IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 24, 2018, accepted July 26, 2018, date of publication August 10, 2018, date of current version September 5, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2864679

AP-STA Association Control for Throughput Maximization in Virtualized WiFi Networks

MAHSA DERAKHSHANI[1], (Member, IEEE), XIAOWEI WANG[2], DANIEL TWEED[2], (Student Member, IEEE), THO LE-NGOC[2], (Fellow, IEEE), AND ALBERTO LEON-GARCIA[3], (Fellow, IEEE)

[1]Wolfson School of Mechanical, Electrical and Manufacturing Engineering, Loughborough University, Loughborough LE11 3TU, UK
[2]Department of Electrical and Computer Engineering, McGill University, Montreal, QC H3A 0E9, Canada
[3]Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 3G4, Canada

Corresponding Author: Mahsa Derakhshani (m.derakhshani@lboro.ac.uk)

This work was supported by various Natural Sciences and Engineering Research Council of Canada (NSERC) grants.

ABSTRACT To manage and enable service customization among multiple internet service providers (ISPs) sharing the common physical infrastructure and network capacity in virtualized Wi-Fi networks, this paper models and optimizes access point-station (STA) association via airtime usage control. More specifically, an optimization problem is formulated on the STAs' transmission probabilities to maximize the overall network throughput, while providing airtime usage guarantees for the ISPs. As the proposed optimization problem is inherently non-convex, an algorithm to reach the optimal solution is developed by applying monomial approximation and geometric programming iteratively. Based on the proposed 3-D Markov-chain model of the enhanced distributed channel access protocol, the detailed implementation of the optimal transmission probability of each STA is also discussed by manipulating medium access control parameters. The performance of the developed association and airtime control scheme is evaluated through numerical results. For both homogeneous and non-homogeneous STA distributions, numerical results reveal performance gains of the proposed algorithm in improving the throughput and keeping airtime usage guarantees.

INDEX TERMS Wireless local area networks (WLANs), IEEE 802.11 standard, AP-STA association, airtime control, throughput, Markov chain, geometric programming, virtualized wireless networks.

I. INTRODUCTION

A. BACKGROUND AND MOTIVATION

To support the growing demand for mobile data traffic, the idea of heterogeneous networks (HetNets) has ensued, aiming to improve the network capacity and coverage. Such improvement can be enabled in HetNets by offloading traffic from the existing macro-cellular network to short-range small cells. To deploy such small-cells, WiFi offloading appears to be a promising candidate due to WiFi's success and its low-cost deployment [1]. However, especially in congested scenarios, IEEE 802.11-based WiFi networks struggle with guaranteeing connectivity and quality-of-service (QoS) provisioning because of inevitable collisions occurring in its contention-based medium access control (MAC). [2].

Recently, software-defined wireless networking (SDWN), along with wireless vitualization, has emerged as an architectural choice for wireless networks [3]–[5]. The principle in this architecture is the separation of control and data planes in the network. This separation provides the capacity to abstract and share resources from the infrastructure level as well as to deploy centralized control, in which all transmission parameters and connections can be actively manipulated based on requested services, user conditions, and available resources.

To overcome the challenges in 802.11 networks, wireless virtualization and delegation of management rights to a centralized controller can be useful to guarantee connectivity, support customized services with finer control over QoS, and balance the traffic load among different access points (APs). There have been several works in the literature proposing *network architectures* and *designs* to enable WiFi virtualization (e.g., [2], [6], [7]).

In virtualized 802.11-based WiFi networks, where physical infrastructure and wireless resources are shared by different internet service providers (ISPs), one key issue is to provide isolation among virtual networks run by different ISPs. Using a random access protocol, i.e., carrier sense multiple access (CSMA), unavoidable collisions act to couple the transmissions of different virtual WLANs (V-WLANs). Moreover, since the network capacity is shared yet constrained, the increase of traffic in one V-WLAN may reduce

TABLE 1. Summary of abbreviations.

ACK	Acknowledgement
AIFS	Arbitration Inter-frame Space
AP	Access Point
BSS	Basic Service Set
CSMA/CA	Carrier Sense Multiple Access with Collision Avoidance
CW	Contention Window
DCF	Distributed Coordination Function
EDCA	Enhanced Distributed Channel Access
GP	Geometric Programming
HetNet	Heterogeneous Network
ISP	Internet Service Provider
MAC	Medium Access Control
QoS	Quality of Service
RAT	Radio Access Technology
SDWN	Software-defined Wireless Networking
SIFS	Short Inter-frame Spacing
SNR	Signal-to-Noise Ratio
STA	Station
TXOP	Transmission Opportunity
WLAN	Wireless Local Area Network
V-WLAN	Virtual WLAN

the available network capacity to another [8]. Thus, an efficient resource allocation among V-WLANs is essential to manage the MAC-layer couplings.

In multi-cell or multi-tier wireless networks, before a STA can access the network, a decision needs to be made about which AP to associate with. Such STA-AP association control could create an opportunity to provide fairness guarantees among different ISPs in V-WLANs. In addition to STA-AP association, the airtime among STAs associated to the same AP can optimally be adjusted. Controlling airtime usage of the STAs provides the opportunity to optimize the V-WLAN performance (e.g., improving throughput by exploiting multi-user diversity) as well as another degree of freedom to guarantee fairness among the ISPs.

To overcome MAC-layer couplings and balance the load, in this paper,[1] we propose a network-originated association control algorithm in virtualized multi-cell WiFi networks. In the context of *resource management* for virtualized WiFi networks, there are a few works addressing only airtime control in the literature [8], [10], [11]. But, in this context, to the best of our knowledge, the *association control* in *virtualized multi-cell* 802.11 networks has not been studied, which we show in this paper can provide another degree of freedom for finer service customization.

More specifically, in this work, AP-STA association and airtime control are *jointly* explored to provide fairness and throughput guarantees for different V-WLANs. Taking into account STA transmission rates and ISP airtime reservations, an optimization problem is formulated to adjust the transmission probability of each STA at each AP. The objective is to maximize the overall network throughput, while keeping a total airtime guarantee for each ISP. To solve the formulated problem which is non-convex and thus computationally intractable, an iterative algorithm is developed through successive geometric programming. This algorithm can achieve an optimal solution with an affordable complexity.

Furthermore, to implement the optimal transmission probabilities, we model the EDCA protocol with a three-dimensional Markov chain and establish the relationship between the transmission probability of a STA and the EDCA parameters. Based on this relationship, we developed a control algorithm to approach the optimal transmission probability by jointly manipulating the EDCA parameters such as contention window (CW) size and arbitration inter-frame space (AIFS). Finally, through numerical results, we verify the performance enhancement provided by the developed AP-STA association and airtime control approach in terms of throughput and airtime usage guarantees.

B. RELATED WORKS

1) AP-STA ASSOCIATION AND AIRTIME CONTROL

There exists a large body of work on the resource allocation (e.g., [12], [13]) and user association problem in wireless networks, with applications to WLAN, V-WLAN, and multi-cell/multi-tier cellular networks, including radio access technology (RAT) selection and association in the offloading scenario. Reference [14] presents a survey of user association techniques in cellular networks, specifically towards fifth generation network technologies such as HetNet, multiple-input multiple-output (MIMO), and millimeter wave. Many works take a distributed approach, wherein each STA determines which AP to associated with, for example, reference [15] dealing with multi-RAT HetNets, [16] addressing user association in HetNet for outage and QoS guarantees, reference [17] and [18] which focus on mm-wave. In 802.11 WiFi, the focus of this work, due to the probabilistic nature of the multiple access mechanism, namely CSMA/CA, different techniques to exploit user association for load balancing and fairness are required.

In most current vendor implementations, 802.11 STAs choose the AP with the highest received signal-to-noise ratio (SNR) to connect with. Since the STA density is often uneven in the network [19], [20], the Max-SNR approach can lead to an unbalanced distribution of STAs among APs, causing unfairness. In order to balance the load of APs, several AP-STA association algorithms have been presented in the literature, mostly by maximizing the minimum throughput of all STAs [21]–[24]. Nevertheless, in a basic service set (BSS) including an AP and its associated multi-rate STAs, it is shown that the throughput is limited by the STA with the lowest data rate. This phenomenon is also to as the *performance anomaly* problem [25]. Thus, comparing with the Max-SNR approach, these load-balancing approaches improve the

[1] Parts of this paper have been presented in [9]

max-min fairness among STAs at the cost of decreasing the aggregate throughput.

To address the *performance anomaly* and balance the trade-off between aggregate throughput and fairness, proportional fair throughput allocation has widely been considered in multi-rate 802.11 WLANs [26]–[30]. In [26], proportional fairness is studied in a single BSS. It is shown that *proportional fairness* leads to an *airtime-fairness*, where equal airtime usage is provided to all STAs. Moreover, in a multi-AP WLAN, [28], [29] study AP-STA association problem with an objective to maximize the proportional fairness. More precisely, association control is implemented in the form of airtime allocation, where the transmission time of STAs at different APs are jointly optimized [28], [29].

Precise control of the STA airtime usage in a 802.11 contention-based WLAN is difficult due to the distributed and random nature of the CSMA-based MAC protocol. Nevertheless, the airtime control algorithms have been proposed in the literature by manipulating the MAC parameters [31]–[34]. These discussions are limited to heuristically adjusting the parameters *separately* and thus can only find a suboptimal solution in a subset of the feasibility region. In this work, we develop an airtime control algorithm by *jointly* manipulating the EDCA parameters.

In a virtualized WLAN serving multiple ISPs, AP-STA association and airtime control become more challenging. The reason is that fairness guarantees and service customization are required for each ISP while there are unavoidable couplings among the STA transmissions of different ISPs in the network. There are a few works addressing only airtime control in the literature. Considering a virtualized single-cell WLAN, in [10], a heuristic airtime control algorithm is proposed to achieve the target airtime usage for each ISP by controlling the minimum CW of each STA. Similarly, [11] addresses optimizing CW using control theory however, as the discussion is limited to controlling minimum CW, the optimality of the result might be sacrificed. In a multi-cell WLAN, [8] presents an analysis on the feasibility region of ISP airtimes and characterizes the ISP airtimes at the rate region boundary. Furthermore, a distributed algorithm is developed to allocate airtime slices among ISPs and flow rates within each slice in a max-min fair manner. Since max-min fairness is used as an objective for rate allocation among the flows in each ISP, the optimality of the achieved total throughput may not be guaranteed in a multi-rate WLAN. In addition, the association control is not discussed in [8].

2) EDCA MODELING AND OPTIMIZATION

To implement the optimal AP-STA association and airtime allocation, it is desirable to control the transmission probability of each STA. However, in a CSMA/CA-based WLAN, the only controllable parameters are the MAC layer parameters. Thus, it is essential to mathematically model the effects of such parameters on the transmission probabilities of STAs.

The EDCA modeling has received considerable attention in the literature. There are several studies, such as [35], [36], that only focus on numerically solving the stationary state, i.e., transmission probability of each STA, given the network configuration. Another group of works, e.g., [37]–[44], provide models for the EDCA protocol with the help of a Markov chain, inspired by the classical work of Bianchi [45]. Using such Markov chain models, it is possible to optimize the network by tuning the EDCA parameters. However, due to the complexity of the proposed EDCA models, it is hard to establish an explicit relationship between the transmission probabilities and the controllable EDCA parameters. Thus, these works mostly provide a numerical, e.g., [37]–[40], or an approximate analytical, e.g., [41]–[44], solution in terms of MAC parameters. Moreover, none of these works has jointly controlled the EDCA parameters.

The proposed Markov chain model is based on [46], with which an explicit relationship between the transmission probability and EDCA MAC parameters can be established. One limitation of the EDCA model proposed in [46] is the accuracy of AIFS differentiation. Unlike in DCF, the time duration each STA has to wait before its backoff process (i.e. AIFS) is different in EDCA. This can lead to the result that the number of contending STAs is not time-homogeneous [36]. Therefore, the collision probability that each STA faces is also not time-homogeneous. But, as pointed out in [42], time-homogeneity assumption greatly simplifies the modeling complexity, and thus, the steady state performance can be characterized as explicit functions of backoff parameters. Furthermore, the accuracy of such model can be effectively improved by setting a relatively large initial CW.

C. STRUCTURE

The rest of this paper is organized as follows. Section II presents an overview of the system configuration and modeling. In Section III, we first analyze the feasibility region of the transmission probabilities based on the proposed Markov chain model for IEEE EDCA. Then, we formulate the transmission probability optimization problem, which maximizes the system throughput and guarantees customized airtime usage for each ISP. In Section IV, the implementation details of MAC parameter control are discussed in order to achieve the optimal transmission probability. Illustrative results are provided in Section V to evaluate the performance of the developed algorithms. Section VI provides some concluding remarks.

II. SYSTEM CONFIGURATION AND MODELING

We consider an IEEE 802.11-based WLAN that consists of a large number of APs. APs operate on non-overlapping frequency channels. Let $\mathcal{A}$ be the set of APs and $N_a = |\mathcal{A}|$ be the total number of APs. Each AP has a limited coverage area and all STAs are randomly distributed in the field. The network carries traffic belonging to a number of different ISPs (also referred to as V-WLANs). Let $\mathcal{K}$ be the set of ISPs using the network. Furthermore, let $\mathcal{S}_k$ be the set of STAs of ISP $k \in \mathcal{K}$ and $N_k = |\mathcal{S}_k|$ be the number of STAs belong to ISP k. Furthermore, let $\mathcal{S}$ be the set of all STAs and $N_s = \sum_{k\in\mathcal{K}} N_k$

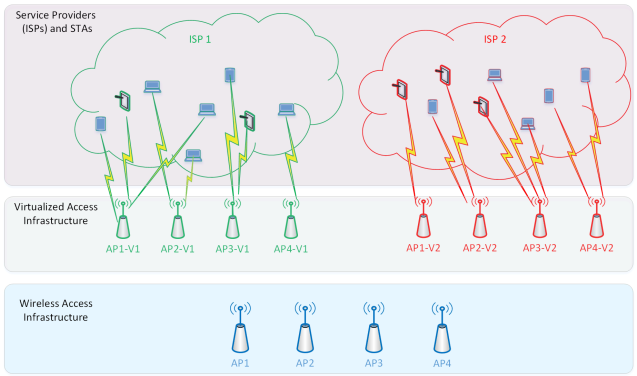


FIGURE 1. Layered system model.

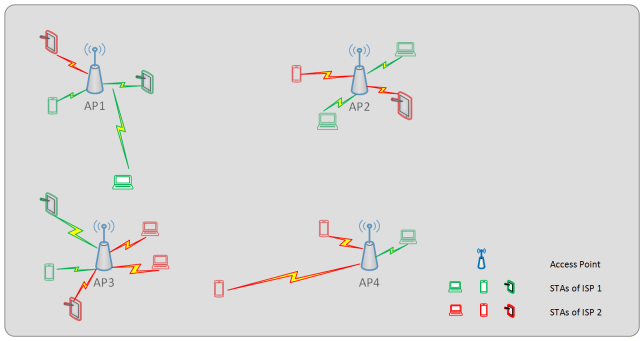


FIGURE 2. System model.

be the total number of STAs in the network. The network is administratively virtualized, i.e., each AP will broadcast multiple different SSIDs, one for each ISP. Figure 1 and 2 illustrate an example of the network architecture with four physical APs and two ISPs.

A. ENHANCED DISTRIBUTED CHANNEL ACCESS (EDCA)

To access the channel, the STAs are assumed to follow the EDCA MAC protocol of 802.11e standard. EDCA protocol is designed to enhance the basic MAC mechanism, i.e., distributed coordination function (DCF), aiming to support service differentiation. Similar to DCF, EDCA is also a contention-based access scheme, based on CSMA/CA using binary exponential backoff rules to manage retransmission of collided packets. However, in contrast to DCF, EDCA provides service differentiation by tuning contention parameters to impact probability of accessing the channel [45], [47], [48].

In DCF, STAs wait until the channel has been sensed idle for a DCF inter-frame space (DIFS) and then back off from the channel for a random number of slots selected from the range $[0, W_i]$ where W_i represents the contention window of STA $i \in \mathcal{S}$ [45]. Initially, at backoff stage $m_i = 0$, W_i is set equal to the minimum contention window size $W_{\min,i}$. Then, after each unsuccessful transmission, STA i moves to the next backoff stage and W_i is doubled. When the maximum backoff stage m_i is reached, W_i is no longer increased and stays at $2^{m_i} W_{\min,i}$. If the STA experiences a collision at the maximum backoff stage, it will retry transmission for at most h_i times. If the data frame is successfully received, the AP waits for a period of time called short inter-frame space (SIFS) and then sends an acknowledgement (ACK).

TABLE 2. Summary of notations and variables.

Variable	Definition
$\mathcal{A}$	Set of APs
N_a	Total number of APs
$\mathcal{K}$	Set of ISPs
$\mathcal{S}$	Set of STAs
$\mathcal{S}_k$	Set of STAs of ISP $k \in \mathcal{K}$
N_k	Number of STAs of ISP k
N_s	Total number of STAs
L_i	Long inter-frame space for STA i
m_i	Maximum backoff stage of STA i
h_i	Retransmission limit of STA i at the maximum backoff stage m_i
r_i^a	Transmission data rate, in bits per second
q_i	Probability that STA i enters the backoff process
p_i	Conditional collision probability faced by STA i
γ	Propagation delay
T_c	Duration of a collision
T_s	Duration of a successful transmission
$T_{air,i}^a$	Total access airtime for STA i at AP a
T_{TXOP}	Duration of a data frame
x_i^a	Expected number of consecutive transmission attempts of STA i at AP a
P_{idle}^a	Stationary probability a time-slot is idle
$P_{\text{succ},i}^a$	Stationary probability of successful transmission for STA i at AP a
$P_{\text{coll},i}^a$	Stationary probability of colliding transmission for STA i at AP a
η_k	Target share of airtime for ISP k
I_g	Transmitted information, in bits
A_i	The AIFS of STA i minus 1
W_i	Contention window size of STA i
$W_{\min,i}$	Minimum contention window size of STA i
τ_i^a	Probability that STA i attempts to transmit to AP a
$\underline{\tau}_i^a$	Lower bound of τ_i^a
$\bar{\tau}_i^a$	Upper bound of τ_i^a
N	Average frozen time, in slots

To support service differentiation, EDCA defines access categories with different service level priorities. Each access category has different inter-frame space waiting times, called arbitration inter-frame spaces (AIFS), and contention parameters, i.e., minimum contention window $W_{\min}$ and maximum backoff stage m_i, which impact the maximum contention window size. Combined, these differing parameters impact the probability of successful transmission. Further, for each transmission opportunity (TXOP) in EDCA a STA can transmit multiple back-to-back packets for a fixed period of time. Figure 3 illustrates an example of the channel-access

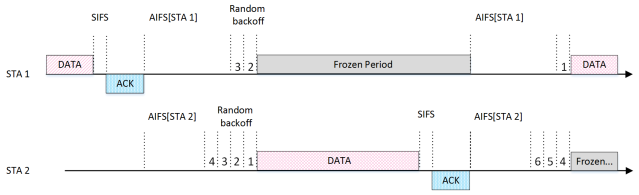


FIGURE 3. 802.11e EDCA channel access procedure.

procedure of two STAs using EDCA [47], [48]. Let T_{TXOP} be the duration of a transmission opportunity. Then, the duration of a successful EDCA transmission becomes

$$T_s = T_{\text{TXOP}} + \text{SIFS} + \gamma + \text{ACK} + \gamma + \text{AIFS} \tag{1}$$

where γ denotes the propagation delay. Similarly, the duration of a collision can be calculated as

$$T_c = T_{\text{TXOP}} + \gamma + \text{AIFS}. \tag{2}$$

Note that when the colliding STAs use different AIFS values, the AIFS value in (2) should take the value of the largest AIFS. But, since the fixed transmission duration T_{TXOP} (in the order of *ms*) is much larger than AIFS (in the order of μs), the difference between the AIFS can be ignored here.

B. ASSOCIATION CONTROL VIA TRANSMISSION PROBABILITIES

In a WLAN with APs densely deployed, STAs need to determine which APs to connect with. We aim to generalize the association control problem by adjusting the transmission probability of each STA at any AP, which can be controlled by configuration of MAC parameters. Thus, we define τ_i^a ($0 \leq \tau_i^a \leq 1$) as the probability that STA i attempts to transmit at AP a in a general time-slot.[2] Consequently, the stationary probability that a time-slot is idle in the BSS including AP a is

$$P_{\text{idle}}^a = \prod\nolimits_{i\in\mathcal{S}}(1-\tau_i^a). \tag{3}$$

In a given BSS, transmitted packets will be received successfully, if exactly one STA transmits on the channel. Thus, the stationary probability of a successful transmission initiated by STA i becomes

$$P_{\text{succ},i}^a = \tau_i^a \prod\nolimits_{i'\in\mathcal{S},i'\neq i}(1-\tau_{i'}^a). \tag{4}$$

Since ACK, SIFS, and also γ are significantly smaller (in the order of μs) than the fixed transmission duration T_{TXOP} (in the order of ms), we approximate T_s and T_c to be of the same size and denote them by T. In particular, the propagation delay (the time for wireless signals to travel from transmitter to receiver) is negligible here since it is typically less than one μs as compared to T_{TXOP}. Consequently, the expected length of a general time-slot becomes

$$\mathbb{E}\{T_g\} = \delta P_{\text{idle}}^a + (1-P_{\text{idle}}^a)T \tag{5}$$

[2]We focus on the uplink scenario. For downlink transmission, the AP acts as a STA and accesses the channel in the same manner, thus the solution developed for the uplink can be directly applied to the AP.

where δ is the duration of an idle time-slot. Furthermore, the expected information (in bits) transmitted by STA i to AP a in a general time-slot can be derived as

$$\mathbb{E}\{I_g\} = P_{\text{succ},i}^a r_i^a T_{\text{TXOP}} \tag{6}$$

where r_i^a represents the transmission data rate of the link between STA i and AP a. As defined in [45], based on the (5) and (6), the throughput of STA i at AP a becomes

$$T_i^a = \frac{\mathbb{E}\{I_g\}}{\mathbb{E}\{T_g\}} = \frac{P_{\text{succ},i}^a r_i^a T_{\text{TXOP}}}{P_{\text{idle}}^a \delta + (1-P_{\text{idle}}^a)T}. \tag{7}$$

Let define a new variable $x_i^a = \frac{\tau_i^a}{1-\tau_i^a}$ ($x_i^a \geq 0$), which represents the expected number of consecutive transmission attempts by STA i at AP a as [8] and [26]. Consequently, P_{idle}^a and $P_{\text{succ},i}^a$ will be transformed into

$$P_{\text{idle}}^a = \frac{1}{\prod_{i\in\mathcal{S}}(1+x_i^a)}, \tag{8}$$

$$P_{\text{succ},i}^a = \frac{x_i^a}{\prod_{i'\in\mathcal{S}}(1+x_{i'}^a)} = x_i^a P_{\text{idle}}^a. \tag{9}$$

Subsequently, from (8) and (9), T_i^a can be represented in terms of x_i^a as

$$T_i^a = \frac{x_i^a P_{\text{idle}}^a r_i^a T_{\text{TXOP}}}{T-(T-\delta)P_{\text{idle}}^a} = \frac{x_i^a r_i^a t}{\prod_{i'\in\mathcal{S}}(1+x_{i'}^a)-t'} \tag{10}$$

where $t = \frac{T_{\text{TXOP}}}{T}$ and $t' = \frac{T-\delta}{T}$.

In addition to the throughput of each STA, the fraction of time that each STA spends for transmission could be considered as another performance metric, specifically in order to measure and preserve service customization among different STAs or ISPs. The total access airtime for STA i, including both successful transmissions and collisions, becomes

$$T_{\text{air},i}^a = \frac{P_{\text{coll},i}^a T + P_{\text{succ},i}^a T}{P_{\text{idle}}^a \delta + (1-P_{\text{idle}}^a)T}, \tag{11}$$

where

$$P_{\text{coll},i}^a = \tau_i^a\Big[1 - \prod_{i'\in\mathcal{S},i'\neq i}(1-\tau_{i'}^a)\Big]$$

is the stationary probability that STA i attempts to transmit at the same time as another STA in a general time-slot and a collision occurs, as defined in [26]. Consequently,

$$T_{\text{air},i}^a = \frac{\tau_i^a}{1-P_{\text{idle}}^a t'} = \frac{x_i^a \prod_{i'\in\mathcal{S},i'\neq i}(1+x_{i'}^a)}{\prod_{i'\in\mathcal{S}}(1+x_{i'}^a)-t'}. \tag{12}$$

In this work, we aim to maximize the overall network throughput, while guaranteeing a minimum requirement on the aggregate airtime of each ISP. To this end, the transmission probability of STAs (τ_i^a) needs to be adaptively optimized by maximizing the aggregate throughput of all STAs at all APs, i.e.,$\sum_{i\in\mathcal{S},a\in\mathcal{A}} T_i^a$. Furthermore, for each ISP (e.g., ISP k), a constraint needs to be set in order to keep the total airtime of all STAs belonging to ISP k larger than a minimum requirement. More specifically, $\sum_{i\in\mathcal{S}_k,a\in\mathcal{A}} T_{\text{air},i}^a \geq \eta_k$ where η_k denotes the target share of the airtime for ISP k.

Accordingly, to formulate such optimization problem, the feasibility region of τ_i^a (or x_i^a) is required. Thus, we study the behavior of a single STA, which is using EDCA, with a three-dimensional Markov chain. With the aid of the proposed Markov model, we can learn how to implement or control τ_i^a in terms of EDCA parameters. As a result, based on the established relationship between τ_i^a and EDCA parameters, we would be able to analyze its feasibility region and also design an algorithm to control EDCA parameters to approach the optimal τ_i^a.

III. OPTIMIZATION PROBLEM

In this section, we present the AP-STA association and airtime control optimization problem based on the system model introduced in Section II. We first use a Markov chain to model the EDCA protocol. With this model, we can analyze the feasibility region of τ_i^a for each STA. Then, the optimization problem is formed and solved by applying monomial approximation and geometric programming.

A. MARKOV CHAIN MODEL FOR EDCA

To study the feasibility region of τ_i^a and its implementation in the EDCA MAC protocol, we use a three-dimensional Markov chain. Such model enables us to estimate the transmission probability of the STAs in a WLAN employing the EDCA protocol. Different from the previous works, which discuss a single parameter in the EDCA protocol (e.g., [32]), we aim at developing a general mathematical model that can show the influences of all the EDCA parameters on the transmission probability.

Figure 4 shows the three-dimensional Markov chain model to describe the behavior of a single STA, which uses the EDCA MAC protocol. Since STA transmissions are only coupled within a BSS, we consider a single BSS where an AP and its associated STAs reside in. Thus, in this subsection without loss of generality, we remove the index of a (e.g., τ_i^a is replaced by τ_i) to keep the notation simple.

Our proposed three-dimensional Markov chain model is an extension to the presented model in [46]. To be able to further control the competition among STAs, and thus the collision probability in a BSS, we introduce two new MAC parameters q_i and L_i for each STA. Similar to [49], after a successful transmission or a packet drop, the STA has to flip a biased coin with successful probability q_i to enter the backoff process. Whenever the STA fails to enter the backoff process, it needs to wait for a period of long interframe space of L_i time-slots before another try. This flip-coin process helps to better control the contention in the system. These two variables are incorporated in the proposed Markov chain.

It should be noted that time in this Markov chain is slotted and the interval between any two adjacent states is a general time-slot. Here, a general time-slot can be an idle time-slot δ, a successful transmission duration (T_s) or a collision duration (T_c). In this Markov chain, the triple $\{s(t), b(t), v(t)\}$

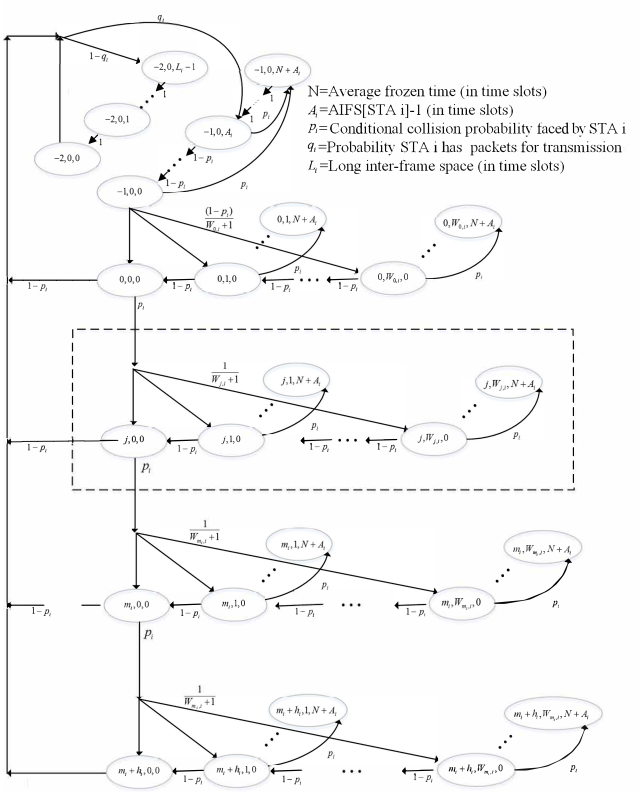


FIGURE 4. Three dimensional Markov chain model for one STA.

denotes the state of the STA at time t. More specifically, $s(t)$ represents the backoff/retransmission stage of STA, where $-2 \le s(t) \le m_i + h_i$ ($s(t) = -2, -1$ will be explained later). Furthermore, $b(t)$ denotes the backoff counter, which takes a value between 0 and the CW of the current backoff stage. Furthermore, $v(t)$ represents the remaining time of a frozen period or the waiting time to the next trial when the STA already failed to enter the backoff process. More specifically, for the latter case, $v(t)$ is set to $L_i - 1$ once the flip coin process fails (with probability $1 - q_i$) to impose extra waiting time until another trial.

The detailed transition probabilities between states of the Markov chain and the stationary condition of the Markov chain can be found in the Appendix. With such stationary conditions, the transmission probability τ of each STA is derived as function of the detailed EDCA parameters as in (13), as shown at the top of the next page, where p_i is the conditional collision probability faced by STA i if a packet were to be transmitted by STA i as defined in (26).

B. FEASIBILITY REGION ANALYSIS

In this subsection, we study the feasibility region of τ_i^a in order to complete the optimization problem formulation. Based on (13), the adjustable variables to tune τ_i^a are m_i^a, h_i^a, L_i^a, q_i^a, A_i^a, and $W_{j,i}^a$ where $0 \le j \le m_i^a + h_i^a$. Recall that $0 \le q_i^a \le 1$, $0 \le L_i^a$, and $0 \le W_{j,i}^a$. Furthermore, suppose that AIFS[STA i] denotes the AIFS of STA i.

$$\tau_i = \sum_{j=0}^{m_i+h_i} b_{j,0,0} = \frac{1-p_i^{m_i+h_i+1}}{1-p_i}\left[L_i\frac{1-q_i}{q_i} + \frac{1+p_iN}{p_i}\frac{1-(1-p_i)^{A_i+1}}{(1-p_i)^{A_i+1}} + \frac{1-p_i^{m_i+h_i+1}}{1-p_i} + \frac{1+Np_i}{2(1-p_i)^{A_i}}\left(W_{\min,i}\left[\frac{1-(2p_i)^{m_i+1}}{1-2p_i} + \frac{2^{m_i}p_i^{m_i+1}(1-p_i^{h_i})}{1-p_i}\right] - \frac{1-p_i^{m_i+h_i+1}}{1-p_i}\right)\right]^{-1} \quad (13)$$

Since A_i = AIFS[STA i] − 1, and AIFS > SIFS in EDCA, we have $A_i \geq$ SIFS. Assuming that SIFS is equal to one time-slot, we have $A_i^a \geq 1$ without loss of generality.

1) LOWER BOUND OF τ_i^a

Let $\underline{\tau}_i^a$ be the lower bound of τ_i^a. From (13), it is clear that $\tau_i^a \to 0$ when $L_i^a \to \infty$, or $q_i^a \to 0$, or $A_i^a \to \infty$. Consequently,

$$\underline{\tau}_i^a = 0. \quad (14)$$

2) UPPER BOUND OF τ_i^a

Let $\bar{\tau}_i^a$ be the upper bound of τ_i^a. To calculate $\bar{\tau}_i^a$, we need to make the denominator of τ_i^a in (13) as small as possible. Thus, we set $L_i^a = 0$ (or $q_i^a = 1$) and $W_{\min,i}^a = 0$, which make the first and fourth terms in the denominator zero. Then, by setting A_i to its lower bound, i.e., $A_i^a = 1$, we can minimize the second term in the denominator of τ_i^a. Consequently, at $L_i^a = 0$, $W_{\min,i}^a = 0$, and $A_i^a = 1$,

$$\tau_i^a = \left[1 + \frac{(1+p_i^aN)(2-p_i^a)}{(1-p_i^a)(1-(p_i^a)^{m_i+h_i+1})}\right]^{-1}. \quad (15)$$

From (15), it is clear that the upper bound of τ_i^a can be achieved when $m_i^a + h_i^a \to \infty$. Thus, $\bar{\tau}_i^a$ can be written as

$$\bar{\tau}_i^a = \left[1 + \frac{(1+p_i^aN)(2-p_i^a)}{1-p_i^a}\right]^{-1}. \quad (16)$$

C. OPTIMIZATION PROBLEM

Taking into account the feasibility region of τ_i^a derived in Section III-B, here, we can mathematically present the transmission probability optimization problem. The objective is to maximize the overall network throughput, while distributing access airtime among different ISPs according to their reservations. More specifically, the optimization can be formulated as

$$\max_{\boldsymbol{X},\boldsymbol{P}} \sum_{i\in\mathcal{S},a\in\mathcal{A}} \frac{x_i^a r_i^a t}{\prod_{i'\in\mathcal{S}}(1+x_{i'}^a) - t'}, \quad (17a)$$

$$\text{subject to,} \sum_{i\in\mathcal{S}_k,a\in\mathcal{A}} \frac{x_i^a \prod_{i'\in\mathcal{S},i'\neq i}(1+x_{i'}^a)}{\prod_{i'\in\mathcal{S}}(1+x_{i'}^a) - t'} \geq \eta_k, \quad \forall k \in \mathcal{K} \quad (17b)$$

$$\frac{x_i^a}{1+x_i^a} \leq \bar{\tau}_i^a, \quad \forall i \in \mathcal{S},\ a \in \mathcal{A} \quad (17c)$$

$$\bar{\tau}_i^a\left(1 + \frac{(1+p_i^aN)(2-p_i^a)}{1-p_i^a}\right) = 1, \quad \forall i \in \mathcal{S},\ a \in \mathcal{A} \quad (17d)$$

$$p_i^a = 1 - \prod_{i'\in\mathcal{S},i'\neq i} 1 - \frac{x_{i'}^a}{1+x_{i'}^a}, \quad \forall i \in \mathcal{S},\ a \in \mathcal{A}. \quad (17e)$$

where $\boldsymbol{X} = [x_i^a]$ ($x_i^a \geq 0$) and $\boldsymbol{P} = [p_i^a]$ ($0 \leq p_i^a \leq 1$). Let us recall that $x_i^a = \frac{\tau_i^a}{1-\tau_i^a}$ and $p_i^a = 1 - \prod_{i'\neq i}(1-\tau_{i'}^a)$. It should be noted that the optimization problem is alternatively formulated with respect to x_i^a instead of τ_i^a, since it will prove useful to solve the problem.

In the preceding optimization problem, the objective function in (17a) represents the overall network throughput, i.e., $\sum_{i\in\mathcal{S},a\in\mathcal{A}} T_i^a$, based on (10). Furthermore, constraints in (17b) guarantee the minimum airtime reservations for all ISPs, i.e., $\sum_{i\in\mathcal{S}_k,a\in\mathcal{A}} T_{\text{air},i}^a \geq \eta_k$, based on (12). This set of constraints enables controlling ISPs' share of access airtime regardless of their number of STAs. To guarantee that the provided solution fits into the feasibility region, in constraint (17c), τ_i^a is limited to its upper bound. Then, the equality constraint (17d) establishes the relationship between the upper bound $\bar{\tau}_i^a$ and p_i^a based on (16). Finally, constraint (17e) establishes the relationship between p_i^a and all $x_{i'}^a$, $\forall i' \in \mathcal{S}$, $i' \neq i$ according to (26).

Due to the non-convex objective function and ISP airtime constraints as well as non-linear equality constraints, the formulated problem is non-convex and thus intractable to solve. However, it potentially looks like an extension of Geometric Programming (GP) (defined in Section VI-B). Thus, by applying successive transformation strategies, we will try to convert the original problem into a series of standard GP problems that can be solved to reach an optimal solution. First, we introduce three auxiliary variables, $y^a = \prod_{i'\in\mathcal{S}}(1+x_{i'}^a) - t'$, $\forall a \in \mathcal{A}$, $u_i^a = 1 - p_i^a$, $\forall i \in \mathcal{S}, \forall a \in \mathcal{A}$, and $t_i^a = 1 + x_i^a$, $\forall i \in \mathcal{S}, \forall a \in \mathcal{A}$. Then, the optimization problem in (17) can be transformed into

$$\min_{\boldsymbol{X},\boldsymbol{T},\boldsymbol{U},\boldsymbol{P},\boldsymbol{Y}} \sum_{i\in\mathcal{S},a\in\mathcal{A}} -\frac{x_i^a r_i^a t}{y^a}, \quad (18a)$$

$$\text{subject to,} \frac{\prod_{i\in\mathcal{S}}(1+x_i^a)}{t'+y^a} = 1, \quad \forall a \in \mathcal{A} \quad (18b)$$

$$\frac{\eta_k+1}{1+\sum_{i\in\mathcal{S}_k,a\in\mathcal{A}} \frac{x_i^a \prod_{i'\neq i} t_i^a}{y^a}} \leq 1, \quad \forall k \in \mathcal{K} \quad (18c)$$

$$\frac{x_i^a}{1+x_i^a} \leq \frac{1}{1+\frac{(1+p_i^a N)(2-p_i^a)}{1-p_i^a}}, \quad \forall i \in \mathcal{S}, a \in \mathcal{A} \tag{18d}$$

$$u_i^a \prod_{i' \in \mathcal{S}, i' \neq i} t_{i'}^a = 1, \quad \forall i \in \mathcal{S}, a \in \mathcal{A} \tag{18e}$$

$$\frac{t_i^a}{1+x_i^a} = 1, \quad \forall i \in \mathcal{S}, a \in \mathcal{A} \tag{18f}$$

$$u_i^a + p_i^a = 1, \quad \forall i \in \mathcal{S}, a \in \mathcal{A} \tag{18g}$$

where $\boldsymbol{T} = [t_i^a]$ ($t_i^a \geq 1$), $\boldsymbol{U} = [u_i^a]$ ($0 \leq u_i^a \leq 1$), $\boldsymbol{Y} = [y^a]$ ($y^a > 0$). Nevertheless, the transformed problem is still not in a GP form. One reason is that the objective function (18b) is not a posynomial because of negative multiplicative coefficients. To deal with such problem, first, we equivalently substitute the objective function by $\sum_{i \in \mathcal{S}, a \in \mathcal{A}} -x_i^a r_i^a t\,(y^a)^{-1} + M$ where M is a sufficiently large positive constant. Adding M makes sure that the objective function is always positive. Then, we introduce an additional auxiliary variable $x_0 \geq 0$. By minimizing x_0 and guaranteeing constraint C11 in (19), we can effectively minimize the objective function in (18b). Consequently,

$$\begin{aligned}
\min_{\boldsymbol{X},\boldsymbol{T},\boldsymbol{U},\boldsymbol{P},\boldsymbol{Y},x_0} \quad & x_0, \\
\text{subject to,} \quad & \text{C11}: \frac{M}{x_0 + \sum_{i \in \mathcal{S}, a \in \mathcal{A}} \left(\frac{x_i^a r_i^a t}{y^a}\right)} \leq 1 \\
& \text{C12}: \frac{\prod_{i \in \mathcal{S}}(1+x_i^a)}{t' + y^a} = 1, \quad \forall a \in \mathcal{A} \\
& \text{C13}: \frac{\eta_k + 1}{1 + \sum_{i \in \mathcal{S}_k, a \in \mathcal{A}} \frac{x_i^a \prod_{i' \in \mathcal{S}, i' \neq i} t_{i'}^a}{y^a}} \leq 1, \quad \forall k \in \mathcal{K} \\
& \text{C14}: \frac{u_i^a x_i^a + (1+N)x_i^a}{u_i^a + N(u_i^a)^2 x_i^a} \leq 1, \quad \forall i \in \mathcal{S}, a \in \mathcal{A} \\
& \text{C15}: u_i^a \prod_{i' \in \mathcal{S}, i' \neq i} t_{i'}^a = 1, \ \forall i \in \mathcal{S}, a \in \mathcal{A} \\
& \text{C16}: \frac{t_i^a}{1+x_i^a} = 1, \quad \forall i \in \mathcal{S}, a \in \mathcal{A}
\end{aligned} \tag{19}$$

In the preceding optimization problem, p_i^a is replaced by $1-u_i^a$ based on constraint in (18g). The optimization problem in (19) belongs to the class of *complementary GP* problems that allow upper bound constraints on the ratio between two posynomials and equality constraints on the ratio between a monomial and a posynomial [50], [51]. By approximating the posynomials in the denominator of such constraints, a *complementary GP* can be turned into a standard form of GP. Consequently, the optimal solution can be achieved by iteratively applying monomial approximations and solving a series of GPs. The arithmetic-geometric mean inequality can be used to approximate a posynomial with a monomial. The details of such monomial approximation are provided in Section VI-B.

Accordingly, we propose an iterative algorithm to reach to an optimal solution of the transmission probability optimization problem. In each iteration, monomial approximations are applied to the denominator of C11, C12, C13, C14, and C16. Then, the resulting GP can be solved for instance by using a standard interior-point algorithm. More specifically, Algorithm 1 presents the steps needed to be performed until convergence.

Algorithm 1 GP-Based Association Control Algorithm

Initialize x_i^a, t_i^a, p_i^a, u_i^a, for all $i \in \mathcal{S}, a \in \mathcal{A}$, y_a for all $a \in \mathcal{A}$, x_0;
Record the current system state as $\boldsymbol{Z} = (\boldsymbol{X}, \boldsymbol{T}, \boldsymbol{U}, \boldsymbol{P}, \boldsymbol{Y}, x_0)$;
repeat
  Compute the ratio α of each monomial term in the denominator of C11, C12, C13, C14, and C16 according to (30), at the current system state $\boldsymbol{Z}$;
  Apply monomial approximation to the denominators mentioned above according to (29);
  Solve the resulting GP problem using cvx;
  Update the current system state $\boldsymbol{Z} = (\boldsymbol{X}, \boldsymbol{T}, \boldsymbol{U}, \boldsymbol{P}, \boldsymbol{Y}, x_0)$;
until all x_i^a converge.
Compute the optimal transmission probabilities ${\tau^*}_i^a = \frac{{x^*}_i^a}{1+{x^*}_i^a}$

It should be noted that the optimal solution achieved by Algorithm 1 might require multiple associations for an STA to different APs. Since the APs operate on non-overlapping channels, connection to multiple APs can be implemented through dynamic channel bonding/aggregation techniques (proposed for IEEE 802.11ax [52]) with only one transceiver required at each STA. Moreover, for scenarios that the multiple association is not feasible, each STA can only associate with the AP with the largest transmission probability, i.e., $\text{argmax}_a {x^*}_i^a$.

D. ASYMPTOTIC COMPLEXITY ANALYSIS

In this subsection, we study the asymptotic complexity and scalability of Algorithm 1 in terms of the number of STAs, i.e., N_s, and the number of APs, i.e., N_a.

In Algorithm 1, for each iteration, the computational complexity is incurred by applying monomial approximations and solving the resulting GP problem. Suppose C_{MA} and C_{GP} denote the required computing efforts for monomial approximations and solving GP problem in each iteration.

More specifically, first in each iteration, the denominators of constraints C11, C12, C13, C14, and C16 need to be approximated according to (29). The required computational complexity for these monomial approximations is proportional to the number of monomial terms in the denominators of constraints C11, C12, C13, C14 and C16. Consequently, the complexity of monomial approximations totals to

$$\begin{aligned}
C_{\text{MA}} = {} & \mathcal{O}(N_a N_s) + \mathcal{O}(N_a N_s) + \mathcal{O}(N_a N_s^2) \\
& + \mathcal{O}(N_a N_s) + \mathcal{O}(N_a N_s) = \mathcal{O}(N_a N_s^2)
\end{aligned} \tag{20}$$

Subsequently, the resulting GP problem needs to be solved by transforming it to a convex problem. It is reported in [53] that the worst-case computational complexity of this approach is $\mathcal{O}(pn^3)$, where n is the number of variables, and p is the total number of terms in all monomials and posynomials in the objective function and constraints. The number of variables of the optimization problem in (19) is

$$n = 4N_aN_s + 2N_a + 1 = \mathcal{O}(N_aN_s) \tag{21}$$

Furthermore, the total number of terms in all monomials and posynomials can be counted as

$$\begin{aligned} p &= 1 + (4N_aN_s + 1) + N_a(N_s + 2) \\ &\quad + \sum_k (1 + N_kN_a(N_s + 1)) \\ &\quad + 7N_aN_s + N_aN_s^2 + 2N_aN_s \\ &= \mathcal{O}(N_aN_s^2) \end{aligned} \tag{22}$$

where each term in (22) is respectively the number of terms in all the monomials and posynomials in the objective function and constraints C11-C16. Based on (21) and (22), to solve the GP problem, the complexity is

$$C_{\text{GP}} = \mathcal{O}(pn^3) = \mathcal{O}(N_a^4N_s^5) \tag{23}$$

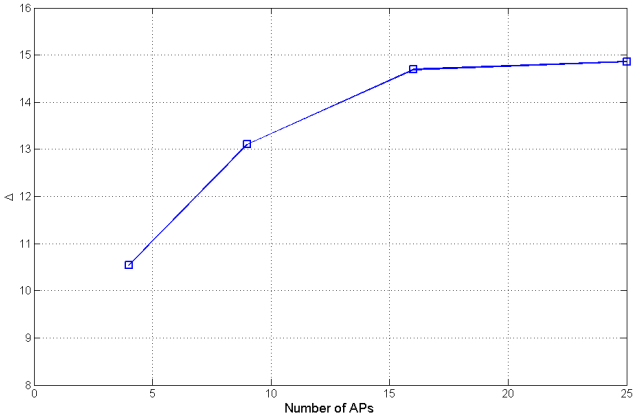
Thus, based on (20) and (23), *per-iteration asymptotic complexity* of Algorithm 1 becomes

$$C_{\text{Alg1}-\text{iteration}} = C_{\text{MA}} + C_{\text{GP}} = \mathcal{O}(N_a^4N_s^5) \tag{24}$$

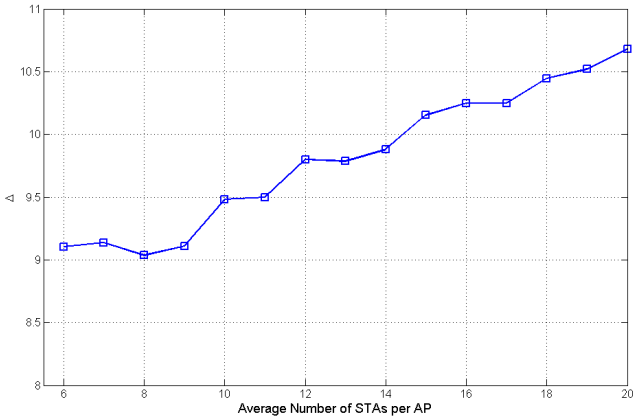
We now turn to studying the number of iterations (denoted by Δ) required for Algorithm 1 to converge. Figures 5a and 5b illustrate numerical results on Δ versus N_a and average number of STAs per AP, i.e., λ_{mean}, respectively. The simulation setup used in these figures is the same as the one presented in Section V. In Figure 5a, λ_{mean} is set to $\frac{10}{N_a}$, so that the expected total number of STAs in the WLAN stays fixed. Figure 5a shows that Δ increases with N_a when N_a is small, and then, it becomes *steady* for a larger range of N_a. In Figure 5b, N_a is fixed equal to 4 and λ_{mean} is varying. Figure 5b shows that Δ grows *linearly* over a typical range of λ_{mean}. It should be noted that the fluctuations in Δ are caused by the randomness in the number of STAs per AP, which follows a Poisson distribution with mean of λ_{mean}.

Therefore, based on (24) and the numerical results in Figures 5a and 5b, it is shown that the overall complexity of the proposed algorithm only grows *polynomially* with the number of STAs and APs, which is a considerable improvement over direct search methods since the user association problem is combinatorial, and the complexity of a direct search is therefore exponential.

In this sense, Algorithm 1 scales well and implementation in large scale networks will depend on both the considered network size and the frequency with which the algorithm will need to be run. To be as adaptive as possible to the system, Algorithm 1 can be run as frequent as any change happens in the network status. However, the frequency on which the proposed algorithm would be run can be decided considering the network dynamics and feasibility issues. Ideally, the proposed algorithm would be run every coherence time in the system.



(a)



(b)

FIGURE 5. Number of iterations required for Algorithm 1 to converge. (a) Number of iterations vs. number of APs (N_a). (b) Number of iterations vs. STA density (λ_{mean}).

IV. IMPLEMENTATION DETAILS

The optimal transmission probability of STA i at AP a, i.e., τ^{*a}_i, can be obtained from Algorithm 1. But, in the EDCA protocol, the transmission probabilities of STAs are not directly controllable. Instead, what we can control are the MAC-layer parameters (e.g., W_{min} and AIFS) to achieve the optimal transmission probability. In this section, we first verify the accuracy of the relationship between τ^a_i and the EDCA parameters provided in (13) and also the validity of the Markov chain model. Then, we develop an algorithm to adjust EDCA parameters aiming to implement the optimal values of τ^a_i obtained from Algorithm 1 in the EDCA protocol.

We first investigate the achievable accuracy by controlling different EDCA parameters through an example. In this example, we consider one BSS with 6 STAs. The transmission probabilities of STA 2 to STA 6 are fixed (e.g., $\tau_2 = \tau_3 = \tau_4 = \tau_5 = \tau_6 = 0.005$), while the transmission probability

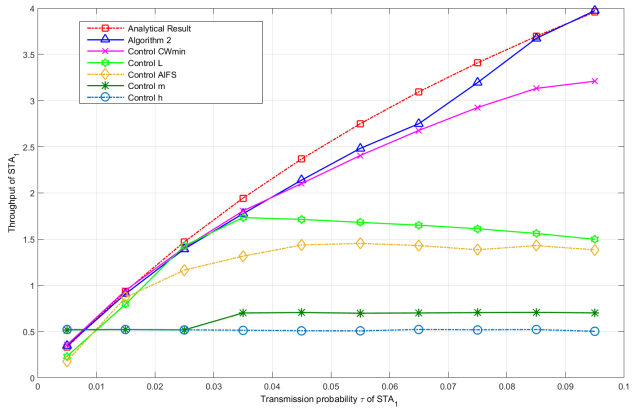


FIGURE 6. Accuracy of controlling different parameters.

of STA 1, i.e., τ_1, is varied from 0.005 to 0.1. Figure 6 compares the throughput of STA 1 analytically derived from (10) and numerically measured using an EDCA simulator. In the numerical results, different EDCA parameters are *separately* adjusted to tune τ_1 to the desired value according to (13). Figure 6 demonstrates that controlling a single parameter might not be adequately effective to adjust τ_i to a desired value. To achieve a broader feasible range of τ_i, we need to jointly adjust these parameters.

Thus, we develop a control algorithm for MAC parameters that can adjust all the 5 parameters to achieve the optimal transmission probability obtained from Algorithm 1 solving the optimization problem (19) in Section III. Algorithm 2 details how we use the derived optimal probabilities τ_i^*, and (13) to determine the required, and implementable, MAC parameters. In other words, Algorithm 2 is to compute the EDCA parameters for a given STA to realize the optimal transmission probabilities obtained from Algorithm 1 based on (13).

In Section III-B, the feasibility region of transmission probability is calculated assuming that all the MAC parameters, i.e., $W_{\min}$, A, q or L, m, h, can be freely varied in their feasible ranges. Considering the feasibility regions and the fact that some MAC parameters must be set to integer values, a search algorithm is required to find a feasible solution of (13) to achieve the optimal transmission probability. Here, we propose an effective heuristic search Algorithm 2 (with affordable complexity) where EDCA parameters are first initialized and then adjusted one-by-one, taking into account the achievable accuracy by controlling different parameters. This algorithm tunes all MAC parameters in a descending order of their effectiveness.

More specifically, for a given STA at a given AP, we initially set values for the parameters to be the largest possible. Then, we start tuning $W_{\min}$, i.e., the best single parameter to be controlled, by solving (13). To compute $W_{\min}$ based on (13), the optimal transmission probability is obtained from Algorithm 1 and other MAC parameters are fixed to their initial values. If the derived $W_{\min}$ is feasible, the algorithm stops and the EDCA parameters to achieve the optimal transmission probability are obtained. Otherwise, if the output for $W_{\min}$ is negative, we set $W_{\min}$ to zero (which is the smallest value possible) to move on to tune L to reach the optimal transmission probability based on (13). This procedure similarly continues with tuning A and m, and finishes with h.

Algorithm 2 MAC Parameter Control Algorithm

For each AP-STA pair (i, a)**:**
$p_i^a = \prod_{i' \in \mathcal{S}, i' \neq i}(1 - \tau^{*a}_{i'})$
Initialization:
Set $A_i^a = 6$, $L_i^a = 100$, $m_i^a = 6$, $h_i^a = 6$, $q_i^a = 0.5$;
Compute $W^a_{\min,i}$ from (13), while other variables are fixed;
Round $W^a_{\min,i}$ to an integer;
if $W^a_{\min,i} < 0$ **then**
  Set $W^a_{\min,i} = 0$; Compute L_i^a from (13), while other variables are fixed;
  if $L_i^a < 0$ **then**
    Set $L_i^a = 0$; Compute A_i^a from (13), while other variables are fixed; Round A_i^a to an integer;
    if $A_i^a < 1$ **then**
      Set $A_i^a = 1$; Compute m_i^a from (13), while other variables are fixed; Round m_i^a to an integer;
      if $m_i^a < 0$ **then**
        Set $m_i^a = 0$; Compute h_i^a from (13), while other variables are fixed; Round h_i^a to an integer;
        if $h_i^a < 0$ **then**
          Set h_i^a=0;
Start transmission using EDCA protocol with parameters $\{W^a_{\min,i}, A_i^a, q_i^a, L_i^a, m_i^a, h_i^a\}$.

The achievable throughput for STA 1 in Algorithm 2 is plotted in Figure 6. As expected, it is shown that Algorithm 2 can approach the desired throughput closely for the total range of τ_1, i.e., the upper bound τ_1 computed from (16). Thus, Algorithm 2 can improve the control accuracy compared with controlling the parameters separately.

Since Algorithm 2 examines each AP-STA pair to determine the required MAC parameters independently for each STA, the time complexity of the algorithm is $\mathcal{O}(N_s)$, i.e, is linear in the number of STAs in the system. Figure 7 compares the average simulation time of Algorithm 2 versus the number of STAs. Without loss generality, for a fixed number of STAs, we examined the complexity of the algorithm while the transmission probability of STA 1, τ_1, was varied from 0.01 to 0.1. For the remaining STAs, $\tau_{i'}$, $i' \in \mathcal{S}$, $i' \neq 1$ were fixed at 0.005 and the average simulation time over 1000 trials for each value of τ_1 is presented.

V. NUMERICAL RESULTS

In this section, we present numerical results to evaluate the performance of the proposed STA-AP association and airtime control algorithm, and also the MAC parameter control algorithm. More specifically, the performance of our GP-based association scheme is compared with the

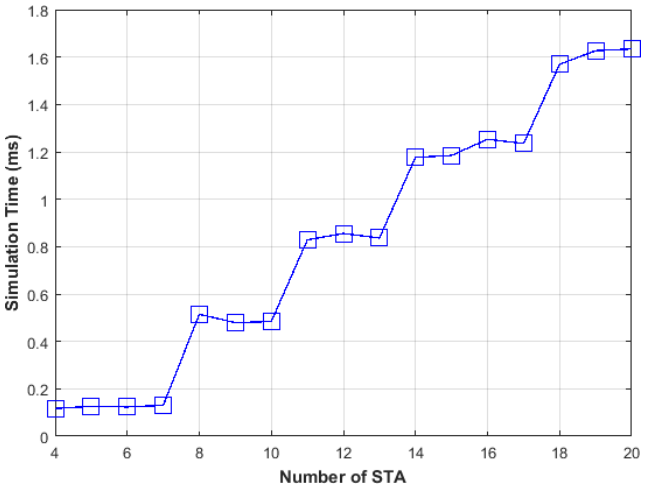


FIGURE 7. Complexity of Algorithm 2.

Max-SNR scheme in terms of throughput and fairness assuming equal airtime share for different ISPs. We implemented a simulator for EDCA (including access probability q and inter-frame space L described in Section VI-A) in Matlab to measure the achieved throughput numerically. Furthermore, we used CVX to solve the GP problems in the association algorithm.

We consider a network in which 4 APs are deployed in a $10 \times 10\, m^2$ area. More specifically, the APs are placed at the centers of four different $5 \times 5\ m^2$ grids to provide seamless coverage. To eliminate interference between the transmission of different APs, non-overlapping 20 MHz channels are assigned to each of the four APs. The STAs are distributed in the entire area according to the two-dimensional Poisson point process (PPP).

The wireless channel model includes path loss and small-scale fading. Generally, the channel gain can be expressed as $h = Ah'd^{-\alpha/2}$, where d is the distance between a STA and an AP, $\alpha \geq 2$ is the path loss exponent, A is a constant dependent on the frequency and transmitter/receiver antenna gain, and h' represents the small scale fading component. In the numerical results, we set $\alpha = 3$ and $A = 1$. Furthermore, h' is randomly generated according to the Rayleigh distribution assuming $\mathbb{E}\{|h'|^2\} = 1$. The received SNR at STA i is equal to $\frac{Pg_i^a}{\sigma^2}$ where P is the transmission power, $g_i^a = |h_i^a|^2$ is the channel power gain from STA i to AP a, and σ^2 is the power of noise. In all the numerical results, P/σ^2 is assumed fixed and set to 10dB.

To determine the transmission rate of each STA-AP pair, the 802.11a physical layer model is used. More specifically, to guarantee a maximum packet error rate, adaptive modulation and coding is used based on the received SNR. Table 3 shows the achievable transmission rates and adaptive modulation and coding schemes standardized in IEEE 802.11a and the SNR range for each scheme used in the simulations.

The MAC layer parameters used in our simulations are summarized in Table 4. Moreover, the target airtime share

TABLE 3. 802.11a adaptive modulation and coding scheme and the SNR ranges used in the numerical results.

Modulation	FEC Rate	Data Rate (Mbps)	SNR (dB)
BPSK	1/2	6	[5,8)
BPSK	3/4	9	[8,10)
QPSK	1/2	12	[10,13)
QPSK	3/4	18	[13,16)
16QAM	1/2	24	[16,19)
16QAM	3/4	36	[19,22)
64QAM	2/3	48	[22,25)
64QAM	3/4	54	[25,∞)

TABLE 4. 802.11e MAC parameters used in the numerical results.

time-slot δ	9 μs
Propagation Delay γ	1 μs
T_{TXOP}	1 ms
SIFS	10 μs
ACK	40 μs

for each ISP k, i.e., η_k, is set equal to the number of APs divided by the number of ISPs. In other words, we assume that the ISPs have the same minimum airtime reservation and share the total airtime in a fair manner. In the Markov chain, the average frozen time N is approximated by T_{TXOP}/δ.

A. EFFECTS OF STA DISTRIBUTION

Here, we investigate the impact of STA distribution on the fairness and throughput achieved by the two association algorithms. More specifically, we set up two examples considering homogeneous and non-homogeneous STA distributions. In both examples, the STAs are randomly associated to the two ISPs with equal probabilities.

1) HOMOGENEOUS DISTRIBUTION

In this example, the STAs are distributed in the square space according to a homogeneous two-dimensional PPP. Accordingly, the number of STAs in each grid (where an AP is centered) follows a Poisson distribution with mean λ_{mean}, which represents the average number of STAs per AP. Figure 8 shows that GP-based association scheme improves both fairness and total throughput compared with Max-SNR scheme.

2) NON-HOMOGENEOUS DISTRIBUTION

Let consider that STAs are distributed according to a non-homogeneous PPP. More specifically, the number of STAs in a grid (where AP a centred at) follows a Poisson distribution with mean λ_a. Here, λ_a is randomly generated between 0 and λ_{mean}. With such STA distribution, Figure 9 shows the achieved throughput of two ISPs versus λ_{mean}. As expected, the performance gap is significantly larger with non-homogeneous distribution compared to the homogeneous case. Max-SNR can hardly guarantee the fairness

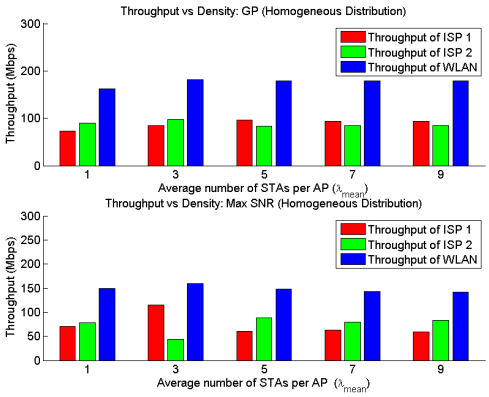


FIGURE 8. Throughput vs. STA density, homogeneous distribution.

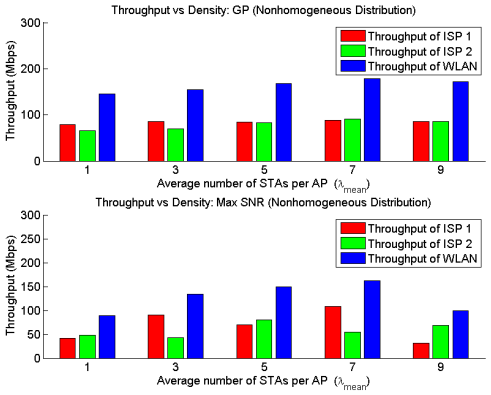


FIGURE 9. Throughput vs. STA density, non-homogeneous distribution.

between different ISPs. While for the same distribution, GP-based STA-AP association can manage to keep the balance between two ISPs. Furthermore, it can offer improvement over Max-SNR in terms of the total throughput due to the load-balancing among APs.

B. EFFECTS OF STA DENSITY AND ISP LOAD

Let define ρ_1 (also referred to as ISP 1 load) as the ratio of number of STAs serving by ISP 1 to the total number of STAs in the network. Here, the performance of the two association approaches are compared under different STA density and ISP load.

Assuming a homogeneous STA distribution with $\lambda_{\text{mean}} = 3$, Figure 10 depicts the achieved throughput of two ISPs versus different values of ρ_1 for both GP-based and Max-SNR association schemes. By Max-SNR association, it is shown that throughput of ISP 1 grows linearly with ρ_1, while the achieved throughput of ISP 2 is decreasing. But, GP-based association can fairly distribute the airtime between ISPs regardless of their ISP loads and thus maintain a balance between the achieved throughput of the two ISPs.

Figure 11 shows the total throughput achieved by the two association algorithms versus λ_{mean} for a homogeneous STA distribution. For a fixed ρ_1, the total throughput by both

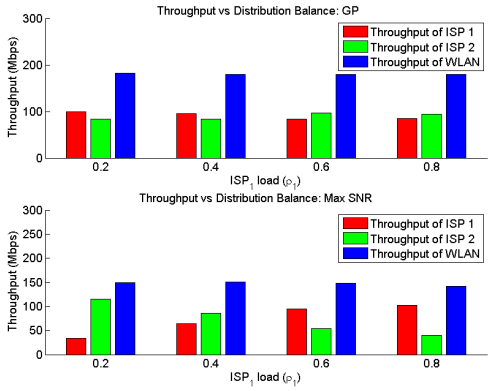


FIGURE 10. Throughput vs. ISP 1 load.

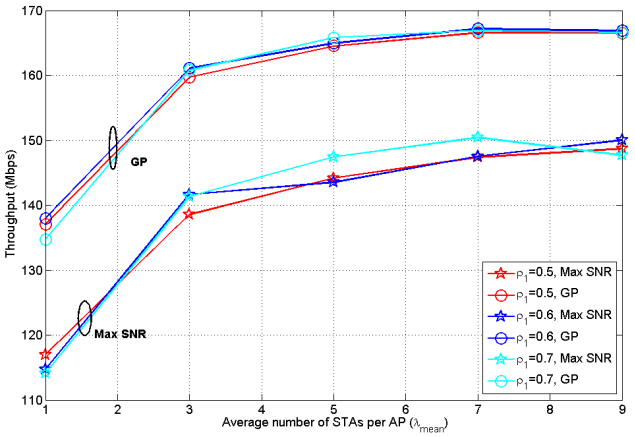


FIGURE 11. Total throughput vs. STA density for different ρ_1.

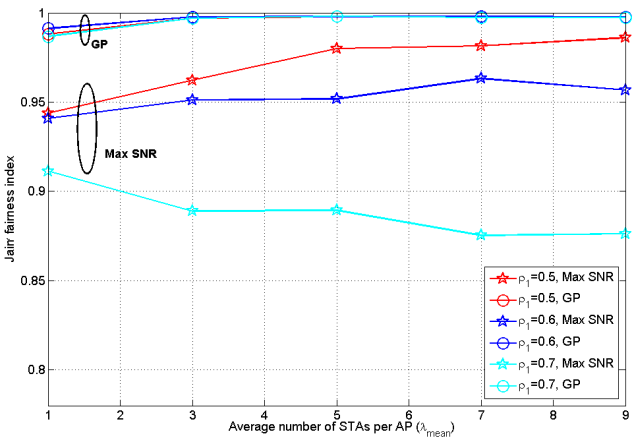


FIGURE 12. Fairness vs. STA density for different ρ_1

algorithms increases with the STA density, i.e., λ_{mean}. But, the throughput increase rate is decreasing with λ_{mean}. This is because the wireless channel is underutilized when the STA density is low. Thus, the increase in the STA density will improve the total throughput. But, when the STA density is large, increasing the STA density further will result in a higher collision probability, and hence, slow down the total

throughput improvement. For any fixed ρ_1, it is shown that GP-based association significantly improves the total throughput compared with the Max-SNR association.

Figure 12 shows the fairness by employing the Jain's fairness index as

$$F = \frac{(\sum_{k\in\mathcal{K}} T_k)^2}{|\mathcal{K}| \sum_{k\in\mathcal{K}} T_k^2} \tag{25}$$

where $T_k = \sum_{i\in S_k, a\in A} T_i^a$ is the achieved throughput for all the STAs of ISP k. From Figure 12, it is clear that the proposed GP-based association approach can always guarantee perfect fairness between the ISPs regardless of the STA density or ρ_1. The achieved fairness level by Max-SNR association is always worse than GP-based, especially when the STA load is highly unbalanced between ISPs, i.e., ρ_1 is not close to 0.5.

VI. CONCLUSION

This paper considers the STA-AP association and airtime control in virtualized 802.11 networks aiming to provide fairness guarantees among ISPs despite the number of STAs per ISP. First, a three-dimensional Markov chain is developed to model a generalized 802.11e EDCA protocol. This model establishes the relationship between the transmission probability of each STA and the detailed parameters in the MAC protocol. Based on this relationship, the feasibility region of the transmission probabilities are derived. Subsequently, an optimization problem is formulated which can maximize the network throughput, while guaranteeing the fairness between different ISPs. The implementation of the optimal transmission probabilities obtained by successive geometric programming is discussed by controlling the MAC parameters. Extensive numerical results confirm that the Markov chain can accurately describe the MAC protocol. Furthermore, it is verified that the proposed association algorithm can improve the throughput and provide fairness guarantees in virtualized 802.11 WLANs.

APPENDIX

A. MARKOV CHAIN MODEL

Here, we detail the Markov chain model for EDCA introduced in Section III-A. Firstly, the transition rules among the states of the Markov chain are given according to the EDCA protocol. Then, the transition probability between any two adjacent states are summarized. Finally, by solving the stationary point of the Markov chain, the relationship between transmission probability and the detailed EDCA parameters are derived.

1) TRANSITION RULES

First, we describe the transitions among the states in the proposed Markov chain according to the EDCA protocol. As shown in Figure 4, after a successful transmission or a packet drop, the STA will flip a biased coin with successful probability q_i. If the coin appears to be tail (with probability $1 - q_i$), the STA will go to the state $\{-2, 0, L_i - 1\}$, and try again after L_i time-slots. Otherwise, if the coin is head (with probability q_i), the STA will move to the state $\{-1, 0, A_i\}$, where $A_i = \text{AIFS[STA } i] - 1$. If the channel stays idle in the following $A_i + 1$ time-slots, the STA will move on to the backoff/retransmission stage 0 and uniformly pick a random backoff time from $[0, W_{0,i}]$.

Otherwise, if the STA senses the channel busy in a state $\{-1, 0, d\}$ where $0 \le d \le A_i$, it will freeze and move to the state $\{-1, 0, N+A_i\}$, where N is the average frozen time (The frozen time is equal to either T_s or T_c and thus N can be approximated to T as discussed before). Afterwards, it will wait until channel becomes idle again, by counting down $v(t)$ from $N + A_i$ to A_i. When the STA gets to the state $\{-1, 0, A_i\}$, it has to keep repeating the explained process until it enters the backoff process. Let denote p_i the probability of collision conditioned on the STA i having counted down to 0 and being to transmit with probability 1, which is equal to the probability of the channel being observed busy (from the point of view of STA i) in the timeslot in which the transmission is initiated. Given the transmission probabilities of all STAs, i.e., τ_i, in the BSS, we have

$$p_i = 1 - \prod_{i'\neq i} (1 - \tau_{i'}). \tag{26}$$

For the backoff process, we use Figure 13 to explain the possible transitions at the backoff/retransmission stage j. When STA i reaches the state $\{j-1, 0, 0\}$, it will start a transmission with probability one. Then, if other STAs happen to access the channel at the same time, STA i will experience a collision. The probability that STA i encounters a collision is equal to p_i. In a case of collision, the STA will enter the next backoff/retransmission stage, i.e., j, pick a random backoff counter b between 0 and $W_{j,i}$, and go to the state $\{j, b, 0\}$. When the STA reaches the state $\{j, b, 0\}$, it will count down to $\{j, b-1, 0\}$ if the channel is idle. Otherwise, if the channel is sensed busy, i.e., other STAs are transmitting in the channel, STA i will freeze its backoff counter and move to the state $\{j, b, N + A_i\}$. In the frozen period, the STA will count down $v(t)$ by 1 each time-slot until it gets to the state $\{j, b, A_i\}$. Then, if the channel is still idle (with probability $(1 - p_i)$), the STA will continue counting down to $\{j, b, 1\}$ and then go to state $\{j, b-1, 0\}$. Otherwise, if the channel becomes busy when STA i is in the state $\{j, b, d\}$ where $1 \le d \le A_i$, it has to freeze its backoff counter again and go back to the state $\{j, b, N + A_i\}$.

Finally, for states $\{j, 0, 0\}$, $0 \le j \le m_i + h_i$, the STA will initiate a transmission. If no other STA transmits at the same time (with probability $(1-p_i)$), the STA will have a successful transmission of duration T_s. Otherwise, if any other STA simultaneously starts a transmission (with probability p_i), STA i will experience a collision of duration T_c. After the collision, if $0 \le j \le m_i + h_i - 1$, the STA will go to the next backoff/retransmission stage. But, if $j = m_i + h_i$ (which means that the maximum backoff/retransmission limit is reached), the STA has to drop this packet. Then, for a new packet transmission, it has to flip a biased coin and start

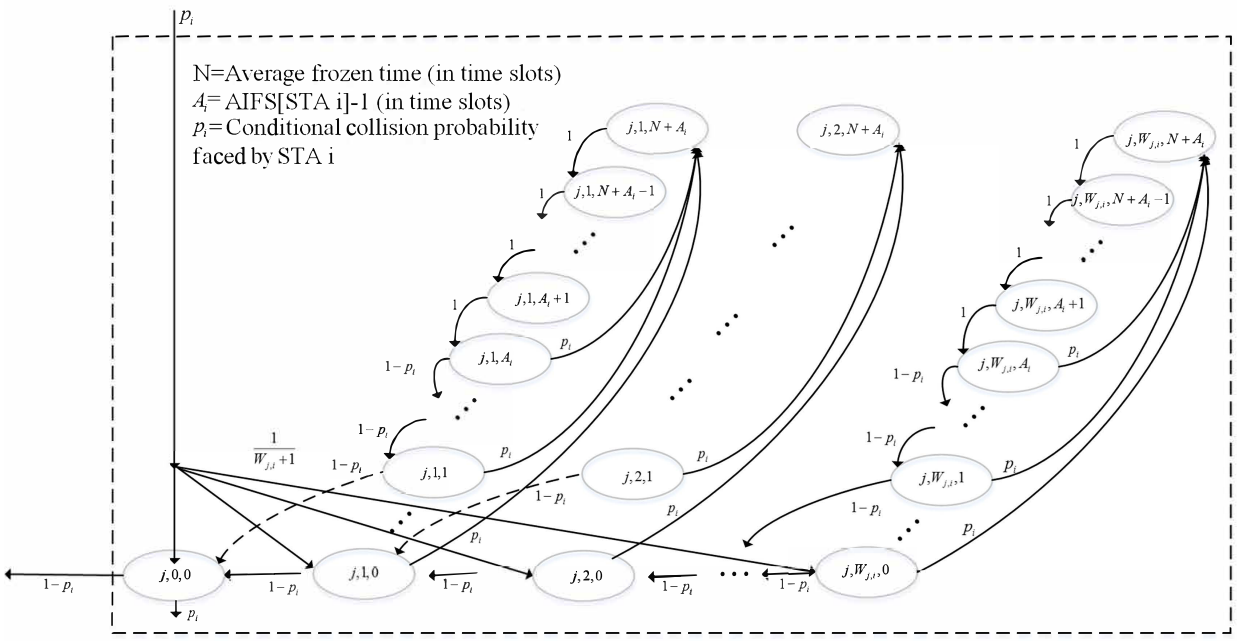


FIGURE 13. Detailed transition for one backoff stage.

over from the state $\{-2, 0, L_i - 1\}$ or $\{-1, 0, A_i\}$ according to the result of the coin toss. The mathematical representations of the transition probabilities are summarized in Section VI-A.2.

2) TRANSITION PROBABILITIES OF THE MARKOV CHAIN

Given the transition rule between the states described in above, we have the following transition probabilities.

1) For states $\{-2, 0, d\},\ 0 \le d \le L_i - 1$, i.e., after getting a tail in a coin toss,

$$\begin{cases} P\{(-2,0,d-1)|(-2,0,d)\} = 1, \ 1 \le d \le L_i - 1 \\ P\{(-2,0,L_i-1)|(-2,0,0)\} = 1 - q_i \\ P\{(-1,0,A_i)|(-2,0,0)\} = q_i \end{cases}$$

2) For states $\{-1, 0, d\},\ 0 \le d \le N + A_i$, i.e., before entering the backoff process,

$$\begin{cases} P\{(-1,0,d-1)|(-1,0,d)\} = 1, & A_i + 1 \le d \le N + A_i \\ P\{(-1,0,d-1)|(-1,0,d)\} = 1 - p_i, & 1 \le d \le A_i \\ P\{(-1,0,N+A_i)|(-1,0,d)\} = p_i, & 0 \le d \le A_i \\ P\{(0,b,0)|(-1,0,0)\} = \dfrac{1-p_i}{W_{0,i}+1}, & 0 \le b \le W_{0,i} \end{cases}$$

3) For states $\{j, b, d\},\ 0 \le j \le m_i + h_i$, i.e., in the backoff process,

a) If $b = 0, d = 0$, i.e., before a successful transmission or a collision,

$$\begin{cases} P\{(-1,0,A_i)|(j,0,0)\} = (1-p_i)q_i, \\ \quad 0 \le j \le m_i + h_i - 1 \\ P\{(-2,0,L_i-1)|(j,0,0)\} = (1-p_i)(1-q_i), \\ \quad 0 \le j \le m_i + h_i - 1 \\ P\{(j+1,b,0)|(j,0,0)\} = p_i \dfrac{1}{W_{j+1,i}+1}, \\ \quad 0 \le j \le m_i + h_i - 1, \ 0 \le b \le W_{j+1,i} \\ P\{(-2,0,L_i-1)|(m_i+h_i,0,0)\} = 1 - q_i \\ P\{(-1,0,A_i)|(m_i+h_i,0,0)\} = q_i \end{cases}$$

b) If $d = 0, 1 \le b \le W_{j,i}$, i.e, in the count down process,

$$\begin{cases} P\{(j,b-1,0)|(j,b,0)\} = 1 - p_i, & 1 \le b \le W_{j,i} \\ P\{(j,b,N+A_i)|(j,b,0)\} = p_i, & 1 \le b \le W_{j,i} \end{cases}$$

4) If $1 \le d \le N + A_i,\ 1 \le b \le W_{j,i}$, i.e., during a frozen period or the AIFS after it,

$$\begin{cases} P\{(j,b,d-1)|(j,b,d)\} = 1, & A_i + 1 \le d \le N + A_i \\ P\{(j,b,N+A_i)|(j,b,d)\} = p_i, & 1 \le d \le A_i \\ P\{(j,b,d-1)|(j,b,d)\} = 1 - p_i, & 2 \le d \le A_i \\ P\{(j,b-1,0)|(j,b,1)\} = 1 - p_i \end{cases}$$

3) STATIONARY CONDITION OF THE MARKOV CHAIN

Based on the transition rule above and the structure of the Markov chain, as in [46] and [45], we have,

$$b_{j,0,0} = p_i^j b_{0,0,0},\ 0 \le j \le m_i + h_i$$

$$b_{j,b,0} = \frac{W_{j,i}+1-b}{W_{j,i}+1} b_{j,0,0}, \quad 0 \le j \le m_i + h_i, 1 \le b \le W_{j,i}$$

$$1 = \sum_{d=0}^{L_i-1} b_{-2,0,d} + \sum_{d=0}^{N+A_i} b_{-1,0,d} + \sum_{j=0}^{m_i+h_i} b_{0,0,0} + \sum_{j=0}^{m_i+h_i} \sum_{b=1}^{W_{j,i}} \sum_{d=0}^{N+A_i} b_{j,b,d}$$
$$= \left[L_i \frac{1-q_i}{q_i} + \frac{1+p_iN}{p_i} \frac{1-(1-p_i)^{A_i+1}}{(1-p_i)^{A_i+1}} + \frac{1-p_i^{m_i+h_i+1}}{1-p_i} + \frac{1+Np_i}{2(1-p_i)^{A_i}} \sum_{j=0}^{m_i+h_i} W_{j,i} p_i^j \right] b_{0,0,0} \quad (27)$$

By exploiting the transition structure within a frozen process and the post AIFS, i.e., the states $\{j, b, d\}$, $0 \le d \le N + A_i$, we have,

$$b_{j,b,d} = \frac{p_i}{(1-p_i)^d} b_{j,b,0}, \quad 0 \le j \le m_i + h_i - 1, \ 1 \le d \le A_i - 1$$
$$b_{j,b,d} = \frac{p_i}{(1-p_i)^{A_i}} b_{j,b,0}, \quad 0 \le j \le m_i + h_i, \ A_i \le d \le N + A_i$$

For states $\{-2, 0, d\}$, by looking at the incoming and outgoing probability, we can have

$$b_{-2,0,d} = \frac{1-q_i}{q_i} b_{0,0,0}, \quad 0 \le d \le L_i - 1$$

For states $\{-1, 0, d\}$, by analyzing the structure of transition between the states $\{-1, 0, d\}$, $0 \le d \le N + A_i$ and the state $\{0, 0, 0\}$, we can have

$$b_{-1,0,d} = \frac{1}{(1-p_i)^{d+1}} b_{0,0,0}, \quad 0 \le d \le A_i$$
$$b_{-1,0,d} = \frac{1-(1-p_i)^{A_i+1}}{(1-p_i)^{A_i+1}} b_{0,0,0}, \quad A_i + 1 \le d \le N + A_i$$

To this point, all the stationary probabilities in the Markov chain are represented in terms of $b_{0,0,0}$. Then, $b_{0,0,0}$ can be derived from the normalization condition in (27), as shown at the top of this page. Consequently, having all stationary probabilities $b_{j,b,d}$, the transmission probability of STA i can be calculated as in (13). To derive (13), the contention window $W_{j,i}$ is set equal to

$$W_{j,i} = \begin{cases} W_{\min,i} 2^j, & 0 \le j \le m_i \\ W_{\min,i} 2^{m_i}, & m_i + 1 \le j \le m_i + h_i \end{cases}$$

according to the exponential backoff rules in 802.11e. From the definition of transmission probability, τ_i can be calculated by summing all the stationary probabilities of the states in which the STA will initiate a transmission, i.e., $b_{j,0,0}$, $\forall j \in \{0, 1, ..., m_i + h_i\}$. The expression in (13) establishes the relationship between τ_i and all the detailed parameters for STA i in the EDCA protocol, which will be useful to characterize the feasibility region of τ_i^a in the optimization and design an algorithm to control the transmission probabilities of the STAs.

B. GEOMETRIC PROGRAMMING AND MONOMIAL APPROXIMATION

An optimization problem is called geometric programming if it follows the following form,

$$\min_{\boldsymbol{x}} f_0(\boldsymbol{x}),$$
$$\text{subject to } f_i(\boldsymbol{x}) \le 1, \quad i = 1, \ldots, n_1$$
$$g_i(\boldsymbol{x}) = 1, \quad i = 1, \ldots, n_2$$

where $f_0, \ldots, f_{n_1}$ are posynomials and $g_1, \ldots, g_{n_2}$ are monomials. In the context of geometric programming, a monomial function f of $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ is defined as,

$$f(\boldsymbol{x}) = c x_1^{a_1} x_2^{a_2} \ldots x_n^{a_n}$$

where $c > 0$ and $a_i \in \mathcal{R}$. Furthermore, a posynomial is defined as the summation of multiple monomials, i.e.,

$$g(\boldsymbol{x}) = \sum_{k=1}^{K} f_k(\boldsymbol{x})$$

The basic idea of monomial approximation is as follows: consider a posynomial function $g(\boldsymbol{x}) = \sum_k f_k(\boldsymbol{x})$ with $f_k(\boldsymbol{x})$ being the monomial terms. By the arithmetic-geometric mean inequality, we have

$$g(\boldsymbol{x}) \ge \hat{g}(\boldsymbol{x}) = \prod_k \left(\frac{f_k(\boldsymbol{x})}{\alpha_k(\boldsymbol{x}_0)} \right)^{\alpha_k(\boldsymbol{x}_0)} \quad (29)$$

where the parameters $\alpha_k(\boldsymbol{x}_0)$ can be obtained by computing

$$\alpha_k(\boldsymbol{x}_0) = \frac{f_k(\boldsymbol{x}_0)}{g(\boldsymbol{x}_0)}, \quad \forall k \quad (30)$$

where $\boldsymbol{x}_0 > 0$ is a fixed point (e.g., the optimal solution from the last round of optimization). It is proved that $\hat{g}(\boldsymbol{x})$ is the best local monomial approximation of $g(\boldsymbol{x})$ near $\boldsymbol{x}_0$ [54].

REFERENCES

[1] K. Lee, J. Lee, Y. Yi, I. Rhee, and S. Chong, "Mobile data offloading: How much can WiFi deliver?" *IEEE/ACM Trans. Netw.*, vol. 21, no. 2, pp. 536–550, Apr. 2013.
[2] K. Nakauchi and Y. Shoji, "WiFi network virtualization to control the connectivity of a target service," *IEEE Trans. Netw. Service Manag.*, vol. 12, no. 2, pp. 308–319, Jun. 2015.
[3] C. Liang and F. R. Yu, "Wireless network virtualization: A survey, some research issues and challenges," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 1, pp. 358–380, 1st Quart., 2015.
[4] M. Yang, Y. Li, D. Jin, L. Zeng, X. Wu, and A. V. Vasilakos, "Software-defined and virtualized future mobile and wireless networks: A survey," *Mobile Netw. Appl.*, vol. 20, no. 1, pp. 4–18, Feb. 2015.
[5] H. Wen, P. K. Tiwary, and T. Le-Ngoc, *Wireless Virtualization*. Cham, Switzerland: Springer, 2013.
[6] G. Bhanage, D. Vete, I. Seskar, and D. Raychaudhuri, "SplitAP: Leveraging wireless network virtualization for flexible sharing of wlans," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Dec. 2010, pp. 1–6.
[7] L. Xia *et al.*, "Virtual WiFi: Bring virtualization from wired to wireless," *ACM SIGPLAN Notices*, vol. 46, no. 7, pp. 181–192, Jul. 2011.
[8] A. Checco and D. J. Leith, "Fair virtualization of 802.11 networks," *IEEE/ACM Trans. Netw.*, vol. 23, no. 1, pp. 148–160, Feb. 2015.
[9] M. Derakhshani, X. Wang, T. Le-Ngoc, and A. Leon-Garcia, "Airtime usage control in virtualized multi-cell 802.11 networks," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2015, pp. 1–6.

[10] K. Nakauchi, Y. Shoji, and N. Nishinaga, "Airtime-based resource control in wireless LANs for wireless network virtualization," in *Proc. 4th Int. Conf. Ubiquitous Future Netw. (ICUFN)*, Jul. 2012, pp. 166–169.

[11] A. Banchs, P. Serrano, P. Patras, and M. Natkaniec, "Providing throughput and fairness guarantees in virtualized WLANs through control theory," *Mobile Netw. Appl.*, vol. 17, no. 4, pp. 435–446, Aug. 2012.

[12] V. Chandrasekhar, J. G. Andrews, T. Muharemovic, Z. Shen, and A. Gatherer, "Power control in two-tier femtocell networks," *IEEE Trans. Wireless Commun.*, vol. 8, no. 8, pp. 4316–4328, Aug. 2009.

[13] E. E. Tsiropoulou, P. Vamvakas, and S. Papavassiliou, "Supermodular game-based distributed joint uplink power and rate allocation in two-tier femtocell networks," *IEEE Trans. Mobile Comput.*, vol. 16, no. 9, pp. 2656–2667, Sep. 2017.

[14] D. Liu *et al.*, "User association in 5G networks: A survey and an outlook," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 1018–1044, 2nd Quart. 2016.

[15] E. Aryafar, A. Keshavarz-Haddad, M. Wang, and M. Chiang, "RAT selection games in HetNets," in *Proc. INFOCOM*, Apr. 2013, pp. 998–1006.

[16] Y. Xu, G. Athanasiou, C. Fischione, and L. Tassiulas, "Distributed association control and relaying in millimeter wave wireless networks," in *Proc. IEEE ICC*, May 2016, pp. 1–6.

[17] G. Athanasiou, P. C. Weeraddana, and C. Fischione, "Auction-based resource allocation in millimeterwave wireless access networks," *IEEE Commun. Lett.*, vol. 17, no. 11, pp. 2108–2111, Nov. 2013.

[18] H. Boostanimehr and V. K. Bhargava, "Joint downlink and uplink aware cell association in HetNets with QoS provisioning," *IEEE Trans. Wireless Commun.*, vol. 14, no. 10, pp. 5388–5401, Oct. 2015.

[19] M. Balazinska and P. Castro, "Characterizing mobility and network usage in a corporate wireless local-area network," in *Proc. 1st Int. Conf. Mobile Syst., Appl. Services*, May 2003, pp. 303–316.

[20] D. Schwab and R. Bunt, "Characterising the use of a campus wireless network," in *Proc. INFOCOM*, Mar. 2004, pp. 862–870.

[21] H. Gong, K. Nahm, and J. Kim, "Distributed fair access point selection for multi-rate IEEE 802.11 WLANs," in *Proc. 5th IEEE Consum. Commun. Netw. Conf.*, Jan. 2008, pp. 528–532.

[22] O. Ercetin, "Association games in IEEE 802.11 wireless local area networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 5136–5143, Dec. 2008.

[23] W. Xu, C. Hua, and A. Huang, "A game theoretical approach for load balancing user association in 802.11 wireless networks," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Dec. 2010, pp. 1–5.

[24] P. BK and J. Kuri, "An estimated delay based association policy for Web browsing in a multirate WLAN," *IEEE Trans. Netw. Service Manag.*, vol. 9, no. 3, pp. 346–358, Sep. 2012.

[25] M. Heusse, F. Rousseau, G. Berger-Sabbatel, and A. Duda, "Performance anomaly of 802.11b," in *Proc. IEEE Int. Conf. Comput. Commun. (INFOCOM)*, Mar./Apr. 2003, pp. 836–843.

[26] A. Checco and D. J. Leith, "Proportional fairness in 802.11 wireless LANs," *IEEE Commun. Lett.*, vol. 15, no. 8, pp. 807–809, Aug. 2011.

[27] X. Chen and D. Leith, "Proportional fair coding for 802.11 WLANs," *IEEE Wireless Commun. Lett.*, vol. 1, no. 5, pp. 468–471, Oct. 2012.

[28] S. C. Liew and Y. J. Zhang, "Proportional fairness in multi-channel multi-rate wireless networks—Part I: The case of deterministic channels with application to AP association problem in large-scale WLAN," *IEEE Trans. Wireless Commun.*, vol. 7, no. 9, pp. 3446–3456, Sep. 2008.

[29] Y. J. Zhang and S. C. Liew, "Proportional fairness in multi-channel multi-rate wireless networks—Part II: The case of time-varying channels with application to OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 7, no. 9, pp. 3457–3467, Sep. 2008.

[30] H. Zhou, P. Fan, and J. Li, "Global proportional fair scheduling for networks with multiple base stations," *IEEE Trans. Veh. Technol.*, vol. 60, no. 4, pp. 1867–1879, May 2011.

[31] T. Joshi, A. Mukherjee, Y. Yoo, and D. P. Agrawal, "Airtime fairness for IEEE 802.11 multirate networks," *IEEE Trans. Mobile Comput.*, vol. 7, no. 4, pp. 513–527, Apr. 2008.

[32] C.-T. Chou, K. G. Shin, and N. S. Shankarn, "Contention-based airtime usage control in multirate IEEE 802.11 wireless LANs," *IEEE/ACM Trans. Netw.*, vol. 14, no. 6, pp. 1179–1192, Dec. 2006.

[33] M. Laddomada, F. Mesiti, M. Mondin, and F. Daneshgaran, "On the throughput performance of multirate IEEE 802.11 networks with variable-loaded stations: Analysis, modeling, and a novel proportional fairness criterion," *IEEE Trans. Wireless Commun.*, vol. 9, no. 5, pp. 1594–1607, May 2010.

[34] P. Lin, W.-I. Chou, and T. Lin, "Achieving airtime fairness of delay-sensitive applications in multirate IEEE 802.11 wireless LANs," *IEEE Commun. Mag.*, vol. 49, no. 9, pp. 169–175, Sep. 2011.

[35] V. Ramaiyan, A. Kumar, and E. Altman, "Fixed point analysis of single cell IEEE 802.11e WLANs: uniqueness and multistability," *IEEE/ACM Trans. Netw.*, vol. 16, no. 5, pp. 1080–1093, Oct. 2008.

[36] I. Tinnirello and G. Bianchi, "Rethinking the IEEE 802.11e EDCA performance modeling methodology," *IEEE/ACM Trans. Netw.*, vol. 18, no. 2, pp. 540–553, Apr. 2010.

[37] J. W. Robinson and T. S. Randhawa, "Saturation throughput analysis of IEEE 802.11e enhanced distributed coordination function," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 5, pp. 917–928, Jun. 2004.

[38] A. Banchs and L. Vollero, "Throughput analysis and optimal configuration of 802.11 e EDCA," *Comput. Netw.*, vol. 50, no. 11, pp. 1749–1768, 2006.

[39] B. M. Parker, J. A. Schormans, and S. G. Gilmour, "Increasing throughput in IEEE 802.11 by optimal selection of backoff parameters," *IET Netw.*, vol. 4, no. 1, pp. 21–29, 2015.

[40] P. Dong, J. Wang, H. Wang, and Y. Pan, "Boosting VoIP capacity via service differentiation in IEEE 802.11e EDCA networks," *Int. J. Distrib. Sensor Netw.*, vol. 11, no. 3, p. 235648, 2015.

[41] J. Y. Lee, H. S. Lee, and J. S. Ma, "Model-based QoS parameter control for IEEE 802.11e EDCA," *IEEE Trans. Commun.*, vol. 57, no. 7, pp. 1914–1918, Jul. 2009.

[42] Y. Gao, X. Sun, and L. Dai, "IEEE 802.11e EDCA networks: Modeling, differentiation and optimization," *IEEE Trans. Wireless Commun.*, vol. 13, no. 7, pp. 3863–3879, Jul. 2014.

[43] B. Li, R. Battiti, and Y. Fang, "Achieving optimal performance by using the IEEE 802.11 MAC protocol with service differentiation enhancements," *IEEE Trans. Veh. Technol.*, vol. 56, no. 3, pp. 1374–1387, May 2007.

[44] Y. Ge, J. C. Hou, and S. Choi, "An analytic study of tuning systems parameters in IEEE 802.11e enhanced distributed channel access," *Comput. Netw.*, vol. 51, no. 8, pp. 1955–1980, 2007.

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

[46] Z.-N. Kong *et al.*, "Performance analysis of IEEE 802.11e contention-based channel access," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 10, pp. 2095–2106, Dec. 2004.

[47] Y. Xiao, "IEEE 802.11e: QoS provisioning at the MAC layer," *IEEE Wireless Commun.*, vol. 11, no. 3, pp. 72–79, Jun. 2004.

[48] M. Derakhshani and T. Le-Ngoc, "Cognitive MAC designs: Background," in *Cognitive MAC Designs for OSA Networks*. Cham, Switzerland: Springer, 2014, pp. 15–31.

[49] M. Derakhshani and T. Le-Ngoc, "Adaptive access control of CSMA/CA in wireless LANs for throughput improvement," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2013, pp. 2951–2955.

[50] M. Chiang, *Geometric Programming for Communication Systems*. Berlin Germany: Hanover, 2005.

[51] G. Xu, "Global optimization of signomial geometric programming problems," *Eur. J. Oper. Res.*, vol. 233, no. 3, pp. 500–510, 2014.

[52] B. Bellalta, "IEEE 802.11ax: High-efficiency WLANS," *IEEE Wireless Commun.*, vol. 23, no. 1, pp. 38–46, Feb. 2016.

[53] M. D. Hershenson, S. P. Boyd, and T. H. Lee, "Optimal design of a CMOS op-amp via geometric programming," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 20, no. 1, pp. 1–21, Jan. 2001.

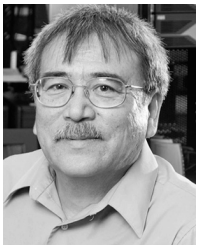
[54] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

MAHSA DERAKHSHANI (S'10–M'13) received the B.Sc. and M.Sc. degrees from the Sharif University of Technology, Tehran, Iran, in 2006 and 2008, respectively, and the Ph.D. degree from McGill University, Montreal, Canada, in 2013, all in electrical engineering. From 2013 to 2015, she was a Post-Doctoral Research Fellow with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, Canada, and a Research Assistant with the Department of Electrical and Computer Engineering, McGill University. From 2015 to 2016, she was an Honorary Natural Sciences and Engineering Research Council of Canada (NSERC) Post-Doctoral Fellow with the Department of Electrical and Electronic Engineering, Imperial College London. She is currently a Lecturer (Assistant Professor) in digital communications with the Wolfson School of Mechanical, Electrical and Manufacturing Engineering, Loughborough University. Her research interests include radio resource management for wireless networks, software-defined wireless networking, applications of convex optimization and game theory for communication systems, and spectrum sensing techniques in cognitive radio networks. She received the John Bonsall Porter Prize, the McGill Engineering Doctoral Award, the Fonds de Recherche du Quebec-Nature et Technologies, and the NSERC Postdoctoral Fellowships. She currently serves as an Associate Editor for the journal of *IET Signal Processing*.

XIAOWEI WANG received the B.Sc. degree from Shanghai Jiao Tong University in 2013, and the M.Sc. degree in electrical engineering from McGill University, Montreal, QC, Canada, under the supervision of Prof. T. Le-Ngoc, in 2016. He received the Academic Excellence Scholarship from Shanghai Jiao Tong University in 2012 and the Graduate Excellence Fellowship from McGill University in 2014.

DANIEL TWEED (S'13) received the B.Sc. degree (Hons.) in computer engineering from the University of Manitoba, Winnipeg, Canada, in 2015, and the M.Eng. degree from McGill University, Montreal, Canada, in 2018. His research interests include machine learning and neural networks, software-defined wireless network architectures, and resource allocation techniques for cognitive and autonomous networks.

THO LE-NGOC (F'97) received the B.Eng. degree (Hons.) in electrical engineering and the M.Eng. degree from McGill University, Montreal, QC, Canada, in 1976 and 1978, respectively, and the Ph.D. degree in digital communications from the University of Ottawa, Canada, in 1983. From 1977 to 1982, he was with Spar Aerospace Ltd., Sainte-Anne-de-Bellevue, Canada, where he was involved in the development and design of satellite communications systems. From 1982 to 1985, he was with SR Telecom Inc., Saint-Laurent, QC, Canada, where he developed the new point-to-multipoint DA-TDMA/TDM subscriber radio system SR500. From 1985 to 2000, he was a Professor with the Department of Electrical and Computer Engineering, Concordia University, Montreal. Since 2000, he has been with the Department of Electrical and Computer Engineering, McGill University. His research interest is in the area of broadband digital communications. He is a fellow of The Engineering Institute of Canada, The Canadian Academy of Engineering, and the Royal Society of Canada. He was a recipient of the 2004 Canadian Award for Telecommunications Research and the IEEE Canada Fessenden Award in 2005. He is currently the Canada Research Chair (Tier I) in Broadband Access Communications.



ALBERTO LEON-GARCIA (S'74–M'77–SM'97–F'99) is currently a Distinguished Professor in electrical and computer engineering with the University of Toronto. He has authored textbooks *Probability and Random Processes for Electrical Engineering* and *Communication Networks: Fundamental Concepts and Key Architecture*. His research is on application platforms for smart applications including smart cities. He has contributions in multiplexing and switching of integrated services traffic.