# A Unified Analysis of IEEE 802.11 DCF Networks: Stability, Throughput and Delay

Lin Dai, *Member, IEEE*, and Xinghua Sun

**Abstract**—In this paper, a unified analytical framework is established to study the stability, throughput and delay performance of homogeneous buffered IEEE 802.11 networks with Distributed Coordination Function (DCF). Two steady-state operating points are characterized using the limiting probability of successful transmission of Head-of-Line (HOL) packets $p$ given that the network is in unsaturated or saturated conditions.

The analysis shows that a buffered IEEE 802.11 DCF network operates at the desired stable point $p = p_L$ if it is unsaturated. $p_L$ does not vary with backoff parameters, and a stable throughput can be always achieved at $p_L$. If the network becomes saturated, in contrast, it operates at the undesired stable point $p = p_A$, and a stable throughput can be achieved at $p_A$ if and only if the backoff parameters are properly selected. The stable regions of the backoff factor $q$ and the initial backoff window size $W$ are derived, and illustrated in cases of the basic access mechanism and the request-to-send/clear-to-send (RTS/CTS) mechanism. It is shown that the stable regions are significantly enlarged with the RTS/CTS mechanism, indicating that networks in the RTS/CTS mode are much more robust. Nevertheless, the delay analysis further reveals that lower access delay is incurred in the basic access mode for unsaturated networks. If the network becomes saturated, the delay performance deteriorates regardless of which mode is chosen. Both the first and the second moments of access delay at $p_A$ are sensitive to the backoff parameters, and shown to be effectively reduced by enlarging the initial backoff window size $W$.

**Index Terms**—Stability, throughput, delay, IEEE 802.11 DCF networks, Binary Exponential Backoff.

## 1 Introduction

IEEE 802.11 Wireless Local Area Networks (WLANs) have gained significant attention in both industry and academia [1]. Fueled by the widespread popularity of commercial WLANs, research activities have been intensified over the last few years, and a major focus has been put on the Medium Access Control (MAC) layer with Distributed Coordination Function (DCF).

DCF is based on the Carrier Sense Multiple Access (CSMA) protocol with two access mechanisms including the basic access mechanism and the request-to-send/clear-to-send (RTS/CTS) mechanism. As a random access protocol, DCF inherits the merits of minimum coordination and distributed control, which, on the other hand, also renders difficulty in modeling and performance evaluation. A widely adopted model of IEEE 802.11 DCF networks was proposed by Bianchi in [2], where a two-dimensional Markov chain was established to characterize the backoff behavior of each single node. It is well supported by simulation results and shown to be a powerful, yet simple, analytical tool to evaluate the throughput performance when the network becomes saturated, i.e., each node always has a packet to transmit.

Bianchi's model has been refined by a series of follow-up papers [3]–[15] to include more practical assumptions such as freezing of backoff counters [3], [5]–[7], finite retransmission attempts [4] [8] and imperfect channel conditions [9]–[12]. A great deal of effort was also made to extend the model to unsaturated networks [16]–[29]. For instance, it was assumed in [16]–[19] that each node has a one-packet buffer and a new packet is generated with a certain probability after the previous one is successfully transmitted. A more general buffered network was considered in [20]–[29], where each node is equipped with a buffer of finite [20] [21] or infinite size [22]–[29]. Most of them followed the analysis in [2] and used the conditional collision probability of each node as a key variable to characterize the network operating point. No consensus, however, has been reached on the fixed-point equation of the conditional collision probability in unsaturated scenarios. Moreover, an explicit expression of the service time distribution is usually too complicated to obtain based on these models. Approximate methods are therefore adopted to simplify the analysis. For example, in [20] and [28], the attempt rate in an unsaturated network is approximated by a scaled version of the saturated attempt rate. Good accuracy was demonstrated via simulations when the traffic is light.

The main focus of the above studies is on the numerical calculation of throughput or delay for given network configuration. Yet how to tune system parameters to optimize the network performance is another interesting problem, which attracts much attention as well. For instance, it has been long observed that IEEE 802.11 DCF networks may suffer from low throughput if

*Manuscript received May 16, 2011; revised January 9, 2012; accepted May 3, 2012. This work was supported by the Research Grants Council (RGC) of Hong Kong under GRF Grant CityU 112810 and City University of Hong Kong under SRG Grant 7008049.*

*L. Dai and X. Sun are with the Department of Electrical Engineering, City University of Hong Kong, 83 Tat Chee Avenue, Kowloon Tong, Hong Kong, China (email: lindai@cityu.edu.hk; xinghusun2@student.cityu.edu.hk).*

the backoff parameters are improperly selected. Various algorithms were therefore developed to estimate the number of active nodes and adaptively adjust the initial backoff window size [30]–[34]. On the other hand, severe delay jitter as well as short-term unfairness was found to arise when an IEEE 802.11 DCF network becomes saturated. In that case, nodes are pushed to large phases with extremely small transmission probabilities such that the node who once succeeds can capture the channel for a long time and produce a continuous stream of packets. To address this issue, a number of modified DCF protocols have been proposed for IEEE 802.11 networks [35]–[39]. Most of these studies aimed at improving the fairness performance or enhancing the throughput. In many cases, however, the improvement is indeed obtained at the cost of delay degradation. It is, therefore, especially desirable to establish a coherent theory of IEEE 802.11 DCF networks, based on which the effect of key system parameters can be evaluated within the same framework.

In this paper, a unified analytical framework is established to study the stability, throughput and delay performance of IEEE 802.11 DCF networks. Consider an $n$-node homogeneous buffered IEEE 802.11 DCF network where each node is equipped with a buffer of infinite size and an arrival rate of $\lambda$. In contrast to the classic model proposed in [2], the behavior of each Head-of-Line (HOL) packet, including backoff, collision and successful transmission, is modeled as a discrete-time Markov renewal process. The corresponding steady-state distribution is shown to be crucially determined by the limiting probability of successful transmission of HOL packets given that the channel is idle, $p$. According to whether the network is unsaturated or saturated, distinct fixed-point equations of $p$ are derived, and two steady-state operating points, i.e., the desired stable point $p_L$ and the undesired stable point $p_A$, are obtained as explicit functions of system parameters. The network throughput and delay performance at the bi-stable points $p_L$ and $p_A$ are further characterized.

Specifically, it is shown that a buffered IEEE 802.11 DCF network operates at the desired stable point $p_L$ if it is unsaturated. $p_L$ does not vary with backoff parameters, and a stable network throughput $\hat{\lambda}_{out}=n\lambda$ can be always achieved at $p_L$. Both the first and second moments of access delay at the desired stable point $p_L$ are shown to be insensitive to 1) the backoff factor $q$, 2) the cutoff phase $K$ and 3) the number of nodes $n$. They, however, grow with the initial backoff window size $W$, indicating that a small $W$ is desirable to improve the delay performance.

If the network becomes saturated, in contrast, it operates at the undesired stable point $p_A$ which is a function of 1) the backoff factor $q$, 2) the initial backoff window size $W$, 3) the cutoff phase $K$ and 4) the number of nodes $n$. It is shown that for saturated IEEE 802.11 DCF networks, the network throughput $\hat{\lambda}_{out}$ may fall below the aggregate input rate $\hat{\lambda}=n\lambda$ if the backoff parameters are not properly selected. The stable region of backoff factor $q$ is further characterized, within which a stable throughput $\hat{\lambda}_{out}=\hat{\lambda}$ can be always achieved at the undesired stable point $p_A$. The delay performance at $p_A$ may also significantly deteriorate and becomes closely dependent on backoff parameters. In this case, the key to reducing the access delay is to enlarge $p_A$, which can be achieved by increasing the initial backoff window size $W$. With a small $W$, the second moment of access delay exponentially grows with the cutoff phase $K$, and eventually becomes infinite as $K\to\infty$.

Our analysis sheds important light on the capability of IEEE 802.11 DCF networks for quality of service (QoS) provisioning, and provides direct guidance on network control and optimization. For instance, the maximum throughput of IEEE 802.11 DCF networks is derived in this paper, and shown to be solely determined by the holding times of HOL packets in successful transmission and collision states, $\tau_T$ and $\tau_F$. The optimal backoff factor and the optimal initial backoff window size to achieve the maximum throughput are obtained as functions of $\tau_F$ and the number of nodes $n$. It is shown that thanks to a drastic reduction of $\tau_F$, a higher maximum throughput is achieved in the RTS/CTS mode, which can be approached with a wide range of backoff factor $q$ or initial backoff window size $W$.

In current IEEE 802.11 DCF networks, Binary Exponential Backoff (BEB) is adopted (i.e., the backoff factor $q$ is fixed to be $1/2$), and small values of initial backoff window size and cutoff phase are selected. The analysis shows that the default standard setting leads to suboptimal performance, and may cause significant degradation when the network size or the traffic level increases. The stable region of initial backoff window size $W$ with BEB is characterized, and the optimal $W$ to achieve the maximum throughput is obtained as a linear function of the number of nodes $n$. The delay analysis further reveals that the fundamental reason behind the observed short-term unfairness of BEB is a high second moment of access delay in both the basic access and the RTS/CTS modes. To improve the delay performance at the saturated point, a large initial backoff window size $W$ should be adopted.

The remainder of this paper is organized as follows. Section 2 establishes the network model and presents the preliminary analysis. The stable regions are characterized in Section 3 and the delay analysis is presented in Section 4. A special focus is given to BEB in Section 5. Finally, conclusions are summarized in Section 6.

## 2 Modeling and Preliminary Analysis

Consider an $n$-node IEEE 802.11 DCF network with packet transmissions over a noiseless channel. A detailed description of the DCF protocol can be found in [2], and is omitted here. In this paper, we focus on a homogeneous buffered network, where each node has an arrival rate of $\lambda$ and is equipped with a buffer of infinite size. We assume that each HOL packet independently follows an identical transition process which will be characterized in the following subsection. Note that this assumption is also widely adopted in previous studies [2]-[29].

### 2.1 State Characterization of HOL Packets

A discrete-time Markov renewal process $(\boldsymbol{X}, \boldsymbol{V}) = \{(X_j, V_j), j = 0, 1, \dots\}$ is established in this subsection to model the behavior of each HOL packet. $X_j$ denotes the state of a tagged HOL packet at the $j$-th transition and $V_j$ denotes the epoch at which the $j$-th transition occurs. Fig. 1 shows the embedded Markov chain $\boldsymbol{X} = \{X_j\}$.

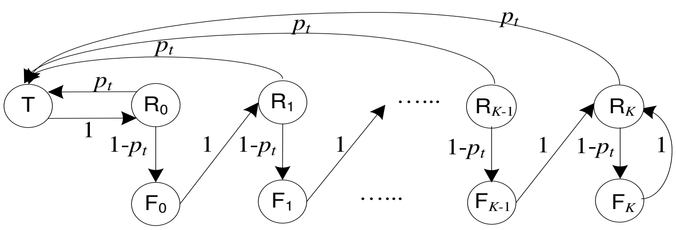



Fig. 1. Embedded Markov chain $\{X_j\}$ of the state transition process of an individual HOL packet in IEEE 802.11 DCF networks.

The states of $\{X_j\}$ can be divided into three categories: 1) waiting to request (State $\mathrm{R}_i$, $i = 0, \dots, K$), 2) collision (State $\mathrm{F}_i$, $i = 0, \dots, K$) and 3) successful transmission (State T). As Fig. 1 illustrates, a HOL packet moves from State $\mathrm{R}_i$ to State T if the request of transmission is successful. Otherwise, it stays at State $\mathrm{F}_i$ until the end of the collision and then shifts to State $\mathrm{R}_{i+1}$. Here $i$ denotes the number of collisions experienced by the HOL packet and is incremented until it reaches the cutoff phase $K$ (which is referred to as the *maximum backoff stage* in [2]).

In IEEE 802.11 DCF networks, a HOL packet can request a transmission only if it senses the channel idle. Let $p_t$ represent the probability of successful transmission of HOL packets at time slot $t$ given that the channel is idle at $t-1$. It can be easily shown that the Markov chain in Fig. 1 is uniformly strongly ergodic if and only if the limit

$$\lim_{t\to\infty} p_t = p \tag{1}$$

exists [40]. The steady-state probability distribution of the embedded Markov chain can be further obtained as

$$\pi_{R_i} = \begin{cases} (1-p)^i \pi_T & i = 0, ..., K-1 \\ \frac{(1-p)^K}{p} \pi_T & i = K \end{cases} \tag{2}$$

and

$$\pi_{F_i} = \pi_{R_i} \cdot (1-p), \quad i = 0, \dots, K. \tag{3}$$

The interval between successive transitions, i.e., $V_{j+1} - V_j$, is called the holding time in State $X_j$, which solely depends on State $X_j, j = 0, 1, \dots$. In IEEE 802.11 DCF networks, the holding time $\tau_T$ in State T and the holding time $\tau_F$ in State $\mathrm{F}_i, i = 0, \dots, K$, vary under different access mechanisms. A graphic illustration of $\tau_T$ and $\tau_F$ in the basic access and RTS/CTS modes can be found in Fig. 5 of [2] (corresponding to $T_s$ and $T_c$, respectively, in unit of time slots), and the typical values are provided in Table V of [2].

The mean holding time $\tau_{R_i}$ in State $\mathrm{R}_i$, $i = 0, \dots, K$, on the other hand, is determined by the backoff protocol. In IEEE 802.11 DCF networks, when a HOL packet enters State $\mathrm{R}_i$, it randomly selects a value from $\{0, \dots, W_i - 1\}$, where $W_i$ is the backoff window size, $i = 0, \dots, K$, and then counts down at each *idle* time slot. It leaves State $\mathrm{R}_i$ and makes a transmission request when the channel is idle and the counter is zero. Let $G_t^i$ denote the state of the backoff counter of a State-$\mathrm{R}_i$ HOL packet at time slot $t$, $i = 0, \dots, K$. The transition process of $\{G_t^i\}$ can be described by the Markov chain shown in Fig. 2, where $\alpha_t$ represents the probability of sensing the channel idle at time slot $t$.

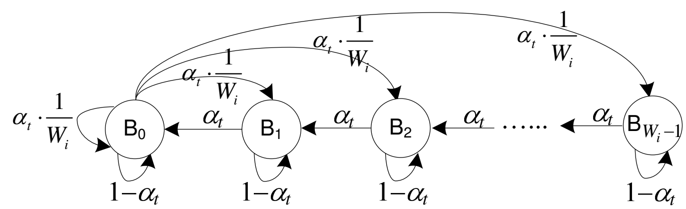



Fig. 2. State transition diagram of a State-$\mathrm{R}_i$ HOL packet in IEEE 802.11 DCF networks, $i = 0, \dots, K$.

Similarly, the Markov chain in Fig. 2 is uniformly strongly ergodic if and only if the limit

$$\lim_{t\to\infty} \alpha_t = \alpha \tag{4}$$

exists [40]. The mean holding time $\tau_{R_i}$ can be then obtained as

$$\tau_{R_i} = \frac{1}{\alpha} \cdot \frac{1 + W_i}{2}, \tag{5}$$

$i = 0, \dots, K$. Appendix A presents the detailed derivation of (5). Appendix B further shows that

$$\alpha = \frac{1}{1 + \tau_F - \tau_F p - (\tau_T - \tau_F) p \ln p}. \tag{6}$$

By substituting (6) into (5), the mean holding time $\tau_{R_i}$ can be written as

$$\tau_{R_i} = \frac{1}{2} (1 + W_i)(1 + \tau_F - \tau_F p - (\tau_T - \tau_F) p \ln p). \tag{7}$$

Finally, the limiting state probabilities of the Markov renewal process ($\boldsymbol{X}$, $\boldsymbol{V}$) are given by [41]

$$\tilde{\pi}_j = \frac{\pi_j \cdot \tau_j}{\sum_{i \in S} \pi_i \cdot \tau_i}, \tag{8}$$

$j \in S$, where $S$ is the state space of $\boldsymbol{X}$. Specifically, the probability of being in State T can be obtained as

$$\tilde{\pi}_T = 1 \Big/ \left( 1 + \frac{\tau_F}{\tau_T} \cdot \frac{1-p}{p} + \left( \frac{1}{\tau_T} + \frac{\tau_F}{\tau_T} \cdot (1-p) - \left(1 - \frac{\tau_F}{\tau_T}\right) \cdot p \ln p \right) \cdot \left( \sum_{i=0}^{K-1} (1-p)^i \cdot \frac{1+W_i}{2} + \frac{(1-p)^K}{p} \cdot \frac{1+W_K}{2} \right) \right) \tag{9}$$

by substituting (2-3) and (7) into (8). Note that $\tilde{\pi}_T$ is also the service rate of each node's queue as each queue has a successful output if and only if the HOL packet stays at State T. The offered load of each node's queue, $\rho$, is then given by

$$\rho = \lambda / \tilde{\pi}_T, \tag{10}$$

where $\lambda$ is the input rate of each node.

### 2.2 Limiting Probability of Success $p$

The analysis in Section 2.1 indicates that the steady-state performance of IEEE 802.11 DCF networks is crucially determined by $p$, the limiting probability of successful transmission of HOL packets given that the channel is idle. In this subsection, the steady-state operating points of IEEE 802.11 DCF networks will be characterized based on the fixed-point equations of $p$.

#### 2.2.1 Steady-state Point of Unsaturated Networks

Let us first consider an unsaturated network. Each node must be in one of the following four states:
$S_1$: the queue is empty;
$S_2$: the HOL packet is in State $R_i = 0, \ldots, K$;
$S_3$: the HOL packet is in State T;
$S_4$: the HOL packet is in State $F_i = 0, \ldots, K$.

For a given HOL packet, its transmission request is successful if and only if the other $n-1$ nodes are either in state $S_1$, or, in state $S_2$ and not requesting any transmission. The limiting probability of successful transmission of HOL packets given that the channel is idle, $p$, is then given by

$$\begin{aligned} p = (&\Pr\{\text{node is in } S_1 | \text{channel is idle}\} \\ &+ \Pr\{\text{node is in } S_2 \text{ with no request} | \text{channel is idle}\})^{n-1}. \end{aligned} \tag{11}$$

If the channel is idle, each node must be in either state $S_1$ or $S_2$. The probabilities that a node is in state $S_j$, $j{=}1,2$, are given by

$$\Pr\{\text{node is in } S_1\} = 1 - \rho, \tag{12}$$

$$\Pr\{\text{node is in } S_2\} = \rho \sum_{i=0}^{K} \tilde{\pi}_{R_i}, \tag{13}$$

where the offered load of each node's queue $\rho < 1$. We then have

$$\Pr\{\text{node is in } S_1 | \text{channel is idle}\} = \frac{1-\rho}{1-\rho+\rho\sum_{i=0}^{K}\tilde{\pi}_{R_i}}, \tag{14}$$

and

$$\begin{aligned} &\Pr\{\text{node is in } S_2 \text{ with no request} | \text{channel is idle}\} \\ &= \frac{\rho\sum_{i=0}^{K}\tilde{\pi}_{R_i}(1-r_i)}{1-\rho+\rho\sum_{i=0}^{K}\tilde{\pi}_{R_i}}, \end{aligned} \tag{15}$$

where $r_i$ is the conditional probability of a State-$R_i$ HOL packet making a transmission request given that the channel is idle, which can be obtained as

$$r_i = \frac{2}{1+W_i}, \tag{16}$$

$i = 0, \ldots, K$. Detailed derivation of (16) can be found in Appendix A. By substituting (14-15) into (11), the limiting probability of success $p$ can be written as

$$p = \left\{ \frac{1-\rho+\rho\sum_{i=0}^{K}\tilde{\pi}_{R_i}(1-r_i)}{1-\rho+\rho\sum_{i=0}^{K}\tilde{\pi}_{R_i}} \right\}^{n-1}. \tag{17}$$

With a large number of nodes $n$, we have

$$p \approx \exp\left\{ -n\rho \sum_{i=0}^{K} \tilde{\pi}_{R_i} r_i \right\} = \exp\left\{ -\hat{\lambda} \sum_{i=0}^{K} \frac{\tilde{\pi}_{R_i}}{\tilde{\pi}_T} r_i \right\}, \tag{18}$$

according to (10), where $\hat{\lambda} = n\lambda$ is the aggregate input rate.

Finally, by combining (2-3), (7-8) and (16), (18) can be written as

$$p = \exp\left\{ \frac{\hat{\lambda}\tau_F/\tau_T}{1-(1-\tau_F/\tau_T)\hat{\lambda}} \right\} \cdot \exp\left\{ -\frac{\hat{\lambda}(1+\tau_F)/\tau_T}{1-(1-\tau_F/\tau_T)\hat{\lambda}} \cdot \frac{1}{p} \right\}. \tag{19}$$

(19) has two non-zero roots:

$$p_L = \exp\left\{ \mathbb{W}_0\left( -\frac{\hat{\lambda}(1+\tau_F)/\tau_T}{1-(1-\tau_F/\tau_T)\hat{\lambda}} \cdot \exp\left\{ -\frac{\hat{\lambda}\tau_F/\tau_T}{1-(1-\tau_F/\tau_T)\hat{\lambda}} \right\} \right) + \frac{\hat{\lambda}\tau_F/\tau_T}{1-(1-\tau_F/\tau_T)\hat{\lambda}} \right\} \tag{20}$$

and

$$p_S = \exp\left\{ \mathbb{W}_{-1}\left( -\frac{\hat{\lambda}(1+\tau_F)/\tau_T}{1-(1-\tau_F/\tau_T)\hat{\lambda}} \cdot \exp\left\{ -\frac{\hat{\lambda}\tau_F/\tau_T}{1-(1-\tau_F/\tau_T)\hat{\lambda}} \right\} \right) + \frac{\hat{\lambda}\tau_F/\tau_T}{1-(1-\tau_F/\tau_T)\hat{\lambda}} \right\}, \tag{21}$$

if the aggregate input rate $\hat{\lambda}$ is no larger than

$$\hat{\lambda}_{\max} = \frac{-\mathbb{W}_0\left(-\frac{1}{e(1+1/\tau_F)}\right)}{\tau_F/\tau_T - (1-\tau_F/\tau_T)\mathbb{W}_0\left(-\frac{1}{e(1+1/\tau_F)}\right)}. \tag{22}$$

$\mathbb{W}_0(\cdot)$ and $\mathbb{W}_{-1}(\cdot)$ in (20-22) are two branches of the Lambert W function [1] [42], and we have $p_S \leq p_L$. It is interesting to note that only the larger root $p_L$ is a steady-state operating point. It is determined by the aggregate input rate $\hat{\lambda}$ and the holding times in successful transmission and collision states, $\tau_T$ and $\tau_F$, and does not vary with the backoff parameters. A stable throughput $\hat{\lambda}_{out} = \hat{\lambda}$ can be always achieved at $p_L$ as the offered load $\rho$ of each node's queue is lower than 1 if the network is unsaturated.

#### 2.2.2 Steady-state Point of Saturated Networks

As the aggregate input rate $\hat{\lambda}$ increases, the network will eventually become saturated, i.e., all the nodes are busy with non-empty queues. In this case, given that the channel is idle, each node must be in state $S_2$. Similar to

1. The defining equation for the Lambert W function $\mathbb{W}(z)$ is $z = \mathbb{W}(z)e^{\mathbb{W}(z)}$ for any complex number $z$.

(11), the limiting probability of success $p$ in a saturated network can be written as

$$p = (\Pr\{\text{node is in } \mathrm{S}_2 \text{ with no request}|\text{channel is idle}\})^{n-1}, \quad (23)$$

which is given by

$$p = \left\{ \frac{\sum_{i=0}^{K} \tilde{\pi}_{R_i}(1-r_i)}{\sum_{i=0}^{K} \tilde{\pi}_{R_i}} \right\}^{n-1} \overset{\text{with a large } n}{\approx} \exp\left\{ -n \sum_{i=0}^{K} \tilde{\pi}_{R_i} r_i \right\}. \quad (24)$$

By substituting (2-3), (5), (8) and (16) into (24), we have

$$p = \exp\left\{ -n \Big/ \left( \alpha(\tau_T \cdot p + \tau_F \cdot (1-p)) + \frac{1}{2}\left(1 + \sum_{i=0}^{K-1} p(1-p)^i \cdot W_i + (1-p)^K \cdot W_K \right)\right)\right\}, \quad (25)$$

where $\alpha$ is the limiting probability of sensing the channel idle which is given in (6).

It is difficult to derive the exact root of (25). Nevertheless, it will be shown in Fig. 4 that the idle probability $\alpha$ is close to zero when the network becomes saturated. By ignoring the first term of the denominator in the right-hand side of (25), (25) can be written as

$$p = \exp\left\{ -\frac{2n}{1 + \sum_{i=0}^{K-1} p(1-p)^i \cdot W_i + (1-p)^K \cdot W_K} \right\}. \quad (26)$$

In IEEE 802.11 DCF networks, the backoff window size $W_i$ is set as

$$W_i = W \cdot q^{-i}, \quad (27)$$

$i = 0, \ldots, K$. $W$ is called the initial (or minimum) backoff window size and $q$ is the backoff factor. By combining (26) and (27), we have

$$p = \exp\left\{ -\frac{2n}{1 + W\left( \frac{qp}{q+p-1} - \left( \frac{qp}{q+p-1} - 1 \right) \cdot \left( \frac{1-p}{q} \right)^K \right)} \right\}. \quad (28)$$

Recall that in [2], the fixed-point equation of the collision probability (i.e., corresponding to $1-p$ in this paper) was derived for saturated IEEE 802.11 DCF networks with Binary Exponential Backoff ($q$=1/2). It can be easily shown that (28) is consistent with Eq. (9) in [2] if $q$ is equal to $1/2$.

In contrast to (19), (28) has a single non-zero root $p_A$ for any cutoff phase $K = 0, \ldots, \infty$. Moreover, the non-zero root $p_A$ is closely dependent on backoff parameters such as the cutoff phase $K$, the backoff factor $q$ and the initial backoff window size $W$. With $K = \infty$, $p_A$ can be explicitly written as

$$p_A^{K=\infty} \overset{\text{for large } W}{\approx} \frac{2n(1-q)/(Wq)}{\mathbb{W}_0\left(2n(1-q)/(Wq) \cdot \exp\left(2n/W/q\right)\right)}. \quad (29)$$

### 2.3 Desired Stable Point versus Undesired Stable Point

So far we have demonstrated that for any aggregate input rate $\hat{\lambda} \leq \hat{\lambda}_{\max}$, an IEEE 802.11 DCF network operates at the steady-state point $p_L$ if the network is unsaturated. As the aggregate input rate $\hat{\lambda}$ increases, however, the network may become saturated and shift to another steady-state point $p_A$. As IEEE 802.11 DCF networks possess the same bi-stable property as Aloha networks [43], we follow the terminology in [43] and refer to $p_L$ and $p_A$ as the *desired stable point* and the *undesired stable point*, respectively.

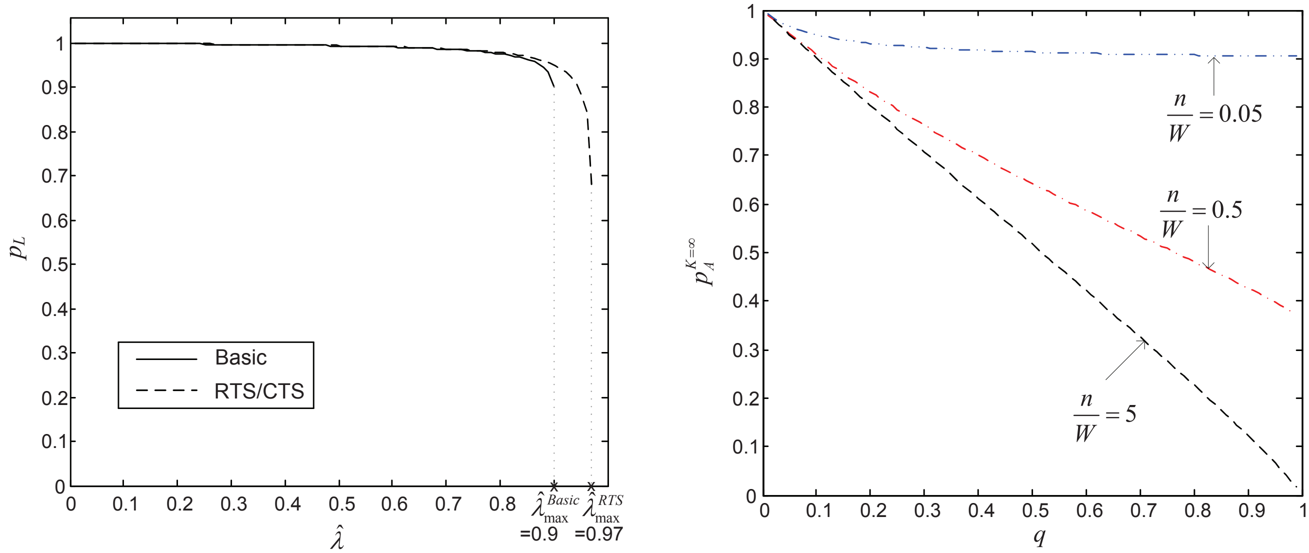
#### 2.3.1 Desired Stable Point $p_L$

(20) indicates that the desired stable point $p_L$ does not vary with backoff parameters, and is crucially determined by the holding times in successful transmission and collision states, $\tau_T$ and $\tau_F$. Typical values of $\tau_T$ and $\tau_F$ in the basic access and RTS/CTS modes have been summarized in Table V of [2]. In particular, with the basic access mechanism, nodes are unaware of whether their transmissions are successful or not until the end of the packet frame, and hence $\tau_T^{Basic}$ and $\tau_F^{Basic}$ are approximately equal to each other. With the RTS/CTS mechanism, collisions occur only on the RTS frames that are much shorter than the payload frames [2]. The holding time in collision states $\tau_F^{RTS}$ is therefore drastically reduced and becomes much smaller than $\tau_T^{RTS}$. Fig. 3a illustrates how the desired stable point $p_L$ varies with the aggregate input rate $\hat{\lambda}$ in the basic access mode (i.e., $\tau_T^{Basic}$=180 time slots and $\tau_F^{Basic}$=175 time slots) and the RTS/CTS mode (i.e., $\tau_T^{RTS}$=192 time slots and $\tau_F^{RTS}$=9 time slots). It can be clearly seen from Fig. 3a that $p_L$ declines as $\hat{\lambda}$ increases in both modes, and for given aggregate input rate $\hat{\lambda}$, a higher $p_L$ is always obtained in the RTS/CTS mode thanks to the reduction of the holding time in collision states.

#### 2.3.2 Undesired Stable Point $p_A$

In contrast, the undesired stable point $p_A$ is a function of 1) the cutoff phase $K$, 2) the backoff factor $q$, 3) the initial backoff window size $W$ and 4) the number of nodes $n$. For instance, (29) indicates that $p_A^{K=\infty}$ monotonically decreases as the backoff factor $q$ or the ratio of the number of nodes $n$ and the initial backoff window size $W$ increases, which can be clearly observed from Fig. 3b. $p_A$ could be much lower than $p_L$ if the backoff parameters are not properly selected, implying that both throughput and delay performance may significantly deteriorate when the network becomes saturated.

The above analysis is verified by the simulation results presented in Fig. 4. In this paper, all the simulations are conducted using the *ns*-2 simulator, and the values of system parameters are in accordance with [2] (which were summarized in Table II of [2]). In the simulations, each node is assumed to have Bernoulli arrivals with rate $\lambda$, i.e., each node has probability $\lambda$ to generate a




(a) (b)

Fig. 3. Stable points of IEEE 802.11 DCF networks. (a) Desired stable point $p_L$ versus aggregate input rate $\hat{\lambda}$. (b) Undesired stable point $p_A^{K=\infty}$ versus backoff factor $q$.

new packet every $\tau_T$ time slots. The new arrival packet is attached to the end of the waiting queue, and the buffer size is assumed to be infinite. It can be clearly seen from Fig. 4 that with a low aggregate input rate $\hat{\lambda} = 0.2$, the network operates at the desired stable point $p_L$, which does not vary with the backoff factor $q$. While the aggregate input rate $\hat{\lambda}$ increases to 0.8, the network becomes saturated at the undesired stable point $p_A$ that declines as $q$ increases. It is also shown in Fig. 4 that the limiting probability of sensing the channel idle $\alpha$ is close to zero when the network operates at $p_A$. It verifies that (26) can serve as a good approximation of (25).

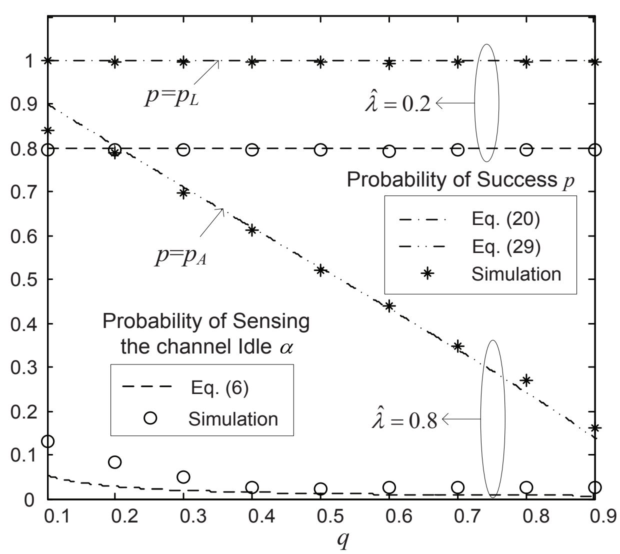



Fig. 4. Desired and undesired stable points of IEEE 802.11 DCF networks with the basic access mechanism. $n = 50$, $K = \infty$ and $W = 16$.

Note that the network exhibits distinct performances at different stable points. A stable throughput $\hat{\lambda}_{out} = \hat{\lambda}$ can be always achieved when it operates at the desired stable point $p_L$. In contrast, the undesired stable point $p_A$ depends on backoff parameters, indicating that the network throughput $\hat{\lambda}_{out}$ may fall below the aggregate input rate $\hat{\lambda}$ unless the backoff parameters are carefully selected. In the next section, we will focus on the throughput performance when the network operates at the undesired stable point $p_A$. We are particularly interested in how to properly choose the backoff parameters to stabilize the network at $p_A$, and how to achieve the maximum stable throughput.

## 3 Stable Region and Maximum Stable Throughput

Section 2 shows that an IEEE 802.11 DCF network operates at the undesired stable point $p_A$ if it is saturated. According to (24), the aggregate service rate at $p_A$ can be obtained as

$$n\tilde{\pi}_T = \frac{-\tau_T p_A \ln p_A}{1 + \tau_F - \tau_F p_A - (\tau_T - \tau_F) p_A \ln p_A}. \tag{30}$$

To achieve a stable throughput $\hat{\lambda}_{out} = \hat{\lambda}$, the aggregate service rate should be no smaller than the aggregate input rate $\hat{\lambda}$:

$$n\tilde{\pi}_T \geq \hat{\lambda}. \tag{31}$$

By combining (30-31) and (19), we can conclude that a stable throughput can be achieved at $p_A$ if and only if

$$p_S \leq p_A \leq p_L, \tag{32}$$

where $p_L$ and $p_S$ are given in (20-21).

Note that the undesired stable point $p_A$ is a function of backoff parameters. In this section, we will demonstrate how to stabilize the network at $p_A$ by properly choosing the backoff factor $q$. Specifically, define $S_A^q$ as the stable region of backoff factor $q$, within which a stable throughput $\hat{\lambda}_{out} = \hat{\lambda}$ is achieved at the undesired stable point $p_A$. According to (32), $S_A^q$ can be written as

$$S_A^q = \{q|p_S \le p_A \le p_L\}. \tag{33}$$

According to (20-21), $p_L$ and $p_S$ are monotonic decreasing and increasing functions of the aggregate input rate $\hat{\lambda}$, respectively. As a result, $S_A^q$ will shrink as $\hat{\lambda}$ increases, and eventually becomes a single point when $\hat{\lambda}$ reaches the maximum stable throughput $\hat{\lambda}_{\max\_S_A^q}$. With $\hat{\lambda}>\hat{\lambda}_{\max\_S_A^q}$, $S_A^q$ is an empty set, implying that the network cannot be stabilized. We will elaborate on the above results in the following subsections.

### 3.1 Stable Region of Backoff Factor $q$ with Cutoff Phase $K{=}\infty$

Let us first assume an infinite cutoff phase $K{=}\infty$. The undesired stable point $p_A^{K=\infty}$ has been given in (29), and the corresponding stable region can be obtained by combining (33) and (29) as

$$S_A^{q,K=\infty} = \left[\frac{1-p_L}{1+\frac{W}{2n}p_L\ln p_L}, \frac{1-p_S}{1+\frac{W}{2n}p_S\ln p_S}\right]. \tag{34}$$

According to (34), the stable region $S_A^{q,K=\infty}$ diminishes as the aggregate input rate $\hat{\lambda}$ increases, and finally shrinks to a single point

$$q_m^{K=\infty}{=}\left(1{+}(1{+}1/\tau_F)\mathbb{W}_0\left(-\tfrac{1}{e(1+1/\tau_F)}\right)\right)\Big/\left(1{-}\frac{W}{2n}(1{+}1/\tau_F)\cdot\mathbb{W}_0\left(-\tfrac{1}{e(1+1/\tau_F)}\right)\ln\left(-(1{+}1/\tau_F)\mathbb{W}_0\left(-\tfrac{1}{e(1+1/\tau_F)}\right)\right)\right) \tag{35}$$

with which the maximum stable throughput of

$$\hat{\lambda}_{\max\_S_A^{q,K=\infty}}{=}\hat{\lambda}_{\max}{=}\frac{-\mathbb{W}_0\left(-\frac{1}{e(1+1/\tau_F)}\right)}{\tau_F/\tau_T{-}(1{-}\tau_F/\tau_T)\mathbb{W}_0\left(-\frac{1}{e(1+1/\tau_F)}\right)} \tag{36}$$

can be achieved. Note that $q_m^{K=\infty}$ in (35) should not exceed 1, which requires that

$$W \le \frac{2n}{\ln\left(-(1+1/\tau_F)\mathbb{W}_0\left(-\frac{1}{e(1+1/\tau_F)}\right)\right)}. \tag{37}$$

Otherwise, the maximum stable throughput $\hat{\lambda}_{\max\_S_A^{q,K=\infty}}$ is lower than $\hat{\lambda}_{\max}$.

Recall that it is shown in Section 2.2.1 that the desired stable point $p_L$ exists if and only if the aggregate input rate $\hat{\lambda}$ is no larger than $\hat{\lambda}_{\max}$. Here we can conclude from (22) and (36) that $\hat{\lambda}_{\max}$ is the maximum throughput that can be achieved by an IEEE 802.11 DCF network irrespective of which stable point it operates at. It is independent of the backoff parameters and solely determined by the holding times in successful transmission and collision states, $\tau_T$ and $\tau_F$. The maximum throughputs of IEEE 802.11 DCF networks with the basic access mechanism (i.e., $\tau_T^{Basic}{=}180$ time slots and $\tau_F^{Basic}{=}175$ time slots) and the RTS/CTS mechanism (i.e., $\tau_T^{RTS}{=}192$ time slots and $\tau_F^{RTS}{=}9$ time slots) can be obtained as $\hat{\lambda}_{\max}^{Basic}{=}0.9$ and $\hat{\lambda}_{\max}^{RTS}{=}0.97$, respectively.

For any aggregate input rate $\hat{\lambda} \le \hat{\lambda}_{\max\_S_A^{q,K=\infty}}$, a stable throughput $\hat{\lambda}_{out} = \hat{\lambda}$ can be achieved at $p_A^{K=\infty}$ if the backoff factor $q$ is selected from the stable region $S_A^{q,K=\infty}$. With $q \notin S_A^{q,K=\infty}$, the network throughput falls below the aggregate input rate $\hat{\lambda}$ and is determined by the aggregate service rate, which can be obtained by combining (29-30) as:

$$\hat{\lambda}_s^{K=\infty}{=}\frac{\tau_T}{(\tau_T{-}\tau_F){+}\frac{(1+\tau_F)\cdot\mathbb{W}_0\left(\frac{2n(1-q)}{Wq}\cdot\exp\left(\frac{2n}{Wq}\right)\right)-\frac{2n\tau_F(1-q)}{Wq}}{\frac{4n^2(1-q)}{W^2q^2}-\frac{2n(1-q)}{Wq}\cdot\mathbb{W}_0\left(\frac{2n(1-q)}{Wq}\cdot\exp\left(\frac{2n}{Wq}\right)\right)}}. \tag{38}$$

In the following subsections, we will demonstrate the above results in the cases of the basic access mode and the RTS/CTS mode, respectively.

#### 3.1.1 Basic Access Mechanism

Fig. 5a presents the stable region $S_A^{q,K=\infty}$ of a 50-node IEEE 802.11 DCF network with the basic access mechanism. It can be observed from Fig. 5a that the stable region $S_A^{q,K=\infty}$ is substantially enlarged as the initial backoff window size $W$ increases from 8 to 128. Intuitively, with a larger $W$, HOL packets should have higher chances to succeed as their requests are spread in a more even way. A closer look at (29) and Fig. 3b further indicates that the probability of success is indeed improved with an increase of the initial backoff window size $W$.

Nevertheless, an excessively large $W$ may impair the throughput performance. With $W{=}8$, 32 and 128, the stable region $S_A^{q,K=\infty}$ rapidly shrinks as the aggregate input rate $\hat{\lambda}$ increases, and finally becomes a single point $q_m^{K=\infty}$ when $\hat{\lambda}$ reaches the maximum $\hat{\lambda}_{\max\_S_A^{q,K=\infty}}{=}\hat{\lambda}_{\max}^{Basic}{=}0.9$. With $W{=}1024$, in contrast, the maximum stable throughput $\hat{\lambda}_{\max\_S_A^{q,K=\infty}}$ is slightly lower than $\hat{\lambda}_{\max}^{Basic}$. In fact, according to (37), $\hat{\lambda}_{\max\_S_A^{q,K=\infty}}$ starts declining from $\hat{\lambda}_{\max}^{Basic}$ if the initial backoff window size $W$ exceeds 971.

#### 3.1.2 RTS/CTS Mechanism

The stable region $S_A^{q,K=\infty}$ with the RTS/CTS mechanism is also illustrated in Fig. 5a. Compared to the basic access mode, the stable region is now greatly enlarged and becomes insensitive to the aggregate input rate $\hat{\lambda}$ and the initial backoff window size $W$. Moreover, the maximum stable throughput $\hat{\lambda}_{\max\_S_A^{q,K=\infty}}$ is boosted to $\hat{\lambda}_{\max}^{RTS}{=}0.97$, which can be approached with a wide range of backoff factor $q$ even with a small $W$. Both facts are attributed to the decrease of the holding time in collision states.

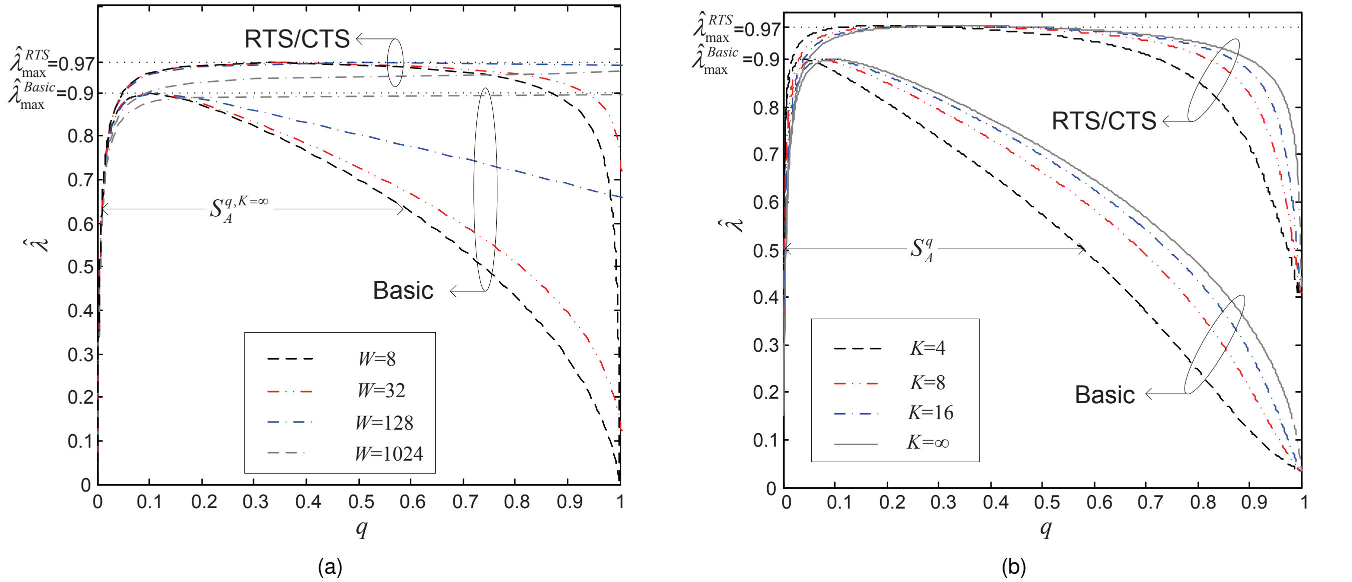



Fig. 5. Stable region of backoff factor $q$ in IEEE 802.11 DCF networks. $n$=50. (a) Stable region with various values of initial backoff window size $W$. $K$=$\infty$. (b) Stable region with various values of cutoff phase $K$. $W$=32.

### 3.2 Stable Region of Backoff Factor $q$ with Cutoff Phase $K$<$\infty$

With a finite cutoff phase $K < \infty$, the undesired stable point $p_A$ is an implicit function of the backoff factor $q$. Numerical results of the stable region $S_A^q$ can be obtained by combining (28) and (33), and are presented in Fig. 5b. It can be seen from Fig. 5b that the maximum stable throughput $\hat{\lambda}_{\max\_S_A^{q,K=\infty}}$=$\hat{\lambda}_{\max}$ can be always achieved with $K \geq 4$. Moreover, a large cutoff phase $K$ leads to an improved stable region in both the basic access and the RTS/CTS modes. Intuitively, HOL packets would back off to deeper phases if they experience collisions, and the transmission probability quickly decreases as the phase number grows, until it reaches the maximum $K$. A large cutoff phase $K$ implies that nodes have more room to reduce their transmission probabilities, and hence the network is better capable of absorbing the mounting contention and remaining stable as the aggregate input rate increases.

### 3.3 Simulation Results and Discussions

So far we have shown that when the network operates at the undesired stable point $p_A$, a stable throughput $\hat{\lambda}_{out} = \hat{\lambda}$ can be achieved if the backoff factor $q$ is selected from the corresponding stable region $S_A^q$. Otherwise, the throughput $\hat{\lambda}_{out}$ falls below $\hat{\lambda}$ and the network becomes unstable. The stable region $S_A^{q,K=\infty}$ and the network throughput with $q \notin S_A^{q,K=\infty}$ have been given in (34) and (38), respectively, and are verified by the simulation results presented in Fig. 6a. A closer look at Fig. 6a also confirms that the stable region can be significantly improved by properly enlarging the initial backoff window size $W$ if the basic access mechanism is adopted. With RTS/CTS, however, the stable region is not sensitive to the value of $W$.

With a finite cutoff phase $K$<$\infty$, the numerical results of the stable region $S_A^q$ under various values of cutoff phase $K$ have been presented in Fig. 5b and are verified by the simulation results presented in Fig. 6b. It can be clearly observed from Fig. 6b that with the basic access mechanism, the stable region is remarkably enlarged even with a slight increment of $K$ when the cutoff phase $K$ is small. The improvement nevertheless becomes marginal as $K$ increases. The stable region with $K$=16 approaches that with $K$=$\infty$. If the RTS/CTS mechanism is adopted, the cutoff phase $K$ can be further reduced.

Figs. 6a and 6b corroborate that the stable region of backoff factor $q$ can be always improved by increasing the initial backoff window size $W$ or the cutoff phase $K$. A proper selection of $W$ and $K$ is especially crucial for networks in the basic access mode. In contrast, with the RTS/CTS mechanism, a stable throughput can be achieved within a wide range of backoff factor $q$ even for small $W$ and $K$, indicating that IEEE 802.11 DCF networks in the RTS/CTS mode are much more robust compared to those in the basic access mode.

## 4 Delay Analysis

In this section, we will characterize the first and second moments of access delay of HOL packets and explore how the moments of access delay vary with system parameters at the desired and undesired stable points.

Let $Y_i$ denote the holding time of a HOL packet in State $R_i$, and $D_i$ denote the time spent from the beginning of State $R_i$ until the service completion, $i =$

$$G''_{D_0}(1)=\sum_{i=0}^{K-1}(1-p)^i G''_{Y_i}(1)+\frac{(1-p)^K}{p}G''_{Y_K}(1)+2\left(p\tau_T+(1-p)\,\tau_F\right)\cdot\left(\sum_{i=0}^{K-1}(1-p)^i G'_{Y_i}(1)+\frac{(1-p)^K}{p}G'_{Y_K}(1)\right)$$
$$+\,2\sum_{i=0}^{K-1}\left(\tau_F+G'_{Y_i}(1)\right)\cdot\left(\sum_{j=i+1}^{K-1}(1-p)^j\left(p\tau_T+(1-p)\tau_F+G'_{Y_j}(1)\right)+\frac{(1-p)^K}{p}\left(p\tau_T+(1-p)\tau_F+G'_{Y_K}(1)\right)\right)$$
$$+\,2\frac{(1-p)^{K+1}}{p^2}\left(\tau_F+G'_{Y_K}(1)\right)\cdot\left(p\tau_T+(1-p)\tau_F+G'_{Y_K}(1)\right)+\tau_T\left(\tau_T-1\right)+\frac{1-p}{p}\tau_F\left(\tau_F-1\right). \tag{42}$$

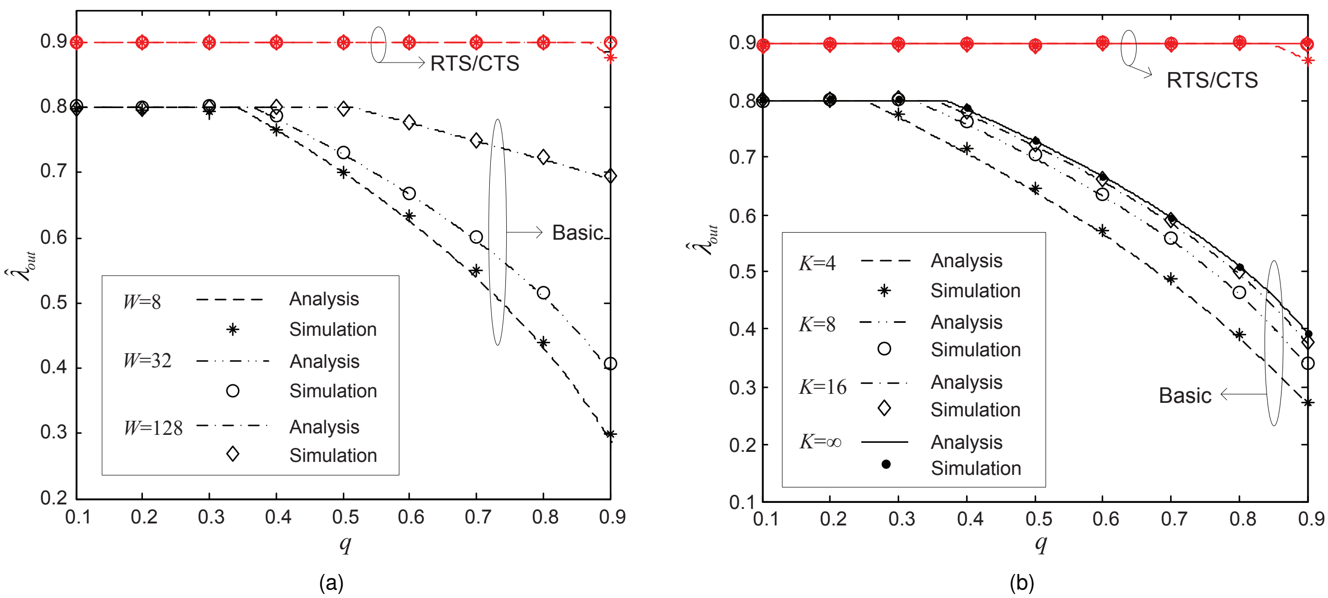



Fig. 6. Network throughput $\hat{\lambda}_{out}$ versus backoff factor $q$ in IEEE 802.11 DCF networks. $\hat{\lambda}$=0.8 with the basic access mechanism and $\hat{\lambda}$=0.9 with the RTS/CTS mechanism. $n$=50. (a) Network throughput with various values of initial backoff window size $W$. $K$=$\infty$. (b) Network throughput with various values of cutoff phase $K$. $W$=32.

$0,\ldots,K$. According to Fig. 1, we have

$$D_i=\begin{cases}Y_i+\tau_T & \text{with probability } p\\ Y_i+\tau_F+D_{i+1} & \text{with probability } 1-p,\end{cases} \tag{39}$$

$i$=$0,\ldots,K-1$, and

$$D_K=\begin{cases}Y_K+\tau_T & \text{with probability } p\\ Y_K+\tau_F+D_K & \text{with probability } 1-p,\end{cases} \tag{40}$$

where $\tau_T$ and $\tau_F$ are holding times in State T and States F$_i$, $i=0,\ldots,K$, respectively.

Note that $D_0$ is the service time of HOL packets (also the access delay). Let $G_{D_0}(z)$ denote its probability generating function. It can be obtained that

$$G'_{D_0}(1)=\tau_T+\frac{1-p}{p}\tau_F+\sum_{i=0}^{K-1}(1-p)^i G'_{Y_i}(1)+\frac{(1-p)^K}{p}G'_{Y_K}(1), \tag{41}$$

and $G''_{D_0}(1)$ is shown at the top of the page, where $G'_{Y_i}(1)$ and $G''_{Y_i}(1)$ are given by

$$G'_{Y_i}(1)=\frac{1}{2\alpha}\left(Wq^{-i}+1\right) \tag{43}$$

and

$$G''_{Y_i}(1)=\frac{1}{3\alpha^2}W^2q^{-2i}+\frac{1-\alpha}{\alpha^2}Wq^{-i}+\frac{2-3\alpha}{3\alpha^2}, \tag{44}$$

respectively, $i$=0, ..., $K$. Appendix C presents the detailed derivation of (41-44).

Accordingly, the mean access delay $E[D_0]$ (in the unit of time slots) can be obtained as

$$E[D_0]=G'_{D_0}(1)=\tau_T+\frac{1-p}{p}\tau_F+\frac{1}{\alpha}\cdot\left(\frac{1}{2p}+\frac{W}{2}\left(\frac{1}{1-\frac{1-p}{q}}+\left(\frac{1}{p}-\frac{1}{1-\frac{1-p}{q}}\right)\cdot\left(\frac{1-p}{q}\right)^K\right)\right). \tag{45}$$

The second moment of access delay $E[D_0^2]$ is given by

$$E[D_0^2]=G''_{D_0}(1)+G'_{D_0}(1), \tag{46}$$

which can be obtained by substituting (41-44) into (46).

Equations (45-46) indicate that the delay performance crucially depends on $p$, the probability of successful transmission of HOL packets given that the channel is idle. Section 2 has revealed that an IEEE 802.11 DCF network operates at the desired stable point $p=p_L$ if it is unsaturated, and shifts to the undesired stable point $p=p_A$ if it becomes saturated. In the following sections, we will analyze the delay performance at the bi-stable points $p_L$ and $p_A$, respectively.

### 4.1 Access Delay at the Desired Stable Point $p_L$

(20) has shown that the desired stable point $p_L$ is determined by the aggregate input rate $\hat{\lambda}$ and the holding times in successful transmission and collision states, $\tau_T$ and $\tau_F$. It is further illustrated in Fig. 3a that $p_L$ is close to 1 within a wide range of aggregate input rate $\hat{\lambda}$ in both the basic access and the RTS/CTS modes. With $p_L \approx 1$, the first and second moments of access delay can be obtained as

$$E[D_{0,p=p_L}] \approx \tau_T + \frac{1+W}{2}, \tag{47}$$

and

$$E[D^2_{0,p=p_L}] \approx \tau_T^2 + (1+W)\tau_T + \frac{1+3W+2W^2}{6}, \tag{48}$$

respectively, according to (45-46). It is clear from (47-48) that both $E[D_{0,p=p_L}]$ and $E[D^2_{0,p=p_L}]$ increase with the holding time in the successful transmission state $\tau_T$ and the initial backoff window size $W$. They are insensitive to the backoff factor $q$, the cutoff phase $K$, and the holding time in collision states $\tau_F$ because the HOL packets hardly encounter any collisions if $p_L$ is close to one. They are also invariant to the number of nodes $n$ as $p_L$ is independent of $n$.

The analysis is verified by simulation results presented in Fig. 7. It can be seen from Fig. 7 that better delay performance is achieved in the basic access mode, because the holding time in the successful transmission state is larger in the RTS/CTS mode due to the overhead of RTS/CTS frames.

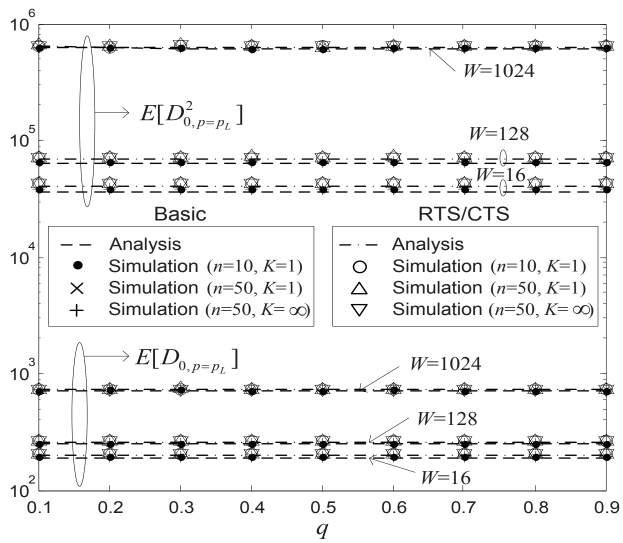



Fig. 7. Mean access delay $E[D_{0,p=p_L}]$ and second moment of access delay $E[D^2_{0,p=p_L}]$ versus backoff factor $q$ in unsaturated IEEE 802.11 DCF networks. $\hat{\lambda} = 0.1$.

### 4.2 Access Delay at the Undesired Stable Point $p_A$

The undesired stable point $p_A$ is a function of 1) the cutoff phase $K$, 2) the backoff factor $q$, 3) the initial backoff window size $W$ and 4) the number of nodes $n$. The first and second moments of access delay at the undesired stable point $p_A$ can be obtained by substituting (28) into (45-46). Specifically, with an infinite cutoff phase $K{=}\infty$, the mean access delay at $p_A$ can be written as

$$E[D^{K=\infty}_{0,p=p_A}] = \tau_T + \frac{1-p_A}{p_A}\tau_F + \left(\frac{1}{2p_A} + \frac{Wq}{2(p_A+q-1)}\right) \cdot \left(1 + (1-p_A)\tau_F + \frac{2n(p_A+q-1)}{Wq}(\tau_T - \tau_F)\right), \tag{49}$$

which is minimized at

$$\min_q E[D^{K=\infty}_{0,p=p_A}]{=}n\left(\tau_T{-}\left(1{+}\frac{1}{\mathbb{W}_0\left(-\frac{1}{e(1+1/\tau_F)}\right)}\right)\tau_F\right). \tag{50}$$

The minimum mean access delay is achieved when the backoff factor is set to be $q{=}q_m^{K=\infty}$ according to (35-36).

Fig. 8a illustrates how the mean access delay performance varies with the backoff parameters when the network is saturated at the undesired stable point $p_A^{K=\infty}$. As we can see from Fig. 8a, with $W$=16 and $n$=50, $E[D^{K=\infty}_{0,p=p_A}]$ sharply increases as the backoff factor $q$ departs from $q_m^{K=\infty}$, indicating that $q$ should be carefully selected to optimize the delay performance. As $n/W$ declines, nevertheless, $p_A^{K=\infty}$ becomes less sensitive to the backoff factor $q$, and hence only a slight increment of $E[D^{K=\infty}_{0,p=p_A}]$ can be observed as $q$ increases if $W$=128 or $n$=10. The minimum mean access delay linearly grows with the number of nodes $n$. Note that $p_A$ is a monotonic increasing function of cutoff phase $K$ according to (28), implying that the mean access delay with a finite $K$ should be higher than $E[D^{K=\infty}_{0,p=p_A}]$. As we can see from Fig. 8b, the mean access delay declines as the cutoff phase $K$ increases, and quickly converges to $E[D^{K=\infty}_{0,p=p_A}]$.

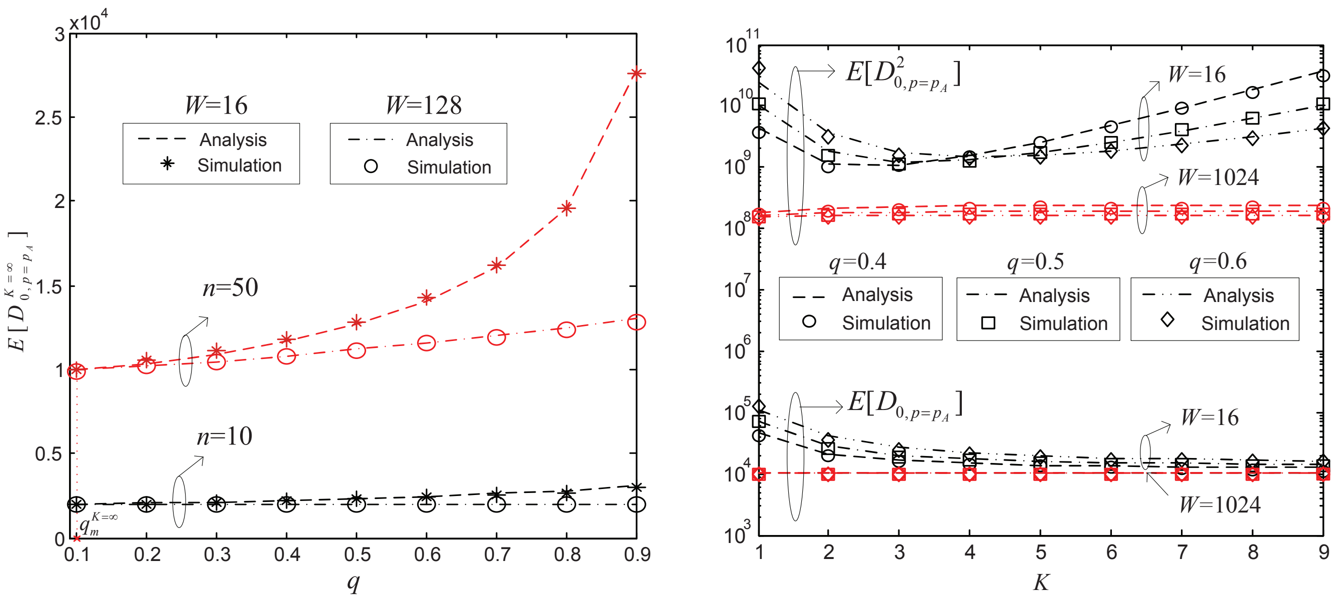
The second moment of access delay $E[D^2_{0,p=p_A}]$ is also closely determined by the above backoff parameters. In contrast to the mean access delay, however, Fig. 8b shows that with $W$=16, $E[D^2_{0,p=p_A}]$ rapidly grows as the cutoff phase $K$ increases. A closer look at (41-42) suggests that a dominating component of the second moment is given by

$$\frac{W^2}{3\alpha^2}\left(\sum_{i=0}^{K-1}\left(\frac{1-p_A}{q^2}\right)^i + \frac{1}{p_A}\cdot\left(\frac{1-p_A}{q^2}\right)^K\right), \tag{51}$$

which sharply increases with the cutoff phase $K$, and eventually becomes infinite as $K{\to}\infty$, if $(1-p_A)/q^2 > 1$. A large second moment indicates that the access delay performance drastically varies from node to node. In that case, some node may capture the channel and produce a continuous stream of packets, while others have to wait for a long time. The short-time unfairness therefore arises. This irregular behavior has long been observed and referred to as the "capture phenomenon" [44] [45].

To prevent the capture phenomenon, backoff parameters should be carefully selected to ensure that $(1 - p_A)/q^2 < 1$. According to (28) and Fig. 3b, $(1-p_A)/q^2$

$$E[D^{2,K=\infty}_{0,p=p_A}]=\frac{W^2}{\alpha^2\left(1-\frac{1-p_A}{q^2}\right)}\cdot\left(\frac{1}{3}+\frac{\frac{1-p_A}{q}}{2\left(1-\frac{1-p_A}{q}\right)}\right)+\frac{W}{\alpha\left(1-\frac{1-p_A}{q}\right)}\cdot\left(\tau_T+\frac{1-p_A}{p_A}\tau_F+\frac{\frac{1-p_A}{q}}{1-\frac{1-p_A}{q}}\left(\tau_F+\frac{1}{2\alpha}\right)-\frac{1}{2}\right.$$
$$\left.+\frac{1+p_A}{2\alpha p_A}\right)+\frac{1}{p_A}\cdot\left(2(1-p_A)\tau_F\left(\tau_T+\frac{1-p_A}{p_A}\tau_F\right)+\frac{1}{\alpha}\left(\tau_T+\frac{2(1-p_A)}{p_A}\tau_F-\frac{1}{2}\right)+\frac{1}{\alpha^2}\left(\frac{1}{2p_A}+\frac{1}{6}\right)\right)+\tau_T^2+\frac{1-p_A}{p_A}\tau_F^2. \tag{53}$$




(a) (b)

Fig. 8. Access delay of saturated IEEE 802.11 DCF networks with the basic access mechanism. (a) Mean access delay $E[D^{K=\infty}_{0,p=p_A}]$ versus backoff factor $q$. $K{=}\infty$. (b) Mean access delay $E[D_{0,p=p_A}]$ and second moment of access delay $E[D^2_{0,p=p_A}]$ versus cutoff phase $K$. $n{=}50$.

can be effectively diminished by choosing a small ratio of the number of nodes $n$ and the initial backoff window size $W$. Specifically, with $K{=}\infty$, the second moment $E[D^{2,K=\infty}_{0,p=p_A}]$ is finite if and only if

$$W > \frac{2n}{-(1+q)\ln(1-q^2)}. \tag{52}$$

As we can see from Fig. 8b, by increasing the initial backoff window size $W$ from 16 to 1024, the second moment of access delay is drastically reduced, and converges to $E[D^{2,K=\infty}_{0,p=p_A}]$ (which is shown on the top of the page), as the cutoff phase $K$ goes to infinity.

We can conclude from Figs. 7-8 that the delay performances at the bi-stable points are drastically different from each other. When the network operates at the desired stable point $p_L$, both the first and the second moments of access delay are insensitive to the number of nodes $n$, the cutoff phase $K$ and the backoff factor $q$, and increase with the initial backoff window size $W$. If the network is saturated at the undesired stable point $p_A$, on the other hand, the delay performance significantly deteriorates, and becomes closely dependent on backoff parameters. With a small initial backoff window size $W$, for instance, the second moment of access delay may grow unboundedly as the cutoff phase $K$ increases, and the capture phenomenon occurs eventually. A large initial backoff window size $W$ is therefore much desirable to improve the delay performance at the undesired stable point $p_A$, which is in sharp contrast to that at the desired stable point $p_L$.

## 5 Performance Analysis of Binary Exponential Backoff

Note that in current IEEE 802.11 DCF networks, Binary Exponential Backoff (BEB) is widely adopted where the backoff factor $q$ is fixed to be $1/2$. In this section, we will apply the analysis presented in Sections 3-4 and demonstrate how to optimize the performance of IEEE 802.11 DCF networks with BEB by properly choosing system parameters. We only focus on the undesired stable point $p_A$, because the throughput and delay performance is insensitive to the backoff factor $q$ when the network operates at the desired stable point $p_L$.

### 5.1 Saturation Throughput

Section 3 demonstrates that for any aggregate input rate $\hat{\lambda} \leq \hat{\lambda}_{\max}$, a stable throughput $\hat{\lambda}_{out} = \hat{\lambda}$ can be achieved at the undesired stable point $p_A$ if the backoff factor $q$ is properly selected. With BEB where $q$ is fixed to be $1/2$, the initial backoff window size $W$ should be carefully tuned to achieve a stable throughput.

Specifically, define $S_A^{W,q=1/2}$ as the stable region of the initial backoff window size $W$ with BEB, within which a stable throughput $\hat{\lambda}_{out} = \hat{\lambda}$ is achieved at the undesired stable point. Similar to (33), it can be written as

$$S_A^{W,q=1/2} = \{W | p_S \leq p_A^{q=1/2} \leq p_L\}. \tag{54}$$

With an infinite cutoff phase $K{=}\infty$, $S_A^{W,q=1/2,K=\infty}$ can be obtained as

$$S_A^{W,q=1/2,K=\infty} = \left[\frac{4np_S - 2n}{-p_S \ln p_S}, \frac{4np_L - 2n}{-p_L \ln p_L}\right] \tag{55}$$

by combining (54) and (29). $S_A^{W,q=1/2,K=\infty}$ decreases as the aggregate input rate $\hat{\lambda}$ increases, and eventually shrinks to a single point

$$W_m^{q=1/2,K=\infty}{=}2n\left(1{+}2\left(1{+}\frac{1}{\tau_F}\right)\mathbb{W}_0\Big({-}\tfrac{1}{e(1+1/\tau_F)}\Big)\right)\Big/\left({-}\left(1 {+}\frac{1}{\tau_F}\right)\mathbb{W}_0\Big({-}\tfrac{1}{e(1+1/\tau_F)}\Big)\ln\left({-}\left(1{+}\frac{1}{\tau_F}\right)\mathbb{W}_0\Big({-}\tfrac{1}{e(1+1/\tau_F)}\Big)\right)\right), \tag{56}$$

with which the maximum stable throughput of $\hat{\lambda}_{\max\_S_A^{W,q=1/2,K=\infty}} = \hat{\lambda}_{\max}$ can be achieved. It can be easily shown from (56) that the optimal initial backoff window sizes in the basic access mode (i.e., $\tau_F^{Basic}{=}175$ time slots) and the RTS/CTS mode (i.e., $\tau_F^{RTS}{=}9$ time slots) are given by

$$W_{m,Basic}^{q=1/2,K=\infty} \approx 17.3n, \tag{57}$$

and

$$W_{m,RTS}^{q=1/2,K=\infty} \approx 2.66n, \tag{58}$$

respectively.

With $W \notin S_A^{W,q=1/2}$, the network throughput falls below the aggregate input rate $\hat{\lambda}$ and is determined by the aggregate service rate, which can be obtained by combining (28) and (30). With an infinite cutoff phase $K{=}\infty$, it can be explicitly written as

$$\hat{\lambda}_s^{q=1/2,K=\infty}{=}\frac{\tau_T}{(\tau_T{-}\tau_F){+}\frac{(1+\tau_F)\cdot\mathbb{W}_0(2n/W\exp(4n/W))-2n\tau_F/W}{8n^2/W^2-2n/W\cdot\mathbb{W}_0(2n/W\exp(4n/W))}}. \tag{59}$$

Note that $\hat{\lambda}_s^{q=1/2}$ is usually referred to as *saturation throughput* in previous studies [2]–[13]. In particular, the aggregate input rate $\hat{\lambda}$ is intentionally raised until the network throughput falls below $\hat{\lambda}$. In that case, the network becomes unstable, and the throughput starts to vary with backoff parameters including the initial backoff window size $W$ and the cutoff phase $K$.

Figs. 9a-9b present the curves of $\hat{\lambda}_s^{q=1/2}$ versus initial backoff window size $W$ under various values of cutoff phase $K$ in the basic access mode and the RTS/CTS mode, respectively. As we can see from both figures, with a small $W$, the throughput with $K{=}6$ is lower than that with $K{=}\infty$. The gap, however, diminishes as $W$ increases, and the throughput in both cases is maximized when $W$ is set to be $W_m^{q=1/2,K=\infty}$, i.e., $W_{m,Basic}^{q=1/2,K=\infty}{=}865$ and $W_{m,RTS}^{q=1/2,K=\infty}{=}133$ for $n{=}50$ according to (57) and (58), respectively. It can be clearly seen from Fig. 9a that an appropriate value of $W$ is of great importance to the throughput performance of BEB if the basic access mechanism is adopted. Considerable throughput loss is incurred if the initial backoff window size $W$ and the cutoff phase $K$ are both small.

Note that numerical and simulation results of the saturation throughput of BEB with a finite cutoff phase $K{<}\infty$ are also presented in [2]. The throughput considered in [2] is the fraction of time that the payload is transmitted, which is slightly different from the one defined in this paper. Nevertheless, let $\xi$ denote the ratio of the packet payload size $L$ and the holding time in the successful transmission state $\tau_T$. The effective throughput used in [2] can be then expressed as $\xi \cdot \hat{\lambda}_s^{q=1/2}$. It can be easily shown that with the packet size $L{=}164$ time slots, $\xi$ is around $91.1\%$ with the basic access mechanism (i.e., $\tau_T^{Basic}{=}180$ time slots) and $85.4\%$ with the RTS/CTS mechanism (i.e., $\tau_T^{RTS}{=}192$ time slots). The curves of effective throughput with $K{=}6$ plotted in Figs. 9a and 9b are consistent with Figs. 9 and 10 in [2], respectively.

Fig. 10 illustrates the effect of cutoff phase $K$ on the throughput performance of BEB. According to (59), with an initial backoff window size of $W{=}32$, the throughput $\hat{\lambda}_s^{q=1/2,K=\infty}$ in the basic access and the RTS/CTS modes are given by $0.73$ and $0.97$, respectively. As we can see from Fig. 10, to approach the throughput with $K{=}\infty$, the cutoff phase $K$ should be large enough, i.e., $K{=}16$ for basic access and $K{=}6$ for RTS/CTS. The effective throughput plotted in Fig. 10 is consistent with Fig. 13 in [2].

In the current IEEE 802.11 FHSS standard, the initial backoff window size $W$ and the cutoff phase $K$ are fixed to be 16 and 6, respectively [46]. Fig. 11 illustrates that in the basic access mode, the throughput achieved with the standardized parameters is significantly lower than $\hat{\lambda}_{\max}$, and the gap is further enlarged as the network size $n$ increases. To achieve the maximum throughput $\hat{\lambda}_{\max}$, our analysis has revealed that one option is to select a large cutoff phase $K$ and set the backoff factor $q$ as $q = q_m^{K=\infty}$ according to (35). If $q$ is fixed to be $1/2$, the initial backoff window size $W$ should be adjusted according to (56). Both cases lead to an optimized throughput, as Fig. 11 shows.

In the RTS/CTS mode, the throughput performance of BEB is substantially enhanced and becomes insensitive to backoff parameters such as the initial backoff window size $W$ and the cutoff phase $K$. It can be clearly seen from Fig. 11 that with the standard setting, $\hat{\lambda}_s^{q=1/2}$ is quite close to the maximum throughput $\hat{\lambda}_{\max}$ even with a large network size $n$. Although the throughput can still be improved by dynamically tuning the backoff factor $q$ or the initial backoff window size $W$, the gains become marginal.

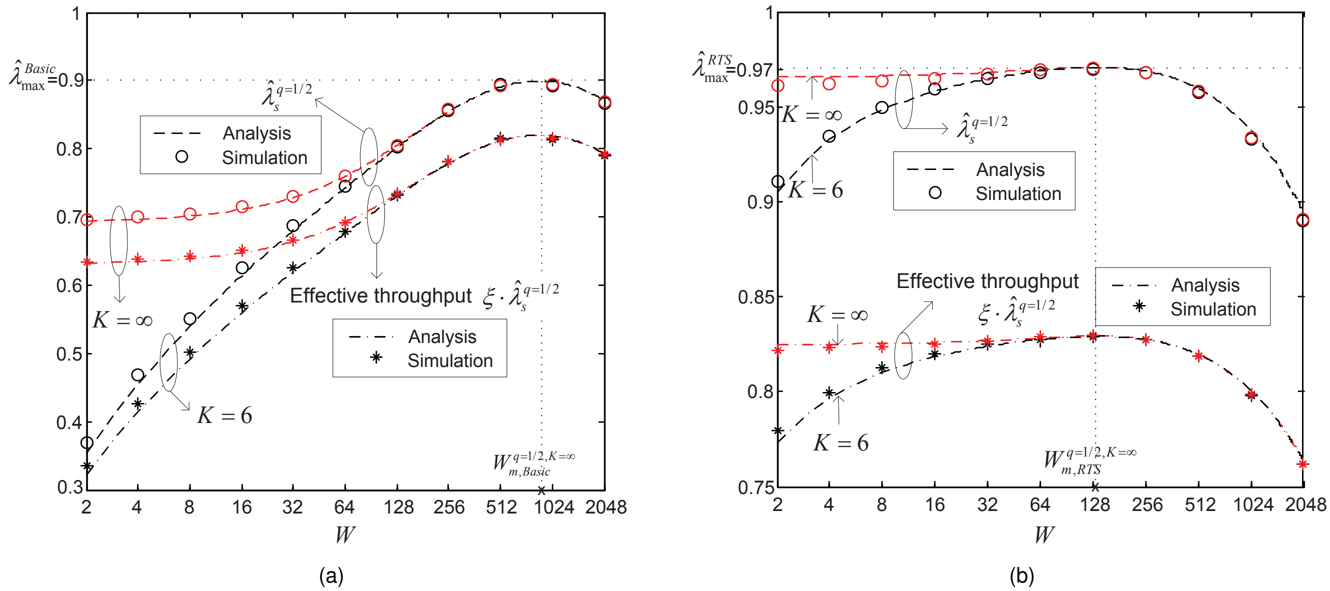



Fig. 9. Saturation throughput $\hat{\lambda}_s^{q=1/2}$ versus initial backoff window size $W$ in IEEE 802.11 DCF networks with BEB ($q$=1/2) under various values of cutoff phase $K$. $n$=50. (a) Basic access mechanism. (b) RTS/CTS mechanism.

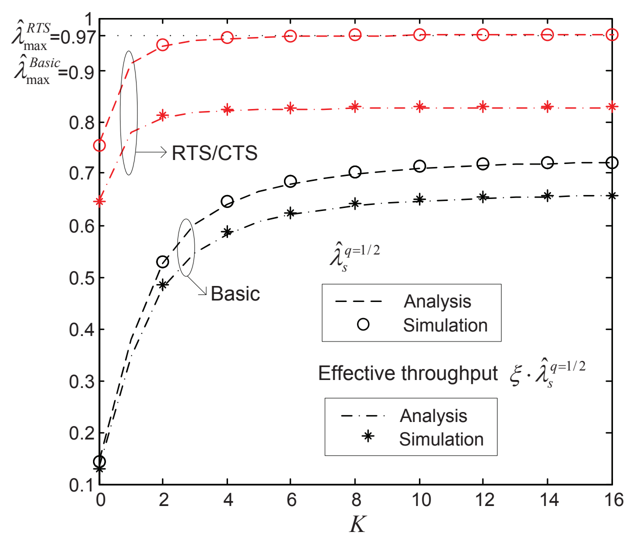



Fig. 10. Saturation throughput $\hat{\lambda}_s^{q=1/2}$ versus cutoff phase $K$ in IEEE 802.11 DCF networks with BEB ($q$=1/2). $n$=50 and $W$=32.

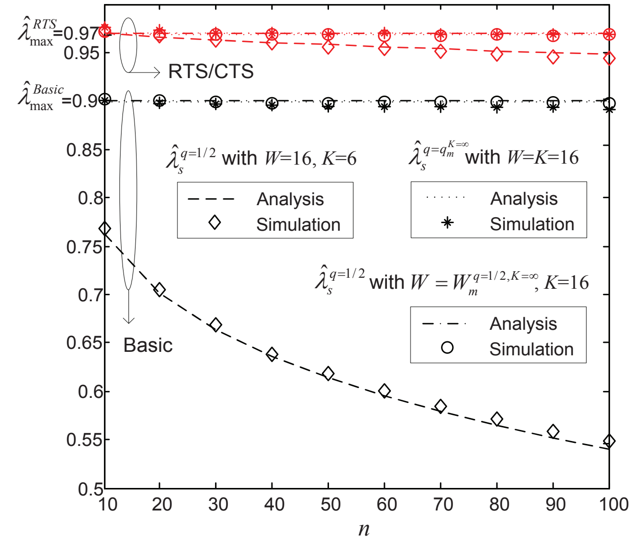



Fig. 11. Saturation throughput versus number of nodes $n$ in IEEE 802.11 DCF networks.

## 5.2 Access Delay

Section 4.2 has shown that when the network shifts to the undesired stable point $p_A$, the access delay performance becomes crucially dependent on backoff parameters. With BEB, the mean access delay at the undesired stable point with an infinite cutoff phase $K$=$\infty$ can be written as

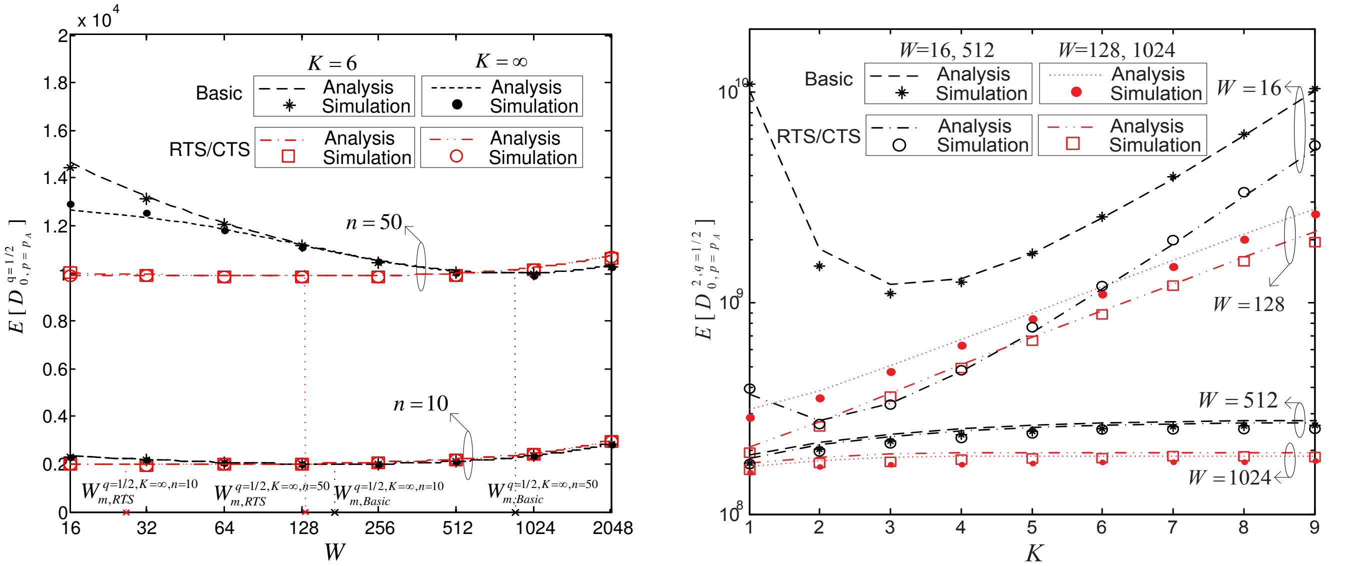
$$E[D_{0,p=p_A}^{q=1/2,K=\infty}] = \tau_T+\frac{1-p_A}{p_A}\tau_F+\left(\frac{1}{2p_A}+\frac{W}{2(2p_A-1)}\right) \cdot\left(1+(1-p_A)\tau_F+\frac{2n(2p_A-1)}{W}(\tau_T-\tau_F)\right), \tag{60}$$

by substituting $q$=1/2 into (49). It can be easily shown that (60) is consistent to Lemma 1 in [15]. The minimum mean access delay can be further obtained as

$$\min_W E[D_{0,p=p_A}^{q=1/2,K=\infty}]=n\left(\tau_T-\left(1+\frac{1}{\mathbb{W}_0\left(-\frac{1}{e(1+1/\tau_F)}\right)}\right)\tau_F\right), \tag{61}$$

which is achieved when the initial backoff window size is set to be $W$=$W_m^{q=1/2,K=\infty}$.

Fig. 12a presents the curves of mean access delay $E[D_{0,p=p_A}^{q=1/2}]$ versus initial backoff window size $W$ under various values of cutoff phase $K$. As we can see from Fig.




(a) (b)

Fig. 12. Access delay of saturated IEEE 802.11 DCF networks with BEB ($q$=1/2). (a) Mean access delay $E[D_{0,p=p_A}^{q=1/2}]$ versus initial window size $W$. (b) Second moment of access delay $E[D_{0,p=p_A}^{2,q=1/2}]$ versus cutoff phase $K$. $n$=50.

12a, with a small $W$, the mean access delay with $K$=6 is larger than that with $K$=$\infty$ due to a lower probability of success of HOL packets. The mean access delay in both cases declines as $W$ increases, and is optimized when $W$=$W_m^{q=1/2,K=\infty}$. A closer look at (61) further indicates that the minimum mean access delay in the basic access mode (i.e., $\tau_T^{Basic}$=180 time slots and $\tau_F^{Basic}$=175 time slots) and the RTS/CTS mode (i.e., $\tau_T^{RTS}$=192 time slots and $\tau_F^{RTS}$=9 time slots) is approximately given by $200n$ and $198n$, respectively. Both of them linearly increase with the number of nodes $n$.

Section 4.2 also demonstrates that the second moment of access delay at the undesired stable point may sharply grow with the cutoff phase $K$ if the backoff factor $q$ or the initial backoff window size $W$ is not properly selected. With BEB, a huge second moment of access delay can be observed from Fig. 12b when $W$ is small, i.e., $W$=16. The second moment is significantly reduced by enlarging $W$ to 128. Nevertheless, it still increases with the cutoff phase $K$ and becomes infinite as $K\to\infty$.

According to (52), the second moment of access delay of BEB, $E[D_{0,p=p_A}^{2,q=1/2,K=\infty}]$, is finite if and only if the initial backoff window size $W$ satisfies

$$W > \frac{4n}{3\ln\frac{4}{3}} \approx 4.63n. \qquad (62)$$

As we can see from Fig. 12b, with a large $W$, i.e., $W$>232 for $n$=50, the second moment of access delay converges as $K$ goes to infinity, and steadily decreases as the initial backoff window size $W$ increases. The second moment in the RTS/CTS mode is slightly better than that in the basic access mode when $W$ is small. The gap, however, diminishes for $W\geq$512.

We can conclude from Fig. 12 that the delay performance of BEB at the undesired stable point significantly deteriorates if the initial backoff window size $W$ is small. As Fig. 13 illustrates, with the current FHSS standard setting ($K$=6 and $W$=16), the second moment of access delay of BEB is much higher than that with a large $W$, i.e., $W$=1024, and the gap is further widened as the network size $n$ grows. Note that a high second moment also indicates serious short-term unfairness among nodes. It is therefore of great importance to choose a large initial backoff window size when the network becomes saturated at the undesired stable point $p_A$.

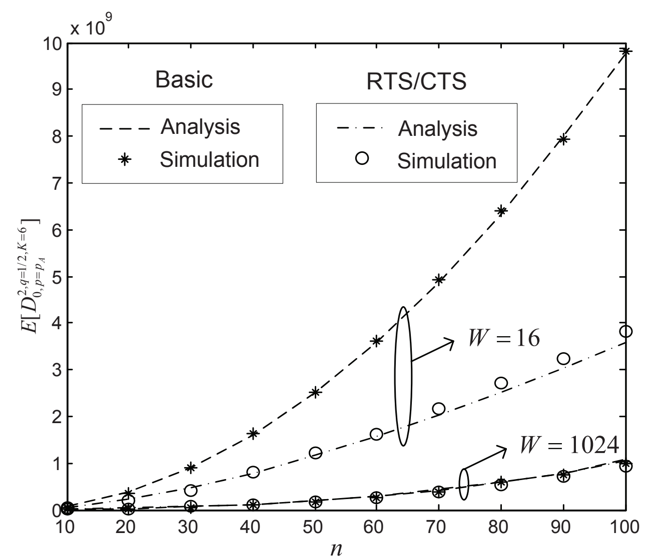



Fig. 13. Second moment of access delay $E[D_{0,p=p_A}^{2,q=1/2,K=6}]$ versus number of nodes $n$ in saturated IEEE 802.11 DCF networks with BEB ($q$=1/2). $K$=6.

## 6 Conclusion

This paper presents the stability, throughput and delay analysis of buffered IEEE 802.11 DCF networks. It is revealed that an IEEE 802.11 DCF network has two steady-state points. It operates at the desired stable point $p_L$ if it is unsaturated, and a stable throughput can be always achieved at $p_L$. If it becomes saturated, it shifts to the undesired stable point $p_A$, and a stable throughput can be achieved if and only if the backoff parameters are properly selected from their corresponding stable regions. Both the maximum stable throughput and the stable region of backoff factor $q$ are derived, and shown to be crucially dependent on the holding times of HOL packets in successful transmission and collision states, $\tau_T$ and $\tau_F$. With a decrease of $\tau_F$, the stable region is enlarged and becomes less sensitive to system parameters, which justifies the observation that an IEEE 802.11 DCF network in the RTS/CTS mode is more robust than that in the basic access mode. The delay analysis further shows that for unsaturated networks, both the first and second moments of access delay grow with $\tau_T$ and the initial backoff window size $W$, implying that better delay performance can be achieved in the basic access mode with a small $W$. If the network is saturated at the undesired stable point $p_A$, on the other hand, a large $W$ becomes desirable in both modes to prevent the capture phenomenon.

Our analysis provides plenty of insight for practical network design. In current IEEE 802.11 DCF networks, the backoff factor $q$ is fixed to be $1/2$, and small values of initial backoff window size $W$ and cutoff phase $K$ are usually selected. It is shown that in the basic access mode, there exists a huge gap between the maximum throughput that can be achieved with the default standard setting and the optimum throughput. To eliminate the gap, a large cutoff phase $K$ should be selected, and the initial backoff window size $W$ should be linearly adjusted with the network size $n$ according to (56). Moreover, the delay analysis shows that with the default standard setting, a saturated IEEE 802.11 DCF network suffers from serious short-term unfairness due to a high second moment of access delay no matter which mode is chosen. To improve the delay performance at the saturated point, a large initial backoff window size $W$ should be adopted, which nevertheless may cause unnecessarily long delay if the network is unsaturated.

It should be noted that throughout the paper, we assume that each node is equipped with a buffer of infinite size, and each HOL packet has no limit on the number of retransmission attempts. In practical networks, however, packets may be dropped due to 1) a finite buffer size, and 2) a limit on the number of retransmissions. Intuitively, the buffer size does not affect the contention process of HOL packets, and hence the analysis presented in this paper remains valid for the finite buffer size case. The effect of retry limit, in contrast, could be prominent when the network becomes saturated. How to refine the proposed analytical framework to include more practical assumptions is an interesting issue, which deserves much attention in the future study.

## Appendix A
## Derivation of (5) and (16)

The Markov chain shown in Fig. 2 illustrates the transition process of $\{G_t^i\}$, where $G_t^i$ denotes the state of the backoff counter of a State-R$_i$ HOL packet at time slot $t$, $i = 0, \ldots, K$.

1) Let $Y_i$ denote the holding time of a HOL packet in State R$_i$. When a HOL packet enters State R$_i$, it randomly selects a number $x$ from $\{0, \ldots, W_i - 1\}$ as the initial value of its backoff counter. According to Fig. 2, $Y_i$ is the sum of the sojourn time at states B$_x$, B$_{x-1}$, …, and B$_0$, which can be written as

$$Y_i = \sum_{k=0}^{x} J_{B_k}, \tag{63}$$

where $J_{B_k}$ is the sojourn time at state B$_k$, which follows a geometric distribution with parameter $\alpha$ as $t \to \infty$, and $x$ follows a uniform distribution with state space $\{0, \ldots, W_i - 1\}$. The mean holding time $\tau_{R_i}$ is then given by

$$\tau_{R_i} = \mathbb{E}[Y_i] = \mathbb{E}[x+1] \cdot \mathbb{E}[J_{B_k}], \tag{64}$$

and (5) can be obtained accordingly.

2) The limiting state probabilities of the Markov chain in Fig. 2 can be obtained as

$$f_{B_k} = \frac{2}{1+W_i} \cdot \frac{W_i - k}{W_i}, \tag{65}$$

$k = 0, \ldots, W_i - 1$. Given that the channel is idle, a transmission request is made if the backoff counter is zero. The conditional probability of a State-R$_i$ HOL packet making a transmission request given that the channel is idle, $r_i$, is then equal to $f_{B_0}$, and (16) can be obtained according to (65).

## Appendix B
## Derivation of (6)

The channel has three states: 1) Idle, 2) Successful Transmission and 3) Collision. Accordingly, the probability of sensing the channel idle at time slot $t+1$, $\alpha_{t+1}$, can be written as

$$\begin{aligned}\alpha_{t+1} &= \Pr\{\text{idle at } t+1 | \text{success at } t\} \cdot \Pr\{\text{success at } t\} \\ &+ \Pr\{\text{idle at } t+1 | \text{collision at } t\} \cdot \Pr\{\text{collision at } t\} \\ &+ \Pr\{\text{idle at } t+1 | \text{idle at } t\} \cdot \Pr\{\text{idle at } t\}.\end{aligned} \tag{66}$$

In IEEE 802.11 DCF networks, a successful transmission and a collision last for $\tau_T$ and $\tau_F$ time slots, respectively. As a result, if the channel is sensed busy at time slot $t$, the probability of sensing the channel idle at the next time slot $t+1$ is given by $1/\tau_T$ if the transmission is successful, and $1/\tau_F$ if a collision occurs. We have

$$\Pr\{\text{idle at } t+1 | \text{success at } t\} = \frac{1}{\tau_T}, \tag{67}$$

and

$$\Pr\{\text{idle at } t+1|\text{collision at } t\} = \frac{1}{\tau_F}. \quad (68)$$

On the other hand, if the channel is sensed idle at time slot $t$, the probability that it is idle at $t$+1 is indeed $p_{t+1}$, which is the probability that all the nodes do not transmit at $t$+1 given that the channel is sensed idle at $t$. We have

$$\Pr\{\text{idle at } t+1|\text{idle at } t\} = p_{t+1}. \quad (69)$$

The unconditional probability of sensing the channel in successful transmission at time slot $t$ can be written as

$$\Pr\{\text{success at } t\} = \sum_{i=1}^{\tau_T} \Pr\{\text{success at } t-i+1|\text{idle at } t-i\} \cdot \Pr\{\text{idle at } t-i\}. \quad (70)$$

Let $\omega_t$ denote the probability that a node has a request at time slot $t$ given that the channel is sensed idle at $t-1$. We have

$$p_t = (1-\omega_t)^{n-1} \stackrel{\text{for large } n}{\approx} \exp(-n\omega_t). \quad (71)$$

The probability that the channel has a successful transmission at time slot $t-i+1$ given that the channel is idle at time slot $t-i$ can be then written as

$$\Pr\{\text{success at } t-i+1|\text{idle at } t-i\} = n\omega_{t-i+1} \cdot p_{t-i+1} \approx -p_{t-i+1} \ln p_{t-i+1}, \quad (72)$$

according to (71). By substituting (72) into (70), we have

$$\Pr\{\text{success at } t\} = -\sum_{i=1}^{\tau_T} \alpha_{t-i} \cdot p_{t-i+1} \ln p_{t-i+1}. \quad (73)$$

Finally, by combining (66-69) and (73), the dynamic equation of $\alpha_{t+1}$ can be obtained as

$$\alpha_{t+1} = -\frac{1}{\tau_T}\sum_{i=1}^{\tau_T} \alpha_{t-i} p_{t-i+1} \ln p_{t-i+1} + \frac{1}{\tau_F}\left(1 + \sum_{i=1}^{\tau_T} \alpha_{t-i} p_{t-i+1} \cdot \ln p_{t-i+1} - \alpha_t\right) + \alpha_t p_{t+1}. \quad (74)$$

As $t \to \infty$, we have

$$\alpha = -\alpha p \ln p + \frac{1}{\tau_F}(1 + \tau_T \alpha p \ln p - \alpha) + \alpha p. \quad (75)$$

(6) can be obtained by solving (75).

## Appendix C
## Derivation of (41-44)

The probability generating functions of $D_i$, $i = 0, \ldots, K$, can be obtained from (39) and (40) as

$$\begin{cases} G_{D_i}(z) = pz^{\tau_T} G_{Y_i}(z) + (1-p)z^{\tau_F} G_{Y_i}(z) G_{D_{i+1}}(z), i<K \\ G_{D_K}(z) = pz^{\tau_T} G_{Y_K}(z) + (1-p)z^{\tau_F} G_{Y_K}(z) G_{D_K}(z), \end{cases} \quad (76)$$

where $Y_i$ is the holding time of a HOL packet in State $\mathrm{R}_i$, $i$=0, …, $K$.

According to (63) in Appendix A, the probability generating function of $Y_i$ can be written as

$$G_{Y_i}(z) = \frac{1}{W_i} \frac{G_{J_{B_k}}(z) - G_{J_{B_k}}(z)^{W_i+1}}{1 - G_{J_{B_k}}(z)}, \quad (77)$$

where

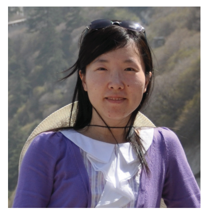
$$G_{J_{B_k}}(z) = \frac{\alpha z}{1-(1-\alpha)z}. \quad (78)$$

(41-42) and (43-44) can be then obtained from (76) and (77-78), respectively.

## References


[1] J. F. Kurose and K. W. Ross, *Computer Networking: A Top-Down Approach Featuring the Internet*, 3rd ed. Addison Wesley, 2004.
[2] 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.
[3] E. Ziouva and T. Antonakopoulos, "CSMA/CA performance under high traffic conditions: throughput and delay analysis," *Comput. Commun.*, vol. 25, no. 3, pp. 313-321, 2002.
[4] 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 INFOCOM*, vol. 2, 2002, pp. 599-607.
[5] C. H. Foh and J. W. Tantra, "Comments on IEEE 802.11 saturation throughput analysis with freezing of backoff counters," *IEEE Commun. Lett.*, vol. 9, no. 2, pp. 130-132, Feb. 2005.
[6] J. Vardakas, M. Sidiropoulos, and M. Logothetis, "Performance behaviour of IEEE 802.11 distributed coordination function," *IET Circuits, Devices & Systems*, vol. 2, no. 1, pp. 50-59, Feb. 2008.
[7] I. Tinnirello, G. Bianchi, and Y. Xiao, "Refinements on IEEE 802.11 Distributed Coordination Function Modeling Approaches," *IEEE Trans. Veh. Technol.*, vol. 59, no. 3, pp. 1055-1067, Mar. 2010.
[8] J. Choi, J. Yoo, and C. Kim, "A novel performance analysis model for an IEEE 802.11 wireless LAN," *IEEE Commun. Lett.*, vol. 10, no. 5, pp. 335-337, May 2006.
[9] Z. Hadzi-Velkov and B. Spasenovski, "Saturation throughput - delay analysis of IEEE 802.11 DCF in fading channel," in *Proc. IEEE ICC*, vol. 1, May 2003, pp. 121-126.
[10] P. Chatzimisios, A. Boucouvalas, and V. Vitsas, "Influence of channel BER on IEEE 802.11 DCF," *Electronics Letters*, vol. 39, no. 23, pp. 1687-1689, Nov. 2003.
[11] X. Dong and P. Varaiya, "Saturation throughput analysis of IEEE 802.11 wireless LANs for a lossy channel," *IEEE Commun. Lett.*, vol. 9, no. 2, pp. 100-102, Feb. 2005.
[12] Y. Zheng, K. Lu, D. Wu, and Y. Fang, "Performance Analysis of IEEE 802.11 DCF in Imperfect Channels," *IEEE Trans. Veh. Technol.*, vol. 55, no. 5, pp. 1648-1656, Sept. 2006.
[13] A. Kumar, E. Altman, D. Miorandi, and M. Goyal, "New insights from a fixed-point analysis of single cell IEEE 802.11 WLANs," *IEEE/ACM Trans. Netw.*, vol. 15, pp. 588-601, June 2007.
[14] A. Zanella and F. De Pellegrini, "Statistical characterization of the service time in saturated IEEE 802.11 networks," *IEEE Commun. Lett.*, vol. 9, no. 3, pp. 225-227, Mar. 2005.
[15] T. Sakurai and H. Vu, "MAC Access Delay of IEEE 802.11 DCF," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1702-1710, May 2007.
[16] K. Duffy, D. Malone, and D. Leith, "Modeling the 802.11 distributed coordination function in non-saturated conditions," *IEEE Commun. Letters*, vol. 9, no. 8, pp. 715-717, Aug. 2005.
[17] E. M. M. Winands, T. J. J. Denteneer, J. A. C. Resing, and R. Rietman, "A finite-source feedback queueing network as a model for the IEEE 802.11 Distributed Coordination Function," *Eur. Trans. Telecommun.*, vol. 16, no. 1, pp. 77-89, Dec. 2005.
[18] C. Foh, M. Zukerman, and J. Tantra, "A Markovian Framework for Performance Evaluation of IEEE 802.11," *IEEE Trans. Wireless Commun.*, vol. 6, no. 4, pp. 1276-1265, Apr. 2007.
[19] D. Malone, K. Duffy, and D. Leith, "Modeling the 802.11 Distributed Coordination Function in Nonsaturated Heterogeneous Conditions," *IEEE/ACM Trans. Networking*, vol. 15, no. 1, pp. 159-172, Feb. 2007.

[20] H. Zhai, Y. Kwon, and Y. Fang, "Performance analysis of IEEE 802.11 MAC protocols in wireless LANs," *Wirel. Commun. Mob. Comput.*, vol. 4, no. 8, pp. 917-931, Dec. 2004.
[21] R. P. Liu, G. J. Sutton, and I. B. Collings, "A New Queueing Model for QoS Analysis of IEEE 802.11 DCF with Finite Buffer and Load," *IEEE Trans. Wireless Commun.*, vol. 9, no. 8, pp. 2664-2675, Aug. 2010.
[22] B. Li and R. Battiti, "Analysis of the IEEE 802.11 DCF with Service Differentiation Support in Non-saturation Conditions," *Lecture Notes on Computer Science*, vol. 3266, pp. 64-73, 2004.
[23] H. Zhai, X. Chen and Y. Fang, "How well can the IEEE 802.11 wireless LAN support quality of service?," *IEEE Trans. Wireless Commun.*, vol. 4, no. 6, pp. 3084-3094, Nov. 2005.
[24] O. Tickoo and B. Sikdar, "Modeling Queueing and Channel Access Delay in Unsaturated IEEE 802.11 Random Access MAC Based Wireless Networks," *IEEE/ACM Tran. Networking*, vol. 16, no. 4, pp. 878-891, Aug. 2008.
[25] Y. Bae, K. Kim, M. Moon, and B. Choi, "Analysis of IEEE 802.11 non-saturated DCF by matrix analytic methods," *Ann. Oper. Res.*, vol. 162, no. 1, pp. 3-18, Sept. 2008.
[26] K. Ghaboosi, B. H. Khalaj, Y. Xiao, and M. Latva-aho, "Modeling IEEE 802.11 DCF Using Parallel Space-Time Markov Chain," *IEEE Trans. Veh. Technol*, vol. 57, no. 4, pp. 2404-2413, July 2008.
[27] E. Karamad and F. Ashtiani, "Performance analysis of IEEE 802.11 DCF and 802.11e EDCA based on queueing networks," *IET Commun.*, vol. 3, no. 5, pp. 871-881, May 2009.
[28] Q. Zhao, D. H. K. Tsang, and T. Sakurai, "A Simple and Approximate Model for Nonsaturated IEEE 802.11 DCF," *IEEE Trans. Mobile Computing*, vol. 8, no. 11, pp. 1539-1553, Nov. 2009.
[29] E. Felemban and E. Ekici, "Single Hop IEEE 802.11 DCF Analysis Revisited: Accurate Modeling of Channel Access Delay and Throughput for Saturated and Unsaturated Traffic Cases," *IEEE Trans. Wireless Commun.*, vol. 10, no. 10, pp. 3256-3266, Oct. 2011.
[30] 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, Sept. 2000.
[31] F. Cali, M. Conti and E. Gregori, "Dynamic tuning of the IEEE 802.11 protocol to achieve a theoretical throughput limit," *IEEE/ACM Trans. Networking*, nol. 8, no. 6, pp. 785-799, Dec. 2000.
[32] G. Bianchi and I. Tinnirello, "Kalman filter estimation of the number of competing terminals in an IEEE 802.11 network," in *Proc. INFOCOM*, vol. 2, Apr. 2004, pp. 844-852.
[33] A. L. Toledo, T. Vercauteren and X. Wang, "Adaptive optimization of IEEE 802.11 DCF based on Bayesian estimation of the number of competing terminals," *IEEE Trans. Mobile Computing*, vol. 5, no. 9, pp. 1283-1296, Sept. 2006.
[34] D. Deng, C. Ke, H. Chen and Y. Huang, "Contention window optimization for IEEE 802.11 DCF access control," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 5129-5135, Dec. 2008.
[35] C. G. Wang, B. Li and L. Li, "A New Collision Resolution Mechanism to Enhance the Performance of IEEE 802.11 DCF,"in *IEEE Trans. Veh. Technol.*, vol. 53, no. 4, pp. 1235-1246, July 2004.
[36] I. Aad, Q. Ni, C. Barakat and T. Turletti, "Enhancing IEEE 802.11 MAC in congested environments," *Elsevier Comput. Commun. J.*, vol. 28, no. 14, pp. 1605-1617, Sept. 2005.
[37] J. Choi, J. Yoo, S. Choi and C. Kim, "EBA: an enhancement of the IEEE 802.11 DCF via distributed reservation," *IEEE Trans. Mobile Computing*, vol. 4, no. 4, pp. 378-390, July/Aug. 2005.
[38] S. Ye and Y. Tseng, "A multichain backoff mechanism for IEEE 802.11 WLANs," *IEEE Trans. Veh. Technol.*, vol. 55, no. 5, pp. 1613-1620, Sept. 2006.
[39] J. Choi, J. Yoo, and C. Kim, "A distributed fair scheduling scheme with a new analysis model in IEEE 802.11 wireless LANs," *IEEE Trans. Veh. Technol.*, vol. 57, no. 5, pp. 3083-3093, Sept. 2008.
[40] M. Iosifescu, *Finite Markov Processes and Their Applications*. Dover Publications, 2007.
[41] E. P. C. Kao, *An Introduction ot Sotchastic Processes*. Duxbury Press, 1997.
[42] R. M. Corlessa, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth, "On the Lambert W function," *Adv. Comput. Math.*, vol. 5, pp. 329-359, 1996.
[43] D. Bertsekas and R. Gallager, *Data Networks*, 2nd ed. Prentice Hall, 1992.
[44] K. K. Ramakrishman and H. Yang, "The Ethernet capture effect: analysis and solution," in *Proc. 19th Local Computer Networks Conf.*, Oct. 1994.
[45] K. Medepalli and F. A. Tobagi, "On Optimization of CSMA/CA based Wireless LANs: Part I - Impact of Exponential Backoff," in *Proc. IEEE ICC*, vol. 5, June 2006, pp. 2089-2094.
[46] *IEEE Std. 802.11-2007, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, June 2007.



**Lin Dai** (S'00-M'03) received the B.S. in electronic engineering from Huazhong University of Science and Technology, Wuhan, China, in 1998 and the M.S. and Ph. D. degrees in electrical and electronic engineering from Tsinghua University, Beijing, China, in 2000 and 2002, respectively. She was a postdoctoral fellow at the Hong Kong University of Science and Technology and University of Delaware. Since 2007, she has been with City University of Hong Kong, where she is an assistant professor. Her research interests include wireless communications, information and communication theory. She received the Best Paper Award at IEEE WCNC 2007 and the IEEE Marconi Prize Paper Award in 2009.

**Xinghua Sun** received the B.S. degree in communications engineering from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 2008, and is currently pursuing the Ph.D. degree at City University of Hong Kong. His current research interests include performance evaluation of wireless networks and distributed systems.