Cognitive Backoff Mechanism for IEEE802.11ax High-Efficiency WLANs

Nurullah Shahin, Rashid Ali, Sung Won Kim, and Young-Tak Kim

Abstract: **Carrier sense multiple access with collision avoidance(CSMA/CA) in current IEEE 802.11 wireless local area networks (WLANs) uses a uniformly distributed binary exponential backoff (BEB) mechanism that is mainly based upon exponential increases of the contention window (CW) to avoid repeated collisions. After each consecutive collision, the CW is doubled until it reaches the maximum value. Under dense conditions, however, the blind selection of the CW greatly reduces throughput, whereas under sparse conditions with smaller number of contending stations, the blind exponential increase of the CW for collision avoidance causes unnecessarily long delays. Therefore, it fails to achieve high efficiency in both dense and sparse environments with the current BEB mechanism. In this paper, we propose a cognitive backoff (CB) mechanism that adaptively determines the CW to provide efficient collision avoidance with high throughput and low delay under both dense and sparse conditions. In the proposed CB mechanism, the measured conditional collision probability and the number of backoff stages determine the CW. A performance analysis with an event-driven simulator, NS3, reveals that the proposed CB can achieve higher throughput and lower delay than the BEB, without much implementation complexity while preserving fairness.**

Index Terms: **Contention Window, CSMA/CA, IEEE 802.11ax, MAC, WLAN.**

I. INTRODUCTION

W HILE the advances in physical-layer (PHY) technologies for next-generation wireless local area networks (WLANs) promise to deliver sufficient bandwidth to serve user demands, the current medium access control (MAC) based on carrier sense multiple access with collision avoidance (CSMA/CA) for WLANs was analyzed as not being efficient for dense environments with a large number of stations (STAs) [1]. The efficiency of the current MAC protocols soon encounters challenges when WLANs are deployed even more densely, such as in stadiums, train stations, or enterprise offices, where the density of WLAN users is very high. To address the inefficiencies in WLANs, especially in dense indoor and outdoor network environments, and to improve robustness against interference, a new IEEE 802.11 task group looking at the IEEE

802.11ax high-efficiency WLAN (HEW) [2] was formed in 2014 to examine the user experience, and consequently to focus on multi-user performance metrics, such as delay, latency, and average per-user throughput for the IEEE 802.11 standard working group. However, several issues pose difficulties for HEW to become four times more efficient than the current WLANs. Most of the challenges come with the efforts to implement MAC-layer resource allocation (MAC- RA) [1].

HEW is not fully prepared to join next-generation technologies such as the Internet of Things (IoT) [3] due to the conventional distributed coordination function (DCF) as its basic MAC-RA scheme [1]. The DCF uses CSMA/CA to resolve contention for channel access [4] and can operate under either the basic access scheme or the optional request-to-send/clear-to-send (RTS/CTS) scheme. Binary exponential backoff (BEB) is used to handle contentions over access to a shared medium and to transmit data. It defines discrete backoff time slots for which the STA has to defer before accessing the medium. Other STAs overhear the transmission from neighboring STAs by carrier sensing, and set up their network allocation vector (NAV) to avoid collisions [1].

BEB is the key mechanism to avoid repeated collisions under CSMA/CA and is used widely due to its simplicity. On the first transmission attempt, the STA also generates a random backoff value, which is uniformly chosen from the contention window (CW) interval [0, CW], where the CW is initially set to the minimum value($CW_{\min}$). After each unsuccessful transmission by collision, the CW is doubled until it reaches the maximum value ($CW_{\max}$). Once the STA successfully transmits a frame, the CW is reset to the minimum value ($CW_{\min}$). If the STA reaches its maximum number of retransmissions, the transmission of a MAC protocol data unit (MPDU) is determined as a failure, and the next MPDU transmission is attempted at $CW_{\min}$. In a network with heavily loaded stations, the $CW_{\min}$ will be relatively smaller, which results in more collisions and poor network performance. Similarly, in a network with a light traffic load, the blind exponential increase of the CW to avoid collisions causes unnecessarily long delays. Due to the BEB, current WLANs fail to achieve high efficiency in highly dense environments.

In this paper, we propose a cognitive backoff (CB) mechanism that can provide enhanced throughput above the maximum achievable amount under the current BEB, and provides low delay. This performance enhancement is achieved mainly based on cognitive channel sensing and runtime measurement of collision probability (p_{ck}). Incrementing the CW is controlled by an adaptively determined backoff factor, rather than an exponential increase. The measurement of p_{ck} can be easily implemented in currently deployed 802.11 wireless networks.

This paper describes the channel access issues to solve the critical medium collision problem incurred by a large number of densely deployed contending STAs, and suggests a practical channel observation–based mechanism [5], [6]. The CB enhances performance in terms of high throughput and low delay in both high-density and low-density environments by reducing the number of collisions during the channel accesses by using the cognitive backoff mechanism.

The remainder of this paper is organized as follows. Section II explains related work. Section III describes the proposed CB algorithm. Performance analyses of the proposed algorithm are provided in Section IV, followed by a performance evaluation in Section V. Finally, in Section VI, a comprehensive conclusion is presented.

II. RELATED WORK

A. Binary Exponential Backoff (BEB)

In the BEB mechanism used in current CSMA/CA networks, when STAs need channel access, they initialize the retry counter (r) and set the backoff stage (i) to zero. The backoff counter (k) is initialized at a randomly chosen value from $[0, CW_{\min}]$, where $CW_{\min}$ is the minimum CW size. After each collision, r and i are incremented. There are maximum backoff stage ($\max B$) and maximum retry limit ($\max R$) specified by the standard. The packet is discarded if its number of transmission attempts reaches $\max R$. For each repeated collision, the CW is exponentially increased (i.e., $2^i(CW_{\min} + 1) - 1$) with respect to the backoff stage, (for $i \in [1, \max B]$). Furthermore, if either retry limit reaches $\max R$, or transmission is successful, the r and i are reset to zero, and a new value of k is selected. Following is a description of updating the CW in BEB when a packet transmission either collides or succeeds:

$$CW = \begin{cases} \min\left(\left(2^i\left(CW_{\min} + 1\right)\right) - 1, CW_{\max}\right), & \text{collide} \\ CW_{\min}, & \text{succeed.} \end{cases} \tag{1}$$

Although the random nature of the BEB algorithm diminishes the probability of collisions, it cannot entirely avoid collisions, and suffers from low throughput in highly dense environments. The main reason is that when the density of contending STAs is high and a packet is successfully transmitted after a number of collisions, resetting the CW value to $CW_{\min}$ increases the probability of collision.

B. Enhanced Collision Avoidance (ECA)

Many researchers have proposed modifications to the BEB-based CSMA/CA to enhance WiFi performance [7]–[16]. In several studies, throughput was increased under saturation conditions by preventing the CW from resetting to its minimum value after each successful transmission [7]–[10]. Those authors referred to the class of algorithms that use deterministic backoff after successful transmission as CSMA with enhanced collision avoidance (CSMA/ECA). ECA was first proposed by Barcelo et al. [7], and later, a more detailed analysis of both saturated and non-saturated traffic conditions was presented [8]–[10]. In the proposed enhanced collision avoidance, a deterministic backoff value, $CW_{\min}/2$, was used instead of resetting CW

to $CW_{\min}$, which reduces the chances of collisions for STAs that were successful in the previous transmission. However, the performance gain from the basic version of ECA is limited to deterministic threshold $CW_{\min}/2$, and starts suffering performance degradation after the threshold. In later works [8]–[10], the basic ECA was extended to ECA-Hysteresis and ECA-Fair Share, which allow CSMA/ECA to enlarge the deterministic cycle length with many more contenders, keeping an even distribution of the available bandwidth. Although, their proposed enhancements increased the efficiency of ECA, this performance enhancement comes at the expense of reduced short-term fairness, since STAs that have recently failed to transmit due to consecutive collisions are forced to stay at a higher backoff stage without knowing the network density, and thus, are further penalized with less frequent transmissions [9].

The ECA behaves exactly the same as the current BEB using the CSMA/CA protocol, except that a deterministic backoff is chosen after each successful transmission. To guarantee fair coexistence with legacy CSMA/CA STAs, the value of the deterministic backoff is chosen from a similar CW, as follows:

$$k = \frac{\lceil CW_{\min} \rceil}{2} + 1. \tag{2}$$

The deterministic value of the backoff after each successful transmission is the key parameter of the ECA, since it determines the maximum number of STAs that can be accommodated in collision-free mode under CSMA/CA. Although the ECA provides a collision-free environment for a WLAN, it is limited to the number of STAs defined by the size of $CW_{\min}$. Another issue with the implementation of ECA is that it only focuses the backoff changes after a successful transmission, while WLAN performance is mainly affected by collisions. The modification to the basic ECA can be described as

$$CW = \begin{cases} \min\left(\left(2^i(CW_{\min} + 1)\right) - 1, CW_{\max}\right), & \text{collide} \\ k = \frac{\lceil CW_{\min} \rceil}{2} + 1, & \text{succeed.} \end{cases} \tag{3}$$

C. Exponential Increase and Exponential Decrease (EIED)

The exponential increase exponential decrease (EIED) backoff algorithm was proposed by Ye et al. [12], where the CW size is exponentially increased after each unsuccessful transmission and exponentially decreased after each successful transmission by backoff factors r_I and r_D, respectively. EIED is as simple to implement as BEB for improved performance of CSMA/CA. The EIED modifies both parts of the current BEB, that is, a change in CW after collision and a change in CW after a successful transmission. In EIED, whenever a packet collides with another STA's transmission, the CW is increased by backoff factor r_I, whereas the CW is decreased by backoff factor r_D if the STA transmits a packet successfully. The increase and decrease of the CW in the EIED backoff mechanism can be described as follows:

$$CW = \begin{cases} \min\left(\left(r_I(CW_{\min} + 1)\right) - 1, CW_{\max}\right), & \text{collide} \\ \min\left(\left((CW + 1)/r_D\right) - 1, CW_{\min}\right), & \text{succeed.} \end{cases} \tag{4}$$

Table 1. Notations in the algorithms.

Notation	Representation of the symbol
r	Number of retransmission attempts
$\max R$	Maximum number of retry limit
i	Number of backoff stages
$\max B$	Maximum number of backoff stages (e.g., 3 or 6)
k	Backoff counter
N_{bo}	Number of backoff slots
N_{busy}	Number of observed busy slots
N_{coll}	Number of experienced collisions
N_{ck}	Number of times busy and in collision
N_{bc}	Number of busy and collision time slot
p_{ck}	Conditional collision probability
CW	Contention window size
$CW_{\max}$	Maximum Contention window
$CW_{\min}$	Minimum contention window
W^s	$W = CW_{\min} + 1$, scaling factor

The performance of the EIED is affected by the choice of the values of r_I and r_D. The authors proposed using $r_I = 2$ and $r_D = \sqrt{2}$ to achieve better performance, compared to BEB [12]. The use of constant parameters like r_I and r_D by the STAs, however, creates hurdles for the EIED algorithm's being a part of future WLANs since upcoming dense WLANs require more adaptive and optimized control of parameters.

All above mentioned proposals for enhancements to the currently implemented BEB try to improve the efficiency of WLANs; on the other hand, since WLANs are rapidly changing to denser scenarios, such schemes need to recognize network congestion, as well. Adaptive adjustment to the size of the CW according to the network traffic must be more optimized. In this paper, we propose a novel CB scheme that provides high performances of WLANs at both high-density and low-density environments.

III. COGNITIVE BACKOFF MECHANISM

In this section, we explain the details of the proposed CB to enhance the performance of the CSMA/CA protocol. Algorithm 1 depicts a simple pseudo-code that compares the essence of BEB (currently implemented in the IEEE 802.11 standard), EIED, and ECA for the CSMA/CA protocol. In addition, Algorithm 2 shows how the proposed CB can be implemented with a few modifications to the CSMA/CA algorithm. The notations in the algorithms are summarized in Table I.

A. Conditional Collision Probability (p_{ck})

In this section, we derive a formula that explicitly relates to the CW of each competing STA to figure out the amount of traffic that can be measured independently during run time. The given analysis provides the modeling of the DCF mechanism, where each time slot corresponds either to an empty slot, or to a transmission or collision slot. We consider a scenario where every contending STA operates under saturation conditions, i.e., at least one packet is always ready in the queue for transmission.

Let p_{ck} be the conditional collision probability that a packet transmission will collide with another. Each individual STA can efficiently measure p_{ck} through physical layer sensing. The channel suffers a collision if a STA tries to use a time slot that is currently used by another STA for transmission. More-
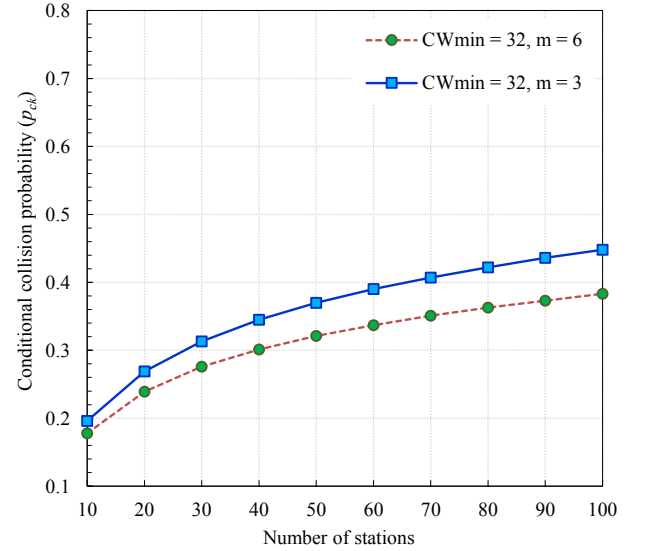
Fig. 1. Conditional collision probability (p_{ck}) versus number of STAs.

over, the STAs experience the collisions after transmission of the frame. Therefore, the conditional collision probability can be derived at each backoff stage by counting the number of busy time slots (N_{busy}), and the current number of time slots with failed transmissions (N_{coll}); then, p_{ck} is calculated from the total number of busy and collision time slots (N_{bc}) and the total number of observed slots that is composed of the number of backoff slots (N_{bo}) and the N_{bc} during the backoff interval, as follows [5]:

$$N_{\mathrm{bc}} = N_{\mathrm{busy}} + N_{\mathrm{coll}}, \tag{5}$$

$$p_{\mathrm{ck}} = \frac{N_{\mathrm{bc}}}{N_{\mathrm{busy}} + N_{\mathrm{coll}}}. \tag{6}$$

Fig. 1 depicts the average channel collision probability observed by n STAs in simulations of up to 100 STAs. It plots the number of STAs (n) contending for channel access versus p_{ck} in a saturated traffic environment with two different backoff parameters (i.e., minimum contention window size and maximum number of backoff stages); that is, ($CW_{\min} = 31$, $\max B = 6$) and ($CW_{\min} = 31$, $\max B = 3$). The rest of the simulation parameters are described in Table II. Fig. 1 shows that the increase in network density has a direct relationship to the channel conditional collision probability; the denser the network, the higher the channel collision probability.

B. Cognitive Backoff (CB)

The proposed CB mechanism increases or decreases the CW size according to the measured p_{ck}. It avoids unnecessary time spent in backoff procedures, and provides a gentle increment for the CW using channel-sensed data-driven intelligence. The advantages of using conditional collision probability p_{ck} in CB are threefold: (i) It permits adaptive adjustment of the CW value to fast-track variations of channel states in the WLAN; (ii) it allows significant adjustment of the optimal CW by exploiting the variance in the measurement of p_{ck}, whereas the measurement of p_{ck} can be independently generated by each STA; and

---

Algorithm 1: BEB, EIED, and ECA mechanisms.

1 **while** *(the device is on)* **do**
2   set $r \leftarrow 0, \max R \leftarrow 7, i \leftarrow 0, \max B \leftarrow 3 \; / \; 6$
3   set $CW_{\min} \leftarrow 16 \; / \; 31 \; / \; 64, CW_{\max} \leftarrow 1023,$
     $CW \leftarrow CW_{\min}$
4   set $r_D \leftarrow \sqrt{2}, r_I \leftarrow 2, k \leftarrow \text{uniform}[0, CW]$
5   **while** *(packet is in $TxQueue$)* **do**
6   repeat:
7   **while** *(k > 0)* **do**
8   **if** *(ChannelState = busy)* **then**
9   freeze backoff
10   **else**
11             $k \leftarrow k - 1$ // decrement backoff counter
            by one idle slot
12       $TxResult \leftarrow TxPacket()$
13   **if** *(TxResult = collision)* **then**
14          $r \leftarrow r + 1$
15          $i \leftarrow \min(i + 1, \max B)$
16   **if** *(protocol = BEB) or*
         (protocol = ECA) **then**
17             $CW \leftarrow$
            $\min(((2^i(CW_{\min} + 1)) - 1), CW_{\max})$
18   **else if** *(protocol = EIED)* **then**
19                $CW \leftarrow \min((r_I(CW_{\min} + 1)) -$
               $1, CW_{\max})$
20          $k \leftarrow \text{uniform}[0, CW]$
21   until $(retry = \max R)$ or
      $(TxResult = success)$
22   set $r \leftarrow 0, i \leftarrow 0$
23   **if** *(TxResult = success) or (r > \max R)* **then**
24   **if** *(protocol = BEB)* **then**
25             $CW \leftarrow CW_{\min}$
26   **else if** *(protocol = EIED)* **then**
27                $CW \leftarrow$
               $\min(((CW + 1)/r_D) - 1, CW_{\min})$
28   **else if** *(protocol = EIED)* **then**
29                $k \leftarrow \frac{[CW_{\min}]}{2} + 1$
30    $k \leftarrow \text{uniform}[0, CW]$

Algorithm 2: Cognitive backoff (CB) mechanism.

1 **while** *(the device is on)* **do**
2   set $r \leftarrow 0, \max R \leftarrow 7, i \leftarrow 0, \max B \leftarrow 3 \; / \; 6$
3   set $CW_{\min} \leftarrow 16 \; / \; 31 \; / \; 64, CW_{\max} \leftarrow 1023,$
     $CW \leftarrow CW_{\min}$
4   set $p_{\text{ck}} \leftarrow 0, N_{\text{bo}} \leftarrow k, N_{\text{busy}} \leftarrow 0, N_{\text{coll}} \leftarrow 0,$
     $N_{\text{bc}} \leftarrow 0$
5   set $k \leftarrow \text{uniform}[0, CW]$
6   **while** *(packet is in $TxQueue$)* **do**
7   repeat:
8   **while** *(k > 0)* **do**
9   **if** *(ChannelState = busy)* **then**
10             $N_{\text{busy}} \leftarrow N_{\text{busy}} + 1$ // count up busy
            slots
11   **else**
12             $k \leftarrow k - 1$ // decrement backoff counter
            one idle slot
13       $TxResult \leftarrow TxPacket()$
14   **if** *(TxResult = collision)* **then**
15          $r \leftarrow r + 1$ // count of re-transmission
16          $i \leftarrow \min(i + 1, \max B)$
17          $N_{\text{coll}} \leftarrow N_{\text{coll}} + 1$ //N_{coll} current
         transmission slots
18          $N_{\text{bc}} \leftarrow N_{\text{busy}} + N_{\text{coll}}$
19          $p_{\text{ck}} \leftarrow \frac{N_{\text{bc}}}{N_{\text{busy}} + N_{\text{coll}}}$
20          $CW \leftarrow$
         $\min\left((2^i(CW_{\min} + 1)^{(p_{\text{ck}}+1)} - 1), CW_{\max}\right)$
21          $k \leftarrow \text{uniform}[0, CW]$
22   until $(retry = \max R)$ or
      $(TxResult = success)$
23   set $r \leftarrow 0, i \leftarrow 0$
24   **if** *(TxResult = success) or (r > \max R)* **then**
25          $CW \leftarrow CW_{\min}$
26          $k \leftarrow \text{uniform}[0, CW]$

(iii) it is less complex to implement, that is compatible with the current standard, and the usage of p_{ck} does not have practical drawbacks.

There are a few basic changes in the CB, compared to the legacy BEB in the CSMA/CA protocol. The update of the CW in BEB is processed according to the pseudo-code in lines 21 and 31 of Algorithm 1 on collision and successful transmission, respectively. The CB is similar to the BEB in that it uses the same procedure after a successful transmission. On the other hand, the update in CB differs from BEB in that it uses an optimally determined backoff factor, $(CW_{\min} + 1)^{(p_{\text{ck}}+1)}$, to update the CW after collision. This procedure is described between lines 16 and

24 in Algorithm 2. The backoff value is uniformly selected as $k \leftarrow \text{uniform}[0, CW]$. If any STA failed in its transmission in previous attempts, the STA must increase its CW at the following attempt according to following:

$$CW = \min\left(\left(2^i(CW_{\min} + 1)^{(p_{\text{ck}}+1)} - 1\right), CW_{\max}\right), \quad (7)$$

where $(CW_{\min} + 1)$ is a constant design parameter to control the optimal size of the contention window, and i is the number of backoff stages.

Algorithm 2 describes the implementation of CB in which an adaptive update mechanism of the CW is used after each collision. The major changes with respect to BEB are in fact in line 10 and lines 17–23, while the busy slot, N_{busy}, is incremented at line 10 in each time slot when a STA observes the channel as busy. The uniform assignment of k from an exponentially increased CW in BEB is replaced by an adaptively adjusted CW determined from the conditional collision probability in the CB
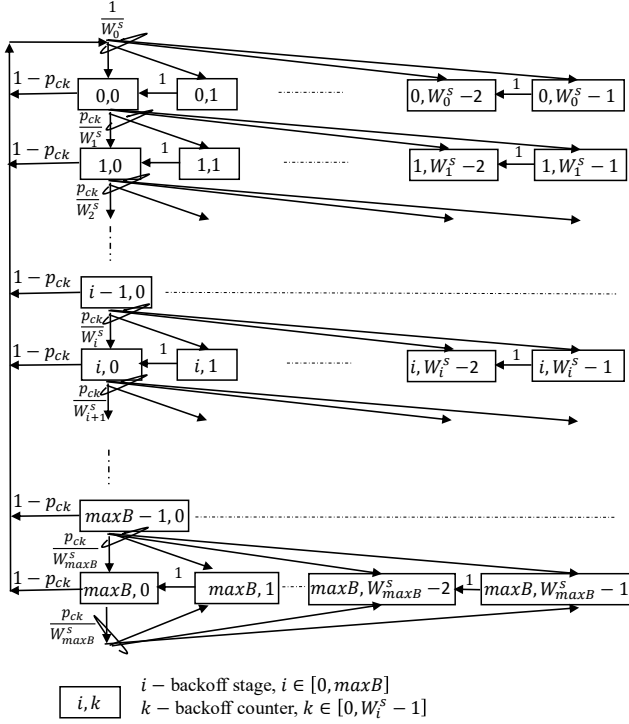
Fig. 2. Discrete-time Markov chain (DTMC) model for adaptive scaling of backoff window size in CB.

Table 2. MAC layer parameters used in simulations.

Operating Frequency	5 GHz
Bandwidth	20 MHz
Physical rate of the channel	6 Mbps
MAC header	24 bytes
MAC payload	1024 bytes
MAC trailer	4 bytes
PHY header	$20\,\mu s$
ACK length	14 bytes + PHY header
RTS length	20 bytes + PHY header
CTS length	14 bytes + PHY header
Transmission range	10 meters
Min. contention window ($CW_{\min}$)	15/31/63
Max. contention window ($CW_{\max}$)	1023
Slot duration (δ)	$9\,\mu s$
SIFS	$16\,\mu s$
DIFS	$60\,\mu s$
Propagation delay (Δ)	$1\,\mu s$
ACK_{timeout} length	DIFS $+\delta + 2\Delta+$ ACK
CTS_{timeout} length	DIFS $+\delta + 2\Delta+$ CTS
Maximum backoff stage (m) and retry limit	3/6
Simulation time	500 s

mechanism.

IV. PERFORMANCE ANALYSES OF CB

In this section, we formulate an analytical model of the proposed CB mechanism for saturated throughput and delay, on the assumption of ideal channel conditions (i.e., no hidden terminal and capture effects). In the analysis, we assume a fixed number of STAs that are in a saturated condition, and thus, the transmission queue of each STA is assumed as always nonempty. Initially, we study the behavior of a STA with a discrete-time Markov chain (DTMC) model [13], and we obtain stationary transmission probability τ of the STA. Later, by knowing the exact state transitions that can occur on the communication channel within a randomly selected time slot, we formulate the normalized throughput and the average delay of the proposed CB mechanism.

A. The Discrete-Time Markov Chain Model

Consider n STAs are competing for a channel in a WLAN. In a saturated WLAN, each STA always has a data frame available to transmit immediately after each successful transmission. Thus, due to the consecutive data frame transmissions, each data frame needs to wait for a random backoff counter before transmission.

Assume $\Omega(t)$ is the function of a stochastic process representing the backoff counter for a STA. Since the counter is discretized in an integer time scale for time slots, t and $t + 1$ correspond to the beginning of two consecutive time slots, and the backoff counter for each STA decrements at the beginning of

each time slot. The backoff counter decrements when the communication channel is sensed as idle, and it stops when the channel is sensed as busy, which is due to a transmission by any other STA. Therefore, the time interval between two consecutive time slot beginnings may be much longer than the idle time slot size (σ).

Since the backoff counter of each STA also depends on its retransmission attempts (referred as the number of backoff stages in the IEEE 802.11 standard), stochastic proces $\Omega(t)$ behaves like a non-Markovian process. Let i be the backoff stage of a STA, and let $\max B$ be the maximum number of backoff stages that i can experience for a data frame (i.e., for $i \in [1, \max B]$), such that $W_i^s = 2^i(CW_{\min} + 1)^{(p_{\text{ck}}+1)}$ for the ith backoff stage. $CW_{\min}$ is the minimum contention window used for the initial backoff. For convenience, we define an adaptively scaled contention window, $W^s = (CW_{\min} + 1)^{(p_{\text{ck}}+1)}$, where $(CW_{\min} + 1)$ is the scaling factor to adaptively increase the backoff counter with the observed channel conditional collision probability (p_{ck}). Let us adopt the notation $W_{\max}^s = 2^{\max B}W^s$ for the $\max B^{\text{th}}$ maximum number of backoff stages, and we adopt $W_{(i+1)}^s = 2^{(i+1)}W^s$ for the adaptively scaled-up contention window for the $(i + 1)$th backoff stage when transmission fails at the ith backoff stage.

Let $\pi(t)$ be the stochastic process representing the backoff stage $(0, 1, 2, \cdots, \max B)$ of the STA at time t. The key point in our DTMC model is that, on each data frame transmission attempt, regardless of the number of retransmission attempts (i.e., backoff counter and backoff stage), the state transition probability is calculated independently with a practically observed conditional collision probability, p_{ck}.

With these assumptions, the stochastic processes of backoff counter and backoff stages (i.e., $\Omega(t)$ and $\pi(t)$) can be modeled as the two-dimensional process $(\Omega(t), \pi(t))$ with the DTMC, as depicted in Fig. 2. In this DTMC, the transition probabilities are described as follows.

(i) The backoff counter decrements when the channel is

sensed as idle with following probability:

$$P_r\{i,k|i,k+1\} = 1, \quad i \in [0,\max B], k \in [0, W_i^s - 2] \tag{8}$$

(ii) The STA moves to the next backoff stage, i, if a data frame transmission fails at the $(i-1)$th stage with following probability:

$$P_r\{i,k|i-1,k\} = \frac{p_{\text{ck}}}{W_i^s}, \quad i \in [0,\max B], k \in [0, W_i^s - 1] \tag{9}$$

(iii) The STA moves to initial backoff stage 0 after a successful transmission at the ith stage with following probability:

$$P_r\{0,k|i,k\} = 1 - p_{\text{ck}}, \quad i \in [0,\max B], k \in [0, W_i^s - 1] \tag{10}$$

(iv) The STA remains at the maximum number of stages (i.e., the $\max B^{\text{th}}$ stage) after an unsuccessful transmission at the $\max B^{\text{th}}$ stage with following probability:

$$P_r\{\max B, k|\max B, k\} = \frac{p_{\text{ck}}^s}{W_{\max}},$$
$$i \in [0,\max B], k \in [0, W_i^s - 1] \tag{11}$$

In particular, for the above transmission probabilities, as considered in (9), when a data frame transmission collides at backoff stage $i-1$, the backoff stage increases, and the new backoff counter value (k) is uniformly chosen from the adaptively scaled contention window, W_i^s. Once the backoff stage reaches value $\max B$, it is not increased in subsequent data frame retransmission attempts.

We assume that $b_{(i,k)} = \lim_{h \to 0} P_r (\Omega(t), \pi(t)) = (i, k), i \in [0, \max B], k \in [0, W_i^s - 1]$ is the steady state distribution of the DTMC. We now explain how to obtain a closed-form solution for this DTMC; that is, from the transitions of the DTMC, we note that

$$b_{1,0} = p_{\text{ck}} b_{0,0}. \tag{12}$$

Similarly, $b_{i,0} = p_{\text{ck}} b_{i-1,0}$, where, $b_{i-1,0} = p_{\text{ck}} b_{i-2,0}$ till $b_{1,0} = p_{\text{ck}} b_{0,0}$, therefore,

$$b_{i,0} = p_{\text{ck}} b_{0,0}. \tag{13}$$

Now, for the stage $b_{\max B,0}$, we can define that

$$b_{\max B,0} = \frac{p_{\text{ck}}^{\max B}}{1 - p_{\text{ck}}} b_{0,0}. \tag{14}$$

Owing to the chain regularities, for each $k \in [0, W_i^s - 1]$, the stationary distribution for $(\Omega(t), \pi(t))$ can be written as

$$b_{i,k} = \begin{cases} \frac{W_i^s - k}{W_i^s} p_{\text{ck}} b_{0,0}, & 0 \leq i \leq \max B \\ \frac{W_{\max B}^s - k}{W_{\max B}^s} p_{\text{ck}}(b_{\max B-1,0} + b_{\max B,0}), & i = \max B. \end{cases} \tag{15}$$

These two equations can be combined and defined as

$$\sum_{i=0}^{\max B} b_{i,0} = b_{0,0} + b_{1,0} + b_{2,0} + \cdots + b_{\max B,0}$$
$$= b_{0,0}(1 + p_{\text{ck}} + \cdots + p_{\text{ck}}^{\max B-1}) + \frac{p_{\text{ck}}^{\max B}}{1 - p_{\text{ck}}} b_{0,0}. \tag{16}$$

By solving (16), finally, we get $\sum_{i=0}^{\max B} b_{i,0}$, and it can be rewritten as

$$\sum_{i=0}^{\max B} b_{i,0} = \frac{W_i^s - k}{W_i^s} p_{\text{ck}} b_{i,0}. \tag{17}$$

Thus, by (15), (16), and (17), all values $b_{i,k}$ are expressed as the function of the $b_{0,0}$ state and of channel observation–based practical conditional collision probability p_{ck}. And $b_{0,0}$ is finally determined by normalizing the DTMC states as follows:

$$1 = \sum_{i=0}^{\max B} \sum_{k=0}^{W_i^s-1} b_{i,k} = \sum_{i=0}^{\max B} b_{i,0} \sum_{k=0}^{W_i^s-1} \frac{W_i^s - k}{W_i^s}$$
$$= \sum_{i=0}^{\max B} b_{i,0} \frac{W_i^s + 1}{W_i^s}. \tag{18}$$

Since $W_i^s = 2^i W^{(\text{Pck}+1)}$, with $W = CW\min + 1$, the above normalization relation can be written as

$$1 = \frac{b_{0,0}}{2} \left[W^{\text{Pck}+1} \left(\sum_{i=0}^{\max B-1} (2p_{\text{ck}})^i + \frac{2p_{\text{ck}}^{\max B}}{1 - p_{\text{ck}}} \right) + \frac{1}{(1 - p_{\text{ck}})} \right], \tag{19}$$

from which we finally obtain $b_{0,0}$ as follows:

$$b_{0,0} = $$
$$\frac{2(1 - 2p_{\text{ck}})(1 - p_{\text{ck}})}{(W^{\text{Pck}+1} + 1)(1 - 2p_{\text{ck}}) + p_{\text{ck}} W^{\text{Pck}+1} (1 - (2p_{\text{ck}})^{\max B})}.$$

Let τ be the probability that a STA transmits in a randomly selected slot time. Since a transmission occurs only when the backoff counter of the STA reaches zero, regardless of the backoff stage, it can be expressed as follows:

$$\tau = \sum_{i=0}^{\max B} b_{i,0} = \frac{1}{1 - p_{\text{ck}}} b_{0,0}. \tag{20}$$

Using $b_{0,0}$ obtained in (15), τ can be redefined as

$$\tau = \frac{2(1 - 2p_{\text{ck}})}{(W^{\text{Pck}+1} + 1)(1 - 2p_{\text{ck}}) + p_{\text{ck}} W^{\text{Pck}+1} (1 - (2p_{\text{ck}})^{\max B})}. \tag{21}$$

Equation (22) can be alternatively written as

$$\tau = \frac{2}{(W^{\text{Pck}+1} + 1) + p_{\text{ck}} W^{\text{Pck}+1} \left(\sum_{i=0}^{\max B-1} (2p_{\text{ck}})^i \right)}. \tag{22}$$

However, in general, τ depends on practical collision probability p_{ck}, which is always unknown until the channel is observed for

busy slots. To find the value of p_{ck} for analytical considerations, it is sufficient to note for conditional collision probability p_{ck} that a transmitted data frame encounters a collision in a time slot if at least one of the $n-1$ remaining STAs transmits. Since each transmission in the system sensesthis collision in the same state, the steady state can be numerically obtained:

$$p_{\text{ck}} = 1 - (1 - \tau)^{n-1}. \tag{23}$$

B. Normalized Throughput Analysis

Let $\hat{S}$ be the normalized throughput of the overall network system, which can be defined as the fraction of the communication channel used for successful transmission of the data payload. To compute $(\hat{S})$, let P_{tr} be the probability that there is at least one transmission in the considered time slot. Since there are n STAs in the system contending for the medium, and each transmission has probability τ, P_{tr} can be obtained by

$$P_{\text{tr}} = 1 - (1 - \tau)^{n}. \tag{24}$$

The probability of successful transmission P_s defined as the probability that only STAs transmit in the considered time slot, then that will be

$$P_s = \frac{n\tau(1-\tau)^{n-1}}{P_{\text{tr}}} = \frac{n\tau(1-\tau)^{n-1}}{1-(1-\tau)^{n}}, \tag{25}$$

and thus, $\hat{S}$ can be defined as the ratio

$$\hat{S} = \frac{E[\text{mean payload transmitted in a time slot}]}{E[\text{total length of a time slot}]}. \tag{26}$$

Assume $E[F]$ is the average data frame size (assuming that all data frames have the same fixed size), then the time slot for transmitting this average payload data successfully can be obtained as $P_{\text{tr}}P_s E[F]$, since $P_{\text{tr}}P_s$ is the probability of the successful transmission of a data frame in a given time slot. The average length of a given time slot, $E[Slot]$ is calculated with the following consideration: (i) If there is no transmission in time slot $(1 - P_{\text{tr}})\sigma$, that is, it is an idle slot, (ii) it can contain a successfully transmitted data frame, $P_{\text{tr}}P_s$, and (iii) it can also contain a collision, that is, $P_{\text{tr}}(1 - P_s)$. Finally, (27) can be written as follows:

$$\hat{S} = \frac{P_{\text{tr}}P_s E[F]}{(1 - P_{\text{tr}})\sigma + P_{\text{tr}}P_s T_s + P_{\text{tr}}(1 - P_s)T_c}, \tag{27}$$

where T_s and T_c are the average time when the communication channel has been busy due to successful transmissions, and the time when the channel was sensed by each STA as busy during a collision, respectively. The values for T_s and T_c depend upon the IEEE 802.11 standard parameters shown in Table II.

Let $F_{\text{hdr}} = PHY_{\text{hdr}} + MAC_{\text{hdr}}$ be the time to transmit a data frame header, ACK the time to receive an acknowledgement, and δ the channel propagation delay. In the basic access mode, T_s and T_c can be obtained as

$$T_s^{\text{basic}} = F_{\text{hdr}} + E[F] + SIFS + DIFS + 2\delta + ACK, \tag{28}$$

$$T_c^{\text{basic}} = F_{\text{hdr}} + E[F] + DIFS + \delta. \tag{29}$$

Since in RTS/CTS access mode, a collision can occur only in RTS frames, the values for T_s and T_c are obtained as

$$T_s^{\text{rts/cts}} = RTS + CTS + 3SIFS + DIFS + 4\delta$$
$$+ F_{\text{hdr}} + E[F] + ACK, \tag{30}$$

$$T_c^{\text{rts/cts}} = RTS + DIFS + \delta. \tag{31}$$

C. Saturation Delay Analysis

In this subsection, we derive the saturation delay $E[D]$ in the proposed CB mechanism for a successfully transmitted data frame. The saturation delay is defined as the average time from the time the data frame is put at the head of its MAC queue, ready for transmission, until its successful reception at the destination. According to [14],

$$E[D] = E[B]E[Slot], \tag{32}$$

where $E[Slot]$ is the total length of the time slot as given in (23); that is,

$$E[Slot] = (1 - P_{\text{tr}})\sigma + P_{\text{tr}}P_s T_s + P_{\text{tr}}(1 - P_s)T_c, \tag{33}$$

where $E[B]$ is the average number of backoff time slots for a successful data frame transmission. It can be calculated by multiplying the number of time slots, (b_i), that the packet is delayed in each retransmission attempt, by probability d_i to reach the backoff stage. $E[B]$ is given as

$$E[B] = b_i d_i, \tag{34}$$

$$b_i = \frac{W_i^s + 1}{2}, \qquad i \in [0, \text{max}B] \tag{35}$$

$$d_i = \begin{cases} p_{\text{ck}}^i, & i \in [0, \text{max}B - 1] \\ \frac{p_{\text{ck}}^{\text{max}B}}{1 - p_{\text{ck}}}, & i = \text{max}B. \end{cases} \tag{36}$$

E[B] can be solved as

$$E[B] =$$
$$\frac{(W^{\text{Pck}+1} + 1)(1 - 2p_{\text{ck}}) + p_{\text{ck}}W^{\text{Pck}+1}\left(1 - (2p_{\text{ck}})^{\text{max}B}\right)}{2(1 - 2p_{\text{ck}})(1 - p_{\text{ck}})}. \tag{37}$$

D. Model Validation

In this sub section, we compare the analytical results with the simulation results for the proposed CB in an event-driven simulator, Network Simulator-3 (NS-3) version 3.24 [15]. To evaluate the performance analysis, a network of n STAs is considered (n is ranging from 10 to 100), where each STA is within the coverage area of the others (no hidden terminals). The channel does not introduce any errors, and the STAs are set to be in saturation state (always ready to transmit). The specific MAC and PHY layer parameters of HEW [2] are considered as listed in Table II.

Fig. 3 depicts the normalized throughput and saturation delay (ms), and show that the analytical model is accurate, since the analytical results almost overlap the simulation results, in both basic (Figs. 3(a) and 3(b)) and RTS/CTS, (Figs. 3(c) and 3(d)) access mechanisms, respectively. All simulation results in the figures are obtained with a 90% confidence interval of ±0.04.
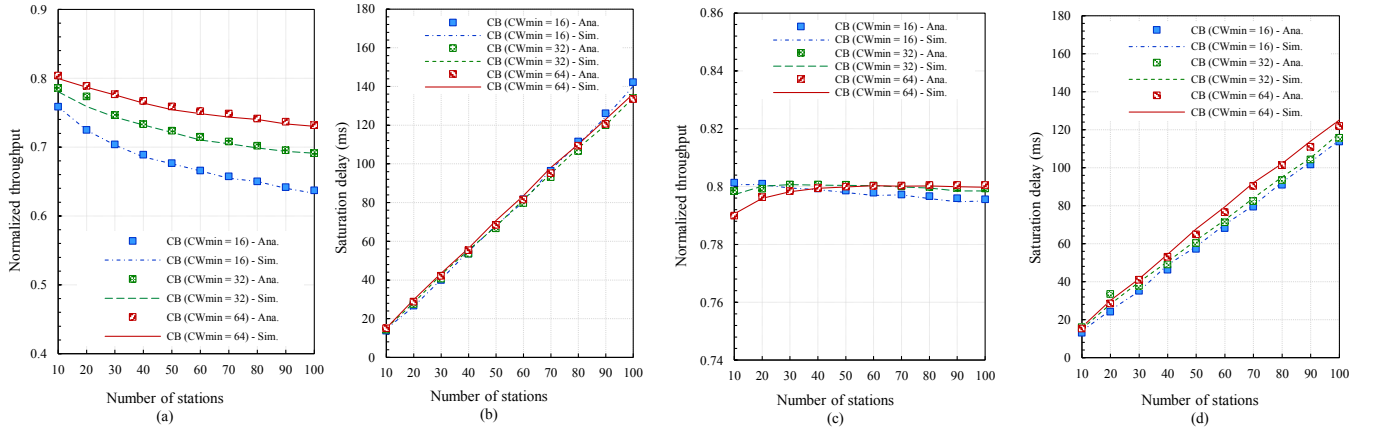
Fig. 3. Performance results of CB in simulation versus analysis with ($CW_{\min} = 16, 32, 64$) and ($\max B = 6$): (a) Normalized throughput, (b) Saturation delay (ms) in basic access mode, (c) Normalized throughput, and (d) Saturation delay (ms) in RTS/CTS mode.
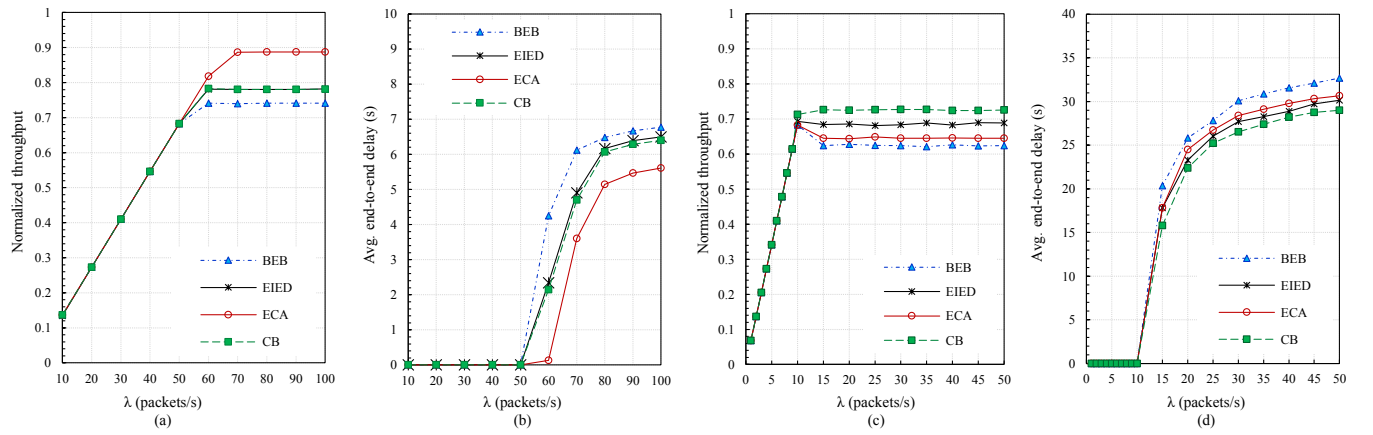

Fig. 4. Performance results of different protocols under saturated traffic with ($CW_{\min} = 32, \max B = 6$)): (a) Normalized throughput, (b) Saturation delay (ms) in basic access mode, (c) Normalized throughput, and (d) Saturation delay (ms) in RTS/CTS mode.
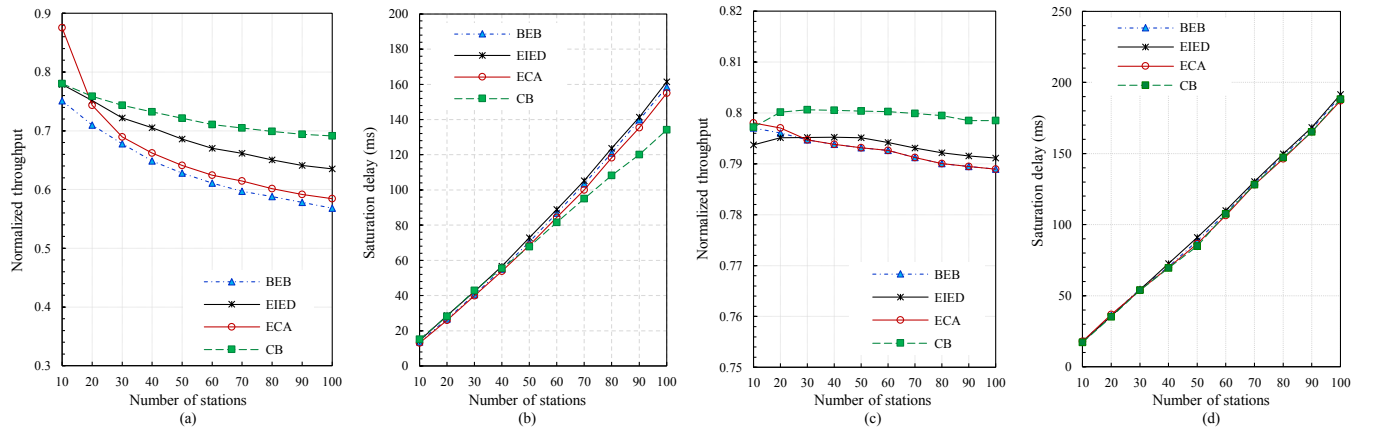

Fig. 5. Performance results of different protocols under non-saturated traffic with an increasing of number of packets per second ($CW_{\min} = 32, \max B = 6$)): (a) Normalized throughput, (b) Average end-to-end delay with $n = 10$, (c) Normalized throughput, and (d) Average end-to-end delay with $n = 50$.

V. PERFORMANCE EVALUATION

In this section, the proposed CB mechanism is evaluated with simulation results. The experimental results and the performance are presented and evaluated by comparing it with BEB, EIED, and ECA algorithms from several aspects, including sat-

urated and non-saturated network environments.

A. Non-Saturated Scenario

Fig. 4(a) shows the normalized throughput of the proposed CB along with compared algorithms (i.e., BEB, ECA, and

EIED) with an increase in the number of packets/s (λ) from 10–100 packets/s when $n = 10$ STAs. In Fig. 4(a), the aggregate throughput linearly increases until the saturation point at around 50 packets/s, and after that saturation point, the algorithms perform differently. The BEB degrades in performance after 60 packets/s, while the ECA manages to perform efficiently up to 70 packets/s due to the low collision environment. Fig. 4(b) describes average delay of ECA is better as compared to BEB, EIED, and the proposed CB. This better performance of ECA is due to fewer contenders in the network. In Fig. 4(b), a rapid increase in the average delay for BEB STAs is observed at the saturation point (that is, around packets/s), whereas with the proposed CB, delay is still low. The performance gain of the proposed CB in dense networks can be observed in Figs. 4(c) and 4(d), where the number of STAs increases to $n = 50$. Figs. 4(c) and 4(d) also provide an important evaluation of the CB for different traffic loads in a non-saturated network environment, when there are 50 fixed STAs with different network traffic loads λ (in packets/s).

B. Saturated Scenario

In CSMA/CA, a large number of saturated STAs is normally the reason for a high collision probability. This effect is in part the result of adaptively scaling the contention window for transmission collisions. However, the saturated scenario provides a condition advantageous to the CB mechanism. In the saturated environment, the CB senses the channel collision probability and more efficiently determines the backoff contention window, and effectively increases the saturation throughput. In this subsection we evaluate CB, BEB, ECA, and EIED in saturated scenarios for both basic and RTS/CTS access modes.

Fig. 5(a) shows the normalized throughput increase of the CB due to more efficient use of the channel. It shows that the normalized throughput decreases with the increased number of STAs. This is because the number of successful STAs in contention increases as the number of contending STAs increases. On the other hand, the throughput of CB highly depends on the channel sensing–based practical collision probability; therefore, even with a large number of contending STAs, the throughput gain is higher than the other mechanisms. As can be observed, the throughput of BEB in CSMA/CA decreases with the increased number of STAs, as the BEB reduces the number of transmission attempts to keep the number of collisions low. The ECA is able to achieve collision-free operation if the number of contending STAs is lower than the deterministic cycle length, i.e., $N < CW_{\min}/2$. This is why, in Fig. 5, a sudden phase transition is observed at $N = 20$, when N is larger than $CW_{\min}/2$. Similarly, the EIED remains higher than the BEB. This is because, instead of resetting the CW to its minimum value on a successful transmission, STAs exponentially decrease the size so that unexpected collisions can be avoided. The proposed CB has a lower delay due to its low collision and optimized contention window adjustment as shown in Fig. 5(b).

Fig. 5(c) shows the normalized throughput with RTS/CTS access mode. The proposed CB performs better than the ECA even with the small number of STAs with RTS/CTS access mode, due to handling collisions more adaptively. Since the ECA is

designed for STAs spending most of the time in successful time slots, with RTS/CTS access mode, the CB handles the percentage of collisions to transmission attempts more efficiently. The CB curve in Fig. 5(d) shows an increased normalized throughput as the channel conditional collision probability grows. The CB makes contending STAs adaptively scale their backoff contention window, effectively increasing throughput.

VI. CONCLUSION

In this paper, we proposed an enhancement of the prevalent binary exponential backoff of the CSMA/CA protocol, which is used in current IEEE 802.11 WLANs. Compared with the BEB algorithm of the CSMA/CA protocol, the proposed cognitive backoff uses a CB mechanism to offer enhanced performance in terms of both throughput and delay, while preserving fairness among the STAs. In fact, with increased contending STAs, a steady state situation can be reached instead of a further decrease in performance in a distributed manner due to adaptive change in the CW. This performance gain is achieved by effectively sensing the channel for a conditional collision probability, and adjusting the backoff contention window intelligently. The results indicate that the performance of the proposed CB algorithm increases relative to the increased number of contending STAs. All these properties make the CB a good candidate for the upcoming densely deployed WLANs, such as HEW.

REFERENCES

[1]   R. Ali, S. W. Kim, B.-S. Kim, and Y. Park, "Design of MAC layer resource allocation schemes for IEEE 802.11ax: Future directions," *IETE Technical Review*, 2016, DOI: 10.1080/02564602.2016.1242387.

[2]   IEEE 802 LAN/MAN Standards Committee, IEEE Standard 802.11ax, 2014.

[3]   M. K. Afzal, M. H. Rehmani, A. Pescape, S. W. Kim, and W. Ejaz, "The new era of smart cities: Sensors, communication technologies, and applications," *IEEE Access*, vol. 5, pp. 27836–27840, Dec. 2017.

[4]   G. P. Joshi, S. Y. Nam, S. W. Kim, and B. Kim, "Adaptive window size-based medium access control protocol for cognitive radio wireless sensor networks," *Hindawi J. Sensors*, vol. 2016, pp. 1–9, Jan. 2016.

[5]   G. Bianchi and I. Tinnirello, "Kalman filter estimation of the number of competing terminals in an IEEE 802.11 network," in *Proc. IEEE INFO-COM*, 2003, pp. 844–852.

[6]   N. Shahin *et al.*, "Adaptively scaled back-off (ASB) mechanism for enhanced performance of CSMA/CA in IEEE 802.11ax high efficiency WLAN," in *Proc. IEEE/IFIP NOMS*, Taiwan, 2018, pp.1–5.

[7]   J. Barcelo *et al.*, "CSMA with enhanced collision avoidance: a performance assessment," in *Proc. IEEE VTC*, Barcelona, 2009, pp. 1–5.

[8]   J. Barcelo *et al.*, "Fairness and convergence of CSMA with enhanced collision avoidance (ECA)," in *Proc. IEEE ICC*, Budapest, 2013, pp. 1274–1279.

[9]   L. Sanabria-Russo *et al.*, "Future evolution of CSMA protocols for the IEEE 802.11 standard," in *Proc. ICC Workshops*, 2013, pp 1274–1279.

[10] L. Sanabria-Russo *et al.*, "A high efficiency MAC protocol for WLANs: Providing fairness in dense scenarios," *IEEE/ACM Trans. Netw.*, Vol. 2005, no. 1, pp 67–78, Mar. 2005.

[11] P. Chatzimisios, A. C. Boucouvalas and V. Vitsas "IEEE 802.11 wireless LANs: Performance analysis and protocol refinement," *EURASIP J. Wireless Commun. Netw.*, vol. 2005, no. 1, pp 67-78, Mar. 2005.

[12] C. Ye, Y. Li and A. Reznik "Performance analysis of exponential increase exponential decrease back-off algorithm," in *Proc. IEEE GLOBECOM*, Miami, 2010.

[13] R. Ali *et al.*, "I-DTMC: An integrated-discrete time Markov chain model for performance analysis in future WLANs," in *Proc. LCN Workshops*, Singapore, 2017.

[14] P. Chatzimisios, A. C. Boucouvalas and V. Vitsas "Packet delay analysis of IEEE 802.11 MAC protocol," *Electron. Lett.*, vol. 39, no. 18, pp. 1358-1359, 2003.

[15] The Network Simulator — ns-3. [Online]. Available: https://www.nsnam.org/.

Nurullah Shahin received his B.Sc. and M.Sc. from the Department of Information and Communication Engineering, Islamic University, Kushtia, Bangladesh, in 2009 and 2010, respectively. He is currently pursuing a combined MS.c. and Ph.D. in the advanced networking technology lab (yuANTL) of the Department of Information and Communication Engineering, Yeungnam University, Gyeongsan, Republic of Korea. He is a Maintenance Engineer (ME) with the ICT Infrastructure Maintenance and Management Department, Bangladesh Bank (Central Bank of Bangladesh), Bangladesh. His research interests include dense wireless networks, vehicular ad-hoc networks (VANETs), and the Internet of Things (IoT).

Rashid Ali received the B.Sc. degree Information Technology from Gomal University, Pakistan, in 2007, the M.Sc. degree in Computer Science (advanced network design), in 2010, under the supervision of Dr. S. Belenki, and the M.Sc. degree in Informatics from University West, Sweden, in 2013, under the supervision of Dr. M. Spante. He is currently pursuing the Ph.D. degree with the Wireless Information Networking Laboratory, Department of Information and Communication Engineering, Yeungnam University, South Korea. Between 2007 and 2009, he was a WiMAX Engineer with the Operations Research Department, Wateen Telecom Pvt. Ltd., Pakistan. From 2013 to 2014, he was a Lecturer with COMSATS University, Vehari, Pakistan. His research interests include enhancement of efficiency and reliability in future WLANs, modeling and analyzing the stochastic process of media access control layer resource allocation in future WLANs, deep learning, blockchain, and the Internet of Things.

Sung Won Kim received his B.S. and M.S. degrees from the Department of Control and Instrumentation Engineering, Seoul National University, Korea, in 1990 and 1992, respectively, and his Ph.D. degree from the School of Electrical Engineering and Computer Sciences, Seoul National University, Korea, in August 2002. From January 1992 to August 2001, he was a researcher at the Research and Development Center of LG Electronics, Korea. From August 2001 to August 2003, he was a researcher at the Research and Development Center of AL Tech, Korea. From August 2003 to February 2005, he was a postdoctoral researcher in the Department of Electrical and Computer Engineering, University of Florida, Gainesville, USA. In March 2005, he joined the Department of Information and Communication Engineering, Yeungnam University, Gyeongsangbuk-do, Korea, where he is currently a professor. His research interests include resource management, wireless networks, mobile networks, performance evaluation, and embedded systems.

Young-Tak Kim (S' 84 – M'90) received a Ph.D. from KAIST in February 1990. He joined Korea Telecom in 1990, where he researched and developed the ATM MAN Switching System (ATM-MSS) and related network operations and management technologies for broadband networking. He participated in the standardization activities of ITU-T Study Group 13 as the representative of Korea Telecom (KT). Currently, he is a professor in the Department of Information and Communication Engineering, College of Engineering, at Yeungnam University, Korea. In 2001, 2008, and 2015 he was a Visiting Scholar at the National Institute of Standards and Technology (NIST), USA, where he joined in the design and implementation of the NIST GMPLS simulator (GLASS) and the Next Generation Routing Architecture (NGRA) project. His research interests include QoS-guaranteed traffic engineering in future Internet, Cloud Computing, QoS-aware Network Operating Systems, OpenFlow, QoS-aware seamless secure mobility, Software Defined Networking (SDN) / Network Function Virtualization (NFV), and related network operations and management. He is a member of the IEEE Communications Society, KICS, KISS, KIPS, and the Korea Multimedia Society. He was the Technical Program Chair of IEEE ComSoc CNOM for 2007-2008, TPC Co-chair of IFIP/IEEE IM2009, the general chair of APNOMS2009, and TPC co-chair of IEEE/IFIP NOMS 2018.