catastrophic effects take place: the throughput drops quickly, the queue starts to build up and the delay and delay variation increase dramatically, and the packet suffers from a large loss rate. Clearly, this turning point is the optimal operating point that we should tune the network to work around, where the throughput is maximized and the delay and delay variation are small.

Second, as shown in Fig. 7, the simulation results verify our analytical study of the IEEE 802.11. The throughput curves obtained from analysis and simulation coincide with each other. Also, as indicated in our analytical study, before the optimal point is reached, the network stays in the nonsaturated case and the queuing delay is almost zero; thus, the packet delay T can be accurately estimated by the service time T_S, which provides the lower bound. Meanwhile, the mean and variation are well bounded by T_{UR} and $\sigma_{T_{UR}}$ before the turning point as shown in (18), (19), (21), and (23).

V. Discussions

A. Impact of Fading Channel

So far, it is assumed that the channel is perfect. However, when channel fading is figured in, packet losses are no longer due to collisions only; they may well be caused by channel fading. Practically, it is extremely difficult to distinguish these two causes. As a matter of fact, 802.11 responds in the same way if the transmitter does not correctly receive its expected frame, which may be either CTS or ACK, no matter whether this is due to collision or channel fading. Based on this observation, we can incorporate the packet error probability into the collision probability as the recent work [10] did, and all the analytical results still hold.

It is important to note that normally channel fading is not a serious problem in WLAN, which features low node mobility and relatively stable channel. However, if the packet error probability due to channel fading becomes significant, i.e., the equivalent collision probability is high in our model, the QoS level will be hurt. Our analytical results show that in this case, as illustrated in Figs. 1, 4, and 5(a) and (b), the normalized throughput decreases, the service time increases, the mean and variation of delay increase along with the service time, and packet loss rate increases as well. However, with our analytical model, we can still calculate the maximum throughput, packet

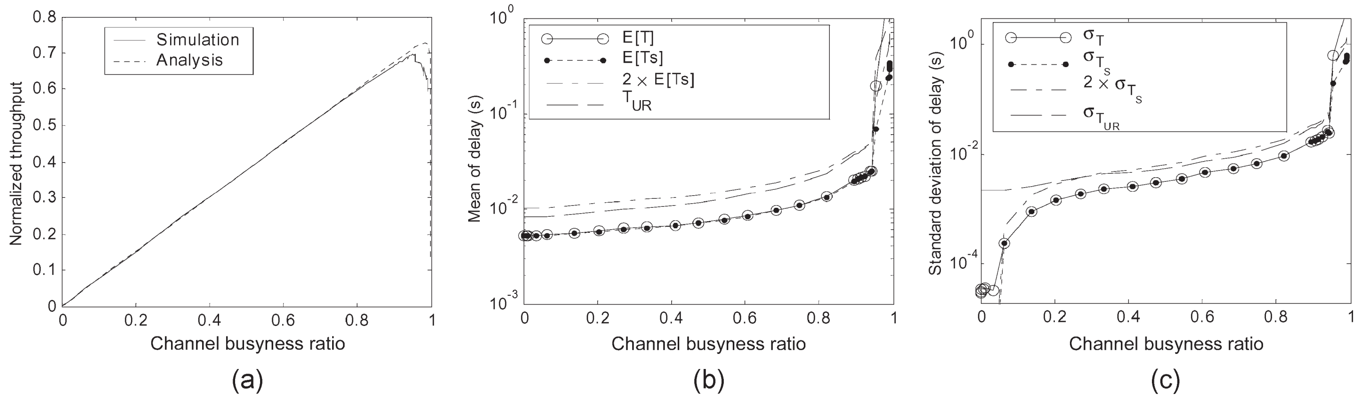


Fig. 7. Simulation results when $n = 50$ and payload size $= 8000$ bits.

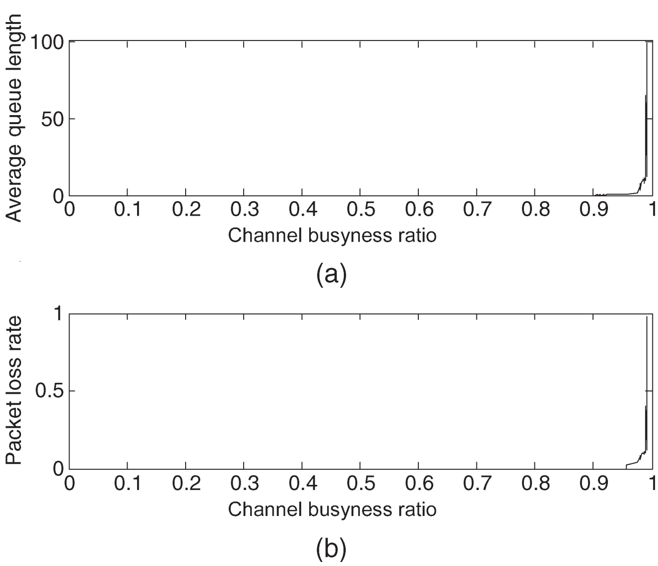


Fig. 8. Simulation results when $n = 50$ and payload size $= 8000$ bits.

loss rate, and give accurate estimates of delay and delay variation according to (5), (18), (19), (21), and (23).

B. Impact of Prioritized MAC

Since our focus is on how well the original IEEE 802.11 DCF can support QoS, we do not change the MAC protocol in the analysis and simulations. Within either the real-time traffic or the best effort traffic, no differentiation is committed. As a result, all the real-time traffic, including CBR and variable bit rate (VBR) traffic, equally shares the delay and delay variation, which sometimes is not flexible enough. If a prioritized 802.11 MAC protocol similar to [1] and [23] is adopted, we are able to provide priority within the real-time traffic. As a result, the high-priority real-time traffic receives smaller delay variation, whereas the low-priority real-time traffic receives higher delay variation [8].

VI. Conclusion

Despite considerable efforts spent on performance analysis and QoS provisioning for the IEEE 802.11 WLAN, the question of how well it can support QoS remains vague. In this paper, we clearly answer this question through thorough studies, which constitute our key contribution.

We have analytically characterized the optimal operating point for 802.11 WLAN and shown that if the network is tuned to work at this point, in addition to achieving theoretical maximum throughput, it can support the major QoS metrics such as throughput, delay and delay variation, and packet loss rate, as required by real-time services. This is further validated via extensive simulations. We therefore clarify that the IEEE 802.11 WLAN can provide statistical QoS guarantees, not just differentiated service, for multimedia services. We also demonstrate that the channel busyness ratio can accurately and timely represent network utilization; hence, it can be used to facilitate the regulation of total input traffic to support QoS.

References

[1] I. Ada and C. Castelluccia, "Differentiation mechanisms for IEEE 802.11," in *Proc. IEEE Int. Conf. Computer Communications (INFOCOM)*, Anchorage, AK, 2001, pp. 209–218.

[2] D. Bertsekas and R. Gallager, *Data Networks*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1992.

[3] V. Bharghavan, "Performance evaluation of algorithms for wireless medium access," in *Proc. IEEE Int. Computer Performance and Dependability Symp.*, Durham, NC, 1998, pp. 86–95.

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

[5] F. Cali, M. Conti, and E. Gregori, "IEEE 802.11 protocol: Design and performance evaluation of an adaptive backoff mechanism," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 9, pp. 1774–1786, Sep. 2000.

[6] ——, "Tuning of the IEEE 802.11 protocol to achieve a theoretical throughput limit," *IEEE/ACM Trans. Netw.*, vol. 8, no. 6, pp. 785–799, Dec. 2000.

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

[8] D. Clark, S. Shenker, and L. Zhang, "Supporting real-time application in an integrated services packet network: Architecture and mechanism," in *Proc. Association Computing Machinery Special Interest Group Data Communication (ACM SIGCOMM)*, Baltimore, MD, 1992, pp. 14–26.

[9] C. H. Foh and M. Zukerman, "Performance analysis of the IEEE 802.11 MAC protocol," in *Proc. Eur. Wireless*, Florence, Italy, 2002, pp. 184–190.

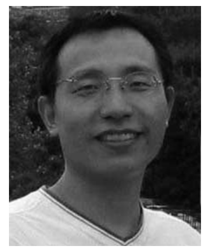
[10] Z. Hadzi-Velkov and B. Spasenovski, "Saturation throughput-delay analysis of IEEE 802.11 DCF in fading channel," in *Proc. IEEE Int. Conf. Communications (ICC)*, Anchorage, AK, 2003, pp. 121–126.

[11] T. S. Ho and K. C. Chen, "Performance analysis of IEEE 802.11 CSMA/CA medium access control protocol," in *Proc. IEEE Personal Indoor Mobile Radio Communications (PIMRC)*, Taipei, Taiwan, 1996, pp. 407–411.

[12] *IEEE Standard for Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications,* ISO/IEC 8802-11: 1999(E), 1999.
[13] *Draft Supplement to Part 11: Medium Access Control (MAC) Enhancements for Quality of Service (QoS),* IEEE Std 802.11e/D8.0, Feb. 2004.
[14] ITU-T G.114, *One-Way Transmission Time*, 1996.
[15] ITU-T G.1010, *End-User Multimedia QoS Categories*, 2001.
[16] V. Kanodia, C. Li, A. Sabharwal, B. Sadeghji, and E. Knightly, "Distributed multi-hop scheduling and medium access with delay and throughput constraints," in *Proc. Association Computing Machinery Mobile Computing and Networks (ACM MobiCom)*, Rome, Italy, 2001, pp. 200–209.
[17] H. Kim and J. Hou, "Improving protocol capacity with model-based frame scheduling in IEEE 802.11-operated WLANs," in *Proc. Association Computing Machinery Mobile Computing and Networks (ACM MobiCom)*, San Diego, CA, 2003, pp. 190–204.
[18] L. Kleinrock, *Queueing Systems*, vol. I. New York: Wiley, 1975.
[19] ——, *Queueing Systems*, vol. II. New York: Wiley, 1975.
[20] Y. Kwon, Y. Fang, and H. Latchman, "A novel MAC protocol with fast collision resolution for wireless LANs," in *Proc. IEEE Int. Conf. Computer Communications (INFOCOM)*, San Francisco, CA, 2003, pp. 853–862.
[21] S.-T. Sheu and T.-F. Sheu, "A bandwidth allocation/sharing/extension protocol for multimedia over IEEE 802.11 ad hoc wireless LANs," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 10, pp. 2065–2080, Oct. 2001.
[22] J. L. Sobrinho and A. S. Krishnakumar, "Real-time traffic over the IEEE 802.11 medium access control layer," *Bell Labs Tech. J.*, vol. 1, no. 2, pp. 172–187, 1996.
[23] A. Veres, A. T. Campbell, M. Barry, and L.-H. Sun, "Supporting service differentiation in wireless packet networks using distributed control," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 10, pp. 2081–2093, Oct. 2001.
[24] M. A. Visser and M. E. Zarki, "Voice and data transmission over an 802.11 wireless network," in *Proc. IEEE Personal Indoor and Mobile Radio Communications (PIMRC)*, Toronto, ON, Canada, 1995, pp. 648-652.
[25] 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," in *Proc. IEEE Int. Conf. Computer Communications (INFOCOM)*, New York, 2002, pp. 599–607.
[26] Y. Xiao, H. Li, and S. Choi, "Protection and guarantee for voice and video traffic in IEEE 802.11e wireless LANs," in *Proc. IEEE Int. Conf. Computer Communications (INFOCOM)*, Hong Kong, 2004, pp. 2152–2162.
[27] J. Y. Yeh and C. Chen, "Support of multimedia services with the IEEE 802-11 MAC protocol," in *Proc. IEEE Int. Conf. Communications (ICC)*, New York, 2002, pp. 600–604.
[28] H. Zhai and Y. Fang, "Performance of wireless LANs based on IEEE 802.11 MAC protocols," in *Proc. IEEE Personal Indoor and Mobile Radio Communications (PIMRC)*, Beijing, China, 2003, pp. 2586–2590.
[29] H. Zhai, Y. Kwon, and Y. Fang, "Performance analysis of IEEE 802.11 MAC protocols in wireless LANs," *Wiley Wireless Commun. Mobile Comput., Special Issue on Emerging WLAN Technologies and Applications*, vol. 4, no. 8, pp. 917–931, Dec. 2004.
[30] H. Zhai, J. Wang, X. Chen, and Y. Fang, "Medium access control in mobile ad hoc networks: Challenges and solutions," *Wiley Wireless Commun. Mobile Comput., Special Issue on Ad Hoc Wireless Networks*, 2004. to be published.

Hongqiang Zhai (S'03) received the B.E. and M.E. degrees in electrical engineering from Tsinghua University, Beijing, China, in July 1999 and January 2002, respectively, and is currently working toward the Ph.D. degree at the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL. His research interests include medium access control, QoS, and cross-layer design in wireless networks.

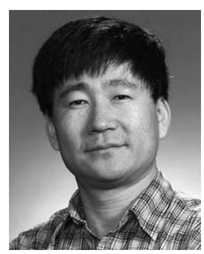
He was a Research Intern at Bell Labs Research, China, from June 2001 to December 2001, and at Microsoft Research, Asia, from January 2002 to July 2002.



Xiang Chen (S'03–M'05) received the B.E. and M.E. degrees in electrical engineering from the Shanghai Jiao Tong University, Shanghai, China, in 1997 and 2000, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Florida, Gainesville, in 2005.

He is currently a Senior Research Engineer at Motorola Labs, Arlington Heights, IL. His research interests include resource management, medium access control, and quality of service (QoS) in wireless networks.

Dr. Chen is a Member of Tau Beta Pi.



Yuguang Fang (S'92–M'93–SM'99) received the Ph.D. degree in systems and control engineering from Case Western Reserve University, Cleveland, OH, in January 1994 and the Ph.D. degree in electrical engineering from Boston University, Boston, MA, in May 1997.

From June 1997 to July 1998, he was a Visiting Assistant Professor at the Department of Electrical Engineering, University of Texas at Dallas. From July 1998 to May 2000, he was an Assistant Professor at the Department of Electrical and Computer Engineering, New Jersey Institute of Technology. In May 2000, he joined the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, where he got the early promotion to an Associate Professor with tenure in August 2003 and a Professor in August 2005. He has published over 100 papers in refereed professional journals and conferences. He received the National Science Foundation Faculty Early Career Award in 2001 and the Office of Naval Research Young Investigator Award in 2002.

He is currently serving as an Editor for many journals including IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE TRANSACTIONS ON MOBILE COMPUTING, and *ACM Wireless Networks*. He is also actively participating in conference organization such as the Program Vice-Chair for IEEE INFOCOM'2005, Program Co-Chair for the Global Internet and Next Generation Networks Symposium in IEEE GLOBECOM'2004 and the Program Vice Chair for 2000 IEEE Wireless Communications and Networking Conference (WCNC'2000).