Harmonious Cross-Technology Coexistence with Heterogeneous Traffic in Unlicensed Bands: Analysis and Approximations

Mohammed Hirzallah, *Student Member, IEEE,* Marwan Krunz, *Fellow, IEEE,* and Yong Xiao, *Senior Member, IEEE,*

Abstract—The dramatic growth in demand for mobile data has prompted mobile network operators (MNOs) to explore spectrum sharing in unlicensed bands. MNOs have been allowed recently to operate their LTE services over the 5 GHz Unlicensed National Information Infrastructure (U-NII) bands, currently occupied by Wi-Fi. The unlicensed LTE operation has been standardized by 3GPP under the name Licensed Assisted Access (LAA). Unlicensed 5G New radio (NR) operation over the U-NII bands, a.k.a., NR-Unlicensed (NR-U), is also being explored. To support applications with diverse quality of service requirements, LAA, NR-U, and Wi-Fi technologies offer multiple priority classes with different contention parameters for accessing an unlicensed channel. How these different classes affect the interplay between coexisting MNOs and Wi-Fi systems is still a relatively under-explored topic. In this paper, we develop a simple yet efficient framework for fair coexistence between LTE MNOs and Wi-Fi systems, each with multiple priority classes. We derive approximate closed-form solutions for the probability of successful transmission (PST), average contention delay, and average throughput under different LAA and Wi-Fi priority classes. LTE and Wi-Fi operators can fit these solutions to offline and/or online measurements, and use them to further optimize their system throughput and latency. Our results reveal that PSTs computed with our approximate models are within 5% of these obtained via simulation under dense network deployments and high traffic loads.

Index Terms—Cross-technology coexistence, licensed assisted access (LAA), 5G New Radio Unlicensed (NR-U), IEEE 802.11, Wi-Fi, Markov analysis, network modeling and simulation.

I. Introduction

To cope with the growing demand for mobile data, mobile network operators (MNOs) have taken steps to secure more spectrum resources. One solution promoted by the Third Group Partnership Project (3GPP) is to allow MNOs to extend their LTE operation into unlicensed spectrum, including the Unlicensed National-Information-Infrastructure (U-NII) bands at 5 GHz, currently used by Wi-Fi [2] among others. In

M. Hirzallah is with the Department of Electrical and Computer Engineering, University of Arizona, Tucson, AZ 85721 USA (e-mail: hirzallah@email.arizona.edu).

M. Krunz is with the Department of Electrical and Computer Engineering, University of Arizona, Tucson, AZ 85721 USA, and also with the School of Electrical and Data Engineering, University of Technology Sydney, Sydney, NSW 2007, Australia (e-mail: krunz@email.arizona.edu).

Y. Xiao is with School of Electronic Information & Communications, Hauzhong University of Science & Technology, Wuhan, Hubei, China (e-mail: yongxiao@hust.edu.cn).
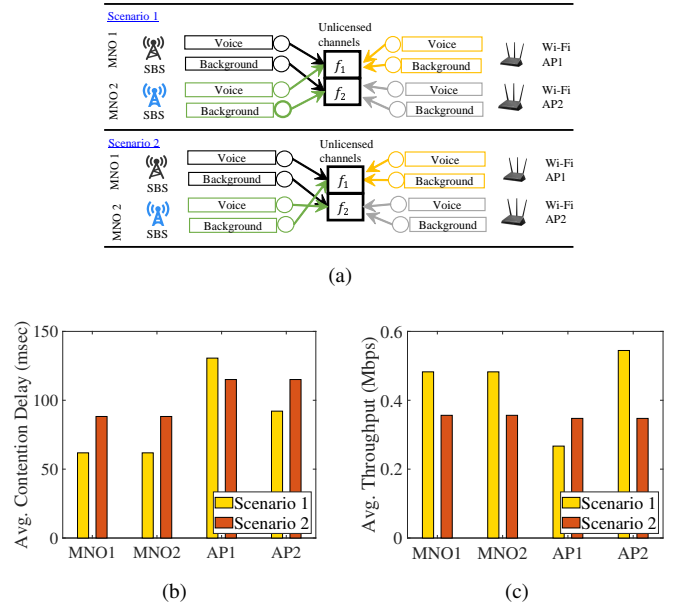
Fig. 1. (a) Example of cross-technology coexistence with heterogeneous traffic classes over two unlicensed channels, (b) average contention delay for downlink (DL) traffic of PC 1, (c) average throughput for DL traffic of PC 4.

Release-13 and Release-14, 3GPP introduced the 'licensed-assisted access' (LAA), 'enhanced LAA' (eLAA), and 'further enhanced LAA' (feLAA) for unlicensed LTE in the U-NII bands. These channel access schemes follow a listen-before-talk (LBT) mechanism, similar to the one adopted in Wi-Fi. Features and procedures introduced for LAA, eLAA, and feLAA are also expected to play a key role in the design of future 5G New Radio Unlicensed (NR-U) [3].

Next-generation networks are expected to support a plethora of services with diverse and often conflicting performance requirements [4]. For example, some of Internet-of-Things (IoT) devices need to support up to 10 years of battery life with high sensitivity to latency. Interactive augmented and virtual reality (AR/VR) require both extremely low latency ($\sim$ 1 ms) and high throughput. Non-interactive applications, such as high-definition video downloading, are more tolerant to delay but have a more stringent throughput requirement. To meet the diverse service requirements of these applications, 3GPP standards introduced four *priority classes* (PCs), labeled as P_1, P_2, P_3, and P_4, for devices that wish to access the U-NII bands using the *Category 4 LBT (CAT4-LBT)* mechanism. [5]. 3GPP is also considering similar PCs for NR-U. Under heavy load,

P_1 and P_2 have the lowest inter-frame waiting time of the four PCs. This time determines the delay between two consecutive channel access attempts. P_1 also has the shortest channel occupancy time (COT), which specifies the duration during which the channel can be used for transmission. These characteristics of PC P_1 and P_2 make them ideal for applications requiring low latency, such as interactive/streaming voice and video services. P_3 and P_4, on the other hand, have higher inter-frame waiting times with longer COTs. Therefore, they are more ideal for applications that require higher throughput but are latency-tolerant. Similarly, Wi-Fi standards, including IEEE 802.11ac, support the Enhanced Distributed Channel Access (EDCA) scheme [6], in which four *access categories* (ACs) (labeled as A_1, A_2, A_3, and A_4) can be used to support different quality of service (QoS) requirements. Although LAA and Wi-Fi technologies follow a similar channel access scheme, they adopt different channel contention parameters, resulting in different channel access probabilities and throughput.

The interaction between LTE PCs and Wi-Fi ACs in a multi-channel setting further complicates the operation of these two technologies. To shed more light on this issue, consider the following example. Suppose that two MNOs, labeled MNO1 and MNO2, share two unlicensed channels with two Wi-Fi access points (APs), labeled AP1 and AP2, as shown in Figure 1(a). Each MNO has one small basestation (SBS) that can access both channels. Each Wi-Fi AP can operate in any of these two channels. Suppose that both SBSs and APs offer two types of services to their respective users: Interactive voice and non-interactive data streaming. Suppose the voice and data streaming traffic has been assigned to PCs P_1 and P_4 by both MNOs, respectively, and to ACs A_1 and A_4 by the Wi-Fi APs, respectively. We conducted discrete-event simulations to investigate the impact of channel assignment on the performance of the coexisting LTE/Wi-Fi systems, as shown in Figures 1(b) and 1(c). In Scenario 1, both MNOs assign the same PC to the same channel, i.e., both SBSs send P_1 traffic over channel 1 and P_4 traffic over channel 2. In Scenario 2, each channel is assigned to different types of services from each MNO. Specifically, MNO1 P_1 and MNO2 P_4 traffic is transmitted over channel 1, while MNO1 P_4 and MNO2 P_1 traffic is transmitted over channel 2. As shown in Figure 1(b), compared to Scenario 2, Scenario 1 achieves lower contention delay and higher throughput for both MNOs, but it results in lower throughput and higher contention delay for the Wi-Fi AP1. Scenario 2, on the other hand, improves fairness over the two channels.

The above example reveals the complexity of LTE/Wi-Fi coexistence when they adopt different priority classes, and shows that MNOs need a tool that allows them to evaluate their expected QoS in different coexistence scenarios. This motivates our work, in which we develop a novel framework for MNOs to evaluate different QoS performance metrics, including the probability of successful transmission (PST), contention delay, and throughput. Our framework incorporates queuing and Markov-based models to derive the above metrics. We further simplify this model and derive closed-form expressions for PST under various PCs and ACs. Furthermore, we develop a discrete-event simulation environment using

C++ and use it to perform fitting and verification of our closed-form expressions. These expressions are simple and can be used to optimize LTE/Wi-Fi coexistence in practical scenarios. We also study the impact of traffic pattern and intensity on LTE/Wi-Fi coexistence, and investigate 3GPP's LBT parameters that need to be adapted to achieve harmonious LTE/Wi-Fi coexistence. Our findings in this paper are also applicable to the study of the coexistence between NR-U and Wi-Fi systems, because the LAA/eLAA/feLAA frameworks are also being considered as baseline for NR-U.

Notational convention - In this paper, we generally use subscripts to denote the priority class and superscripts to denote the technology, i.e., LTE or Wi-Fi. We drop the subscripts and superscripts when they are not needed.

II. RELATED WORK

Previous works on LTE/Wi-Fi coexistence mainly focused on designing collision detection schemes [7], optimizing resource allocation among MNOs [8], [9], and optimizing the selection of priority classes [10]. Authors in [11] focused on addressing challenges of operating next-generation networks over unlicensed bands using asymmetric full-duplex communications. Other works focused on modeling LTE/Wi-Fi coexistence using Markov techniques [12]–[18]. Valls et al. modeled LTE/Wi-Fi coexistence using Markov chains and investigated maximizing the capacity of LTE in unlicensed bands by specifying the maximum airtime for LTE [12]. Zhang et al. modeled LTE/Wi-Fi coexistence using a Markov-based framework and derived the optimal size of contention window that maximizes LTE and Wi-Fi throughput [13]. Han et al. introduced a MAC design for harmonious LTE operation in unlicensed bands [15]. Abdelfattah and Malouch used random walks to model the duty-cycle-based LTE-U operation [16]. Sutton et al. focused on analyzing the delay of LAA [17]. Mehrnoush et al. modeled the impact of energy detection in LTE/Wi-Fi coexistence [18]. Bitar et al. considered modeling LTE/Wi-Fi coexistence using Markov models [19]. Caleffi et al. considered an analytic framework based on the nest-site selection behavior in honeybee swarms to model the coexistence between heterogeneous networks [20]. Xu et al. considered an experimental setup to study and analyze LTE/Wi-Fi coexistence [21]. Chen et al. considered adding a clear-to-send-to-self frame in eLAA to enhance LTE/Wi-Fi coexistence [22]. Yuan et al. investigated the design and modeling of the random access procedure to support initial access in standalone LTE-U systems [23]. Rastegardoost focused on modeling and minimizing Wi-Fi latency in LTE/Wi-Fi coexistence [24]. Although previous works have interesting analysis and provide insightful results, they still have limitations. Some of these works only focus on a single priority class and/or fail to address parameters distinguishing key differences between LTE and Wi-Fi priority classes, such as the arbitration inter-frame space (AIFS). Bianchi et al. alluded to the importance of including the AIFS parameter while modeling performance of the EDCA scheme [25]. The AIFS value used by Wi-Fi systems (a.k.a., 'deferment period' in LAA) decides who can access the channel earlier. Some other works included

the AIFS in their models, including [14], [26], [27], but require expensive computations that limits their applicability in practical scenarios. In our paper, we model the impact of AIFS and provide a simple yet accurate closed-form solution for modeling prioritized channel access for LTE and Wi-Fi networks over unlicensed bands. We also investigate how traffic intensity and pattern affects the performance of LTE/Wi-Fi coexistence and compare their throughput and delay under different traffic loads. We find that 3GPP should change the AIFS setting of PCs P_1 and P_2 to ensure fair coexistence with Wi-Fi. In addition, we find that setting of AIFS affect LTE/Wi-Fi coexistence more than setting of contention window. The fairness issues are also found to be less significant when the number of LTE and Wi-Fi devices increases beyond a certain limit.

III. UNLICENSED CHANNEL ACCESS SCHEMES IN IEEE 802.11 AND 3GPP LAA

A. IEEE 802.11

IEEE 802.11 standards support several channel access schemes in which the DCF and EDCA are the most frequently used ones. EDCA builds on CSMA/CA algorithm to improve QoS provisioning. EDCA defines four ACs (A_1 - A_4): Voice (AC_VO), video (AC_VI), best effort (AC_BE), and background (AC_BK), as shown in Table I. The duration of AIFS, T_{AIFS}, is computed as $T_{\text{AIFS}} = T_{\text{SIFS}} + d_i T_{\text{slot}}$ [6], where $T_{\text{SIFS}} = 16$ μsec is the short inter-frame space, $T_{\text{slot}} = 9$ μsec is the MAC time slot, and d_i is the AIFS number (AIFSN). In addition, EDCA scheme limits the transmission time T_i, a.k.a, transmission-opportunity (TXOP) period, for ACs A_1 and A_2. The TXOP times for ACs A_3 and A_4 are not restricted. During each TXOP period, it is possible to aggregate multiple frames.

The EDCA scheme works as follows. In first transmission attempt, each station (e.g., Wi-Fi AP or device trying to initiate channel access) first senses the channel for an AIFS period and can only start transmission if the channel is sensed idle during the AIFS. If the channel is sensed busy during the AIFS, a backoff mechanism is triggered in which the transmitter randomly picks an integer k between 0 and $K - 1$ for

$$K \in \min\{W_{i,j}, W_{i,\max}\} \tag{1}$$

where $W_{i,j} = 2^j W_{i,0}$, j is the index of the retransmission attempt, $W_{i,0}$ is minimum size of contention window (CW$_{\min}$), and $W_{i,\max}$ is the maximum size of contention window (CW$_{\max}$).

The station counts down for k successive time slots as long as the channel is idle. Whenever the channel is sensed to be busy during the countdown, the station should freeze its counter until it becomes idle again. After every busy period, a station should ensure the channel is idle for an AIFS period prior to pursuing to process its counter. Notice that different ACs have different AIFS periods, and thus stations with smaller AIFS period start processing their counters earlier than others. Once the counter becomes zero, the station can then start its transmission for a duration that is less than or equal its TXOP period. An acknowledgment (ACK) or a block acknowledgement (BA) frame is sent back upon successful

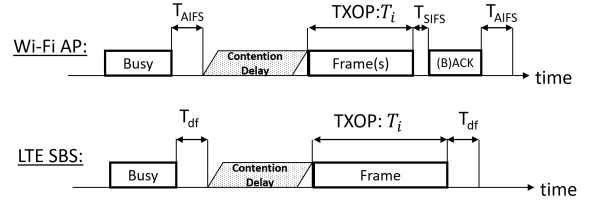| AC A_i (EDCA)	d_i/T_{AIFS}	$W_{i,0}$	$W_{i,\max}$	Max TXOP T_i
A_1:AC_VO	2/ 34 μsec	4	8	2.080 msec
A_2:AC_VI	2/ 34 μsec	8	16	4.096 msec
A_3:AC_BE	3/ 43 μsec	16	1024	-
A_4:AC_BK	7/ 79 μsec	16	1024	-
Legacy DCF	2/ 34 μsec	16	1024	-
PC P_i (CAT4)	d_i/T_{df}	$W_{i,0}$	$W_{i,\max}$	T_i
P_1	1 or 2/ 25 or 34 μsec	4	8	2 msec
P_2	1 or 2/ 25 or 34 μsec	8	16	3 or 4 msec
P_3	3/ 43 μsec	16	64	6, 8, or 10 msec
P_4	7/ 79 μsec	16	1024	6, 8, or 10 msec

Fig. 2. Examples of EDCA (top) and CAT4-LBT (down) channel access procedures.

reception of one or more frames. A station can infer a failure transmission or a collision if it does not receive the ACK/BA frame within a certain period of time. Failed or collided data frames should be retransmitted for at most R_i times before being discarded, and for each retransmission the contention window needs to be doubled. For new frame transmissions, a station should first wait for an AIFS period and then trigger a new backoff process, as discussed above. The smaller the values of AIFS, CW$_{\min}$, and CW$_{\max}$, the higher the probability for a station to successfully capture the channel [25]. An example of the EDCA operation is shown in Figure 2.

B. Licensed Assisted Access (LAA)

For LAA operation, 3GPP adopts the CAT4-LBT scheme, which is similar to EDCA but considers different parameters for accessing the unlicensed spectrum [5]. LAA defines four priority classes, P_1-P_4, as shown in Table I, which, in some sense, can be considered as the equivalent to ACs A_1-A_4 in Wi-Fi. The deferment period T_{df} in LAA is equivalent to AIFS in Wi-Fi, and therefore, in this paper, we use AIFS to refer for both LAA 'deferment period' and Wi-Fi AIFS. We also use TXOP to refer to Wi-PC P_1 has the smallest AIFS, CW$_{\min}$, and CW$_{\max}$ among all PCs, which is equivalent to AC A_1 in EDCA. During the TXOP period, the SBS sends an OFDMA frame, where it schedules resource blocks (distributed across time and frequency) to user equipments (UEs). An example of CAT4-LBT operation over time is shown in Figure 2. In LAA, SBS infers the failure of transmission by monitoring the HARQ-ACK feedback messages sent by UEs over the licensed channel [5]. By comparing the entries of EDCA and CAT4-LBT schemes in Table I, we notice that LAA supports smaller AIFS values and hence LAA devices are expected to capture channels faster than those with Wi-Fi, resulting in an unfair situation.
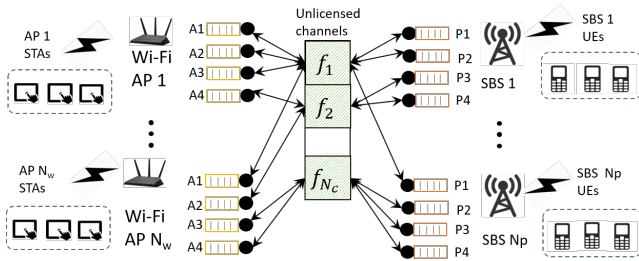
Fig. 3. Example of MNOs and Wi-Fi networks coexistence with prioritized traffic over a set of N_c channels.

IV. NETWORK COEXISTENCE MODEL

We consider a set $\mathcal{N}$ of N_p MNOs, each of which has deployed a set $\mathcal{S}$ SBSs that operate over unlicensed bands using LAA. MNOs share a set $\mathcal{F}$ of N_c unlicensed channels with a set $\mathcal{M}$ of N_w Wi-Fi APs. To simplify our discussion, we focus on the downlink transmission for both LTE and Wi-Fi networks. Our model, however, can be directly extended to uplink scenarios. In this paper, we consider a general model in which each Wi-Fi AP can support a set of ACs, denoted as $C = \{A_1, \cdots, A_{N_{ac}}\}$, where N_{ac} is the number of supported ACs (e.g., $N_{ac} = 4$ in IEEE 802.11ac Wi-Fi technology). Each SBS can also support a set of PCs labeled as $\mathcal{L} = \{P_1, \cdots, P_{N_{pc}}\}$, where N_{pc} is the number PCs (e.g., $N_{pc} = 4$ in LTE-LAA Releases 13 and 14). Let $m_{j,h}$ be the number of Wi-Fi APs that serve traffic of AC A_j over channel h. Let $n_{j,h}$ be the number SBSs that serve traffic of PC P_j over channel h. Wi-Fi devices access the channel using the EDCA mechanism while SBSs access the channel using the CAT4-LBT mechanism, as discussed in Section III.

We define the following channel sharing structure (CSS) to represent the mapping of SBSs and APs traffic to the N_c unlicensed channels. Formally, CSS **cs** is defined as a tuple:

$$\mathbf{cs} = \Big\langle \langle \mathbf{n}(1), \mathbf{m}(1) \rangle, \cdots, \langle \mathbf{n}(N_c), \mathbf{m}(N_c) \rangle \Big\rangle \quad (2)$$

where $\langle \mathbf{n}(h), \mathbf{m}(h) \rangle$ specifies the number of SBSs and Wi-Fi APs operating over channel h; $\mathbf{n}(h) = \langle n_{1,h}, \cdots, n_{N_{pc},h} \rangle$ and $\mathbf{m}(h) = \langle m_{1,h}, \cdots, m_{N_{ac},h} \rangle$. We next introduce a Markov-based framework to help MNOs estimate their performance metrics under each possible channel assignment.

V. MODELING DIFFERENTIATED CHANNEL ACCESS

IEEE 802.11 and 3GPP standards adopt a similar channelization approach with non-overlapping channels at the U-NII bands [6] [2]. For ease of illustration, we drop the index denoting the channel number and focus on a single channel shared by a number of SBSs and APs. Although our analysis and findings are focused on a single channel operation, they can be easily extended to multiple channels.

Channel access dynamics of EDCA and CAT4-LBT schemes can be characterized using Markov processes, under the assumption that frames of the same priority class have the same probability of collision [25], [28], [29]. We next derive a Markov-based model for LTE and Wi-Fi channel access, and discuss how this model can be approximated using closed-form expressions to be used in practical and real-time

operation. Analysis and derivations derived in this section apply similarly to priority classes defined in 3GPP and IEEE 802.11 standards. Therefore, we opt to drop the superscripts that denote the type of technology, but still bring it back occasionally when needed.

A. Queuing Model

In line with [25], we consider a queue model to describe the traffic behavior at an LTE or Wi-Fi transmitter. At a Wi-Fi transmitter, we have N_{ac} parallel queues, modeling traffic of the N_{ac} ACs . For a network with m Wi-Fi transmitters, there are mN_{ac} queues. Similarly, at each SBS, we have N_{pc} queues for the N_{pc} PCs. In a network of n SBSs, there are nN_{pc} queues (see Figure 3 with $N_{ac} = N_{pc} = 4$). We use 'priority class' to refer to PCs defined in 3GPP standards and also ACs defined in IEEE 802.11 standards, and distinguish between them when necessary. Frames arriving at each queue are served on a First-Come-First-Served (FCFS) basis. We follow a commonly adopted setting and assume frame arrivals follow a Poisson distribution [30] (see Annex A.2.1.3.1 'Traffic models'). Our model can be directly extended to more general settings, e.g., frame arrivals follow other distributions. Let λ_i be the arrival rate of frames that are associated with priority class i, $i \in \mathcal{L} \cup C$, measured in frames per second. We consider a general setting where the service time of the ith queue follows an arbitrary distribution Z_i with mean $1/\mu_i$ and variance σ_i^2. The service time of each queue depends on channel occupancy conditions, number of coexisting transmitters, as well as their channel access parameters and traffic profiles. Therefore, we model queues in the system as an M/G/1 queue.

1) Probability of Frame Arrivals: The probability of having y frames arriving during a slot T_{slot} can be written as $(\lambda_i T_{\text{slot}})^y \exp^{-\lambda_i T_{\text{slot}}}/y!$. We set T_{slot} to the duration of a MAC time slot (i.e., 9 microseconds) [5] [6]. We define the *probability of frame arrivals* g_i as the probability of having one or more frames of class i arriving at the queue during a time slot T_{slot}, i.e., $y \geq 1$. g_i is given by:

$$g_i = 1 - \exp^{-\lambda_i T_{\text{slot}}}. \quad (3)$$

2) Average Contention Delay: In CSMA/CA, a device backs off for a random period of time before it starts transmission, resulting in *contention delay* $D_i^{(l)}$ that is different for the different priority classes. Contention delay is a function of the AIFS period, random number of idle time slots, and the time during which the channel is busy during the countdown process. We prove the following result regarding the contention delay for frames of PC P_i, similar result applies for Wi-Fi frames of AC A_i:

Proposition 1. *The average contention delay for frames of PC P_i is given by:*

$$D_i^{(l)} = \lceil 1/p_{s,i}^{(l)} \rceil T_{slot} - T_i^{(l)} \quad (4)$$

where the $p_{s,i}^{(l)}$ is the PST for frames of PC P_i (to be defined and formally formulated in Equations (10)-(11), Section V-B4), and $T_i^{(l)}$ is the TXOP pf PC P_i.

Proof: See Appendix A. ∎

Transmitting a frame over the air requires a time duration that is equivalent to TXOP period $T_i^{(l)}$. The sum of contention delay $D_i^{(l)}$ and TXOP period $T_i^{(l)}$ constitutes the *service time* of the queue; i.e., the time required for a data frame reaching at the head of the queue to be successfully delivered to its destination, which is given by $S_i = D_i^{(l)} + T_i^{(l)}$. The queue service rate $\mu_i = 1/S_i$ represents the average number of frames that can be served in a second.

3) Probability of Saturation: We define the *probability of saturation* q_i for the queue of class i as the probability of having a non-empty queue, i.e., $q_i = 1$ means that there are always data frames available in the queue waiting to be served. To compute the probability of saturation, we consider two queue states, 'Idle' and 'Occupied', with arrival rate of λ_i and service rate μ_i. The probability of saturation is equivalent to the long-term probability of being in state 'Occupied' [31]:

$$q_i = \lim_{t \to \infty}(1 - P_{\text{Idle}}(t)) = \frac{\lambda_i}{\mu_i + \lambda_i} = \frac{\lambda_i(D_i^{(l)} + T_i^{(l)})}{1 + \lambda_i(D_i^{(l)} + T_i^{(l)})}. \quad (5)$$

B. Markov-based Model

Dynamics of EDCA and CAT4-LBT channel access schemes can be modeled using a 3-dimensional Markov chain by following a similar line as in [28] [29]. We discuss the modeling for an arbitrary priority class, say class i, $i \in C$ or $i \in \mathcal{L}$. We define the stochastic process $\{\xi_i(n) : n \in \mathbb{N}^+\} = \{(J_i(n), K_i(n), L_i(n)) : n \in \mathbb{N}^+\}$ that is composed of three tuples: $\{J_i(n) : n \geq 0\}$, $\{K_i(n) : n \in \mathbb{N}^+\}$, and $\{L_i(n) : n \in \mathbb{N}^+\}$. $\{\xi_i(n) : n \in \mathbb{N}^+\}$ describes the channel access and contention behavior for a transmitter serving traffic of class i as follows. The process $\{J_i(n) : n \in \mathbb{N}^+\}$ models the number of transmission attempts of a frame. When collision happens, the contending transmitter should back off and contend again. Therefore, for a given n, $J_i(n)$ takes an integer value in $[0, R_i]$, where R_i is the maximum retransmission limit for class i. When $J_i(n) = j$, we say that the transmitter is at the jth *backoff stage* at time n. We also use $J_i(n)$ to model some special cases. For instance, $J_i(n) = -1$ indicates that the transmitter at time n has already finished frame transmission and it is ready to start serving a new frame. Also, $J_i(n) = -2$ indicates that the transmitter is currently engaged in a successful frame transmission, while $J_i(n) = -3$ denotes that the transmitter at time n has no frames to be transmitted, i.e., awaiting for new frames to arrive from upper layers.

The process $\{K_i(n) : n \in \mathbb{N}^+\}$ models the length of the backoff period. At time n, the random variable $K_i(n)$ takes an integer value in $[0, W_{i,j} - 1]$, where $W_{i,j}$ is the minimum contention window in (1), and j is the retransmission attempt. $K_i(n)$ tracks the value of the backoff counter; $K_i(t) = k$ means that at time n, the transmitter's backoff counter is k.

The process $\{L_i(n) : n \in \mathbb{N}^+\}$ models the time remaining before the channel becomes idle after a busy period. Recall that when the channel becomes busy, the transmitter freezes its counting and waits for the channel to become idle again. The transmitter deems the channel idle if it remains so for d_i idle slots, i.e., AIFS duration. $L_i(n) = l$ indicates that at time

n the transmitter should wait for l slots before the channel is considered idle. The channel busy duration is random, and depends on the TXOP lengths adopted by coexisting transmitters. It is quite unlikely that the channel remains busy continuously for a time period that exceeds the maximum TXOP period of all coexisting transmitters. Therefore, we set the busy time to $B = \max\{\lceil T_i/T_{\text{slot}} \rceil : i \in C \cup \mathcal{L}\}$, i.e., the maximum duration of TXOP period of all priority classes listed in Table I. We also utilize $\{L_i(n) : n \in \mathbb{N}^+\}$ process to model other special cases. For instance, when $L_i(n)$ takes a value in $[1, \lceil T_i/T_{\text{slot}} \rceil]$, it models the time spent in successful transmission. A value for $L_i(n)$ in $[1, C_i]$ is used to model the time spent in a collision, where C_i is the number of time slots spent in collision.

We are interested in modeling the channel access at the steady state, i.e., $\lim_{n \to \infty} \xi_i(n) = \{(J_i, K_i, L_i)\}$, hereafter, we drop the time index n. We introduce a Markov chain in which each state takes a value in the 3-dimensional stochastic process $\xi_i = \{(J_i, K_i, L_i)\}$, as shown in Figure 4. For simplicity of illustration, we categorize the states of ξ_i into the following:

- *Contention-with-Idle-Channel:* Set of states in which the channel is idle and the transmitter backs off, i.e., $\{(J_i, K_i, 0) : J_i \in [0, R_i], K_i \in [1, W_{i,j} - 1]\}$.
- *Contention-with-Busy-Channel:* Set of states in which the channel is busy and the transmitter freezes its counter, waiting for the channel to become idle again, i.e., $\{(J_i, K_i, L_i) : J_i \in [0, R_i], K_i \in [1, W_{i,j} - 1], L_i \in [1, d_i + B]\}$. In Figure 5, we show a detailed illustration of these states at the jth backoff stage. b_i is the *probability of counter freezing* (to be formulated later), and p_i is the *probability of collision p_i*.
- *Channel-Access-Attempt:* Set of states for which the transmitter can access the channel and start transmission, i.e., $\{(J_i, 0, 0) : J_i \in [0, R_i]\}$. The sum of the steady state probabilities of these states constitutes the *probability of channel access (PCA)* denoted as τ_i.
- *Transmission-with-Collision:* Set of states that represent a failed transmission due to collision and/or bad channel conditions, i.e., $\{(J_i, 0, L_i) : J_i \in [0, R_i], L_i \in [1, C_i]\}$. After a collision, the transmitter doubles its contention window, as in (1), and initializes the backoff counter with a new integer value in $\{0, 1, \cdots, W_{i,j} - 1\}$. This triggers a transition from state $(j - 1, 0, C_i)$ to one of the states $\{(j, K_i, 0) : K_i \in [0, W_{i,j} - 1]\}$, which happens with probability $1/W_{i,j}$.
- *Transmission-with-Success:* Set of states in which the transmitter is involved in a successful transmission, i.e., $\{(-2, 0, L_i) : L_i \in [1, \lceil T_i/T_{\text{slot}} \rceil]\}$. Once state $(j, 0, 0)$ is reached, a transmitter can proceed. The *probability of successful transmission* $(p_{s,i})$ for frames of priority class i can be expressed as $p_{s,i} = \tau_i(1 - p_i)$.
- *Idle-Queue:* This state, i.e., state $(-3, 0, 0)$, models the case when the transmitter completes its transmission but finds the queue idle, so it has to wait for a new frame to arrive. This probability of having an idle queue is $1 - q_i$, where q_i is computed in (5). New frames arrive with probability g_i, causing a transition from state $(-3, 0, 0)$ to
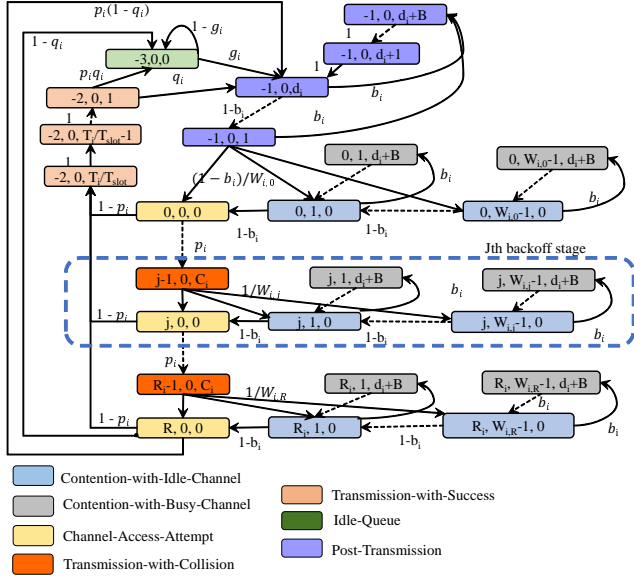
5

Fig. 4. Markov chain that models channel access and contention for the ith priority class, $i \in \mathcal{L}$ or $i \in \mathcal{C}$.

the Post-Transmission states.

- *Post-Transmission:* Set of states in which the transmitter starts serving a new frame, i.e., $\{(-1, 0, L_i) : L_i \in [1, d_i + B]\}$.
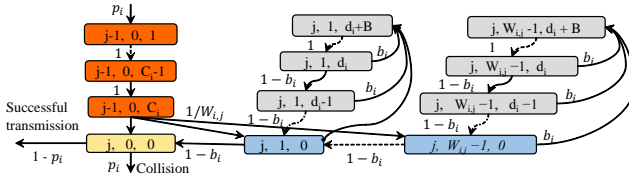

Fig. 5. Illustration of the sub-chain that corresponds to the jth backoff stage in Figure 4.

We next formulate the probability of collision p_i, counter freezing b_i, channel access τ_i, and successful transmission $p_{s,i}$. We bring back the superscript notation to differentiate between CAT4-LBT (LTE) and EDCA (Wi-Fi) priority classes.

1) Probability of Collision:

Proposition 2. *CAT4-LBT frames of PC P_i collide with probability $p_i^{(l)}$ and EDCA frames of AC A_i collide with probability $p_i^{(w)}$, where $p_i^{(l)}$ and $p_i^{(w)}$ can be expressed as follows:*

$$p_i^{(l)} = 1 - (1 - \tau_i^{(l)})^{n_i - 1} \prod_{j \in \mathcal{L}, j \neq i} (1 - \tau_j^{(l)})^{n_j} \prod_{k \in \mathcal{C}} (1 - \tau_k^{(w)})^{m_k} \quad (6)$$

$$p_i^{(w)} = 1 - (1 - \tau_i^{(w)})^{m_i - 1} \prod_{j \in \mathcal{C}, j \neq i} (1 - \tau_j^{(w)})^{m_j} \prod_{k \in \mathcal{L}} (1 - \tau_k^{(l)})^{n_k}. \quad (7)$$

Proof: See Appendix B. ∎

2) Probability of Counter Freezing: The channel is sensed to be busy when at least one transmitter becomes active. In this case, other devices who can hear the transmission must freeze their counters. The probability of counter-freezing (b_i) for a frame of PC P_i can be written as:

$$b_i = 1 - \prod_{j \in \mathcal{L}} (1 - \tau_j^{(l)})^{n_j} \prod_{k \in \mathcal{C}} (1 - \tau_k^{(w)})^{m_k}. \quad (8)$$

A similar expression can be formulated for AC A_i. We can observe that the probability of counter-freezing relates to transmitter's capability of detecting transmissions of other colocated devices. For simplicity, we assume a dense network model in which all colocated devices from different technologies are in close proximity, and thus their signals can be detected by each others. Under this assumption, the dissimilarity between the energy detection thresholds adopted by different coexisting technologies can be ignored.

3) Probability of Channel Access (PCA): A device attempts to start transmission whenever the channel becomes clear and its backoff counter reaches zero. We prove the following result for the PCA of PC i (similar expression can be formulated for AC A_i and PC P_i):

Proposition 3. *The PCA for a frame of PC i can be written as:*

$$\tau_i = \frac{1 - p_i^{R_i+1}}{1 - p_i} \cdot \left[\overbrace{\frac{1 + b_i B}{b_i} \cdot \frac{1 - (1 - b_i)^{d_i}}{(1 - b_i)^{d_i}}}^{(a)} \right.$$
$$+ \overbrace{\lceil T_i/T_{slot} \rceil (1 - p_i^{R_i+1})}^{(b)} + \overbrace{(1 + p_i C_i) \frac{1 - p_i^{R_i+1}}{1 - p_i}}^{(c)}$$
$$\left. + \overbrace{\frac{(1 - q_i)(1 + p_i^{R_i} - p_i^{R_i+1})}{g_i}}^{(d)} + \overbrace{\frac{1 + b_i B}{2(1 - b_i)^{d_i}} \sum_{j=0}^{R_i} W_{i,j} p_i^j}^{(e)} \right]^{-1}$$
$$(9)$$

Proof: See Appendix C. ∎

Equation (9) includes several terms, labeled (a) to (e), which can be mapped to the state diagram in Figure 4. Terms (a) and (e) correspond to the states of Contention-with-Idle-Channel, Contention-with-Busy-Channel, and Post-Transmission. These terms include the key parameters controlling the backoff process for PC P_i, i.e., size of contention window $W_{i,j}$ and AIFS number d_i. Term (b) corresponds to the states of Transmission-with-Success, and includes the time spent in successful transmission T_i. Term (c) corresponds to the states of Transmission-with-Collision. Term (d) corresponds to the state of Idle-Queue and depends of the probabilities g_i and q_i, characterizing the traffic intensity and queue dynamics.

4) Probability of Successful Transmission (PST): Successful transmission happens when a device initiates a channel access and does not collide with any other transmissions. The probability of this collision-free event can be formulated by taking the product of probability of channel access τ_i and the complementary for the probability of collision, i.e., $1 - p_i$. The

PST for PC P_i can be expressed as $p_{s,i}^{(l)} = \tau_i^{(l)}(1 - p_i^{(l)})$ and the PST of AC A_i can be expressed as $p_{s,i}^{(w)} = \tau_i^{(w)}(1 - p_i^{(w)})$:

$$p_{s,i}^{(l)} = \tau_i^{(l)}(1 - \tau_i^{(l)})^{n_i-1}\prod_{j \in \mathcal{L}, j \neq i}(1 - \tau_j^{(l)})^{n_j}\prod_{k \in C}(1 - \tau_k^{(w)})^{m_k}, \quad (10)$$

$$p_{s,i}^{(w)} = \tau_i^{(w)}(1 - \tau_i^{(w)})^{m_i-1}\prod_{j \in C, j \neq i}(1 - \tau_j^{(w)})^{m_j}\prod_{k \in \mathcal{L}}(1 - \tau_k^{(l)})^{n_k}. \quad (11)$$

In homogeneous coexistence with all devices operating according to the same set of parameters, Equation (11) reduces to the PST expression stated in [29].

C. Complexity in Optimizing Performance Measures

Computing the PSTs in (10) and (11) requires computing $\tau_i^{(x)}$ for different PCs and ACs, where the superscript index x denotes the channel access technology, i.e., LTE or Wi-Fi. From Equation (9), we observe that $\tau_i^{(x)}$ also depends on $p_i^{(x)}$ and probabilities governing queue dynamics (i.e., $g_i^{(x)}$ and $q_i^{(x)}$). As observed from Equations (3), (5), (6), and (7), these quantities also depend on the τ_j's of all coexisting devices. Let $\bar{\tau}^{(w)} = \langle \tau_{A_1}^{(w)}, \cdots, \tau_{A_{N_{ac}}}^{(w)} \rangle$ and $\bar{\tau}^{(l)} = \langle \tau_{P_1}^{(l)}, \cdots, \tau_{P_{N_{pc}}}^{(l)} \rangle$. The formulation of $\tau_i^{(x)}$, of $i \in C \cup \mathcal{L}$, can be written as:

$$\tau_i^{(x)} = f_i(\bar{\tau}^{(w)}, \bar{\tau}^{(l)}) \quad (12)$$

where $f_i(.)$ is a complicated nonlinear function. The interdependence among $\tau_i^{(x)}$, $p_i^{(x)}$, $g_i^{(x)}$, and $q_i^{(x)}$ makes it difficult to have closed-form expression for $f_i(\cdot)$, which can be manipulated and optimized. To enable optimization of PST, we provide an approximate yet accurate expression for PST.

D. Approximation of PST

Our target here is to simplify the exact expression of PST by avoiding the complicated terms linking $\tau_i^{(x)}$, $g_i^{(x)}$, $q_i^{(x)}$, and $p_i^{(x)}$. To facilitate such simplification, we conducted detailed simulations of LTE and Wi-Fi networks with variable traffic loads (refer to Section VI-A for more details on the simulation setup). Based on these simulations, we observed that under heavy traffic loads, the PST values experienced by LTE and Wi-Fi networks tend to decrease exponentially with the number of coexisting devices. The rate of this reduction is different for different PCs and ACs. In light of our simulation experiments, we studied the relation between the terms (a) to (e), in Equation (9), and the different channel access parameters, i.e., contention window $W_{i,j}$, AIFS number d_i, TXOP duration T_i, and the number of LTE/Wi-Fi devices. Under high-intensity traffic, we prove the following result (similar result can be formulated for PST of AC A_i):

Proposition 4. *The average PST of PC P_i can be approximated using the following approximation model:*

$$p_{s,i}^{(l)} \approx c_{i,0}^{(l)} +$$
$$\sum_{k \in \mathcal{L}} c_{i,k}^{(l)}(n_k + 1)\log\left(\beta_{i,k}^{(l)}(W_{k,0}^{(l)}d_k^{(l)} + \lceil T_k^{(l)}/T_{slot}\rceil)n_k + e_{i,k}^{(l)}\right)$$
$$+ \sum_{j \in C} h_{i,j}^{(l)}(m_j + 1)\log\left(\gamma_{i,j}^{(l)}(W_{j,0}^{(w)}d_j^{(w)} + \lceil T_j^{(w)}/T_{slot}\rceil)m_j + \epsilon_{i,j}^{(l)}\right)$$
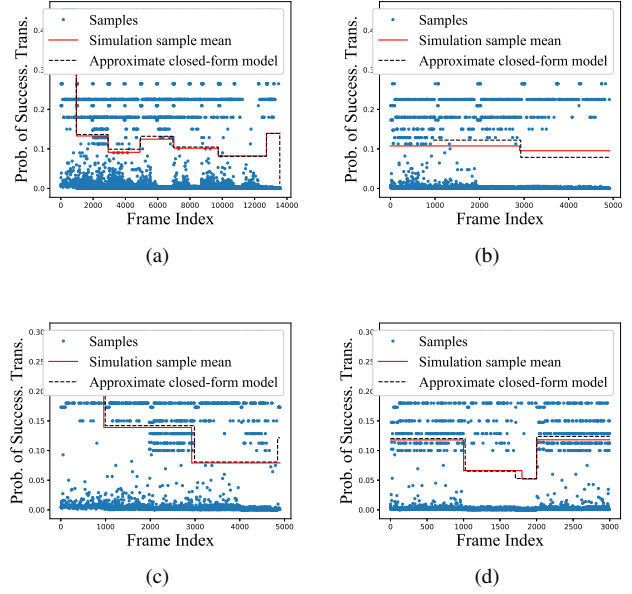
$$(13)$$

Fig. 6. Probability of successful transmission vs. frame index (frames with the same simulation sample mean are collected for the same scenario) (a) CAT4-LBT PC P_1, (b) CAT4-LBT PC P_2, (c) EDCA AC A_1, and (d) EDCA AC A_2.

where the coefficients $c_{i,k}^{(l)}$, $\beta_{i,k}^{(l)}$, $h_{i,j}^{(l)}$, $\gamma_{i,j}^{(l)}$, as well as the constants $c_{i,0}^{(l)}$, $e_{i,k}^{(l)}$ and $\epsilon_{i,j}^{(l)}$, are obtained by fitting the approximate expression to PST samples collected from real networks or from system-level simulations.

Proof: See Appendix D. ∎

The approximate expression in (13) can also be used to model the PSTs for differentiated channel access with homogeneous technology by setting the corresponding constants of the other technology to zero.

1) Fitting Approximate PST Expression: We performed extensive simulations using a discrete-event simulator and collected traces of frames sent by MNOs and Wi-Fi networks. We measure the PST for each frame by taking the inverse of the number of MAC time slots spent in contention as in Equation (4). We scrambled the measured PSTs and divided them into two disjoint sample sets. The first set is used to fit the model in (13) using the 'curve_fit' tool in Python [32], while the second set is used to test our model. Figure 6 shows samples of measured PSTs for a large number of transmitted frames. We plot the sample mean of measured PSTs and their approximate values obtained using the approximate closed-form model in (13) for PCs P_1 and P_2, and ACs A_1 and A_2. The piece-wise constant plots correspond to different scenarios where each scenario corresponds to a network setup with given numbers of LTE and Wi-Fi transmitters. We report the average absolute testing error for these in Figure 7. The approximate model in (13) provides a highly accurate estimate of PST.

2) Exploitation of Closed-form PST Expression: The expression in (13) has many exciting implications:

- By solving for the PSTs, we could evaluate the average throughput $\tilde{r}_i^{(l)}$, for PC P_i, as follows (a similar expression
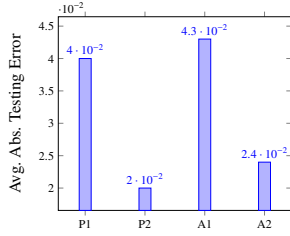
7

Fig. 7. Average absolute testing error for the closed-form approximate model in Equation (13).
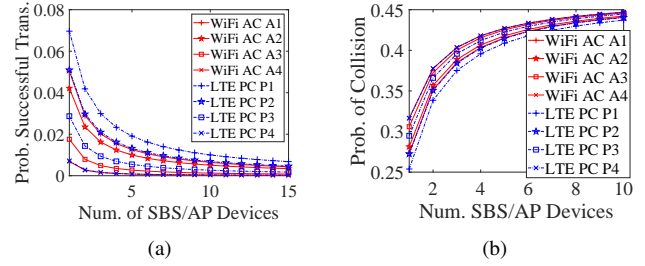

Fig. 8. (a) PST vs. number of transmitters, (b) collision probability vs. number of transmitters.
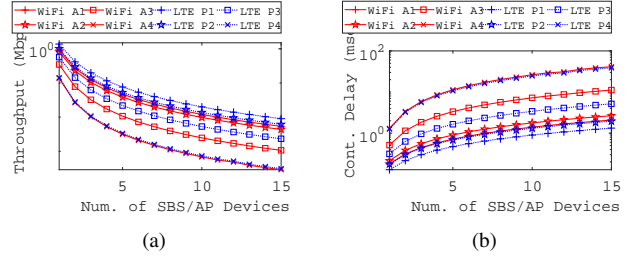

Fig. 9. (a) Average throughput vs. number of transmitters, (b) average contention delay vs. number of transmitters.

can be formulated for AC A_i frames):

$$\tilde{r}_i^{(l)} = \frac{p_{s,i}^{(l)} \mathbb{E}[Q_i^{(l)}]}{D_i^{(l)} + p_{s,i}^{(l)} T_i^{(l)} + (1 - p_{s,i}^{(l)}) C_i^{(l)}} \tag{14}$$

where $Q_i^{(l)}$ is the payload size (in bits) of PC P_i frames, and $D_i^{(l)}$ is the corresponding average contention delay, expressed in Equation (4). The payload size is variable and depends on the selected transmission rate. We report some of typical values for $Q_i^{(l)}$ in Section VI-A.

- PST has a proportional relationship with the average throughput in (14) and an inverse relationship with the average contention delay in (4). These relationships can be leveraged to optimize the performance for different PCs and ACs traffic.
- The expression in (13) can be used to optimize channel assignment and resource allocation among systems with heterogeneous traffic. However, it should not be used to optimize channel access parameters, such as CW_{min}, CW_{max}, AIFS, and TXOP duration that is because the coefficients and constants shown in Equation (13) are subject to fixed values of these parameters.

VI. SIMULATION RESULTS AND DISCUSSION

A. Simulation Setup

To capture the interplay between LTE and Wi-Fi, we developed a discrete-event-based system-level simulator based on CSIM, a C++ library that supports process-oriented discrete-event simulations [33]. CSIM provides functions for creating parallel processes, and capabilities for enabling control and signaling between them. We implemented the most recent CAT4-LBT and EDCA schemes, as specified by 3GPP and IEEE 802.11ac standards [5] [6]. Our simulator has a timing granularity of a nanosecond, and can capture all the timing details listed in Table I. We run the simulator for 10 seconds and collect traces and logs from all LTE and Wi-Fi transmitters, including timestamps of frame arrivals, time spent in various queues, time spent during contention, and time spent during transmission. Our simulation setup is based on a dense random topology, with all SBSs/UEs and APs/STAs are uniformly distributed over a square area of length 40 meters. All devices share a common 20 MHz channel, centered at 5.18 GHz. We consider the 3GPP InH office indoor path loss model. Other simulation parameters are in line with 3GPP parameters [2].

We set the payload size $Q_i^{(l)}$ for frames of PCs P_1, P_2, P_3, and P_4 to 28.824, 57.684, 144.12, and 144.12 Kbits, respectively. For Wi-Fi frames of ACs A_1, A_2, A_3, and A_4,

we set the frame size $Q_i^{(w)}$ to 18.605, 38.156, 129.052, and 129.052 Kbits, respectively. Unless stated otherwise, we adopt a Poisson frame arrival process of rate $\lambda = 1000$ frames per second for all PCs and ACs. We investigate the impact of other traffic patterns in Section VI-E. Payload size per frame has been set while assuming modulation and coding scheme (MCS) of BPSK with 1/2 code rate. We scale these frame sizes according to the used MCS.

B. Number of LTE/Wi-Fi Transmitters

Increasing the number of LTE and Wi-Fi transmitters adds more congestion over unlicensed channels. We evaluate the PST, probability of collision, average throughput, and average contention delay versus the number of transmitter for the different PCs and ACs, as shown in Figures 8(a), 8(b), 9(a), and 9(b) (where the numbers of LTE and Wi-Fi transmitters are equal). The performance of all PCs and ACs degrades exponentially with the increase in the number of transmitters. We noticed that PCs P_1 and P_2 achieve higher PST and average throughput, as well as lower collision probability and average contention delay than Wi-Fi ACs A_1 and A_2. Recall that P_1 and P_2 have smaller AIFS duration than those of ACs A_1 and A_2. Coexistence between devices serving P_1 and P_2 traffic and those serving A_1 and A_2 traffic is not fair to the latter.

C. Size of Contention Window and Fairness Tradeoff

We investigate how changing the CW_{min} value of LTE PC P_1 affects the performance of coexisting networks. We evaluate the average throughput and delay achieved by a transmitter over the four priority classes it supports. These metrics are plotted in Figures 10(a) and 10(b) versus CW_{min} value of PC P_1. Although increasing CW_{min} of PC P_1 improves fairness between LTE and Wi-Fi networks, this improvement becomes negligible when the number of transmitters becomes relatively high (see LTE and Wi-Fi plots with $m = n = 8$).
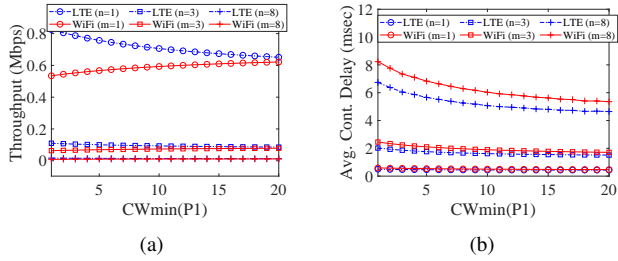
Fig. 10. (a) Average throughput per transmitter vs. size of minimum contention window of PC P_1, (b) average contention delay per transmitter vs. size of minimum contention window of PC P_1.
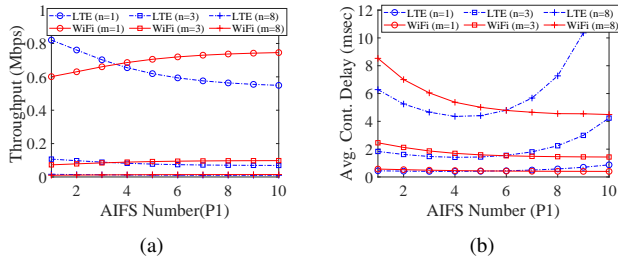

Fig. 11. (a) Average throughput per transmitter vs. AIFS number of PC P_1, (b) average contention delay per transmitter vs. AIFS number of PC P_1.

D. AIFS and Fairness Tradeoff

In Figures 11(a) and 11(b), we investigate how fairness between LTE and Wi-Fi networks is affected by changing the AIFS number (AIFSN) of PC P_1. We take the mean of the contention delay and throughput for all priority classes at each transmitter. We increase the AIFSN for PC P_1 and record the average delay for all transmitters from both technologies. When we increase the AIFSN, frames of PC P_1 wait for a longer AIFS duration, which reduces their chance to occupy the channel faster than others. Although this degrades the performance of PC P_1 traffic, it positively affects other PCs. Therefore, we notice that the average delay experienced by each LTE transmitter reduces as we increase AIFSN for PC P_1. This does not last forever, because the contention delay experienced by P_1 traffic overwhelms the overall LTE contention delay. This justifies the non-monotonic behavior in the delay for LTE nodes. We notice that changing the AIFSN impacts LTE and Wi-Fi networks more than CW_{min}. We also notice that the fairness issue becomes less visible when more devices are present (compare the plots for $m = n = 1$, $m = n = 3$, and $m = n = 8$ in Figures 11(a) &11(b)).

E. Impact of Traffic Pattern and Intensity

To study the impact of the traffic pattern and its intensity, we simulate the network, considering different traffic patterns for various PCs and ACs. In particular, we use a periodic traffic pattern with a fixed frame inter-arrival time for LTE PCs P_1 and P_2, as well as for Wi-Fi ACs A_1 and A_2. This traffic pattern exemplifies voice and video streams. For other traffic classes, we use a Poisson packet arrival process. We vary the frame arrivals rate for all classes and record the average contention delay and throughput at LTE SBS and Wi-Fi AP. LTE and Wi-Fi adopt three MCSs: QPSK with 1/2 code rate, 16-QAM with

(a) QPSK with 1/2 code rate.

(b) 16-QAM with 1/2 code rate.

(c) 64-QAM with 2/3 code rate.

(d) Wi-Fi is 64-QAM and LTE is QPSK with 2/3 and 1/2 code rate.
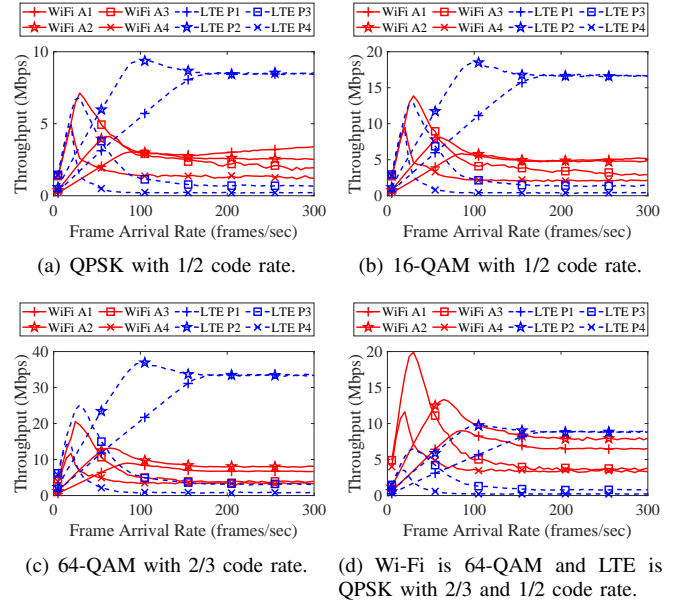
Fig. 12. Network throughput vs. traffic intensity.

1/2 code rate, as well as 64-QAM with 2/3 code rate. In practice, LTE and Wi-Fi receivers have different designs, and thus, they usually associate different SINR thresholds with these MCSs. In Figure 12, we plot the average network throughput for the downlink versus traffic intensity under different MCSs. In Figure 13, we plot the average contention delay versus traffic intensity as experienced by the two technologies. In line with our previous observations, LTE priority classes P_1 and P_2 achieve higher throughput and lower contention delay than Wi-Fi ACs A_1 and A_2. In addition, under the same MCS, LTE PCs P_1 and P_2 achieve at least two orders of magnitude higher throughout than Wi-Fi ACs A_1 and A_2. Traffic assigned as P_1 and P_2 classes also experiences smaller contention delay than A_1 and A_2 traffic. This improvement happens because LTE PCs P_1 and P_2 have a smaller AIFS than Wi-Fi ACs A_1 and A_2, which gives them higher chances to access the channel and, hence, improved throughput and delay. Under higher modulation schemes, we notice that Wi-Fi devices are more prone to outage and starvation, leading to an increase in their average delay, as shown in Figures 13(c) and 13(d).

VII. Conclusion

We develop simple yet accurate approximate closed-form solutions that allow MNOs to evaluate their operation over unlicensed bands with heterogeneous channel access priority classes. These approximate closed-form solutions can be used in practical scenarios to decide on resource allocation for LTE and Wi-Fi systems. Extensive simulation results reveal that some of the channel access parameters adopted by 3GPP to operate over unlicensed channels are not fair to Wi-Fi systems. We find that changing the values of AIFSN and CW_{min} of LTE priority classes P_1 and P_2 improves fairness between LTE and Wi-Fi systems. We also find that changing the AIFS values for LTE/Wi-Fi priority classes impact their coexistence more than changing their contention window.
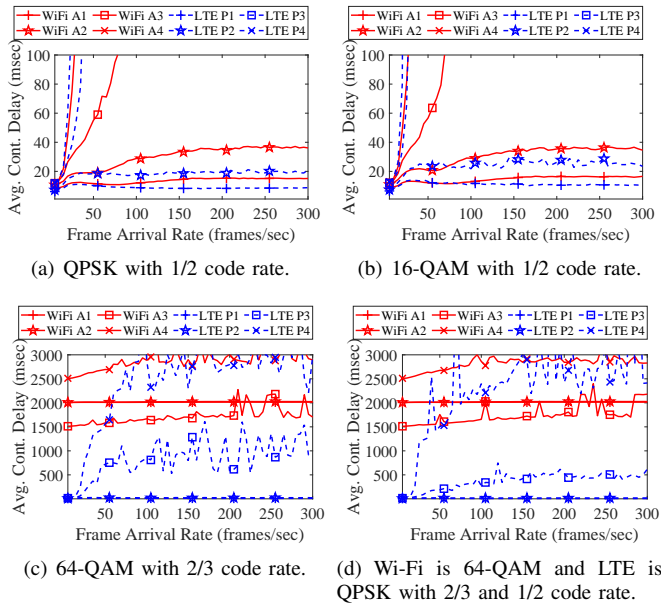
(a) QPSK with 1/2 code rate.

(b) 16-QAM with 1/2 code rate.

(c) 64-QAM with 2/3 code rate.

(d) Wi-Fi is 64-QAM and LTE is QPSK with 2/3 and 1/2 code rate.

Fig. 13. Network throughput vs. traffic intensity.

REFERENCES

[1] M. Hirzallah, Y. Xiao, and M. Krunz, "On modeling and optimizing lte/wi-fi coexistence with prioritized traffic classes," in *Proc. of IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*, Oct 2018, pp. 1–10.

[2] 3GPP, "Study on licensed-assisted access to unlicensed spectrum," 3GPP TR. 36.889 v13.0.0., Jun. 2015.

[3] ——, "Study on NR–based access to unlicensed spectrum," no. 3GPP TR 38.889 v16.0.0, Dec. 2018.

[4] "IMT vision - "framework and overall objectives for the future development of IMT for 2020 and beyond"," Recommendation ITU-R M.2083, Sep. 2015.

[5] 3GPP, "Physical layer procedures," 3GPP TR. 36.213 v15.1.0., Mar. 2018.

[6] IEEE, "IEEE–part 11: Wireless LAN MAC and PHY layer specifications," pp. 1–3534, 2016.

[7] M. Hirzallah, W. Afifi, and M. Krunz, "Full-duplex-based rate/mode adaptation strategies for Wi-Fi/LTE-U coexistence: A POMDP approach," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 1, pp. 20–29, Jan 2017.

[8] Y. Xiao, M. Hirzallah, and M. Krunz, "Distributed resource allocation for network slicing over licensed and unlicensed bands," *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 10, pp. 2260–2274, Oct 2018.

[9] M. Hirzallah, Y. Xiao, and M. Krunz, "Matchmaker: An inter-operator network sharing framework in unlicensed bands," in *to appear in proc. of IEEE SECON'19 Conference*, June 2019, pp. 1–9.

[10] I. Samy and L. Lazos, "Optimum priority class selection under Wi-Fi/LTE coexistence," in *Proc. of IEEE ICC'19 Conference*, May 2019, pp. 1–7.

[11] M. Hirzallah, W. Afifi, and M. Krunz, "Provisioning QoS in Wi-Fi systems with asymmetric full-duplex communications," *IEEE Transactions on Cognitive Communications and Networking*, vol. 4, no. 4, pp. 942–953, Dec 2018.

[12] V. Valls, A. Garcia-Saavedra, X. Costa, and D. J. Leith, "Maximizing LTE capacity in unlicensed bands (LTE-U/LAA) while fairly coexisting with 802.11 WLANs," *IEEE Communications Letters*, vol. 20, no. 6, pp. 1219–1222, June 2016.

[13] Q. Zhang, Q. Wang, Z. Feng, and T. Yang, "Design and performance analysis of a fairness-based license-assisted access and resource scheduling scheme," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 11, pp. 2968–2980, Nov 2016.

[14] H. Lee, H. Kim, H. J. Yang, J. T. Kim, and S. Baek, "Performance analysis of license assisted access LTE with asymmetric hidden terminals," *IEEE Transactions on Mobile Computing*, vol. 17, no. 9, pp. 2141–2154, Sep. 2018.

[15] S. Han, Y. C. Liang, Q. Chen, and B. H. Soong, "Licensed-assisted access for LTE in unlicensed spectrum: A MAC protocol design," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 10, pp. 2550–2561, Oct 2016.

[16] A. Abdelfattah and N. Malouch, "Modeling and performance analysis of wi-fi networks coexisting with lte-u," in *In Proc. of IEEE INFOCOM 2017*, May 2017, pp. 1–9.

[17] G. J. Sutton, R. P. Liu, and Y. J. Guo, "Delay and reliability of load-based listen-before-talk in LAA," *IEEE Access*, vol. 6, pp. 6171–6182, 2018.

[18] M. Mehrnoush, V. Sathya, S. Roy, and M. Ghosh, "Analytical modeling of Wi-Fi and LTE-LAA coexistence: Throughput and impact of energy detection threshold," *IEEE/ACM Transactions on Networking*, vol. 26, no. 4, pp. 1990–2003, Aug 2018.

[19] N. Bitar, M. O. Al Kalaa, S. J. Seidman, and H. H. Refai, "On the coexistence of LTE-LAA in the unlicensed band: Modeling and performance analysis," *IEEE Access*, vol. 6, pp. 52 668–52 681, 2018.

[20] M. Caleffi, V. Trianni, and A. S. Cacciapuoti, "Self-organizing strategy design for heterogeneous coexistence in the sub-6 GHz," *IEEE Transactions on Wireless Communications*, vol. 17, no. 11, pp. 7128–7143, Nov 2018.

[21] S. Xu, Y. Li, Y. Gao, Y. Liu, and H. Gaanin, "Opportunistic coexistence of LTE and WiFi for future 5G system: Experimental performance evaluation and analysis," *IEEE Access*, vol. 6, pp. 8725–8741, 2018.

[22] Q. Chen, G. Yu, and Z. Ding, "Enhanced LAA for unlicensed LTE deployment based on TXOP contention," *IEEE Transactions on Communications*, vol. 67, no. 1, pp. 417–429, Jan 2019.

[23] J. Yuan, A. Huang, H. Shan, T. Q. S. Quek, and G. Yu, "Design and analysis of random access for standalone LTE-U systems," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 10, pp. 9347–9361, Oct 2018.

[24] N. Rastegardoost and B. Jabbari, "Minimizing Wi-Fi latency with unlicensed LTE opportunistic white-space utilization," *IEEE Transactions on Wireless Communications*, vol. 18, no. 3, pp. 1914–1926, March 2019.

[25] G. Bianchi, I. Tinnirello, and L. Scalia, "Understanding 802.11e contention-based prioritization mechanisms and their coexistence with legacy 802.11 stations," *IEEE Network*, vol. 19, no. 4, pp. 28–34, July 2005.

[26] Y. Ma, "Analysis of channel access priority classes in LTE-LAA spectrum sharing system," in *Proc. of International Conference on Computer Communication and Networks (ICCCN)*, July 2018, pp. 1–8.

[27] A. M. Baswade, L. Beltramelli, F. A. Antony, M. Gidlund, B. R. Tamma, and L. Guntupalli, "Modelling and analysis of Wi-Fi and LAA coexistence with priority classes," in *Proc. of WiMob'18 Conference*, Oct 2018, pp. 1–8.

[28] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 3, pp. 535–547, Mar. 2000.

[29] Z. ning Kong, D. H. K. Tsang, B. Bensaou, and D. Gao, "Performance analysis of IEEE 802.11e contention-based channel access," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 10, pp. 2095–2106, Dec. 2004.

[30] "Further advancements for E-UTRA physical layer aspects," 3GPP TS. 36.814 V9.2.0, Mar. 2017.

[31] L. Kleinrock, *Queueing systems, volume 2: Computer applications*. wiley New York, 1976, vol. 66.

[32] "Scipy optimization and root finding," [Available online] https://docs. scipy.org/doc/scipy/reference/generated/scipy.optimize.curve_fit.html.

[33] "Csim20," [http://www.mesquite.com], accessed: 2017-07-15.

[34] S. M. Ross, *Stochastic processes*, 2nd ed. John Wiley & Sons New York, 1996.

APPENDIX A
PROOF OF PROPOSITION 1

Proof: Our proof is based on the expected 'return time' for discrete time Markov chain (DTMC) [34]. The average return time to an arbitrary state, say state *s*, is the average number of transitions until we revisit state *s*. For irreducible and recurrent DTMC, the expected return time to any state can be obtained by taking the inverse of its steady state probability [34]. The DTMC shown in Figure 4 describes the behavior of channel access for CAT4-LBT PCs and EDCA ACs. We

can think of the average contention delay as the expected return time to one of the states of successful transmission, e.g., $(-2, 0, \lceil T_i^{(l)}/T_{\text{slot}} \rceil)$, minus the time consumed in successful transmission, i.e., TXOP duration $T_i^{(l)}$. Let $\pi_{i,(-2,0,\lceil T_i^{(l)}/T_{\text{slot}} \rceil)}$ be the steady state probability of state $(-2, 0, \lceil T_i^{(l)}/T_{\text{slot}} \rceil)$. Then, the average contention delay for frames of the PC P_i can be written as:

$$D_i^{(l)} = \frac{1}{\pi_{i,(-2,0,\lceil T_i^{(l)}/T_{\text{slot}} \rceil)}} T_{\text{slot}} - T_i^{(l)} \qquad (15)$$

By chain regularity, we can express the steady state probability $\pi_{i,(-2,0,\lceil T_i^{(l)}/T_{\text{slot}} \rceil)}$ as a function of the steady state probability for the Channel-Access-Attempt states (i.e., $\pi_{i,(j,0,0)}$ for $j \in [0, R_i]$), as follows:

$$\pi_{i,(-2,0,\lceil T_i^{(l)}/T_{\text{slot}} \rceil)} = \sum_{j=0}^{R_i} (1 - p_i^{(l)}) \pi_{i,(j,0,0)} = (1 - p_i^{(l)}) \tau_i^{(l)} \qquad (16)$$

where $\tau_i^{(l)} = \sum_{j=0}^{R_i} \pi_{i,(j,0,0)}$ is the probability of channel access. PST can be expressed as $p_{s,i}^{(l)} = (1 - p_i^{(l)}) \tau_i^{(l)}$. Substituting Equation (16) in (15) results in the average contention delay of Equation (4). ∎

APPENDIX B
PROOF OF PROPOSITION 2

Proof: We follow the same approach to derive Equations (6) and (7). For brevity, we explain the derivation of Equation (6). We focus on the collision probability $p_i^{(l)}$ experienced by an SBS, say SBS s_i, that wishes to send a frame of priority class P_i in slot t. Consider an event e_i in which all devices, other than SBS l_i, abandon accessing the channel. Let x_i be the probability of event e_i. Let n_i and m_j be the numbers of LTE and Wi-Fi devices that wish to transmit frames of priority class P_i and A_j, respectively, during slot t. An LTE device, aiming to send a frame of the priority class P_i, defers accessing the channel during slot t with probability $1 - \tau_i^{(l)}$, and, similarly, a Wi-Fi device, serving a frame of the priority class A_j, defers accessing the channel with probability $1 - \tau_j^{(w)}$. Therefore, we can express x_i by multiplying the probabilities of deferring access for all SBSs and APs devices, other than SBS s_i, as follows:

$$x_i = (1 - \tau_i^{(l)})^{n_i - 1} \prod_{j \in \mathcal{L}, j \neq i} (1 - \tau_j^{(l)})^{n_j} \prod_{j \in C} (1 - \tau_j^{(w)})^{m_j} \qquad (17)$$

A collision happens at time slot t when at least one or more devices, irrespective of their channel access technology, i.e., LTE or Wi-Fi, transmits at the same time with SBS s_i. This event is also complementary to the event of having no collision, which happens when SBS s_i is the only device accessing the channel or no other device accesses the channel, i.e., channel remains idle for this time slot. Therefore, the probability of collision $p_i^{(l)}$ experienced by SBS s_i can be expressed as:

$$p_i^{(l)} = 1 - (\overbrace{\tau_i^{(l)} x_i}^{\text{Only SBS } s_i \text{ accesses channel}} + \overbrace{(1 - \tau_i^{(l)}) x_i}^{\text{No device accesses channel}})$$

$$= 1 - (1 - \tau_i^{(l)})^{n_i - 1} \prod_{j \in \mathcal{L}, j \neq i} (1 - \tau_j^{(l)})^{n_j} \prod_{j \in C} (1 - \tau_j^{(w)})^{m_j} \quad (18)$$

and this expression matches Equation (6). ∎

APPENDIX C
PROOF OF PROPOSITION 3

Proof: Irreducible and aperiodic discrete time Markov chain (DTMC) has a unique steady state distribution, and the sum of its steady state probabilities should be one. In addition, the steady state probability for an arbitrary state can be written as a function of the steady state probabilities of its connecting states using chain regularity. Probability of channel access τ_i can be computed by adding the steady state probability of Channel-Access-Attempt states, i.e., $\{(J_i, 0, 0) : J_i \in [0, R_i]\}$:

$$\tau_i = \sum_{j=0}^{R_i} \pi_{i,(j,0,0)} \qquad (19)$$

where

$$\pi_{i,(j,0,0)} = p_i^j \pi_{i,(0,0,0)} \qquad (20)$$

By substituting (20) in (19), we get:

$$\tau_i = \frac{1 - p_i^{R_i+1}}{1 - p_i} \pi_{i,(0,0,0)} \qquad (21)$$

The steady state probability for all other states can be expressed as (based on chain regularity):

$$\pi_{i,(-3,0,0)} = \pi_{i,(R_i,0,0)} p_i (1 - q_i) + \pi_{i,(-2,0,1)} (1 - q_i) + \pi_{i,(-3,0,0)} (1 - g_i)$$

$$= \frac{(1 - q_i)(1 + p_i^{R_i} - p_i^{R_i+1})}{g_i} \pi_{i,(0,0,0)} \qquad (22)$$

and similarly we derive the probability for other states:

$$\pi_{i,(j,0,l)} = p_i \pi_{i,(j,0,0)}, j \in [0, R_i], l \in [1, C_i] \qquad (23)$$

$$\pi_{i,(j,k,0)} = \frac{W_{i,j} - k}{W_{i,j}} \pi_{i,(j,0,0)}, j \in [0, R_i], k \in [1, W_{i,j} - 1] \quad (24)$$

$$\pi_{i,(j,k,l)} = \frac{b_i}{(1 - b_i)^{d_i}} \pi_{i,(j,k,0)},$$
$$j \in [0, R_i], k \in [1, W_{i,j} - 1], l \in [d_i, d_i + B] \qquad (25)$$

$$\pi_{i,(j,k,l)} = \frac{b_i}{(1 - b_i)^l} \pi_{i,(j,k,0)},$$
$$j \in [0, R_i], k \in [1, W_{i,j} - 1], l \in [1, d_i - 1] \qquad (26)$$

$$\pi_{i,(-1,0,l)} = \frac{1}{(1 - b_i)^l} \pi_{i,(0,0,0)}, l \in [1, d_i] \qquad (27)$$

$$\pi_{i,(-1,0,l)} = \frac{1 - (1 - b_i)^{d_i}}{(1 - b_i)^{d_i}} \pi_{i,(0,0,0)}, \quad l \in [d_i + 1, d_i + B] \quad (28)$$

$$\pi_{i,(-2,0,l)} = (1 - p_i^{R_i+1})\pi_{i,(0,0,0)}, \, l \in [1, \lceil T_i/T_{\text{slot}}\rceil], \quad (29)$$

We sum the steady state probability of all states in (20) and (22)-(29) to one [34], and solve for the steady state probability $\pi_{0,0,0}^{(i)}$ of state $(0,0,0)$:

$$1 = \pi_{i,(-3,0,0)} + \sum_{d=1}^{\lceil T_i/T_{\text{slot}}\rceil} \pi_{i,(-2,0,d)} + \sum_{l=1}^{d_i+B} \pi_{i,(-1,0,l)}$$

$$+ \sum_{j=0}^{R_i}\sum_{l=1}^{C_i} \pi_{i,(j,0,l)} + \sum_{j=0}^{R_i}\sum_{k=1}^{W_{i,j}-1}\sum_{l=0}^{d_i+B} \pi_{i,(j,k,l)}, \quad (30)$$

$$\pi_{i,(0,0,0)} = \left[\frac{1 + b_i B}{b_i}\frac{1 - (1-b_i)^{d_i}}{(1-b_i)^{d_i}} + \lceil T_i/T_{\text{slot}}\rceil(1 - p_i^{R_i+1}) \right.$$

$$+ (1 + p_i C_i)\frac{1 - p_i^{R_i+1}}{1 - p_i} + \frac{(1-q_i)(1 + p_i^{R_i} - p_i^{R_i+1})}{g_i}$$

$$\left. + \frac{1 + b_i B}{2(1-b_i)^{d_i}}\sum_{j=0}^{R_i} W_{i,j} p_i^j \right]^{-1}$$
$$(31)$$

and finally substituting (31) in (21) results in (9). ∎

APPENDIX D
PROOF OF PROPOSITION 4

Proof: We take the logarithm of PST expression in (10):

$$\log(p_{s,i}^{(l)}) = \log(\tau_i^{(l)}) + (n_i - 1)\log(1 - \tau_i^{(l)})$$
$$+ \sum_{j\in\mathcal{L}, j\neq i} n_j \log(1 - \tau_j^{(l)}) + \sum_{k\in C} m_k \log(1 - \tau_k^{(w)}) \quad (32)$$

The logarithm is monotonically increasing on the period $[0,1]$, therefore, we write the following approximate expression based on (32):

$$p_{s,i}^{(l)} \approx c_{i,0}^{(l)} + \sum_{j\in\mathcal{L}} c_{i,j}^{(l)}(n_j + 1)\log(1 - \tau_j^{(l)})$$
$$+ \sum_{k\in C} h_{i,k}^{(l)}(m_k + 1)\log(1 - \tau_k^{(w)}) \quad (33)$$

where $c_{i,0}^{(l)}$ is a constant, $c_{i,j}^{(l)}$ and $h_{i,k}^{(l)}$ are proportionality coefficients that we can obtain by fitting the formula in (33) with PST measurements. We add ones in $(n_j + 1)$ and $(m_k + 1)$ to ensure that the fitting process is well-behaved and does not diverge for scenarios in which n_j or m_k could be zeros. Based on extensive results that we obtained by solving for $\tau_i^{(l)}$ in (9) numerically, as well as carrying system-level simulations, we notice that the increase in the number of coexisting nodes reduces the probability of channel access, i.e., $n_i \uparrow$ leads to $\tau_i^{(l)} \downarrow$. Similarly, increasing the size of the minimum contention window, AIFSN, and TXOP, also reduces the probability of channel access, i.e., $W_{i,0} \uparrow, d_i^{(l)} \uparrow, T_i^{(l)} \uparrow$ leads to $\tau_i^{(l)} \downarrow$. In light of the these observations, we formulate the following expression to approximate $1 - \tau_i^{(l)}$ (a similar expression can be formulated for $1 - \tau_k^{(w)}$):

$$1 - \tau_i^{(l)} \approx \beta_i^{(l)}(W_{i,0}^{(l)}d_i^{(l)} + \lceil T_i^{(l)}/T_{\text{slot}}\rceil)n_i + e_i^{(l)}, \quad (34)$$

where $\beta_i^{(l)}$ is a proportionality coefficient and $e_i^{(l)}$ is a constant. Proper values for $\beta_i^{(l)}$ and $e_i^{(l)}$ can be obtained by fitting the expression with PST measurements. We have also noticed that setting of AIFSN affects τ_i more than the size of the minimum contention window. A similar observation has also been captured by author in [25]. Our choice of multiplying d_i with $W_{i,0}^{(l)}$, as well as multiplying $(W_{i,0}^{(l)}d_i^{(l)} + \lceil T_i^{(l)}/T_{\text{slot}}\rceil)$ with n_i in (34) is based on making a better distinction between PCs and ACs with respect to their parameters and the number of their serving transmitters. By substituting (34) in (33), we obtain the approximate expression in (13). ∎