sustainability

MDPI

Article

A Self-Scrutinized Backoff Mechanism for IEEE 802.11ax in 5G Unlicensed Networks

Rashid Ali [1], Nurullah Shahin [1], Rojeena Bajracharya [1], Byung-Seo Kim [2] and Sung Won Kim [1,*]

[1] Department of Information and Communication Engineering, Yeungnam University, Gyeongsan 38541, Korea; rashid@ynu.ac.kr (R.A.); nurullah@ynu.ac.kr (N.S.); rozbajracharya@gmail.com (R.B.)

[2] Department of Computer and Information Communication Engineering, Hongik University, Seoul 04066, Korea; jsnbs@hongik.ac.kr

* Correspondence: swon@yu.ac.kr; Tel.: +82-53-810-2483

Received: 28 March 2018; Accepted: 12 April 2018; Published: 16 April 2018

check for updates

Abstract: The IEEE 802.11ax high-efficiency wireless local area network (HEW) is promising as a foundation for evolving the fifth-generation (5G) radio access network on unlicensed bands (5G-U). 5G-U is a continued effort toward rich ubiquitous communication infrastructures, promising faster and reliable services for the end user. HEW is likely to provide four times higher network efficiency even in highly dense network deployments. However, the current wireless local area network (WLAN) itself faces huge challenge of efficient radio access due to its contention-based nature. WLAN uses a carrier sense multiple access with collision avoidance (CSMA/CA) procedure in medium access control (MAC) protocols, which is based on a binary exponential backoff (BEB) mechanism. Blind increase and decrease of the contention window in BEB limits the performance of WLAN to a limited number of contenders, thus affecting end-user quality of experience. In this paper, we identify future use cases of HEW proposed for 5G-U networks. We use a self-scrutinized channel observation-based scaled backoff (COSB) mechanism to handle the high-density contention challenges. Furthermore, a recursive discrete-time Markov chain model (R-DTMC) is formulated to analyze the performance efficiency of the proposed solution. The analytical and simulation results show that the proposed mechanism can improve user experience in 5G-U networks.

Keywords: 5G network; 5G unlicensed; IEEE 802.11ax; channel contention

1. Introduction

Fifth-generation (5G) wireless systems are becoming a priority for telecom operators, as 5G comes with the promise of unseen services and a broad range of new use cases and business models ranging from smart transport systems to smart agriculture and factories. 5G is expected to push the digitization of the economy further due to its ability to handle large volumes of data with low latency in real time. The evolution of the 5G era has promised an age of boundless connectivity and intelligent automation. 5G wireless networks will support 1000-fold gains in capacity, connections for at least 100 billion devices, and 10 Gb/s individual user experiences capable of extremely low latency [1,2]. To support massive capacity and connectivity, the IEEE 802.11ax high-efficiency wireless local area network (HEW) is promising as the foundation to evolve the 5G radio access systems to unlicensed band (namely described as, 5G-U) [3]. However, wireless local area networks (WLAN) will face huge challenges to access this unlicensed band, especially for highly dense user deployments. WLAN medium access control (MAC) protocol mainly focuses on maximizing the communication radio utilization using fair MAC layer resource allocation (MAC-RA) [4,5] employing a carrier sense multiple access with collision avoidance (CSMA/CA) scheme of distributed coordination function (DCF) for the Wi-Fi user equipment's (WEs) competition to access the medium. To achieve the maximum communication radio utilization through fair MAC-RA in the WLANs with the ever-increasing density of contending WEs, the CSMA/CA scheme of the current DCF is of great importance as a part of 5G-U [5].

The binary exponential backoff (BEB) scheme is the typical and traditional CSMA/CA mechanism, which was introduced in IEEE 802.11 DCF [6]. A randomly generated backoff value for the contention procedure is used. At the first transmission attempt, the WE generates a uniform random backoff value B, from the contention window interval $[0, W_{cur}]$, where W_{cur} is initially set to the minimum value W_{min}. After each unsuccessful transmission, W_{cur} is doubled until it reaches the maximum value $W_{max} = 2^m W_{min}$, for m maximum number of backoff stages (b), that is, $b \in (0, m)$. Once a WE successfully transmits its data frame, W_{cur} is reset to the minimum value W_{min}. For a network with a heavy load, resetting b to zero and contention window (W) to its W_{min} value after successful transmission will result in more collisions and poor network performance due to an increase in probability to select similar backoff value B. Similarly, for fewer contending WEs, the blind exponential increase of W_{cur} for collision avoidance causes an unnecessarily long delay due to the wider range for selecting B. Besides, this blind increase/decrease of the backoff window is more inefficient in the highly dense networks proposed for IEEE 802.11ax enabling for 5G-U, because the probability of contention collision increases with the increasing number of WEs. Thus, the current MAC-RA protocol does not allow WLANs to achieve high efficiency in highly dense environments and become a part of future 5G-U, whereas the upcoming HEW will suffer from such unresolved issues as they will be required to achieve four times higher network efficiency, even in highly dense network deployments [3]. Hence, to withstand this challenge, WLAN needs a more efficient and self-scrutinized backoff mechanism to promise enhanced user quality of experience (QoE).

This paper describes some of the use cases of HEW deployments for 5G-U networks. To solve the critical medium collision problem incurred by a large number of densely deployed contending WEs, a practical channel observation-based scaled backoff (COSB) mechanism [4] is presented in this paper. The COSB guarantees enhanced QoE in terms of high throughput and low delay in a high-density environment by reducing the number of collisions during the channel access mechanism by using a self-scrutinized backoff mechanism.

The remainder of the paper is organized as follows: Section 2 describes some of the use cases of HEW deployments in 5G-U. In Section 3, the proposed COSB is described in detail. An analytical model is formulated in Section 4 to affirm the performance enhancement using COSB in highly dense environments. Section 5 describes the performance evaluation of COSB compared with state-of-the art BEB along with two similar backoff scaling protocols. Finally, in Section 6, a comprehensive conclusion and future works are presented.

2. IEEE 802.11ax Use Cases in Fifth-Generation Radio Access Network on Unlicensed Bands (5G-U)

Based on current research trends [5] and developments within IEEE 802.11ax HEW [3], several future use cases for 802.11ax can be identified. Figure 1 shows some of the use cases proposed for HEW in 5G-U [7]. These include a high throughput HEW in the form of a gigabit ethernet connection replacement, improved network capacity with multi-user multiple-input and multiple-output (MU-MIMO) transmissions, using HEW as a backhaul for local area networks (LAN), and supporting highly dense scenarios (such as an office building, stadium, train, etc.).

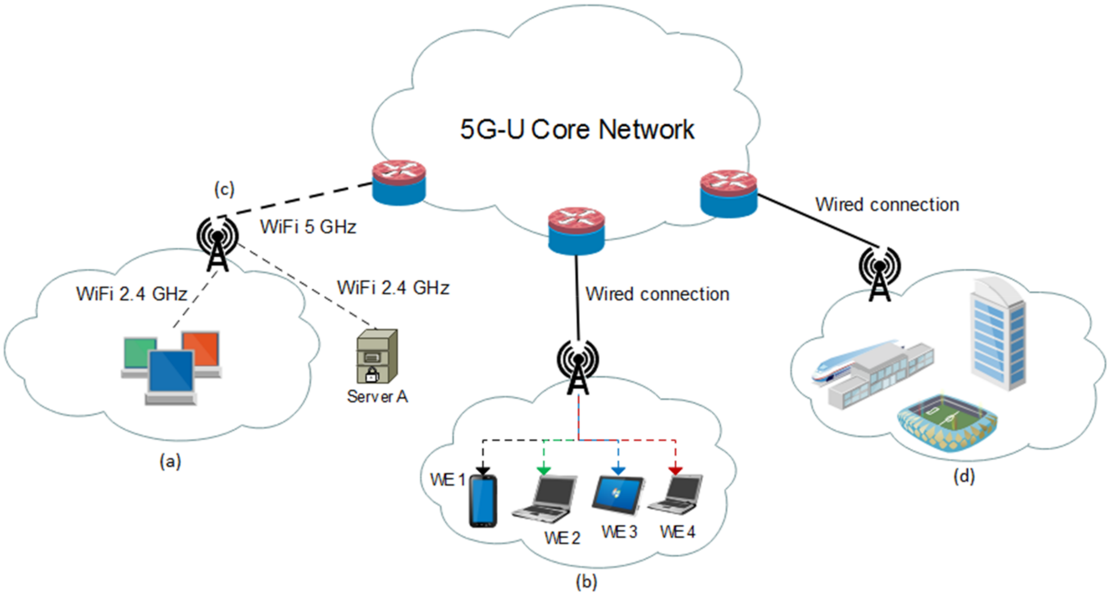


Figure 1. Possible high-efficiency wireless local area network (HEW) use cases in fifth-generation radio access network on unlicensed bands (5G-U) deployments; (**a**) gigabit ethernet connection replacement; (**b**) improved network capacity using multi-user multiple-input and multiple-output (MU-MIMO); (**c**) HEW as a backhaul for local area network (LAN); (**d**) support for highly dense scenarios.

2.1. Gigabit Ethernet Connection Replacement

The most upfront use case of the latest gigabit Wi-Fi amendment IEEE 802.11ax HEW is the opportunity to replace ancient gigabit ethernet connections, such as a connection to a server (e.g., server A in Figure 1a), with HEW radio links. With this use case, it is likely to serve either more WEs (servers) with the same throughput as before or the same number of WEs with improved throughput. The former benefit is particularly significant from the perspective of densely deployed networks, and the latter from the perspective of backhaul radio links supporting dense networks.

2.2. Improved Network Capacity Using Multi-User Multiple-Input and Multiple-Output (MU-MIMO)

HEW-enabled base stations (BSs) typically have more radio antennas than WEs. Therefore, a BS can utilize downlink MU-MIMO (the 802.11ax working group is also proposing uplink MIMO [3]), which allows a single BS to transmit parallel beam-formed transmission streams to different WEs, such as WE 1, WE 2, WE 3 and WE 4 in Figure 1b, on the same frequency. Beam-forming has previously been used in single-user (SU) transmissions to achieve higher data rates, and can now be used to increase overall WLAN network capacity. As a result, a WE equipped with a smaller number of antennas (that is, only one) does not affect network performance by occupying the whole radio channel with its lower transmission rate. Moreover, since MU-MIMO transmissions are realized in parallel and possibly with different transmission rates, the distance from BS is less important than in legacy SU networks in which WEs on the edge of the coverage cell (using low rates) could severely affect the performance of others by deferring their transmissions.

2.3. High-Efficiency Wireless (HEW) as a Backhaul for Local Area Network (LAN)

Since the IEEE 802.11ax HEW amendment is a successor to IEEE 802.11ac very high throughput (VHT), Wi-Fi radio links can effectively replace wired backhaul connectivity, especially for those deployment infrastructures where wired connection deployment is a challenging task [7]. HEW, as a backhaul connection as shown in Figure 1c, is surely the most straightforward application of this concept. However, an emerging idea is to use directed beam-forming (point-to-point) backhaul links using IEEE 802.11ax for small cells to reduce the backhaul cost. While operating in an unlicensed radio spectrum is subject to interference from other unlicensed networks or WEs, interference can be avoided by using highly directional antennas. With the emergence of new MIMO directional antennas, point-to-point links could profit from 802.11ax beam-forming transmissions.

2.4. Support for Highly Dense Scenarios

In most of the WLAN configurations, each BS serves only a limited number of WEs. Traditionally, to analyze the best-case scenario, only a single WE per BS is assumed. The worst-case scenario arises when multiple Wi-Fi networks are deployed in the same area as densely deployed networks. Examples of such scenarios are an office floor/building, train stations, or stadiums (as shown in Figure 1d), in which each end WE connects to its own BS but the maximum system throughput is limited due to interference coming from neighboring devices (BSs and WEs). If more WEs were served by each BS or more Wi-Fi networks were present in the same area, the results would be worse due to higher overheads and increased interference. HEW expects to handle the interference due to highly dense network deployments using more intelligent and optimized MAC-RA schemes, such as dynamic sensitivity thresholds for BS and WE traffic differentiation, and adaptive transmit power [5].

3. A Self-Scrutinized Backoff Mechanism for HEW

3.1. Problem Statement

All of the use cases described above indicate that the performance of a Wi-Fi system can be severely degraded with an increase in the number of contenders, as the collision in the network is directly proportional to the density of the network. This problem statement is assured by the simulation results shown in Figure 2. Figure 2 plots the number of WEs (n) contending for channel access versus the average channel collision probability (p_{obs}) in a saturated (always willing to transmit) network environment with $W_{min} = 32$ and $W_{min} = 64$. The other simulation parameters are described in Table 1. The figure shows that increased network density has a direct relationship with the average channel collision probability; the denser the network, the higher the channel collision probability. In such a troublesome situation, a more adaptive and self-scrutinized MAC-RA is required by the HEW networks to maintain the performance so that it can serve 5G-U.

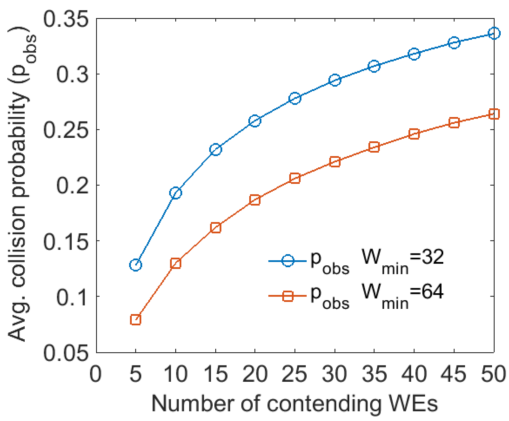


Figure 2. Number of contending Wi-Fi user equipments (WEs) vs. p_{obs} for $W_{min} = 32$, and $W_{min} = 64$.

Table 1. Medium access control (MAC) layer parameters used in simulation and analysis.

Parameter	Value	Parameter	Value
Operating Frequency	5 GHz	Physical rate of channel	54 Mbps
Bandwidth	20 MHz	MAC header	24 Bytes
MAC payload	1024 Bytes	PHY header	20 µs
Acknowledgment (ACK)	14 Bytes + PHY header	Transmission range	10 meters
W_{min}	32	W_{max}	1024
ω	32	σ	9 µs
SIFS	16 µs	DIFS	60 µs
Propagation delay (δ)	1 µs	Max. backoff stages (m)	6

3.2. Channel Observation-Based Scaled Backoff (COSB)

In the proposed COSB protocol, after the communication medium has been idle for a distributed inter-frame space (DIFS), all the WEs competing for a channel proceed to the backoff procedure by selecting a random backoff value B as shown in Figure 3. The time immediately following an idle DIFS is slotted into observation time slots (α). The duration of α is either a constant slot time σ during an idle period or a variable busy (successful or collided transmission) period. While the channel is sensed to be idle during σ, B decrements by one. A data frame is transmitted after B reaches zero. In addition, if the medium is sensed to be busy, the WE freezes B and continues sensing the channel. If the channel is again sensed to be idle for DIFS, B is resumed. Each individual WE can proficiently measure channel observation-based conditional collision probability p_{obs}, which is defined as the probability that a data frame transmitted by a tagged WE fails. We discretize the time in B_{obs} observation time slots, where the value of B_{obs} is the total number of α observation slots between two consecutive backoff stages as shown in Figure 3. A tagged WE updates p_{obs} from B_{obs} of backoff stage b_i at the i^{th} transmission as,

$$p_{obs} = \frac{1}{B_{obs}} \times \sum_{k=0}^{B_{obs}-1} S_k, \tag{1}$$

where for an observation time slot k, $S_k = 0$ if α is empty (idle) or the tagged WE transmits successfully, while $S_k = 1$ if α is busy or the tagged WE experiences collision as shown in Figure 3. In the figure, WE 1 randomly selects its backoff value $B = 9$ for its b_i backoff stage. Since WE 1 observes nine idle slot times, two busy periods, and one collision ($B_{obs} = 9 + 2 + 1 = 12$), p_{obs} is updated as $\frac{2+1}{B_{obs}} = \frac{3}{12} = 0.25$ in the next backoff stage b_{i+1}.

According to the channel observation-based conditional collision probability p_{obs}, the adaptively scaled contention window value is $W_{b_{i+1}}$ at backoff stage b_{i+1} of the transmission time $i+1$, where $b_{i+1} \in (0, m)$ for the maximum m number of backoff stages, and i is the discretized time for the data frame transmissions of a tagged WE. More specifically, when a transmitted data frame has collided, the current contention window W_{b_i} of backoff stage b_i at the i^{th} transmission time slot is scaled-up according to the observed p_{obs} at the i^{th} transmission, and when a data frame is transmitted successfully, the current contention window W_{b_i} is scaled-down according to the observed p_{obs} at the i^{th} transmission. Unlike the BEB (where backoff stage is incremented for each retransmission and resets to zero for new transmission as shown in Figure 4a), the backoff stage b_i in COSB at the i^{th} transmission has the following property of increment or decrement:

$$b_{i+1} = \begin{cases} min[b_i + 1,\ m], collision\ at\ i^{th}\ transmit \\ max[b_i - 1,\ 0],\ success\ at\ i^{th}\ transmit \end{cases}. \tag{2}$$

Figure 4b shows that the backoff stage does not reset after a successful transmission. Since the current backoff stage represents the number of collisions or successful transmissions of a tagged WE, it helps to scale the size of W efficiently. The incremented or decremented backoff stage b_i results in scaling-up or scaling-down of the current contention window, respectively. The scaling-up and scaling-down of the contention window operates as follows:

$$W_{b_{i+1}} = \begin{cases} min\left[2^{b_{i+1}} \times W_{min} \times \omega^{p_{obs}},\ W_{max}\right],\ collision\ at\ i^{th}\ transmit \\ max\left[2^{b_{i+1}} \times W_{min} \times \omega^{p_{obs}},\ W_{min}\right],\ success\ at\ i^{th}\ transmit \end{cases}, \quad (3)$$

where ω is a constant design parameter to control the optimal size of the contention window and is expressed as $\omega = W_{min}$.

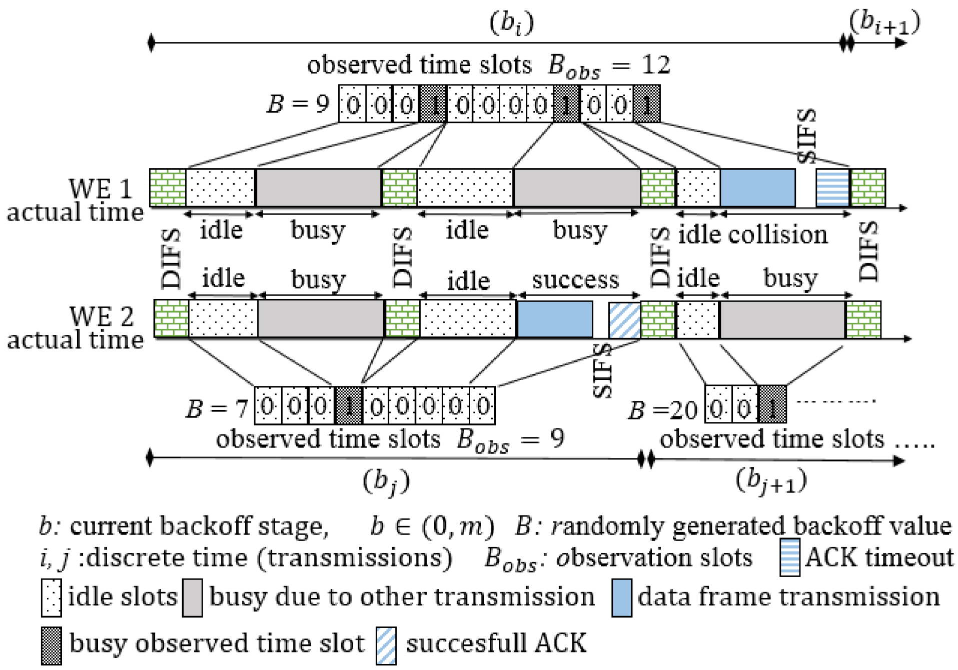


Figure 3. Channel observation mechanism of channel observation-based scaled backoff (COSB) during the backoff procedure.

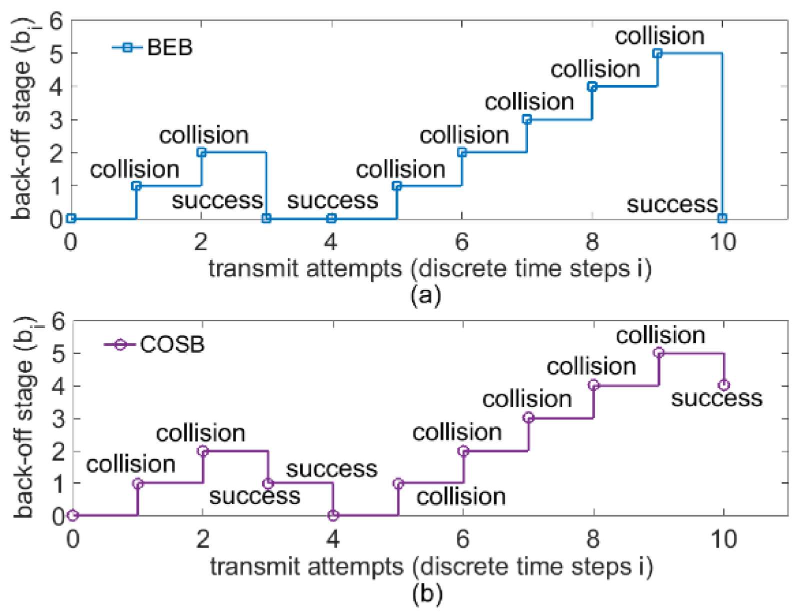


Figure 4. Backoff stage after collision/successful transmission; (**a**) backoff stage increment/reset in binary exponential backoff (BEB); and (**b**) backoff stage increment/decrement in COSB.

4. Analytical Model

We formulate the analytical evaluation of proposed COSB mechanism with saturation throughput and average delay, on the assumption of ideal channel conditions, i.e. no hidden terminal and capture effects. In the analysis, we assume the fixed number of WEs, each of which is always willing to transmit the data frame, i.e., the network is assumed as a saturated traffic environment. Initially, we study the behavior of a tagged WE with a discrete-time Markov chain model (DTMC) [8,9], and we obtain the stationary transmission probability γ for the tagged WE. Since the proposed COSB does not reset the backoff stage to its initial value (that is to zero) after successful transmission, the transmission attempt for every new data frame remains recursive within the backoff stage state dimension. To accurately analyze the performance of COSB, we formulate a recursive discrete-time Markov chain model (R-DTMC). Later, by knowing the exact events that can occur on the communication channel within a randomly selected slot-time, we formulate the normalized throughput and average delay of the proposed COSB mechanism.

4.1. Recursive Discrete-Time Markove Chain (R-DTMC) Model

Consider there are n number of WEs competing for the channel in a WLAN. In the saturated condition, each WE has immediately a data frame available for transmission after each successful transmission. Thus, due to the consecutive data frame transmission, each data frame needs to wait for a random backoff time before transmitting.

Let b be the backoff stage counter for a tagged WE and m be the maximum number of backoff stages b can experience for a data frame, that is $b \in (0, m)$, such that $W_b = 2^b \times W_{min} \times \omega^{p_{obs}}$ for b^{th} backoff stage and $W_{max} = 2^m \times W_{min} \times \omega^{p_{obs}}$ for the m^{th} backoff stage contention window, where W_b is the contention window size at b^{th} backoff stage and p_{obs} is the observed channel collision probability. Let us adopt the notation $W_{b+1} = 2^{b+1} \times W_{min} \times \omega^{p_{obs}}$, for the adaptively scaled-up contention window for $b+1$ backoff stage, when transmission is failed at the b^{th} backoff stage. Similarly, let $W_{b-1} = 2^{b-1} \times W_{min} \times \omega^{p_{obs}}$ be the adaptively scaled-down contention window for the $b-1$ backoff stage, when successfully transmitted at the b^{th} backoff stage.

Assume $\Omega(t)$ is the function for the stochastic process representing the backoff counter u for a tagged WE, where $u \in (0, W_{cur} - 1)$. Since time is discretized as an integer time scale, t and $t+1$ correspond to the beginning of two consecutive transmission time slots, and the backoff time counter of each WE decreases at the beginning of each slot time. Figure 3 illustrates that the backoff time decreases when the communication channel is sensed as idle (σ), and it stops when the channel is sensed as busy, which may be due to a successful or unsuccessful transmission of any other WE. Therefore, the time interval between two consecutive slot time beginnings may be much longer and different from the idle slot time size, i.e., σ. Let $\pi(t)$ be the stochastic process representing the backoff stages $(0,\ 1,\ 2,\ \ldots,\ m)$ of the tagged WE at time t. The key articulation in our R-DTMC model is that, at each data frame transmission attempt regardless of the number of retransmission attempts, each data frame collides with a practically observed and independent collision probability p_{obs}. With these assumptions, COSB can be modeled as the two dimensional process $\{\pi(t),\ \Omega(t)\}$ with the R-DTMC as depicted in Figure 5. In this R-DTMC, the transition probabilities are described as follows.

The tagged WE remains at the first backoff stage after a successful transmission on the first backoff stage with the probability,

$$P\{(0,u)|(0,0)\} = (1 - p_{obs})/W_0,\ u \in (0, W_0 - 1) \tag{4}$$

The backoff counter decreases when the channel is sensed as idle with the probability,

$$P\{(b,u)|(b,u+1)\} = 1,\ b \in (0,m), u \in (0, W_b - 2) \tag{5}$$

The tagged WE scales-up the current contention window and moves to the next stage b if a data frame transmission failed on backoff stage $b-1$ with the probability,

$$P\{(b,u)|(b-1,0)\} = p_{obs}/W_b,\ b \in (0,m), u \in (0, W_b - 1) \tag{6}$$

The tagged WE scales-down the current contention window and decreases its backoff stage for the next transmission attempt to $b-1$ after a successful transmission on backoff stage b with the probability,

$$P\{(b-1,u)|(b,0)\} = (1-p_{obs})/W_{b-1},\ b \in (0,m), u \in (0, W_b - 1) \tag{7}$$

The tagged WE remains at the m^{th} backoff stage after an unsuccessful transmission with the probability,

$$P\{(m,u)|(m,0)\} = p_{obs}/W_m,\ u \in (0, W_m - 1) \tag{8}$$

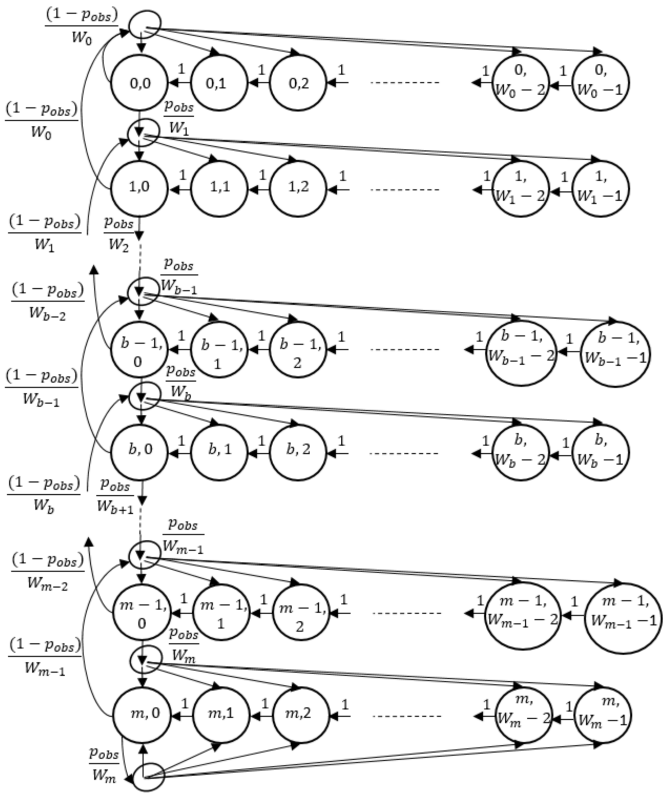


Figure 5. Recursive discrete-time Markov chain model (R-DTMC) for COSB mechanism.

In particular, to the above transition probabilities, as considered in Equation (6), when a data frame transmission is collided at backoff stage $b-1$, the backoff stage increases to b, and the new backoff value is uniformly chosen from the adaptively scaled-up contention window W_b. On the other hand, Equation (7) describes how when a data frame transmission is successful at backoff stage b, the backoff stage decreases to $b-1$, and the new backoff value is uniformly chosen from the adaptively scaled-down contention window W_{b-1}. In case the backoff stage reaches the value m (that is the maximum backoff value), it is not increased in the subsequent data frame transmission attempt.

Let us assume that $d_{b,\ u} = \lim_{t\to\infty} P\{\pi(t) = b,\ \Omega(t) = u\}, b \in (0, m),\ u \in (0, W_b - 1)$ be the stationary distribution of the R-DTMC. From Figure 5, each state transition probability can be written as,

$$d_{1,0} = \frac{p_{obs}}{1 - p_{obs}} d_{0,0}. \tag{9}$$

If $\beta = \frac{p_{obs}}{1-p_{obs}}$, the above equation can be written as, $d_{1,0} = \beta \times d_{0,0}$. Similarly, $d_{b,0} = \beta \times d_{b-1,0}$ where $d_{b-1,0} = \beta \times d_{b-2,0}$ till $d_{1,0} = \beta \times d_{0,0}$. Therefore, we can write:

$$d_{b,0} = \beta^b \times d_{0,0},\ 0 < b < m. \tag{10}$$

Now for the backoff stage m, the $d_{m,0}$ can be written as,

$$d_{m,0} = \beta^m \times d_{0,0} \tag{11}$$

Owing to the Markov process-based chain regularities, for each $u \in (1, W_b - 1)$, the stationary distribution for $\{\pi(t),\ \Omega(t)\}$ can be written as,

$$d_{b,u} = \frac{W_b - u}{W_b} \times \left\{ \begin{array}{c} (1 - p_{obs}) \times d_{b+1,0}\ b = 0 \\ p_{obs} \times d_{b-1,0} + (1 - p_{obs}) \times d_{b+1,0}\ 0 < b < m \\ p_{obs} \times (d_{m-1,0} + d_{m,0})\ b = m \end{array} \right. \tag{12}$$

The recursive characteristic of state transition probabilities can be combined as,

$$\sum_{b=0}^{m-1} (b+1) \times d_{b,0} + m \times d_{m,0}\ = d_{0,0} \left(\frac{1 - \beta^m - m\beta^m(1-\beta)}{(1-\beta)^2} \right) \tag{13}$$

From Equations (9)–(11) and (13), Equation (12) can be re-written as,

$$d_{b,u} = \frac{(W_b - u)(b+1)}{W_b}\ d_{b,0},\ b \in (0, m), u \in (0, W_b - 1) \tag{14}$$

From Equations (9)–(11) and (14), all the values $d_{b,u}$ are expressed as a function of $d_{0,0}$ and channel observation-based practical conditional collision probability p_{obs}. $d_{0,0}$ is finally determined by normalizing the R-DTMC states as follows,

$$1 = \sum_{b=0}^{m-1} (b+1) \times d_{b,0} \sum_{u=0}^{W_b - 1} \frac{W_b - u}{W_b} + \frac{m \times (W_m - u)}{W_m}\ d_{m,0}, \tag{15}$$

From $W^* = W_{min} \times \omega^{p_{obs}}$ and few mathematical steps, the above normalization relation can be written as,

$$1 = \frac{d_{0,0}}{2} \left[W^* \left(\frac{1 - (2\beta)^m - m(2\beta)^m(1 - 2\beta)}{(1 - 2\beta)^2} \right) + \left(\frac{1 - \beta^m - m\beta^m(1-\beta)}{(1-\beta)^2} \right) \right] \tag{16}$$

Finally, we get $d_{0,0}$ as follows,

$$d_{0,0} = \frac{2(1-\beta)^2(1-\ \beta^m)}{(1 - \beta^m - m\beta^m(1-\beta))} \times \frac{1}{((1-2\beta)(1-\ \beta^m)(W^*+1) + \beta W^*(1-\beta)(1-\ (2\beta)^m))}. \tag{17}$$

Since, a transmission occurs only when the backoff counter of the WE reaches zero regardless of the backoff stage, transmission probability γ can be expressed as follows,

$$\gamma = \sum_{b=0}^{m-1}(b+1)d_{b,0} + md_{m,0} = d_{0,0}\left(\frac{1-\beta^m - m\beta^m(1-\beta)}{(1-\beta)^2}\right). \tag{18}$$

Furthermore, after performing a few mathematical steps to Equation (18) using the value of $d_{0,0}$ from Equation (17) we get,

$$\gamma = \frac{2}{\left(W^* + \beta W^*\left(\frac{\sum_{b=0}^{m-1}(2\beta)^b}{\sum_{b=0}^{m-1}(\beta)^b}\right) + 1\right)}. \tag{19}$$

However, in general γ depends on the practical collision probability p_{obs}, which is always unknown until the channel is observed for the busy slots. A transmitted data frame encounters the collision if at least one of the $n-1$ remaining WEs transmit. Since each of the transmissions in the system sees this collision in the same state, a steady state can easily be yielded as [9],

$$p_{obs} = 1 - (1-\gamma)^{n-1} \tag{20}$$

These two (γ and p_{obs}) are monotonic non-linear systems which can be numerically solved for each other.

4.2. Normalized Throughput

Let θ be the normalized throughput of the network and be defined as the fraction of the communication channel used for successful transmission of the data payload. To compute θ, let γ_{tr} be the probability that there is at least one transmission in the considered slot time. Since there are n number of WEs in the system contending for the medium and each transmits with probability γ, the transmission probability γ_{tr} can be defined as,

$$\gamma_{tr} = 1 - (1-\gamma)^n. \tag{21}$$

If the probability γ_s that a transmission is successful is given by the probability that only one WE transmits in the considered slot time, γ_s can be obtained as,

$$\gamma_s = \frac{n\gamma(1-\gamma)^{n-1}}{\gamma_{tr}} = \frac{n\gamma(1-\gamma)^{n-1}}{1-(1-\gamma)^n}. \tag{22}$$

Thus, θ can be expressed as the ratio,

$$\theta = \frac{E[mean\ payload\ transmitted\ in\ a\ slot\ time]}{E[total\ lenght\ of\ a\ slot\ time]} \tag{23}$$

Assume $E[P]$ is the average data frame payload size (assuming that all the data frames have the same fixed size), then the slot time for transmitting average payload data successfully can be obtained as $\gamma_{tr}\gamma_s E[P]$, since $\gamma_{tr}\gamma_s$ is the probability for the successful transmission of a data frame in a given slot time. The average length of a given slot time is the sum of three cases; no transmission in a slot time that is $(1-\gamma_{tr})\sigma$, a successfully transmitted data frame that is $\gamma_{tr}\gamma_s$, and a collision that is $\gamma_{tr}(1-\gamma_s)$. Finally, the relation (23) can be written as follows:

$$\theta = \frac{\gamma_{tr}\gamma_s E[P]}{(1-\gamma_{tr})\cdot\sigma + \gamma_{tr}\gamma_s\cdot T_s + \gamma_{tr}(1-\gamma_s)\cdot T_c} \tag{24}$$

where T_s and T_c are the average time the communication channel has been busy due to successful transmission and collision, respectively. For analytical evaluation, the values of $E[P]$, T_s, T_c, and idle slot time σ must be expressed with the same time unit. Let $P_{hdr} = PHY_{hdr} + MAC_{hdr}$ be the time to transmit a data frame header, and δ be the channel propagation delay. If ACK is the time to receive an acknowledgement, T_s and T_c can be obtained as,

$$T_s = P_{hdr} + E[P] + SIFS + \delta + ACK + DIFS + \delta, \tag{25}$$

$$T_c = P_{hdr} + E[P] + DIFS + \delta. \tag{26}$$

The corresponding values for T_s and T_c depend upon the 802.11 standard. The PHY and MAC layer parameters to compute T_s and T_c are shown in Table 1.

4.3. Average Delay

We derive the average delay $E[D]$ for a COSB mechanism for a successfully transmitted data frame. The saturation average delay is defined as the average time between the time data frame at the head of its MAC queue ready for transmission and its successful reception at the destination.

According to [10],

$$E[D] = E[B] \times \ E[Slot] \tag{27}$$

where $E[Slot]$ is the total length of slot time as given in Equation (24), that is,

$$E[Slot] = (1 - \ \gamma_{tr})\sigma + \gamma_{tr}\gamma_s T_s + \gamma_{tr}(1 - \gamma_s)T_c. \tag{28}$$

$E[B]$ is the average number of backoff slot times for a successful data frame transmission and is given by,

$$E[B] = \sum\nolimits_{b=0}^{m-1} (b+1)\beta^b \times \frac{W_b + 1}{2} + m\beta^m \times \frac{W_m + 1}{2}, b \in (0, m) \tag{29}$$

After some algebraic calculations, $E[B]$ reduces to,

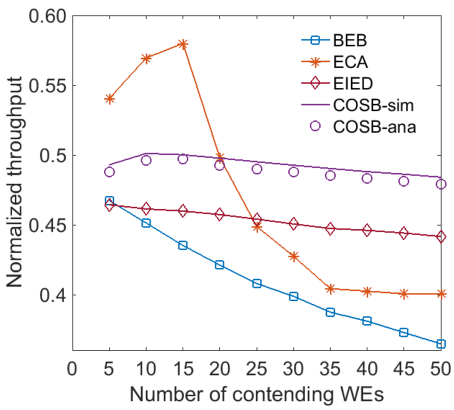
$$E[B] = \frac{(W^*(1-\beta^m) + \beta W^*(1-(2\beta)^m) + 1)}{2(1-2\beta)(1-\beta)(1-\beta^m)} \tag{30}$$

5. Performance Evaluation

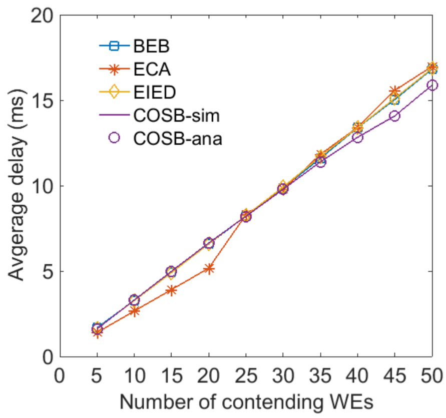
Analytical results of the proposed COSB formulated from R-DTMC are validated with that obtained from the simulation for the proposed COSB in an event-driven simulator, network simulator-3 (NS-3) version 3.24 [11]. To evaluate the performance of COSB, we compare simulation results with BEB, and two of the related contention window-scaling algorithms; enhanced collision avoidance (ECA) mechanism [12], and exponential increase-exponential decrease (EIED) backoff algorithm [13]. These comparison protocols (ECA and EIED) are selected for their characteristic of not resetting the value of W to its minimum value W_{min}. ECA uses a deterministic backoff value $B = W_{min}/2$ instead of resetting W to W_{min} after successful transmission. The W value is exponentially increased after each unsuccessful transmission and is halved after each successful transmission in the EIED mechanism. The limitations of these protocols is that the performance of ECA is limited to the number of contenders below the deterministic cycle length $W_{min}/2$, while EIED increases/decreases the contention window value without knowing the channel collision probability. The analytical model is validated with a network of n WEs ranges from 5 to 50 (a typical office floor deployment of IEEE 802.11ax standard [3]), where each WE is within the coverage area of the others (no hidden terminal). The WEs are set to be in saturation state (always willing to transmit). The specific MAC and PHY layer parameters are listed in Table 1.

5.1. Normalized Throughput and Average Delay

Figure 6a describes the normalized throughput for various numbers of WEs in an indoor HEW system. A monotonic decrease can be observed for the performance of BEB and EIED. The reason for performance degradation of BEB is resetting its contention window value to minimum value after successful transmission, which causes higher collision with the increases in the number of WEs. In spite of increased throughput, the performance of EIED also degrades with the increase of WEs due to a blind decrease of the contention window. In the figure, ECA performs better until $n < 15$, where the number of contenders is less than the deterministic cycle length $W_{min}/2$ due to the collision-free deterministic environment. In the beginning, COSB also has a curved performance (increase and decrease) in normalized throughput. The curved performance of COSB shows that after observing the channel, it adjusts the contention window adaptively and results in increased normalized throughput. The performance degradation of COSB is slower than the compared protocols when the number of the WE increases. However, COSB provides the best throughput (Figure 6a) and average delay (Figure 6b) in a high density network as compared to the BEB, ECA and EIED. ECA has the best performance at low density, but the throughput decreases drastically in high density. COSB has acceptable performance at low density. This performance enhancement of COSB comes from the adaptive channel observation-based scaling of W. Figure 6 shows that the analytical model is accurate because analytical results (COSB-ana) match with the simulation results (COSB-sim) in both normalized throughput (Figure 6a) and average delay (Figure 6b).



(**a**)



(**b**)

Figure 6. (**a**) Normalized throughput and (**b**) average delay (ms).

5.2. Maximum Approximate Saturation Throughput

Bianchi in [9] determined the maximum achievable throughput of a WLAN by formulating approximate solutions for optimal transmission probability γ_{opt} as follows,

$$\gamma_{opt} = \frac{1}{n\sqrt{\frac{T_c}{2\sigma}}}. \tag{31}$$

As described in Section 4.2, T_c is the average time during which the communication channel has been busy due to the collision, and the value T_c depends upon the IEEE 802.11 standard under consideration. Equation (31) has a fundamental theoretical importance to approximate the maximum saturation throughput of a DCF network, which mainly depends on the density of the network (that is, size of n). Thus γ_{opt} is the transmission probability that each WE should adopt in order to achieve the maximum throughput performance. Table 2 shows the maximum approximate throughput achieved theoretically compared with that achieved by BEB and COSB algorithms in DCF for different network sizes. The table shows that the maximum approximate saturation throughput is very smooth, even a very small difference in the estimate of the γ_{opt} leads to similar throughput values. The interesting result is that with the increase of density of the network, throughput achieved by the proposed COSB is closer to the maximum approximate saturation throughput as compared to the BEB. Moreover, the maximum approximated saturation throughput is practically independent of the number of contending WEs in the WLAN.

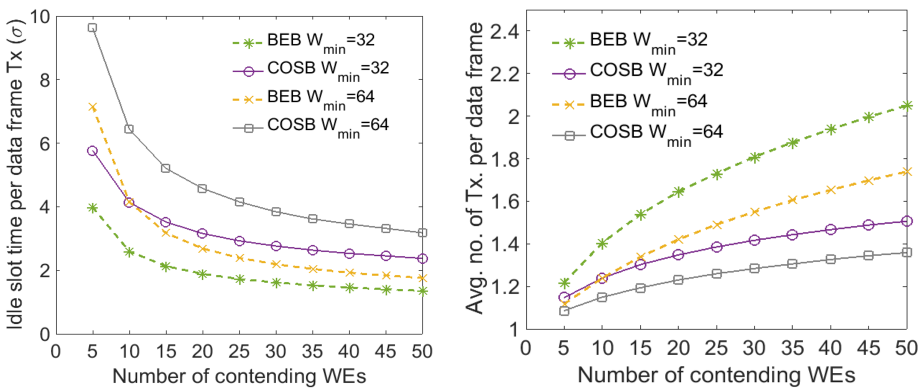
Table 2. Comparison of maximum approximate saturation throughput achieved from γ_{opt} with throughputs of COSB and BEB.

WEs (n)	Max. Approx. Throughput	COSB	BEB
5	0.510 (γ_{opt} = 0.0573)	0.493 (γ = 0.034)	0.468 (γ = 0.048)
10	0.502 (γ_{opt} = 0.0287)	0.501 (γ = 0.024)	0.452 (γ = 0.037)
20	0.498 (γ_{opt} = 0.0143)	0.498 (γ = 0.016)	0.421 (γ = 0.026)
30	0.497 (γ_{opt} = 0.0095)	0.493 (γ = 0.012)	0.401 (γ = 0.020)
40	0.497 (γ_{opt} = 0.0072)	0.488 (γ = 0.010)	0.381 (γ = 0.017)
50	0.497 (γ_{opt} = 0.0057)	0.484 (γ = 0.008)	0.365 (γ = 0.015)

5.3. Average Channel Utilization Per Data Frame Transmission

In order to have a successful data frame transmission, a tagged WE spends $1/(\gamma_{tr}\gamma_s)$ average number of slot times on the communication channel. Of those average slot times spent, $(1 - \gamma_{tr})$ is the time when channel is observed as idle and each idle slot time lasts for σ. Figure 7a plots the average number of idle slot times per data frame transmission for two different W_{min} values with varying network density. The results show that for $W_{min} = 32$ and $W_{min} = 64$, the idle slot time per data frame transmission is very low for both COSB and BEB. This idle channel utilization is more significant when the initial contention window is greater, as shown for $W_{min} = 64$ in Figure 7a. More specifically, for BEB mechanisms, the idle slot time per data frame transmission drastically decreases for a dense network environment even for a large number of initial contention windows that is $W_{min} = 64$. COSB does not reset its initial contention window to W_{min} after a successful transmission and it recursively remains near the adaptive values of the contention window size according to the channel observation-based p_{obs}.

Another important measure to discuss is the average number of transmissions that a tagged WE must perform for a successful transmission of a data frame which is given by $1/(1 - p_{obs})$. Figure 7b shows that the number of transmission attempts per data frame transmission considerably increases as the size of the initial contention window decreases, that is $W_{min} = 32$, especially for BEB when the network size is n = 50. The figure shows that for $W_{min} = 32$, BEB WEs suffer an average 2.1 retransmissions (collisions), and COSB WEs suffer an average 1.5 retransmissions (collisions) when network density is n = 50. The increased idle slot times (Figure 7a) and reduced number of retransmissions (Figure 7b) for dense networks are the reason for efficient saturation throughput achievement of COSB.



(**a**) (**b**)

Figure 7. (**a**) Average number of idle slot times per successful data frame transmission; (**b**) average number of transmissions per data frame.

6. Conclusions

In future technologies, the IEEE 802.11ax HEW hopes to become part of the 5G-U networks, as they promise four times higher network efficiency even in highly dense network deployments. However, currently WLAN itself faces huge challenge of efficient channel access due to its distributed contention-based nature. Currently, CSMA/CA is based on a BEB mechanism, which blindly increases and decreases the contention window for collided and successful transmissions, respectively. In this paper, we highlight some of the use-case scenarios for HEW deployments in 5G-U networks. Furthermore, to handle the performance degradation challenge caused by increasing density of the WLANs in those use cases, a channel observation-based scaled backoff (COSB) mechanism based on practical channel collision probability is proposed. COSB overcomes the limitation of BEB to achieve high efficiency and robustness in highly dense networks. A practical channel collision probability observed by the contending Wi-Fi equipment (WE) is presented to adaptively scale-up and scale-down the size of W during the backoff mechanism for collided and successfully transmitted data frames, respectively. The proposed COSB enhances the performance of CSMA/CA in dense networks. Furthermore, a recursive discrete-time Markov chain model (R-DTMC) is also presented to analyze the performance of COSB. The validated model helps to determine the maximum achievable throughput in highly dense networks. The analytical model helps to investigate the interesting importance of the practical channel observation-based collision probability. The analytical and simulation results show that the proposed protocol compared to the state-of-the-art BEB protocol offers a performance boost in terms of throughput and delay when the number of contending WEs increases. Furthermore, this protocol is designed with very few modifications to the existing BEB mechanism, which makes the COSB protocol a good candidate for upcoming high-efficiency WLANs (HEW) that are the IEEE 802.11ax standard.

Future research considerations include the implementation of the COSB mechanism for long term evolution (LTE)-based 5G-U networks, known as licensed assisted access (LAA). The addictiveness and practical collision probability measurement in COSB motivate us to integrate the algorithm in LAA.

Acknowledgments: This work was supported in part by a 2018 Yeungnam University Research Grant and in part by a National Research Foundation of Korea (NRF) grant funded by the Korea government (No. 2018R1A2B6002399).

Author Contributions: Rashid Ali conceived and designed the algorithm. Rashid Ali and Nurullah Shahin designed and performed the experiments. Rashid Ali and Rojeena Bajracharya wrote the paper. Rashid Ali, Byung-Seo Kim and Sung Won Kim formulated the mathematical model. Byung-Seo Kim and Sung Won Kim supervised and finalized the manuscript for submission.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Al-Dulaimi, A.; Al-Rubaye, S.; Ni, Q.; Sousa, E. 5G Communications Race: Pursuit of More Capacity Triggers LTE in Unlicensed Band. *IEEE Veh. Technol. Mag.* **2015**, *10*, 43–51. [CrossRef]
2. Bajracharya, R.; Shrestha, R.; Zikria, Y.B.; Kim, S.W. LTE in the unlicensed spectrum: A survey. *IETE Tech. Rev.* **2018**, *35*, 78–90. [CrossRef]
3. Bellalta, B. IEEE 802.11ax: High-efficiency WLANs. *IEEE Wireless Communications.* **2016**, *23*, 38–46. [CrossRef]
4. Ali, R.; Shahin, N.; Kim, Y.T.; Kim, B.; Kim, S.W. Channel Observation-based Scaled Backoff Mechanism for High Efficiency WLANs. *Electron. Lett.* **2018**. [CrossRef]
5. Ali, R.; Kim, S.W.; Kim, B.; Park, Y. Design of MAC Layer Resource Allocation Schemes for IEEE 802.11ax: Future Directions. *IETE Tech. Rev.* **2018**, *35*, 28–52. [CrossRef]
6. IEEE Standard for Information Technology. Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. *ANSI/IEEE Std 802.11* **2007**, i-513. [CrossRef]
7. Kosek-Szott, K.; Gozdecki, J.; Loziak, K.; Natkaniec, M.; Prasnal, L.; Szott, S.; Wagrowski, M. Coexistence Issues in Future WiFi Networks. *IEEE Netw.* **2017**, *31*, 86–95. [CrossRef]

8. Ali, R.; Zikria, Y.B.; Amin, F.; Kim, B.; Kim, S.W. I-DTMC: An Integrated-Discrete Time Markov Chain Model for Performance Analysis in Future WLANs. In Proceedings of the 2017 IEEE 42nd Conference on Local Computer Networks Workshops, Singapore, 9–12 October 2017.
9. Bianchi, G. Performance analysis of the IEEE 802.11 distributed coordination function. *IEEE J. Sel. Areas Commun.* **2000**, *18*, 535–547. [CrossRef]
10. Chatzimisios, P.; Boucouvalas, A.C.; Vitsas, V. Packet delay analysis of IEEE 802.11 MAC protocol. *Electron. Lett.* **2003**, *39*, 1358–1359. [CrossRef]
11. The Network Simulator-ns-3. Available online: https://www.nsnam.org/ (accessed on 1 March 2018).
12. Sanabria-Russo, L.; Faridi, A.; Bellalta, B.; Barcelo, J.; Oliver, M. Future Evolution of CSMA Protocols for the IEEE 802.11 Standard. In Proceedings of the 2013 IEEE International Conference on Communications Workshops, Budapest, Hungary, 9–13 June 2013.
13. Ye, C.; Li, Y.; Reznik, A. Performance Analysis of Exponential Increase Exponential Decrease Back-Off Algorithm. In Proceedings of the 2010 IEEE Global Telecommunications Conference (GLOBECOM), Miami, FL, USA, 6–10 December 2010.