Joint Client Association and Random Access Control for MU-MIMO WLANs

Wangkit Wong, Kam-Wa Chau, and S.-H. Gary Chan, *Senior Member, IEEE*

Abstract—In interfering WLANs, clients are usually non-uniformly distributed. Therefore, their throughput can be improved by association control. WLANs adopt DCF (Distributed Coordinating Function) for media access control. The transmit probability of an AP is governed by its CW (Contention Window). Random Access control is to determine the CW size and hence the transmit probability of each AP. Client association control and random access control are coupled. APs associated with more clients require higher channel access opportunities. On the other hand, APs with higher transmit probability should serve more clients. We study the novel problem of maximizing the proportional fairness of client throughput by joint association control and random access control. As the joint problem is non-convex, we first consider that client association is given and propose an optimal policy for random access control. Given the CWs of APs, we propose a factor-2 approximation algorithm, Greedy-Asso to address the client association problem. Inspired by the optimal access control policy and Greedy-Asso, we further propose a distributed algorithm, termed CARA to tackle the joint optimization problem. Extensive simulation and experimental studies show that it outperforms comparison schemes by a wide margin in terms of throughput and fairness.

Index Terms—WLANs, distributed optimization, load balancing, association control, approximation algorithm

1 Introduction

We consider general MU-MIMO WLANs formed by multiple access points (APs). Each AP has a certain, possibly heterogeneous, number of antennas, with some assigned power range and channel. Fig. 1 shows an example of the WLANs under consideration. The AP locations may be random (due to inflexible installation locations, the need to reduce blind spots, the requirement on signal strengths, etc.), and client distribution may be non-uniform.

In the network, AP coverage may overlap. *Association control* is to assign or associate clients in the overlapping regions to one of the APs for service. The conventional approach is to selfishly pick the AP with the strongest signal. Such an approach leads to imbalanced AP load and unsatisfactory client throughput. As shown in Fig. 1, AP A operates on Channel 2, while AP B and AP C operate on Channel 1. Using the strongest signal association, Clients 2, 3, 4 and 5 would associate with AP B, creating congestion at the AP. To improve network performance, we can optimize client association by migrating (i.e., re-associating) some connected clients from the congested APs to those lightly loaded neighboring APs.

In interfering multi-cell WLANs, client association plays a bigger role than just load balancing. Proper client association decisions can substantially reduce channel contention. We take Client 3 in Fig. 1 as an example. It suffers from heavy channel contention if it is associated with AP B or AP C because it is within the overlapping coverage of these co-channel APs. We can reduce channel contention by associating clients in the overlapping coverage to APs operating on different channels. For example, migrating Client 3 to AP A leads to better performance.

An MU-MIMO AP is able to simultaneously transmit packets to different clients. The number of concurrent packets of an AP depends upon the number of antennas it has. That is an AP with K antennas is able to transmit to K clients simultaneously. In other words, APs with different numbers of antennas would have different serving capabilities. As shown in Fig. 1, AP A has more antennas than AP B. Therefore, a good client association solution assigns Clients 2 and 3 to AP A. Association optimization hence needs to take the heterogeneity of AP serving capability into consideration.

802.11 WLANs adopt distributed coordinated function (DCF) MAC to coordinate channel access. Each node maintains a contention window with size CW. When a node has pending packets to send, it first waits for a DIFS (DCF interframe spacing) period. Then, the node draws a random timer t from $[1, CW]$. It continually senses the medium at each time slot. If the medium is idle during the current slot, it decreases t. Otherwise, it suspends the decrement. When the timer t counts down to zero, the node transmits.

Clearly, the transmit probability of a node is governed by CW. If CW is fixed, the node transmits in a randomly chosen idle time slot with probability $p = 2/(CW+1)$ [1], [2]. In default DCF, CW is adjusted by Binary Exponential Backoff (BEB) mechanism. Initially, CW is set to be CW^{min}, which is referred to as minimum contention window size. BEB doubles the contention window whenever collisions occur.

- *The authors are with the Department of Computer Science and Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China.*
E-mail: wwongaa@ust.hk, kelvinwkspace@outlook.com, gchan@cse.ust.hk.

Manuscript received 28 Apr. 2018; revised 15 Jan. 2019; accepted 2 Aug. 2019. Date of publication 14 Aug. 2019; date of current version 3 Nov. 2020. (Corresponding author: Wangkit Wong.)
Digital Object Identifier no. 10.1109/TMC.2019.2935197

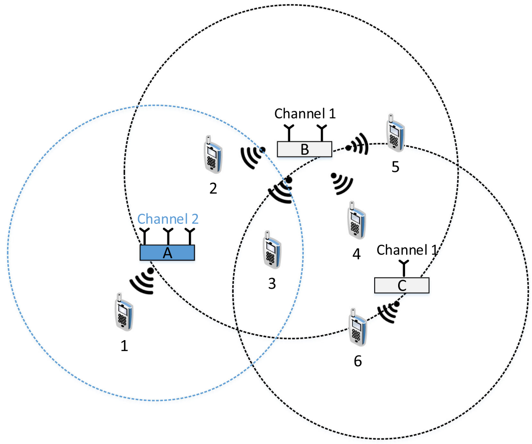


Fig. 1. A wireless network formed by APs with heterogenous numbers of antennas. The channel assigned to each AP is labelled at the top of the AP.

Consequently, the transmit probability is a function of CW^{min} (which can be found in [2]).

When an AP decides to transmit, it selects a subset of clients to serve. The probability that a client been selected is called the service probability of the client. Besides AP association, the throughput of a client is also determined by its service probability. *Random Access control* is to determine the transmit probability of each AP and the service probability assigned to its associated clients.

Previous research on client association [3], [4], [5], [6] has not considered media access control. They assume all APs are transmitting simultaneously. The Signal-to-Interference-and-Noise Ratio (SINR) received at clients degrades severely as the number of interfering APs increases. We take the network shown in Fig. 1 as an example. The distances from Client 4 to both AP B and AP C are similar. If AP B and AP C are transmitting at the same time, the SINR received at Client 4 is low (which is close to 1 in this case), no matter which AP it associates with. In contrast, DCF MAC backs off the transmission of AP C when Client 4 is receiving data from AP B. Therefore, association control needs to consider media access control.

Different decisions of client association form different communication links, and hence different network topologies. Different network topologies lead to different interference patterns, and hence different access control solutions. On the other hand, given transmit probabilities of APs, client association can be optimized. APs with higher transmit probabilities should serve more clients. Therefore, client association and access control are coupled, i.e., the decision of one affects the decision of the other. We then need to jointly optimize them.

Prior art [3], [4], [5], [7] on client association has mostly focused on downlink traffic. However, uplink traffic demand grows rapidly in the next-generation mobile network due to the popularity of user-generated content (UGC). Therefore, client association should address both downlink and uplink traffic.

In this paper, we jointly optimize client association and random access control for MU-MIMO WLANs to maximize the proportional fairness of client throughput, which provides a good tradeoff between fairness and network throughput [8], [9], [10]. Our study is novel because previous works have not considered the joint problem of such nature. The contributions of this work are as follows:

- *Problem formulation:* We formulate the optimization problem of joint client association and random access control to maximize the proportional fairness of client throughput. The optimization problem considers a practical and general wireless model that captures inter-AP interference and heterogeneity of antenna number. We optimize association and random access for both downlink and uplink throughput.
- *Efficient algorithms:* As the joint problem is non-convex and difficult to solve exactly, we first consider the case where client association is given, and propose an optimal access control policy to solve the access control problem. Given the transmission probability of each AP, we proposed a factor-2 approximation algorithm, termed Greedy-Asso to tackle the client association problem. Motivated by the optimal access control policy and Greedy-Asso, we proposed a distributed algorithm termed CARA (Joint **C**lient **A**ssociation and **R**andom **A**ccess Control) to tackle the joint problem. Clients and APs in CARA do not need global knowledge of the network; each node exchanges messages with its neighbors to compute the solution.
- *Extensive simulations and experimental studies:* We demonstrate the effectiveness and practicability of CARA with extensive simulation and real experimental study. We show that CARA outperforms recent schemes by a wide margin in terms of throughput and fairness.

The rest of the paper is organized as follows. After discussion on related work in Section 2, we present the problem formulation in Section 3. In Section 4, we present the optimal access control policy for the random access control problem. In Section 5, we present a factor-2 approximation algorithm for the case where transmission probabilities of APs are given. In Section 6, we propose a distributed algorithm, termed CARA for joint client association and access control. We present illustrative simulation results in Section 7 and experimental study in Section 8. Finally, we conclude in Section 9.

2 Related Works

Client Association in Networks with Single User Transmission Schemes. Many works [7], [11], [12], [13], [14], [15], [16] have studied user association for networks using single user transmission schemes. In these networks, APs are associated with multiple clients. However, at each time slot, an AP can only transmit to one of its clients. These works are impressive, however they cannot be directly applied to MU-MIMO networks because they have not considered the heterogeneity of AP serving power (due to heterogeneous numbers of antennas at APs).

Some research has studied on-line AP association [17], [18], [19], [20]. The association decisions made in these schemes are *irrevocable*, i.e., cannot be changed once the connection is established. Therefore, they are not adaptive to dynamic traffic when users may join and leave anytime, and hence their performance may degrade over time. We consider here re-optimizing association over time by migrating clients to achieve better performance, either periodically or upon detecting a drop in network performance.

There have been works [8], [9], [21], [22], [23] optimizing user association based on the assumption of no inter-AP interference. However, in reality, APs may interfere with each other. It is hence important to consider associating a user with an AP in the presence of interference. To address interference, previous research [3], [4], [5], [6] considers that all APs are transmitting simultaneously. The channel quality/SINR at a client is hence greatly degraded due to interference. On the other hand, we consider APs may interfere with each other and the channel access of interfering transmissions are scheduled by DCF MAC.

Client Association in Multi-User Networks. Another line of research studies user association in MU-MIMO networks. An MU-MIMO AP is capable of transmitting to multiple users simultaneously. Some of these schemes [24], [25], [26], [27], [28] optimize user association using instantaneous channel state information. These works jointly design cell association and beamformer to optimize some utility, such as weighted sum rate or energy consumption. They incur high overhead as optimal cell association changes per coherent time (which is in the order of a few ms). To reduce overhead, user association schemes based on long-term channel condition have recently been proposed [5], [29], [30], [31]. These schemes optimize ergodic client performance using channel state information (CSI) statistics instead of instantaneous CSI.

Some works consider the case of a large number of antennas, i.e., the number of antennas are much larger than the number of clients [32], [33]. Assuming that all clients can be served continuously with abundant antenna elements, these works formulate an optimization model to maximize sum client rate. Client association for massive MIMO networks has been investigated in [34], [35]. A nice survey on client association in 5G is given in [36]. The work in [37] presents an optimal joint antenna assignment and admission control policy for MIMO mesh networks.

Random Access Control: Considering network topologies are fixed, works in [38], [39] propose distributed algorithms to maximize concave utilities of client throughput. Based on locally available channel state, i.e., the average number of consecutive idle slots between two transmissions, Xia et al. [40] design a dynamic contention window adjustment algorithm using control theory. In [41], a random access control protocol, termed PoCMAC is proposed for full-duplex APs. PoCMAC jointly optimizes transmit power and contention window sizes. Although, these works are impressive, they have not considered optimizing network topologies via client association control.

A large body of work has been done on centralized access scheduling for enterprise WLANs [42], [43], [44], [45], [46]. Centralized access scheduling brings significant throughput gain. However, it is difficult to be implemented in a large network due to heavy control overhead.

A preliminary version of this work has been reported in [47]. The work considers the case of no AP interference. The current work advances from it in several major ways: 1) It considers the more realistic scenario where APs may interfere. 2) It considers channel access control, and formulates the problem of joint client association and access control. 3) The previous algorithm considers only single antenna APs. The current work considers a more general scenario of MU-MIMO APs with heterogeneous serving capability. 4) The previous work focuses on centralized algorithms. This work, on the other hand, presents a highly distributed and scalable algorithm, CARA.

3 System Model and Problem Formulation

3.1 System Setting

We consider general MU-MIMO WLANs consisting of a number of APs, each AP is wired to the Internet. Let $\mathbf{A}$ be the set of APs and m be its cardinality, i.e., $|\mathbf{A}| = m$. The frequency channel on which AP i operates is denoted as $c(i)$. We denote the set of clients as $\mathbf{U}$, where $|\mathbf{U}| = n$.

Both APs and clients are referred to as nodes. Denote $\mathbf{N} = \mathbf{A} \cup \mathbf{U}$ as the set of nodes in the network. The number of antennas at node i is denoted as K_i, which is also referred to as its DoF (Degree of Freedom).

Transmission from node i to node j is represented by a link (i, j). We assume a i.i.d. Rayleigh fading model. The wireless channel from node i to node j is denoted as $H_{ji} \in \mathbb{C}^{N_j \times N_i}$. Denote the element at the mth row and nth column of channel matrix H_{ji} as $[H_{ji}]_{mn}$. Each element of H_{ji} is composed by distance based path-loss and fast fading channel coefficient. That is $[H_{ji}]_{mn} = [h_{ji}]_{mn}\sqrt{\gamma_{ji}}$, where γ_{ji} is the long-term channel, determined by large-scale factors such as path loss and shadowing between i and j. Complex number $[h_{ji}]_{mn}$ is the fast fading channel coefficient between the mth receive antenna and the nth transmit antenna. $[h_{ji}]_{mn}$ changes quickly in the order of milliseconds. $[h_{ji}]_{mn} \sim \mathcal{CN}(0, 1)$ and $[H_{ji}]_{mn} \sim \mathcal{CN}(0, \gamma_{ji})$. The long-term channel, i.e., γ_{ji} can be estimated at the receiver side. For example in [5], [34], [35], [48], the long-term channel/large-scale fading is assumed given and easy to obtain.

A client accesses the Internet by associating with an AP. We denote the set of APs that client j potentially associates with as $\mathbf{A}(j)$. Specifically, $\mathbf{A}(j)$ is the set of APs from which the channel quality at client j is larger than some threshold τ, i.e., $\gamma_{ij} \geq \tau, \forall i \in \mathbf{A}(j)$. Following convention, traffic sent from an AP to a client is referred to as downlink traffic. While traffic sent from a client to an AP is referred to as uplink traffic.

Each node has some transmit probability (the probability it transmits after sensing an idle time slot). The transmit probability of node n is denoted as P^n. In this work, we consider optimizing the transmit probability of APs. However, we assume that the transmit probabilities of clients are given and cannot be modified. We assume packet are always backlogged at the transmitter. We assume whenever an AP acquires the channel, it transmits for a fixed number of slots. Packet fragmentation and aggregation are allowed.

3.2 System Operation

We show in Fig. 2 the schematic for the system operation. It consists of two time scales:

- Long-term time scale: Our user association and random access control are carried out over long-term time scale based on CSI statistics (i.e., γ_{ji}).

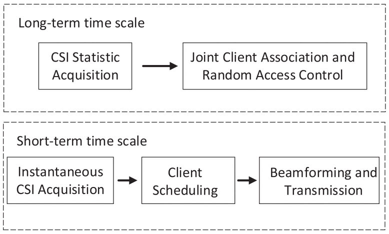


Fig. 2. Block diagram of long-term and short-term operation of the system under study.

- Short-term time scale: The short-term time scale is determined by the coherence time of the instantaneous CSI (i.e $[h_{ji}]_{mn}$). Each AP performs client selection and beamforming design over the short-term time scale.

We consider each AP transmits concurrent packets to its clients with zero-forcing beamforming. In the ideal case, when AP i decides to transmit, it selects $B \leq K_i$ clients according to the instantaneous client channel to maximize a certain objective, such as weighted sum-rate [49], [50] or logarithm utility of client rate [51]. Then, it designs beamformer and optimize power allocation according to the channel matrix of these B clients.

Since the instantaneous CSI of client channel change quickly (in the order of milliseconds), optimizing client association and AP transmit probability based on instantaneous CSI will incur heavy overhead. Therefore, we decoupled user association and random access optimization from client scheduling or bemaformer design.

The joint user association and random access optimization is carried out based on CSI statistics over long-term time scale. As we do not optimize for a particular channel realization, we need to make approximations on the client scheduling and beamforming design procedures. Hence, we assume AP i always schedules K_i clients and allocates power evenly to all clients. Similar assumption has been made in [5], [48], [52]. We assume that each client receives only one stream at a time.

The sender and receiver of a link (i, j) apply linear processing to mitigate interference. AP i uses $V_{ji} \in \mathbb{C}^{N_i \times 1}$ as a zero-forcing precoder for client j. Similarly, client j uses linear receiver combine vector $U_j \in \mathbb{C}^{1 \times N_j}$ as the receive filter.

Let $\mathbf{B}(i)$ be the set of clients selected for downlink transmission. The instantaneous received signal at client j of link (i, j) is given by

$$y_{ji} = \underbrace{U_j H_{ji} V_{ji} x_j}_{\text{desired signal}} + \underbrace{\sum_{m \in \mathbf{B}(i), m \neq j} U_j H_{ji} V_{mi} x_m}_{\text{inter-stream interference}} + \underbrace{I_{ji}}_{\text{inter-AP interference}} + \underbrace{U_j z}_{\text{backgroup noise}}, \tag{1}$$

where x_j is the symbol for client j. In order to eliminate inter-stream interference, precoding vector V_{ji} and receiver combiner are designed to satisfy the following conditions:

1) $U_m H_{mi} V_{ji} = 0, \forall m \in \mathbf{B}(i), m \neq j$;
2) $U_j H_{ji} V_{ji} > 0, \forall j \in \mathbf{B}(i)$.

It has been shown that these conditions can be satisfied almost for sure [53], [54], [55]. Clearly, signal $U_j H_{ji} V_{ji} x_j$ received at client j is hence affected by precoding vector V_{ji}. We consider the precoding vector and the receive filter are semi-unitary i.e., $|V_{ji}|^2 = 1$ and $|U_j|^2 = 1$. We assume that V_{ji} is generated according to zero-forcing that does not consider H_{ji}. That is V_{ji} is statistically independent of H_{ji}. Due to the bi-unitary invariance of H_{ji}, each element of vector $H_{ji} V_{ji}$ is i.i.d. $\mathcal{CN}(0, \gamma_{ji})$ distributed Gaussian random variable. $U_j H_{ji} V_{ji}$ is the linear combination of K_j i.i.d Gaussian random variables. Therefore, the effective desired channel for client j is (Proof can be found in Lemma 1 of [52] and Lemma 1 of [56])

$$g_{ji} = U_j H_{ji} V_{ji} \sim \mathcal{CN}(0, \gamma_{ji}). \tag{2}$$

Since interfering transmitters are coordinated by DCF, the interference I_{ji} of successful transmission is negligible. AP i allocates its power evenly for all clients. Therefore, the received SNR (signal-to-noise-ratio) at client j is $\frac{\gamma_{ji} Power_i}{K_i \sigma^2}$, where σ is the background noise density. $Power_i$ is the transmit power of node i. For uplink traffic, a client allocates all its power for the transmission as it can only transmit to one AP at a time. Therefore, the received SNR is $\frac{\gamma_{ij} Power_j}{\sigma^2}$. Then ergodic capacity of link (i, j) is computed as [52]

$$C(i, j) = \mathbb{E}\left[\log\left(1 + \frac{|g_{ji}|^2 Power_i}{K_i \sigma^2}\right)\right], = e^{1/\rho_{ij}} E_1(1/\rho_{ij}), \tag{3}$$

where $\rho_{ij} = \frac{\gamma_{ji} Power_i}{\sigma^2 K_i}$ and $E_1(\eta) = \int_\eta^\infty t^{-1} e^{-t} dt$ is the exponential integral. Similarly, uplink capacity is computed as $C(j, i) = e^{1/\rho_{ji}} E_1(1/\rho_{ji})$, where $\rho_{ji} = \frac{\gamma_{ij} Power_j}{\sigma^2}$.

3.3 Media Access Model

In this section, we present the media access model under consideration. Two-way error free connectivity is assumed in our model. The transmission of link (i, j) is interfered by node n if the received power of n's signal at receiver j or sender i is above some sensing threshold. This is because data are transmitted from i to j and the ACK is sent from j to i. For successful transmission, both data and ACK need to be received correctly.

We assume there are no hidden terminals. This is possible if the range of carrier-sensing is large enough [57], [58], [59]. During the backoff stage, node i senses the channel to perform clear channel assessment (CCA). If any node in its sensing range is transmitting, the node backs off its transmission. This implies that the channel access of node i is conflicting with the channel accesses of nodes within its sensing range. We denote by $\mathbf{CF}(i)$ the set of nodes conflicting with node i. We assume a symmetric conflict graph. That is if node $n \in \mathbf{CF}(i)$, then node i is also in the set $\mathbf{CF}(n)$. We show an example in Fig. 3. AP a, AP c, client 1, client 2 and client 3 are within the sensing range of AP b. Consider that AP c is transmitting to client 3. Then AP b senses a busy channel.

Denote the size of the contention window of node n as CW^n, which determines the transmission probability of node n. In this paper, we consider using fixed/stationary CW^n which is optimized according to the network topology. Node n with contention window size CW^n randomly selects a

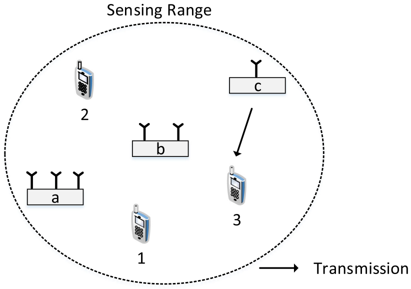


Fig. 3. A network with interfering nodes.

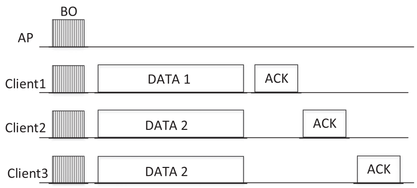


Fig. 4. Downlink transmission scheme with fixed duration

number t from $[1, CW]$ and backs off for t idle time slots. Since the average value of t is $(CW^n + 1)/2$, the node transmits with probability $P^n = 2/(CW^n + 1)$ after sensing an idle time slot.

We show the channel access of an AP in Fig. 4. Similar to the scheme presented in [60] (Fig. 11), when an AP transmits, it selects K_i clients and transmits for a duration of L_i number of time slots. The reason we use fixed transmit duration for all scheduled clients is shown in Fig. 5. In this figure, AP transmit to each scheduled client a packet with the same size. Since the link rates from the AP to different clients are different. Some clients finish receiving faster than the others. As shown in the figure, after client 1 completes receiving the packet, it has to wait for the completion of other clients with slower link rates. The blocks in blue color indicate the waste of channel air time. During to the overhead DCF control frames (such as SIFS, AIFS and ACK), only β fraction of time is used for payload transmission.

3.4 Problem Formulation

In this section, we formulate the joint association and access control problem as an integer programming problem. Notations we use are summarized in Table 1.

We first consider downlink traffic. Given that AP i decides to transmit, the probability that it transmits to client j is denoted as p_{ij}. That is flow (i, j) transmit with probability $p_{ij}P^i$ in an idle slot. Let x_{ij} be the association decision variable indicating whether client j associates with AP i. Clearly, $x_{ij} \in \{0, 1\}$ is a binary variable. Since each client can associate with only one AP, we must have the following constraints.

$$\sum_{i \in \mathbf{A}} x_{ij} = 1, \forall j \in \mathbf{U}. \tag{4}$$

Previous analytic frameworks [58], [61], [62] show that the channel activities of CSMA wireless networks can be modeled by a continues Markov chain (CMC). Each node in the network can be and only can be in two states: active or inactive. Clearly, a state $\mathbf{s}$ of the network and also a state of

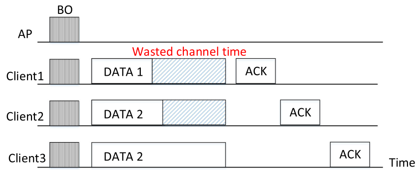


Fig. 5. Downlink transmission scheme with fixed size packets

TABLE 1
Major Symbols Used in this Paper

Notation	Definition
$A(j)$	Set of APs that client j is potential to associate with
g_{ji}	Long-term channel from node i to node j
K_i	DoF (number of antennas) of AP i
$Power_i$	Total power budge of node i
$C(i, j)$	Data rate of link from node i to node j
$\mathbf{CF}(i)$	Set of nodes conflicting with node i
P^n	Node n's transmission probability
p_{ij}	The probability that AP i transmit to client j given that it transmits
x_{ij}	Binary variable indicating whether user j is associated with AP i
β	The fraction of time used for payload transmission during transmission duration L_i

the CMC is represented by the set of all active nodes. For example, $s = \emptyset$ means no node is transmitting in the state, while $\mathbf{s} = \{i, n\}$ implies that only node i and n are active.

If the system is in state $\mathbf{s}$ where node i and its conflicting nodes (i.e nodes in $\mathbf{CF}(i)$) are inactive, the system transits from state s to state $\mathbf{s} + i$ with the transition rate P^i. Similarly, the system transits from state $\mathbf{s} + i$ to s with the transition rate $1/L_i$. A state $\mathbf{s}$ is feasible iif there are no conflicting nodes.

Let $\mathcal{S}$ be the set of all feasible states. It has been shown that the stationary probability that the network is in state s, $\pi(s) = \frac{\prod_{n \in \mathbf{s}} P^n L_n}{C_0}$, where C_0 is a constant equal to $\sum_{s \in \mathcal{S}} \pi(s)$ [58], [63]. We show an example in Fig. 6. As shown in Fig. 6 A), there is a network formed by node 1 and node 2. The edge between them indicates that they conflict with each other. Therefore, the feasible states of the network are $\emptyset$, $\{1\}$ and $\{2\}$. The numbers labeled at the arcs of Fig. 6 B) show the transition rates between states. The steady probability of state $\{1\}$ is hence $\frac{P^1 L_1}{1 + P^1 L_1 + P^2 L_2}$. Hence, the throughput of flow (i, j), T_{ij} is given by

$$\beta C(i, j) p_{ij} \sum_{\mathbf{s} \in \mathcal{S}, i \in \mathbf{s}} \pi(s), \tag{5}$$

where $\sum_{\mathbf{s} \in \mathcal{S}, i \in \mathbf{s}} \pi(s)$ is interpreted as the fraction of active time of node i.

However, the throughput modeled in Equation (5) is mathematically untractable as it requires enumerating all the independent sets and global information. It can be approximated by the following equation (See Equation (21) in [61]). It has been shown that this approximation is close to the exact model (See Fig. 9 in [61]).

$$T_{ij} = \frac{\beta C(i, j) p_{ij} P^i L_i}{(1 + P^i L_i) \prod_{n \in \mathbf{CF}(i)} (1 + P^n L_n)}, \tag{6}$$

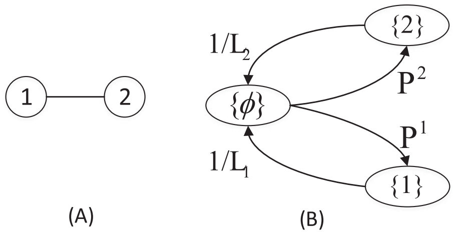


Fig. 6. A network with two nodes and its corresponding continuous Markov chain.

Note that we assume there is no hidden terminals and perfect carrier sensing. Therefore collision due to hidden nodes is not captured in the throughput calculation. Let $\tau_i = P^i L_i/(1+P^i L_i)$. We can transfer it into a form similar to the throughput model of Aloha networks. $T_{ij} = \beta C(i,j) p_{ij} \tau_i \prod_{n \in \mathbf{CF}(i)}(1-\tau_n)$. Thereby, we denote $\pi(i) = \frac{P^i L_i}{(1+P^i L_i)\prod_{n\in\mathbf{CF}(i)}(1+P^n L_n)}$ as the approximated fraction of time that node i is active.

Denote the downlink throughput of client j as T_j^D. As client j is possible to associate with any AP, its downlink throughput can be calculated as follows:

$$T_j^D = \sum_{i\in\mathbf{A}} x_{ij} T_{ij}.$$

We consider a client scheduler with which each AP selects clients for downlink transmission with probabilities proportional to their weights, i.e. $p_{ij} = \min\{\frac{w_j K_j}{\sum_{k\in\mathbf{U}(i)} w_k}, 1\}$, where $\mathbf{U}(i)$ is the set of clients associated with AP i. The scheduler can be achieved as follows. Consider M transmit opportunities (TXOP) of AP i. M is a sufficiently large number, for example $M = 100$. AP i maintains a transmission counter M_j for each client j. M_j is the number of time (frequency) that AP i selects client j. When AP i transmits, it randomly selects K_i clients which with counter $M_j < Mp_{ij}$. In this way, AP i guarantees to transmit to client j for $p_{ij}M$ times. Therefore, the probability that AP i selects j is approximately equal to p_{ij}. During M TXOPs, AP i can schedule MK_i clients. Since $\sum_{j\in\mathbf{U}(i)} p_{ij} M_j = MK_i$, the scheduler is feasible.

We next consider the uplink throughput of client j. Client j associates with only one AP, whenever it accesses the channel, it transmits uplink packets to its associated AP. Therefore $p_{ji} = 1$. The uplink throughput is given by

$$T_{ji} = \frac{\beta C(j,i) p_{ji} P^j L_j}{(1+P^j L_j)\prod_{n\in\mathbf{CF}(j)}(1+P^n L_n)}. \tag{7}$$

Similarly, the uplink throughput of client j can be calculated as $T_j^U = \sum_{i\in\mathbf{A}} x_{ij} T_{ji}$. Usually, downlink and uplink traffic in WLANs are not symmetric. A WLAN often has more downlink traffic requirement than uplink traffic requirement. Moreover, different clients in the network may have different traffic requirement or priorities. We, therefore, associate the downlink and uplink transmission of each client with some weight which indicates the importance of the transmission. Denote the weight of downlink transmission of client j as w_j^D and the weight of its uplink transmission as w_j^U.

To tradeoff between fairness and network throughput, we seek to maximize the proportional fairness of all traffic flows (including downlink and uplink traffic), given by the following client throughput utility function:

$$V(\mathbf{U}) = \sum_{j\in\mathbf{U}} \Big(w_j^D \log T_j^D + w_j^U \log T_j^U \Big). \tag{8}$$

The *J*oint *C*lient *A*ssociation and *R*andom *A*ccess control *P*roblem (JCARAP) can be formally expressed as

$$\begin{aligned} \max\ & V(\mathbf{U}), \\ \text{s.t.}\ & \text{Constraints } (4), \\ & x_{ij} \in \{0,1\}, \forall i \in A, j \in \mathbf{U}, \\ & P^i \geq 0, \forall i \in A. \end{aligned} \tag{9}$$

The optimization problem is an integer programming, where x determines association decisions and P determines random access control decisions. Due to integer constraint $x_{ij} \in \{0,1\}$, it is non-convex and difficult to solve in general. We show in Section 1 of the supplementary materials, which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/TMC.2019.2935197 that the client scheduler is optimal for any given P^i.

4 An Optimal Algorithm for Access Control

In this section, we consider the case where client association is given. Our joint problem then reduces to the random access control problem. We only need to decide the transmit probability of each AP. We present the optimal solution to the problem.

As client association is given, we denote the AP associated by client j as a_j. The set of clients associated with AP i is denoted as $\mathbf{U}(i)$. For the sake of simplicity in notations, we further define w_n as the weight of node n. If n is a client, w_n is defined as its uplink weight, i.e., $w_n = w_n^U$. If n is an AP, w_n is defined as the sum of its clients' downlink weight, i.e., $w_n = \sum_{k\in\mathbf{U}(n)} w_k^D$.

The joint optimization problem is hence reduced to the following ACP (**A**cess **C**ontrol **P**roblem).

$$\begin{aligned} \text{ACP:}\ & \max\ V(\mathbf{U}), \\ & s.t.\ \ P^i \geq 0, \forall i \in \mathbf{A}. \end{aligned} \tag{10}$$

The optimal solution to the ACP can be obtained in closed-form given by Theorem 1, available in the online supplemental material.

Theorem 1. *Optimal Access Control Policy:*

$$P^i = \left[\frac{\sum_{k\in\mathbf{U}(i)} w_k^D}{L_i \sum_{n\in\mathbf{CF}(i)} w_n} \right]_{P^{min}}^{P^{max}}, \forall i \in \mathbf{A}, \tag{11}$$

Notably, the optimal transmit probability of AP i only depends on the aggregated downlink weight of its servicing clients, $\sum_{k\in\mathbf{U}(i)} w_k^D$ and the aggregated weight of nodes conflicting with the AP, $\sum_{n\in\mathbf{CF}(i)} w_n$.

$[P]_{Pmin}^{Pmax} = \max\{\min\{P, P^{max}\}, P^{min}\}$. It requires the transmit probability of each AP to be less than P^{max} and greater than P^{min}. Due to practical issues, most channel access protocols have such constraint. For example, in the EDCF, the maximum allowed channel access probability for voice traffic is $1/3$ (with a minimum CW size 2).

5 An Online Algorithm for Client Association

In this section, we consider that APs' transmit probabilities are given. To gain some insight into the association problem, we consider a special case where all clients have the same weight and focus on downlink traffic only. It is worthy of being studied as a large body of work focuses on similar cases [3], [5]. We propose a greedy algorithm, termed Greedy-Asso to tackle the problem.

As we have shown in Section 3.4, when all clients have the same weight, the optimal channel access policy is to serve clients with the equal probability. Therefore, the joint optimization problem is reduced to the following Client *A*ssociation *P*roblem (CAP).

CAP:

$$\max \sum_{j \in \mathbf{U}} \log \left(\sum_{i \in \mathbf{A}} x_{ij} \beta C(i,j) p_{ij} \pi_i \right),$$
$$p_{ij} = \begin{cases} \frac{K_i}{|\mathbf{U}(i)|}, & \text{if} |\mathbf{U}(i)| > K_i; \\ 1, & \text{otherwise}, \end{cases}$$
$$x_{ij} \in \{0,1\}, \forall i \in A, j \in \mathbf{U},$$
$$\sum_{i \in \mathbf{A}} x_{ij} = 1, \forall j \in \mathbf{U}.$$

Consider that clients in the network associate with APs in order. Before associating client j, a set of clients have arrived at the network. We abuse the notation a little bit here by also representing the set of clients which arrived before j as $\mathbf{U}$.

Recall that the optimization objective is maximizing utility $V(\mathbf{U})$. We define the marginal utility of associating a new client j to AP i given the set of previously associated clients $\mathbf{U}$ as $\Delta V(i,j|\mathbf{U})$. Marginal utility measures the change in the objective value V.

$$\Delta V(i,j|\mathbf{U}) = \begin{cases} \log(\beta C(i,j)\pi_i) & \text{if} |\mathbf{U}(i)| + 1 \leq K_i; \\ \log(\beta C(i,j)\pi_i K_i) + |\mathbf{U}(i)| \log |\mathbf{U}(i)| \\ -(|\mathbf{U}(i)|+1) \log(|\mathbf{U}(i)|+1), & \text{otherwise}. \end{cases} \tag{12}$$

We next present an online approximation algorithm, termed Greedy-Asso to tackle the CAP.

Algorithm 1. Greedy-Asso

Input: A, **U**, **P**
Outpur: x
1: **for** $j = 1, 2, \ldots n$
2: $i^* = \arg\max_i \Delta V(i,j|\mathbf{U})$
3: Associate client j to AP i^*
4: $\mathbf{U} = \mathbf{U} \cup j$
5: **end**

Greedy-Asso is an online algorithm. A joining client makes association decision without knowing any information about the clients arriving after it. It computes the marginal utility of associating with an AP i and associates with the AP such that the marginal utility is maximized.

Theorem 2. *Greedy-Asso is a factor-2 approximation algorithm. (Please refer to Section 3 of the Supplementary Material, available in the online.)*

6 Distributed Joint Client Association and Random Access Control

Motivated by the optimal access control policy and the approximation algorithm for online client association, we present in this section a distributed algorithm, termed Joint *C*lient *A*ssociation and *R*andom *A*ccess Control (CARA) to tackle the joint optimization problem.

Clients may join and leave the network at any time. To accommodate client joining and leaving, CARA consists of two operations. They are AP selection at the joining phase and association re-optimization. When a client is joining the network, it selects the best AP to associate with. As the network evolves with client departure, AP load becomes imbalanced even if we start with optimal AP selection. Therefore, CARA adopts re-optimization to continuously balance AP load.

6.1 Periodic Transmit Probability Adjustment

CARA optimizes the transmit probabilities of APs periodically according to the optimal access control policy (Formula (11)). To realize this, an AP i needs information on the downlink weight of clients associated with it (i.e., $w_j^D, \forall j \in \mathbf{U}(i)$) , the set of nodes (both APs and clients) conflicting with AP i (i.e., $\mathbf{CF}(i)$) and the weight of each node in $\mathbf{CF}(i)$. The weight of clients in $\mathbf{U}(i)$ can be obtained naturally at the AP. However, the weight of nodes in $\mathbf{CF}(i)$ are not naturally available at the AP. We propose the following approach to learn such information.

The nodes that are conflicting with AP i is at most two-hop away from i in the connectivity graph. Using the client reports standardized in the 802.11k standard, AP i can learn the local network topology two-hop away from it. In a client report, APs require its clients to perform active/passive channel scanning on the used channel. Clients report all nodes observed in its vicinity to its associated AP. Channel scanning can be done in around hundreds of milliseconds. Periodically, each AP exchanges client reports and client weights with its neighbour APs via the wired backhaul. In this way, each AP can find out all nodes conflicting with it and the weights of these nodes. Similar approaches have been studied in [64], [65] for building network conflict graphs.

6.2 AP Selection at the Joining Phase

Similar to the Greedy-Asso algorithm, each client j selects the AP to associate with such that the client throughput utility $V(\mathbf{U})$ (given by Equation (8)) is maximized. While Greedy-Asso only focuses on downlink throughput, CARA optimizes for both downlink and uplink throughput.

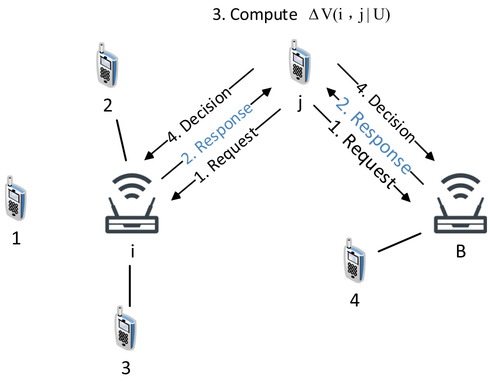


Fig. 7. An illustration of the AP selection algorithm.

Consider that client j is joining the network. $\mathbf{U}$ is the set of clients which have arrived at the network before j. To select the best AP, client j needs to compute the marginal utility $\Delta V(i, j|\mathbf{U})$ of associating to AP i for all i in $\mathbf{A}(j)$.

CARA jointly optimizes client association and random channel access. APs adopt the optimal access control policy, shown in Equation (11), hence the association of client j with AP i may change the channel access policies of APs in the vicinity. For reducing complexity, the joining client assumes that only AP i needs to increase its transmit probability to accommodate the new client when computing $\Delta V(i, j|\mathbf{U})$. While the transmit probabilities of other APs remain the same. Therefore,

$$\begin{aligned}
&\Delta V(i, j|\mathbf{U}) = \\
&\underbrace{\sum_{k\in\mathbf{U}(i)} w_k \log\left(\frac{\hat{p}_{ik}\hat{P}^i(1+P^iL_i)}{p_{ik}P^i(1+\hat{P}^iL_i)}\right) - \sum_{n\in\mathbf{CF}(i)} w_n \log\left(\frac{1+\hat{P}^iL_i}{1+P^iL_i}\right)}_{\text{utility change due to AP } i} \\
&\underbrace{+ w_j^D \log\left(\frac{\beta C(i,j)\hat{p}_{ij}\hat{P}^iL_i}{(1+\hat{P}^iL_i)\prod_{n\in\mathbf{CF}(i)}(1+P^nL_n)}\right)}_{\text{downlink throughput utility of client } j}, \\
&\underbrace{+ w_j^U \log\left(\frac{\beta C(j,i)p_{ji}P^jL_j}{(1+P^jL_j)\prod_{n\in\mathbf{CF}}(1+P^nL_n)}\right)}_{\text{uplink throughput utility of client } j}, \\
&\underbrace{- \sum_{n\in\mathbf{CF}(j)} w_n \log(1+P^jL_j)}_{\text{utility change due to client } j},
\end{aligned} \tag{13}$$

where $\{P^i, p_{ij}\}$ is the access control policy of AP i before the joining of client j. Given that client j associates with AP i, AP i increases its transmit probability to $\hat{P}^i$ and serves its client k with probability $\hat{p}_{ik}$ according to the client scheduler. This might change the throughput utility of its associated clients and nodes which can sense its transmission.

We show an example in Fig. 7. Given that client j associates with AP i, there are more clients sharing the service time of the AP. Therefore the throughput of its clients (i.e., client 2 and client 3) may change. Moreover, due to the change of AP i's transmit probability, the transmission opportunities of nodes which conflicting with AP i (i.e., client 1, client 2, client 3, client 4 and AP B) are affected. In Equation (13), the term $\sum_{k\in\mathbf{U}(i)} w_k \log(\frac{\hat{p}_{ik}\hat{P}^i(1+P^iL_i)}{p_{ik}P^i(1+\hat{P}^iL_i)})$ computes the throughput change of AP i's associated clients. Similarly, the term $-\sum_{n\in\mathbf{CF}(i)} w_n \log(\frac{1+\hat{P}^iL_i}{1+P^iL_i})$ measures the throughput change of nodes in $\mathbf{CF}(i)$.

The term $w_j^D \log(\frac{\beta C(i,j)\hat{p}_{ij}\hat{P}^iL_i}{(1+\hat{P}^iL_i)\prod_{n\in\mathbf{CF}(i)}(1+P^nL_n)})$ estimates the downlink throughput utility achieved by client j if associating with AP i. The term $\prod_{n\in\mathbf{CF}(i)}(1+P^nL_n)$ can be precomputed at the AP side because AP i is able to learn the transmit probabilities and transmission duration of nodes in its sensing range using 802.11k standard. Such infomation can be sent to the joinning client with probe responses.

The term $w_j^U \log(\frac{C(j,i)p_{ji}P^jL_j}{(1+P^jL_j)\prod_{n\in\mathbf{CF}(j)}(1+P^nL_n)})$ estimate the uplink throughput of client j. However, the term $\prod_{n\in\mathbf{CF}(j)}(1+P^nL_n)$ is not easy to obtain during the joining phase. Therefore, we approximate $\mathbf{CF}(j)$ by all the APs that discovered by client j. Transmit probability P^n and duration L_n of these APs can be sent to j with probe responses.

The last term, $-\sum_{n\in\mathbf{CF}(j)} w_n \log(1+P^jL_j)$ measures how much other nodes are affected due to the uplink traffic of client j. However, at the joining phase, it is difficult to determine which nodes conflicting with client j. Therefore, the set $\mathbf{CF}(j)$ only contains APs which receive the probe request of the client.

The detail of the AP selection algorithm is shown in Algorithm 2. We also show an illustrative example in Fig. 7. As shown in Line 2 of Algorithm 2, the client first scans the channel to discover nearby APs by broadcasting probe requests. We consider an AP successfully receiving the request as a candidate for j to associate with. Therefore, it is in set $\mathbf{A}(j)$.

Algorithm 2. AP Selection Algorithm

1 **Client** j**:**
2 Broadcasts probe requests
3 Receives probe responses from APs
4 Computes $\Delta V(i, j|\mathbf{U})$ for each AP i
5 Associates with AP i^*, such that $i^* = \arg\max_i \Delta V(i, j|\mathbf{U})$
6 Broadcasts the association decision
7 **Each AP** i **in** $\mathbf{A}(j)$**:**
8 On receiving probe requests, sends a probe response
9 Receives the association decision from j
10 Updates transmit probability according to the access control policy (Formula (11))

On receiving the request, each AP replies with a probe response frame. The response consists 1) the utility change due to AP i; 2) the serving probability of j, $\hat{p}_{ij}$; 3) $\prod_{n\in\mathbf{CF}(i)}(1+P^nL_n)$; 4) its transmit probability P^i, its weight w_i and transmission duration L_i.

With the information obtained from probe responses, client j computes the marginal utility $\Delta V(i, j|\mathbf{U})$ for each AP i in $\mathbf{A}(j)$ distributively according to Equation (13). Then it selects the AP that maximizes the marginal utility (from Line 4 to Line 5). Then, it broadcasts its association decision to notify nearby APs (Line 6). Each AP l in $\mathbf{A}(j)$ receive a notification from j and updates topology information $\mathbf{U}(l)$ and $\mathbf{CF}(l)$. According to the topology information, each AP updates its transmit probability (Line 10).

TABLE 2
SINR-MCS Mapping

SINR (dB)	<4	[4, 5)	[5, 9)	[9, 11)	[11, 15)	[15, 18)	[18, 20)	[20, 23)	≥ 23
MCS index	None	0	1	2	3	4	5	6	7

6.3 Association Re-Optimization

When the network evolves with client departure, client association on APs becomes imbalanced even with optimal AP selection. To address this, CARA adopts association re-optimization to balance AP load continuously.

Ideally, association re-optimization can be conducted whenever there is client departure. However, it might be costly. Therefore, CARA conducts association re-optimization periodically.

In association re-optimization, each client periodically tries to improve current association solution locally. It considers re-associating to a new AP to improve client throughput utility $V(\mathbf{U})$. In re-association, j first disconnects from the current AP and associates with a new AP.

Define $\Delta V(i,j|\mathbf{x})$ as the change of utility due to re-associating client j with AP i given current association x. Each client j searchs all APs in $\mathbf{A}(j)$ and computes $\Delta V(i,j|\mathbf{x})$ for all $i \in \mathbf{A}(j)$. Then, it re-associates to the AP such that the network utility V is increased the most. Let a_j be the AP currently associated with j. For Implementability and network stability, during the re-association, client j consider that the transmit probabilities of AP a_j and i remain the same. Therefore, $\Delta V(i,j|\mathbf{x})$ can be calculated distributively.

Clearly, the association re-optimization converges. This is because whenever re-optimization is made, client throughput utility increases and the utility is upper bounded by some positive value.

7 Illustrative Simulation Results

In this section, we present our simulation studies on CARA. We first discuss the simulation environment and performance metrics in Section 7.1, and then present illustrative results in Section 7.2.

7.1 Simulation Environment

We conduct simulation on a MAC-level simulator to evaluate CARA. In our simulation, APs are randomly deployed in an area (of size $200\ \mathrm{m} \times 200\ \mathrm{m}$). To create nonuniform user distribution, 70% clients are uniformly distributed in a square with side 120 m centered at (100 m, 100 m). While the other users are uniformly distributed in the whole area.

The transmit power of each AP is set to be 20 dBm. Each AP operates on a channel chosen from a predefined set of orthogonal channels. Channel selection is optimized using the Tabu-search based algorithm presented in [66]. To bring heterogeneity to the network, APs are equipped with heterogeneous numbers of antennas. APs have saturated downlink traffic. Each client has the same weight. Packet are always back-logged at the transmitter. Whenever an AP acquires the channel, it transmits for 10 time slots.

Unless otherwise stated, we use the following as our baseline parameters. We use the log-distance path loss model with reference distance 1 meter, reference loss 46.678 dBm and loss exponent 3. The noise power of the environment is -101 dBm. There are 4 orthogonal spectrum channels in the network, each has a bandwidth of 20 MHz. The number of antennas at each AP is drawn from a normal distribution with mean of 2 and variance of 1. The number of APs and Clients are 20 and 100 respectively. Each client has only one antenna. The slot duration is 10 milliseconds.

We are interested in the following performance metrics: 1) *Throughput:* It is client throughput, calculated as the number of bits received divided by transmission duration. We are interested in both average client throughput and worst-case client throughput. 2) *Proportional fairness utility:* An analytical metric measuring client throughput utility. It is calculated as shown in Equation (8). We compare CARA with the following approaches:

- *Strongest signal based association (MaxSNR)*: In the scheme, clients are always associated with the AP from which they receive the best signal.
- *Proportional fairness association (PFairness) [9]:* In the scheme, user association is optimized by an off-line algorithm to achieve proportional fairness in client throughput. It does not consider co-channel interference.
- *Proportional fairness association based on SINR (PFairness-SINR) [3]:* It represents a large body of work in the literature. Assuming APs are transmitting simultaneously, it factors in inter-AP interference using the SINR model. The channel quality at a client is determined by the SINR received at the client. Then, each client selects the best AP to associate with in an online manner. 802.11 DCF MAC is not considered in the scheme.

In the physical layer, we use zero-forcing beamforming for multi-user beamforming. In our formulation, the link data rate is estimated using long-term channel state. To get more accurate results, in the simulation, we consider instantaneous data rate, which is calculated as follows. First, $SINR_{ij}$ is calculated using

$$SINR_{ij} = \frac{|h_{ij}V_{ij}|^2}{WN_0 + I},$$

where V_{ij} is the zeroforcing beamformer used by AP i for precoding j's signal. I is the interference due to interfering APs not detected by client j. Then, SINR is mapped to a MCS value according to Table 2 [67]. Each MCS value is then mapped to a certain data rate.

For the MAC layer, we simulate the 802.11 DCF MAC. maxSNR, PFairness and PFairness-SINR do not consider optimizing APs' contention windows. Therefore, we set the minimum contention window (CWmin) of APs in these schemes to be 15 according to the 802.11ac standard. Contention windows are adjusted using the Binary Exponential Backoff (BEB) mechanism.

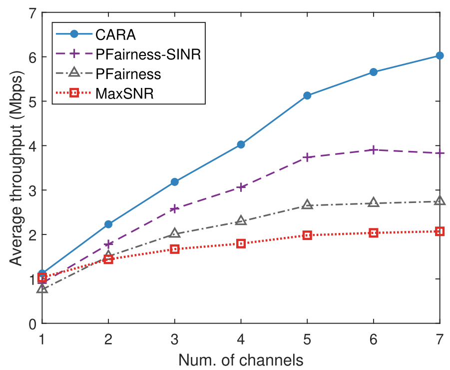


Fig. 8. Throughput versus channel number.

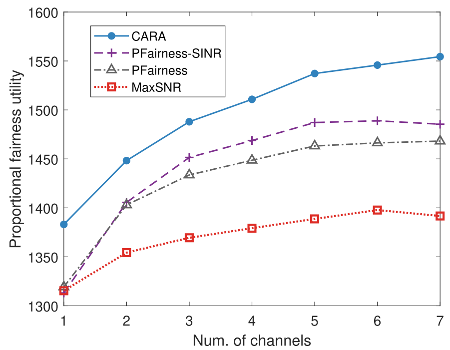


Fig. 9. Proportional fairness versus channel number.

7.2 Illustrative Results

We plot the average throughput of different schemes versus the number of available channels in Fig. 8. The throughput of all schemes increases with the number of channels because channel contention/interference reduces greatly as frequency resources increase. Even though PFairness balances traffic load between APs, it achieves little gain over maxSNR. This is because PFairness fails to balance traffic load among different channels as it does not consider co-channel interference. The throughput of PFairness-SINR outperforms PFairness as it captures inter-AP interference using the SINR model. CARA outperforms PFairness-SINR mainly due to the following reasons: 1) PFairness-SINR fails to consider media access control. It assumes all APs transmit simultaneously and signal power from neighbour APs is modeled as interference. When the network is dense, PFairness-SINR tends to associate a client with the closest AP to improve the received SINR, and hence can not achieve load balancing. However, in the system under study, interfering transmission is scheduled with CSMA/CA. This justifies that client association needs to be aware of the media access control mechanism. 2) APs in PFairness-SINR use the same contention window size in spite of the differences in the number of serving clients. On the contrary, CARA jointly optimizes client association and AP transmission persistence to maximize proportional fairness. APs associated with more clients are able to access the channel with more opportunity.

In Fig. 9, we compare the proportional fairness utility of CARA under different channel resources with other comparison schemes. Client throughput proportional fairness increases with the increase in channel resources. Initially, the fairness metric of PFairness-SINR overlaps with that of PFairness. When the number of channels is beyond a certain point (3 in our study), PFairness-SINR begins to outperform PFairness. This justifies the need for considering interference when making association decisions. CARA outperforms comparison schemes by a wide margin showing that joint optimization achieves better fairness in client throughput allocation.

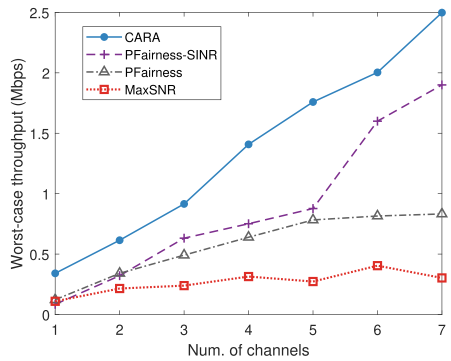


Fig. 10. Worst throughput versus channel number.

We show in Fig. 10 the worst-case throughput versus the number of channels. The throughput of different schemes improves with the increase of available frequency resources. MaxSNR has the lowest worst-case client throughput. This is because APs in the hotspot area are congested with client traffic. Clients associated with congested AP usually have very poor throughput. Although PFairness balances client traffic across APs, it provides very limited performance gain over MaxSNR. This is because without considering interference, it cannot balance client traffic across different channels. Similarly, PFairness-SINR achieves limited improvement over MaxSNR when there are only a few channels (1, 2 or 3 channels in our study). As channel resources increase beyond a certain point, PFairness-SINR offers a significant gain over maxSNR for the worst-case clients. CARA achieves the best performance in both network throughput and minimum client throughput by designing client association and random channel access jointly.

We plot the average client throughput versus the number of available antennas in Fig. 11. Client throughput increases almost linearly with the number of available antennas. This is because APs equipped with multiple antennas are capable of sending multiple concurrent packets to clients with MU-beamforming. The more antennas an AP has, the more clients it can serve simultaneously. The throughput in CARA and PFaireness-SINR increase significantly with the increase in the number of antennas. In contrast, the throughput of maxSNR and PFairness improves slowly with the number of antennas. This is because the client association decisions of maxSNR and PFairness are made without considering the channel assigned to APs and the heterogeneity of antenna numbers at APs.

We study the proportional fairness utility of each scheme with respect to the average number of antennas in Fig. 12. Client throughput fairness of all schemes improves with antenna number. CARA performs the best because it makes better use of spatial degree-of-freedom due to joint optimization.

Fig. 13 shows the worst-case client throughput against the number of available antennas. The throughput of the worst-case client increases with the increase in antenna

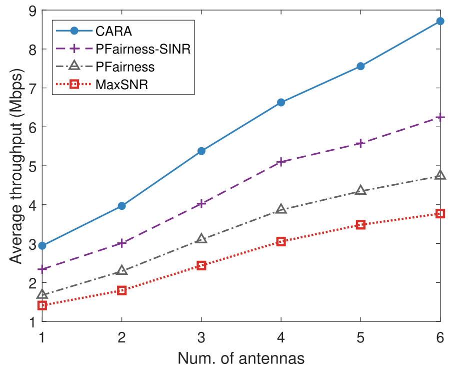


Fig. 11. Throughput versus antenna number.

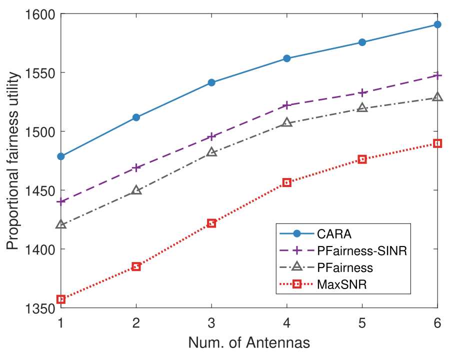


Fig. 12. Proportional fairness versus antenna number.

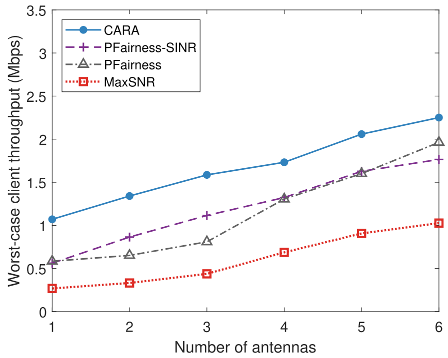


Fig. 13. Worst throughput versus antenna number.

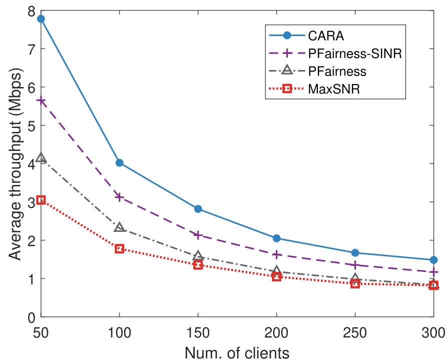


Fig. 14. Throughput versus client number.

number. The slop of CARA is larger than the comparison schemes, showing that throughput in CARA increases faster with antenna number. This is because CARA makes better use of spatial degree-of-freedom for load balancing.

We show the average client throughput versus the number of clients of different schemes in Fig. 14. Average client throughput decreases with the number of serving clients in the networks. Clearly, the total channel resources in the network are fixed. Each client has less chance to access the channel as each AP serves more clients. CARA achieves substantially higher throughput than the comparison schemes.

We show the proportional fairness utility of each scheme with respect to the number of clients in Fig. 15. Proportional fairness utility increases almost linearly with the number of clients. This is due to the definition of proportional fairness, which is $\sum_{j\in\mathbf{U}} \log(T_j)$. Clearly, it gets larger as the size of $\mathbf{U}$ becomes larger. This shows that client throughput proportional fairness is usually a non-decreasing function of the number of clients.

Fig. 16 plots the minimum client throughput versus the number of clients in the network. The minimum client throughput of each scheme drops with the increase in client number. More requesting clients implies more contention for accessing the media. Therefore, each client shares lower throughput. As a result, the throughput of the worst-case client degrades severely.

We plot the average client throughput versus the number of APs in the network in Fig. 17. Initially, client throughput increases with the number of APs in the network because by deploying more APs, we bring closer the average distance between APs and clients. Then, client throughput flats off after a certain point as AP-client distance is not the only factor limiting client throughput. When the AP-client distance is close enough, client throughput is mainly constrained by frequency resources (available channels) in the network. The performance of CARA and PFairness-SINR improves more significantly at the beginning because the received strength of the intended signal increases with the decrease in AP-client distance. Client throughput in maxSNR and PFairness do not scale with the number of deployed APs. This is because more APs are assigned to the same channels as the number of orthogonal channels is limited. Association decisions made in maxSNR and PFairness have not considered the channel selection on APs. Therefore, clients might be assigned to APs using the same channel causing immense channel contention.

Fig. 18 shows the client throughput proportional fairness against the number of APs. While CARA maintains better performance than comparison schemes consistently, its fairness does not scale with the number of APs. Interestingly, the fairness of MaxSNR improves with AP number. This is because MaxSNR always assigns clients to the closest APs. When there are only a few APs, APs in the hotspot area are congested with user traffic. As we deploy more and more APs, user traffic load is potentially offloaded to other nearby APs. The throughput fairness utility of CARA does not improve significantly because client throughput is not only determined by AP-client distance. After proportional fairness reaching a certain point, client throughput is mainly limited by channel resources.

In Fig. 19, we compare the worst-case client throughput of CARA under different numbers of APs with other comparison schemes. The improvement of minimum client throughput due to decreased AP-client distance is

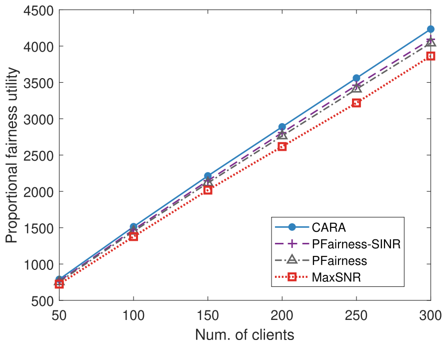


Fig. 15. Proportional fairness versus client number.

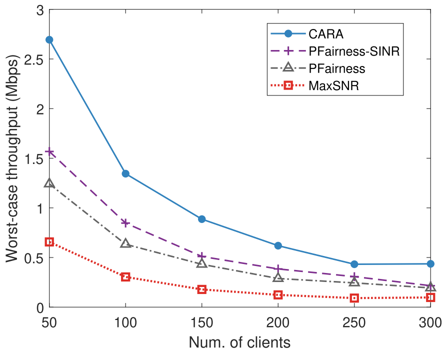


Fig. 16. Worst throughput versus client number.

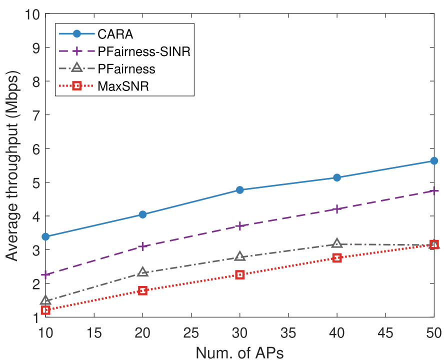


Fig. 17. Throughput versus AP number.

negligible. This is because the minimum client throughput is mainly limited by channel resources and antenna number.

8 Experiment Validation

8.1 Experiment Setup

In this section, we validate the performance of CARA by implementing and testing it in a 802.11n testbed. As shown in Fig. 20, we deploy 4 APs in a 10 meters times 15 meters room. Each AP is implemented by running a Hostapd daemon on a Qualcomm JA76PF0 development board. Each AP is equipped with a Qualcomm Atheros AR9331 network interface card (NIC), which supports IEEE 802.11n. AR9331 NIC does not support MU-MIMO beamforming. APs are labeled by A, B and C and D. All APs operates on 2.4 Ghz channels. We set the channels of AP A and AP B to be 1. The channel of AP C is set to be 6. AP D is set to operate on channel 11. We also place 6 laptops close to the APs to act as clients. The reason why we use laptops instead of mobile

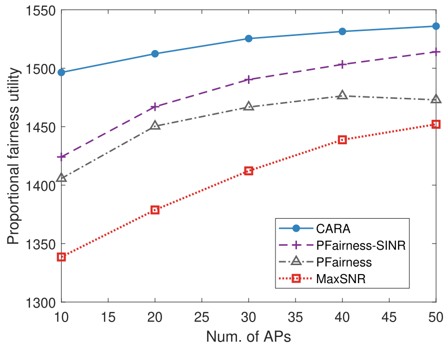


Fig. 18. Proportional fairness versus AP number.

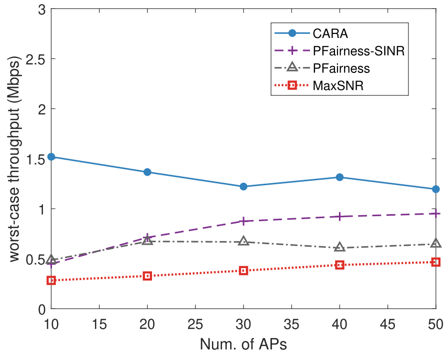


Fig. 19. Worst throughput versus AP number.

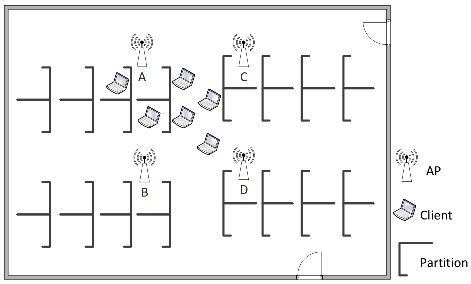


Fig. 20. Locations of APs and clients for the 802.11n testbed.

phones is that active scanning is not allowed by most phone operating systems. Passive scanning too slow for association re-optimization. The model of all the laptops is Lenovo ThinkPad T430s.

To realize the random access control of CARA, we modify hostapd to adapt contention windows according to CARA. We achieve this by setting the cwMin and cwMax of hostapd to be the contention window size optimized by CARA.

To test the performance of CARA, we use network bandwidth measurement tool iperf to evaluate application-layer client throughput. The iperf speed test tool consists of a iperf server and iperf clients. We install iperf server on a PC in the same Ethernet with APs. We also install an iperf client APP to each of the clients. To measure client throughput, the iperf server transmits saturated TCP/UDP traffic to each of the clients.

8.2 Experimental Results

In the first experiment, we have saturated downlink UDP traffic (generated using iperf) from each AP to its clients. The average and worst-case client throughput are shown

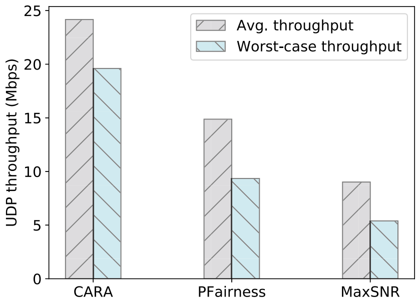


Fig. 21. UDP throughput.

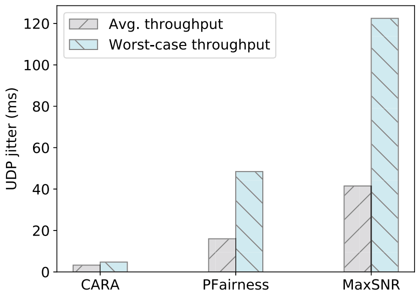


Fig. 22. UDP jitter.

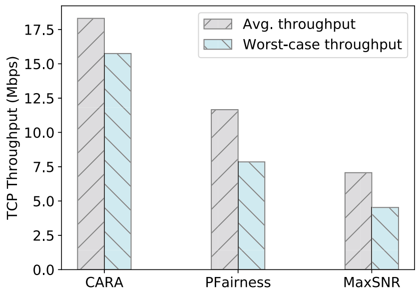


Fig. 23. TCP throughput.

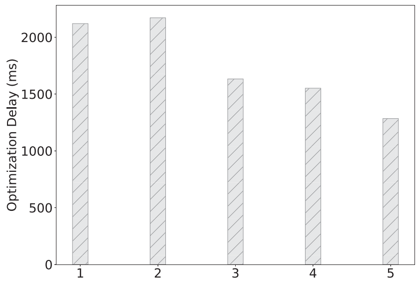


Fig. 24. Optimization.

in Fig. 21. CARA substantially outperforms comparison schemes in terms of aggregate throughput and fairness (almost $1\times$ average throughput increase and $4\times$ worst-case throughput increase over MaxSNR). We find that PFairness tends to offload AP A's client traffic to AP B. However, AP A and B use the same channel. Therefore, PFairness offers limited gain over MaxSNR.

We show the UDP jitter (defined as the variance of UDP packet delay) in Fig. 22. CARA outperforms comparison schemes by a wide margin. PFairness and MaxSNR adopt the BEB mechanism to adjust contention window sizes. When collisions occur, BEB doubles the contention windows of transmit nodes. Evaluation studies [68] show that BEB may lead to unfair throughput allocation in the short term, which in turn results in large delay and jitter. Our result confirms these results and shows that using optimal fixed contention window sizes can potentially reduce jitter.

In the second experiment, we have saturated downlink TCP traffic (generated using iperf) from each AP to its clients. Fig. 23 plots the average and worst-case client throughput of different schemes. The result is qualitatively the same as that shown in Fig. 21.

In the third experiment, we study the re-optimization delay of CARA after client departures. Fig. 24 shows the time needed for convergence against the number of departures. In the experiment, we disconnect different numbers of clients from the network and measure the time that CARA needs to re-optimize the new topology. After clients leaving the network, CARA re-optimizes client association to improve performance. The re-optimization overhead is mainly due to active channel scanning delay and client re-associate delay. In the Figure, re-optimization delay is measured as the time that has elapsed between the first client re-association event and the last client re-association event.

9 Conclusion

Client association control is important in multi-cell WLANs as it can not only balance traffic load among APs but also reduce the number of cell-edge clients (clients in the overlapping coverage of two interfering APs). Moreover, random access control is also an important factor determining network performance. It determines the transmit probabilities of APs and serving probabilities of clients.

Clearly, client association and random access control are coupled, thus need for joint consideration. Prior arts have only focused one of them. We study the novel problem of jointly optimizing client association and random access control to maximize client through utility. Our optimization considers a practical wireless model that well captures interference and AP heterogeneity. Instead of focusing on only downlink traffic as previous works, we optimize for both downlink and uplink throughput.

Since the joint problem is shown to be non-convex and hard to be solved, we first consider the case where client association is given, and propose an optimal random access control policy. Fixing the CWs of APs, we propose a factor-2 approximation algorithm, Greedy-Asso to address the client association problem. Inspired by the optimal access control policy and Greedy-Asso, we further propose a distributed algorithm, termed CARA (Joint Client Association and Random Access Control) to tackle the joint optimization problem. Extensive simulation and experimental studies show that CARA is highly efficient, and it outperforms comparison schemes by a wide margin in terms of throughput and fairness.

Acknowledgments

This work was supported, in part, by the Guangzhou Science and Technology Innovation Commission and NasSoft Technology Ltd. (GZSTI16EG14)

REFERENCES

[1] A. Kumar, E. Altman, D. Miorandi, and M. Goyal, "New insights from a fixed-point analysis of single cell IEEE 802.11 WLANs," *IEEE/ACM Trans. Network.*, vol. 15, no. 3, pp. 588–601, Jun. 2007.

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

[3] W. C. Ao and K. Psounis, "An efficient approximation algorithm for online multi-tier multi-cell user association," in *Proc. 17th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, 2016, pp. 281–290.

[4] H. Zhou, S. Mao, and P. Agrawal, "Approximation algorithms for cell association and scheduling in femtocell networks," *IEEE Trans. Emerging Top. Comput.*, vol. 3, no. 3, pp. 432–443, Sep. 2015.

[5] D. Bethanabhotla, O. Y. Bursalioglu, H. C. Papadopoulos, and G. Caire, "Optimal user-cell association for massive MIMO wireless networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 3, pp. 1835–1850, Mar. 2016.

[6] D. Liu, L. Wang, Y. Chen, T. Zhang, K. K. Chai, and M. Elkashlan, "Distributed energy efficient fair user association in massive MIMO enabled HetNets," *IEEE Commun. Lett.*, vol. 19, no. 10, pp. 1770–1773, Oct. 2015.

[7] M. Hong, A. Garcia, and J. Barrera, "Joint distributed access point selection and power allocation in cognitive radio networks," in *Proc. IEEE INFOCOM*, 2011, pp. 2516–2524.

[8] L. Li, M. Pal, and Y. R. Yang, "Proportional fairness in multi-rate wireless LANs," in *Proc. IEEE INFOCOM*, 2008, pp. 1004–1012.

[9] W. Li, S. Wang, Y. Cui, X. Cheng, R. Xin, M. Al-Rodhaan, and A. Al-Dhelaan, "AP association for proportional fairness in multirate WLANs," *IEEE/ACM Trans. Netw*, vol. 22, no. 1, pp. 191–202, Feb. 2014.

[10] G. Tan and J. V. Guttag, "Time-based fairness improves performance in multi-rate WLANs," in *Proc. USENIX Annu. Tech. Conf. General Track*, 2004, pp. 269–282.

[11] Z. Chen, Q. Xiong, Y. Liu, and C. Huang, "A strategy for differentiated access service selection based on application in WLANs," in *Proc. IEEE Conf. Comput. Commun. Workshops*, 2014, pp. 317–322.

[12] X. Chen, W. Yuan, W. Cheng, W. Liu, and H. Leung, "Access point selection under QoS requirements in variable channel-width WLANs," *IEEE Wireless Commun. Lett.*, vol. 2, no. 1, pp. 114–117, Feb. 2013.

[13] S. Miyata, T. Murase, and K. Yamaoka, "Novel access-point selection for user QoS and system optimization based on user cooperative moving," *IEICE Trans. Commun.*, vol. 95, no. 6, pp. 1953–1964, 2012.

[14] Y. Zhang, D. Bethanabhotla, T. Hao, and K. Psounis, "Near-optimal user-cell association schemes for real-world networks," in *Proc. Inf. Theory Appl. Workshop*, 2015, pp. 204–213.

[15] G. Xue, Q. He, H. Zhu, T. He, and Y. Liu, "Sociality-aware access point selection in enterprise wireless LANs," *IEEE Trans. Parallel Distrib. Syst.*, vol. 24, no. 10, pp. 2069–2078, Oct. 2013.

[16] D. Gong and Y. Yang, "AP association in 802.11 n WLANs with heterogeneous clients," in *Proc. IEEE INFOCOM*, 2012, pp. 1440–1448.

[17] F. Xu, C. C. Tan, Q. Li, G. Yan, and J. Wu, "Designing a practical access point association protocol," in *Proc. IEEE INFOCOM*, 2010, pp. 1–9.

[18] F. Xu, X. Zhu, C. C. Tan, Q. Li, G. Yan, and J. Wu, "Smartassoc: Decentralized access point selection algorithm to improve throughput," *IEEE Trans. Parallel Distrib. Syst.*, vol. 24, no. 12, pp. 2482–2491, pp. 2482–2491, Dec. 2013.

[19] H. Kim, W. Lee, M. Bae, and H. Kim, "Wi-Fi seeker: A link and load aware AP selection algorithm," *IEEE Trans. Mobile Comput.*, vol. 16, no. 8, pp. 2366–2378, Aug. 2016.

[20] T. Sun, Y. Zhang, and W. Trappe, "Improving access point association protocols through channel utilization and adaptive probing," *IEEE Trans. Mobile Comput.*, vol. 15, no. 5, pp. 1157–1167, May 2016.

[21] L.-H. Yen, J.-J. Li, and C.-M. Lin, "Stability and fairness of AP selection games in IEEE 802.11 access networks," *IEEE Trans. Veh. Technol.*, vol. 60, no. 3, pp. 1150–1160, Mar. 2011.

[22] M. Y. Arslan, K. Pelechrinis, I. Broustis, S. Singh, S. V. Krishnamurthy, S. Addepalli, and K. Papagiannaki, "Acorn: An auto-configuration framework for 802.11 n WLANs," *IEEE/ACM Trans. Netw.*, vol. 21, no. 3, pp. 896–909, Jun. 2013.

[23] Y. Bejerano, S.-J. Han, and L. E. Li, "Fairness and load balancing in wireless LANs using association control," in *Proc. ACM MobiCom*, 2004, pp. 315–329.

[24] Y.-C. Hsu, K. C.-J. Lin, and W.-T. Chen, "Client-AP association for multiuser mimo networks," in *Proc. IEEE Int. Conf. Commun.*, 2015, pp. 2154–2159.

[25] S. Corroy, L. Falconetti, and R. Mathar, "Cell association in small heterogeneous networks: Downlink sum rate and min rate maximization," in *Proc. IEEE Wireless Commun. Netw. Conf.*, 2012, pp. 888–892.

[26] T. Van Chien, E. Björnson, and E. G. Larsson, "Joint power allocation and user association optimization for massive mimo systems," *IEEE Trans. Wirel. Commun.*, vol. 15, no. 9, pp. 6384–6399, Sep. 2016.

[27] R. Madan, J. Borran, A. Sampath, N. Bhushan, A. Khandekar, and T. Ji, "Cell association and interference coordination in heterogeneous LTE-A cellular networks," *IEEE J. Select. Areas Commun.*, vol. 28, no. 9, pp. 1479–1489, Dec. 2010.

[28] K. Lin, T.-W. Kuo, P.-J. Yan, W.-J. Cheng, and S.-K. Jeng, "Beam configuration and client association for access points with switched beam antennas," *IEEE Trans. Mobile Comput.*, vol. 9, no. 2, pp. 8–23, Sep. 2016.

[29] G. Athanasiou, P. C. Weeraddana, C. Fischione, and L. Tassiulas, "Optimizing client association for load balancing and fairness in millimeter-wave wireless networks," *IEEE/ACM Trans. Netw.*, vol. 23, no. 3, pp. 836–850, Jun. 2015.

[30] N. Wang, E. Hossain, and V. K. Bhargava, "Joint downlink cell association and bandwidth allocation for wireless backhauling in two-tier HetNets with large-scale antenna arrays," *IEEE Trans. Wirel. Commun.*, vol. 15, no. 5, pp. 3251–3268, May 2016.

[31] J. Yoon, K. Sundaresan, M. Khojastepour, S. Rangarajan, and S. Banerjee, "Joint multicell beamforming and client association in OFDMA small-cell networks," *IEEE Trans. Mobile Comput.*, vol. 9, no. 2, pp. 8–23, Sep. 2016.

[32] M. Hong and Z.-Q. Luo, "Distributed linear precoder optimization and base station selection for an uplink heterogeneous network," *IEEE Trans. Signal Process.*, vol. 61, no. 12, pp. 3214–3228, Jun. 2013.

[33] M. Sanjabi, M. Razaviyayn, and Z.-Q. Luo, "Optimal joint base station assignment and downlink beamforming for heterogeneous networks," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2012, pp. 2821–2824.

[34] Y. Xu and S. Mao, "User association in massive mimo hetnets," *IEEE Syst. J.*, vol. 11, no. 1, pp. 7–19, Mar. 2017.

[35] Q. Ye, O. Y. Bursalioglu, H. C. Papadopoulos, C. Caramanis, and J. G. Andrews, "User association and interference management in massive mimo hetnets," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 2049–2065, May 2016.

[36] D. Liu, L. Wang, Y. Chen, M. Elkashlan, K.-K. Wong, R. Schober, and L. Hanzo, "User association in 5g networks: A survey and an outlook," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 1018–1044, Apr.-Jun. 2016.

[37] D. Niyato, E. Hossain, and D. I. Kim, "Joint admission control and antenna assignment for multiclass qos in spatial multiplexing mimo wireless networks," *IEEE Trans. Wireless Commun.*, vol. 8, no. 9, pp. 4855–4865, Sep. 2009.

[38] J.-W. Lee, M. Chiang, and A. R. Calderbank, "Utility-optimal random-access control," *IEEE Trans. Wireless Commun.*, vol. 6, no. 7, pp. 2741–2751, Jul. 2007.

[39] T. Nandagopal, T.-E. Kim, X. Gao, and V. Bharghavan, "Achieving MAC layer fairness in wireless packet networks," in *Proc. 6th Annu. Int. Conf. Mobile Comput. Netw.*, 2000, pp. 87–98.

[40] Q. Xia and M. Hamdi, "Contention window adjustment for ieee 802.11 WLANs: A control-theoretic approach," in *Proc. IEEE Int. Conf. Commun.*, 2006, vol. 9, pp. 3923–3928.

[41] W. Choi, H. Lim, and A. Sabharwal, "Power-controlled medium access control protocol for full-duplex WiFi networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 7, pp. 3601–3613, Jul. 2015.

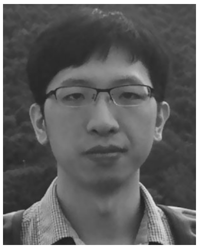
[42] M. Kim, S. Han, and M. Lee, "Demand-aware centralized traffic scheduling in wireless LANs," in *Proc. IFIP Netw. Conf. Workshops*, 2016, pp. 144–152.

[43] Z. Yang, J. Zhang, K. Tan, Q. Zhang, and Y. Zhang, "Enabling TDMA for today's wireless LANs," in *Proc. IEEE Conf. Comput. Commun.*, 2015, pp. 1436–1444.

[44] V. Shrivastava, N. Ahmed, S. Rayanchu, S. Banerjee, S. Keshav, K. Papagiannaki, and A. Mishra, "Centaur: Realizing the full potential of centralized WLANs through a hybrid data path," in *Proc. 15th Annu. Int. Conf. Mobile Comput. Netw.*, 2009, pp. 297–308.

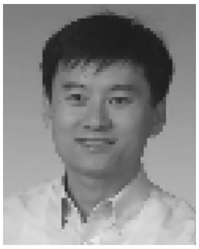
[45] J. Lee and C. Kim, "An efficient multiple access coordination scheme for OFDMA WLAN," *IEEE Commun. Lett.*, vol. 21, no. 3, pp. 596–599, Mar. 2017.

[46] D. Zhao, M. Zhu, M. Xu, and J. Cao, "Downlink packets scheduling in enterprise WLAN," in *Proc. IEEE Wireless Commun. Netw. Conf.*, 2013, pp. 1333–1338.
[47] W. Wong, A. Thakur, and S.-H. G. Chan, "An approximation algorithm for AP association under user migration cost constraint," in *Proc. 35th Annu. IEEE Int. Conf. Comput. Commun.*, 2016, pp. 1–9.
[48] D. Bethanabhotla, O. Y. Bursalioglu, H. C. Papadopoulos, and G. Caire, "User association and load balancing for cellular massive mimo," in *Proc. Inf. Theory Appl. Workshop*, 2014, pp. 1–10.
[49] E. Castaneda, A. Silva, A. Gameiro, and M. Kountouris, "An overview on resource allocation techniques for multi-user mimo systems," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 1, pp. 239–284, Jan.-Mar. 2017.
[50] G. Lee and Y. Sung, "A new approach to user scheduling in massive multi-user mimo broadcast channels," *IEEE Trans. Commun.*, vol. 66, no. 4, pp. 1481–1495, Apr. 2018.
[51] L. Deng, W. Zhang, Y. Rui, and Y. C. Kiat, "Utility maximization for uplink mu-mimo: Combining spectral-energy efficiency and fairness," in *Proc. Eur. Conf. Netw. Commun.*, 2016, pp. 79–83.
[52] R. Brandt, R. Mochaourab, and M. Bengtsson, "Distributed long-term base station clustering in cellular networks using coalition formation," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 2, no. 3, pp. 362–375, Sep. 2016.
[53] C. M. Yetis, T. Gou, S. A. Jafar, and A. H. Kayran, "On feasibility of interference alignment in MIMO interference networks," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4771–4782, Sep. 2010.
[54] M. Razaviyayn, M. Sanjabi, and Z.-Q. Luo, "Linear transceiver design for interference alignment: Complexity and computation," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 2896–2910, May 2012.
[55] T. Liu and C. Yang, "On the feasibility of linear interference alignment for MIMO interference broadcast channels with constant coefficients," *arXiv:1207.1517*, 2012.
[56] O. El Ayach, A. Lozano, and R. W. Heath, "On the overhead of interference alignment: Training, feedback, and cooperation," *IEEE Trans. Wireless Commun.*, vol. 11, no. 11, pp. 4192–4203, Nov. 2012.
[57] L. B. Jiang and S. C. Liew, "Improving throughput and fairness by reducing exposed and hidden nodes in 802.11 networks," *IEEE Trans. Mobile Comput.*, vol. 7, no. 1, pp. 34–49, Jan. 2008.
[58] L. Jiang and J. Walrand, "A distributed CSMA algorithm for throughput and utility maximization in wireless networks," *IEEE/ACM Trans. Netw.*, vol. 18, no. 3, pp. 960–972, Jun. 2010.
[59] J. Ni, B. Tan, and R. Srikant, "Q-csma: Queue-length-based CSMA/CA algorithms for achieving maximum throughput and low delay in wireless networks," *IEEE/ACM Trans. Netw.*, vol. 20, no. 3, pp. 825–836, Jun. 2012.
[60] R. Liao, B. Bellalta, M. Oliver, and Z. Niu, "Mu-mimo MAC protocols for wireless local area networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 1, pp. 162–183, Jan.-Mar. 2016.
[61] X. Wang and K. Kar, "Throughput modelling and fairness issues in CSMA/CA based ad-hoc networks," in *Proc. IEEE 24th Annu. Joint Conf. IEEE Comput. Commun. Societies*, 2005, pp. 23–34.
[62] M. Garetto, T. Salonidis, and E. W. Knightly, "Modeling per-flow throughput and capturing starvation in CSMA multi-hop wireless networks," *IEEE/ACM Trans. Netw.*, vol. 16, no. 4, pp. 864–877, Aug. 2008.
[63] B. Bellalta, A. Zocca, C. Cano, A. Checco, J. Barcelo, and A. Vinel, "Throughput analysis in CSMA/CA networks using continuous time markov networks: A tutorial," in *Wireless Networking for Moving Objects*. New York, NY, USA: Springer, 2014, pp. 115–133.
[64] X. Yue, C.-F. Wong, and S.-H. G. Chan, "CACAO: Distributed client-assisted channel assignment optimization for uncoordinated WLANs," *IEEE Trans. Parallel Distrib. Syst.*, vol. 22, no. 9, pp. 1433–1440, Sep. 2011.
[65] M. Abusubaih and A. Wolisz, "Interference-aware decentralized access point selection policy for multi-rate IEEE 802.11 wireless LANs," in *Proc. 19th Int. Symp. Personal Indoor Mobile Radio Commun.*, 2008, pp. 1–6.
[66] A. P. Subramanian, H. Gupta, S. R. Das, and J. Cao, "Minimum interference channel assignment in multiradio wireless mesh networks," *IEEE Trans. Mobile Comput.*, vol. 7, no. 12, pp. 1459–1473, Dec. 2008.
[67] "White paper: IEEE 802.11ac migration guide," NETSCOUT, Tech. Rep., 2017.
[68] K. H. Almotairi, "Inverse binary exponential backoff: Enhancing short-term fairness for IEEE 802.11 networks," in *Proc. 10th Int. Symp. Wireless Commun. Syst.*, 2013, pp. 1–5.



Wangkit Wong is currently working toward the the PhD degree in the Department Computer Science and Engineering, HKUST, under the supervision of Prof. S.-H. Gary Chan. His research interest includes wireless networking, convex optimization, and combinatorial optimization.

Kam-Wa Chau received the BEng degree in computer science and the MSc degree in computer science and engineering from The Hong Kong University of Science and Technology, Hong Kong, in 2015 and 2017, respectively. His research interest includes wireless networking and indoor localization.



S.-H. Gary Chan (S'89-M'98-SM'03) received the BSE degree (highest honor) in electrical engineering from Princeton University, Princeton, New Jersey, in 1993, with certificates in applied and computational mathematics, engineering physics, and engineering and management systems, the MSE and PhD degrees in electrical engineering from Stanford University, Stanford, California, in 1994 and 1999, respectively, with a minor in business administration. He is currently a professor with the Department of Computer Science and Engineering, The Hong Kong University of Science and Technology (HKUST), Hong Kong. He is also the director of the Entrepreneurship Center, and chair of the Committee on Entrepreneurship Education Program, Center for Education Innovation, HKUST. His research interest includes multimedia networking, mobile computing, data analytics, and IT entrepreneurship. He has been an associate editor of the *IEEE Transactions on Multimedia* (2006-2011), and a vice-chair of the Peer-to-Peer Networking and Communications Technical Sub-Committee of IEEE Comsoc Emerging Technologies Committee (2006-2013). He is and has been guest editor of the *Elsevier Computer Networks* (2017), the *ACM Transactions on Multimedia Computing*, the *Communications and Applications* (2016), the *IEEE Transactions on Multimedia* (2011), the *IEEE Signal Processing Magazine* (2011), the *IEEE Communication Magazine* (2007), and the *Springer Multimedia Tools and Applications* (2007). He was the TPC chair of the IEEE Consumer Communications and Networking Conference (IEEE CCNC) 2010, Multimedia Symposium of IEEE Globecom (2006 and 2007), IEEE ICC (2005 and 2007), and Workshop on Advances in Peer-to-Peer Multimedia Streaming in ACM Multimedia Conference (2005). He has co-founded several startups deploying his research results. Due to their innovations and commercial impacts, his projects have received local and international ICT awards (2012-2015). He is the recipient of Google Mobile 2014 Award (2010 and 2011) and Silver Award of Boeing Research and Technology (2009). He has been a visiting professor and researcher with Microsoft Research (2000-2011), Princeton University (2009), Stanford University (2008-2009), and the University of California at Davis (1998-1999). He was Undergraduate Programs coordinator with the Department of Computer Science and Engineering (2013-2015), the director of the Sino Software Research Institute (2012-2015), codirector of the Risk Management and Business Intelligence program (2011- 2013), and the director of the Computer Engineering Program (2006-2008) at HKUST. He was a William and Leila Fellow at Stanford University (1993-1994), and the recipient of the Charles Ira Young Memorial Tablet and Medal, and the POEM Newport Award of Excellence at Princeton (1993). He is a senior member of the IEEE.