

Client-AP Association for Multiuser MIMO Networks

Yu-Cheng Hsu, and Kate Ching-Ju Lin, Wen-Tsuen Chen, *Fellow, IEEE*,

Abstract

In a Wireless Local Area Network (WLAN), clients typically associate with the AP that offers the maximal signal strength. Later works on client-AP association then further take load balancing and fairness into consideration. Those schemes however are not directly applicable in a multiuser MIMO (MU-MIMO) WLAN since different combinations of clients result in different throughput of each individual client. We show that the AP association problem for maximizing sum-rate in multi-cell MU-MIMO downlink networks is NP-hard. Therefore, in this paper, we present two client-AP association algorithm customized for MU-MIMO WLANs. The proposed algorithm jointly solves the problems of client-AP association and MU-MIMO client grouping with consideration of channel correlation among clients. It hence allows a good group of clients, i.e., those with low channel correlation, to together associate with the same proper AP and achieve a high sum-rate. We further extend the second algorithm to a distributed scheme. The simulation results show that our MU-MIMO AP association algorithm improves the aggregate throughput by about 11%-28% and 26%-45%, as compared to two common association approaches, i.e., RSSI-based and load-based schemes, respectively.

Index Terms

MU-MIMO, AP association, Downlink, Sum Capacity, Multiuser, MIMO,

Yu-Cheng Hsu is with the Department of Computer Science, National Tsing Hua University, Hsinchu, Taiwan (email:king11604@hotmail.com).

Kate Ching-Ju Lin is with the Research Center for IT Innovation, Academia Sinica, Taipei, Taiwan (email:katelin@citi.sinica.edu.tw).

Wen-Tsuen Chen is with the Institute of Information Science, Academia Sinica, Taipei, Taiwan (email:wtchen@iis.sinica.edu.tw).

I. INTRODUCTION

The massive number of mobile devices and the emergence of numerous applications, such as video streaming and social networks, together result in large demand for wireless network bandwidth. In order to satisfy this urgent demand, IEEE specifies 802.11n/ac in recent years to exploit MIMO technologies to provide a higher transmission rate for wireless networks. In a MIMO system, each device, i.e., an access point (AP) or client, can be equipped with multiple antennas, which can be used to deliver the spatial multiplexing gain or the diversity gain, or both [1]. MIMO technologies can hence substantially boost the network throughput by not only enabling simultaneous transmissions but also reducing the effect of signal fading.

Recently, 802.11ac is further proposed to enable multi-user MIMO (MU-MIMO) transmissions. In a MU-MIMO system, the AP equipped with multiple antennas can simultaneously transmit (beamform) packets to different clients (with either single antenna or multiple antennas). Concurrent transmissions can be realized by using a signal precessing technique, called zero-forcing beamforming (ZF-BF) [2]. By explicitly precoding the multiple packets, ZF-BF eliminates the interference among simultaneous transmissions so that each client only receives its targeted signal. However, such precoding might affect the receiving SNR at each client, and thereby the achievable rate. As shown in previous work [3] [4], the sum-rate of a MU-MIMO network closely depends on channel correlation among concurrent beamformed clients. Therefore, the achievable throughput of a client in a MU-MIMO network not only depends on its receiving signal strength but also the channels of clients that associate with the same AP and could be formed as the same beamforming group.

The above property naturally introduces a new challenge: *How should a client associate with a proper AP with consideration of channel correlation among clients?* In a conventional wireless network, clients usually connect to the AP with the strongest signal strength. Some previous works then proposed to associate clients with the APs such that the load can be distributed evenly among the APs [5] or the overall network throughput can be maximized [6] [7]. However, those approaches do not consider the MU-MIMO scenario where clients can be served simultaneously. Since channel correlation among a group of clients might be different if those clients associate with different APs, AP association in a MU-MIMO WLAN should not only consider the signal strength and traffic load, but also the channels, i.e., channel state information (CSI), from different

APs. Therefore, this work aims at investigating the client-AP association problem in multi-cell MU-MIMO networks.

We formulate the joint client-AP association and client grouping problem in a sum-rate maximization model. Due to the NP-hardness of the joint problem, we propose a greedy algorithm to associate each client with a proper AP that allows it to have higher opportunities to find concurrent clients with uncorrelated channels, as a result producing a higher aggregate throughput. Then we propose a second algorithm that dynamically makes some clients associate with different APs for adapting various network conditions (e.g., the change of channel quality, client joining and leaving due to mobility) and also further extend the second algorithm to a distributed scheme. We consider a multi-cell downlink system that all coordinated APs are connected to a centralized controller so that it can be efficient to dynamically change the network topology as illustrated in Fig.1. This scenario is actually appropriate for an enterprise network ,i.e., an airport, university, and conference center. Our simulation results show that the proposed algorithm improves the overall throughput by 11%-28% and 26%-45%, as compared to RSSI-based and load-based schemes, respectively, and have 22%-38% improvement gain on BPF scheme [8].

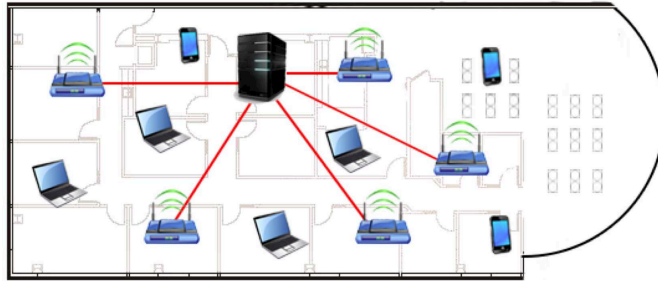


Fig. 1: Enterprise network for multi-cell MU-MIMO LANs. The multiple coordinated APs connected to a centralized controller dynamically change network topology for client association.

In summary, our main contributions are as follows:

- We first propose a greedy algorithm based on centralized scheme such that the APs have the global knowledge of the network and make the proper client association. Therefore, the AP can select the optimal subset of clients to transmit concurrently data to them, and the sum rate of clients can be maximized. The load of APs achieve balance, the total throughput of

clients also can be maximized.

- We propose an adaptation algorithm to execute client reassociation for the mobility of clients and channel variety to keep the maximum total throughput for our MU-MIMO system, i.e., a client arrivals/departs in/from the network.
- We show that the adaptation algorithm can be extend to a distributed scheme. Thus, this algorithm can be exactly adopted in the network without the coordination with a centralized controller.

The rest of our paper is organized as follows. In Section II, we discuss the related works on MU-MIMO client grouping and client-AP association. In Section III, we provide an overview of the ZFBF technique and formulate our system model, and show that our problem is NP-hard. In Section IV, we present the proposed greedy algorithms and extend adaptation algorithm to a distributed scheme. Section V gives the evaluation results of our algorithm. Finally, Section VI concludes this paper.

II. RELATED WORKS

Much attention has been devoted to the effect of adopting MIMO technologies in wireless networks [1], [9]. The advantages of MU-MIMO has been verified theoretically.(e.g., the MU-MIMO systems can achieve higher channel capacity than the conventional networks where the AP only serves as a single client at each transmission) [10]–[12], and implemented practically [3], [4], [13]–[20]. A major problem focused in this network system is how to select a optimal subset of clients to communicate with an AP in order to maximize the sum rate.

For Gaussian broadcast MU-MIMO channel, Dirty Paper Coding (DPC) has been proven to be able to achieve the sum capacity [1], [21]. However, the implementation of DPC in MIMO systems is not practical due to its high computational complexity in encoding and decoding schemes, especially when the number of clients is large.

Another linear precoding technique with low complexity is Block Diagonalization (BD) [14], [17], [18], which uses linear precoding matrix to achieve inter-free interference between every receiver. In [18], under the same work of BD, the authors propose two suboptimal algorithm: c-algorithm and n-algorithm. The c-algorithm greedily selects the clients which maximizes the total throughput, whereas the n-algorithm does greedy selection based on channel Frobenius

norm. The overall performance of BD can achieve about 95% of the total throughput of the complete search.

ZFBBF is a practical precoding scheme similar to BD that can produce the beamforming weight to completely eliminate the interference among the clients concurrently communicating with an AP. With ZFBBF, the beamforming weight can be chosen with ease by inverting the channel gain matrix of clients, and the maximal number of clients that the AP simultaneously serve is exactly the number of antenna of transmitter [2]–[4], [13], [15] [22] [23].

However, selecting the best subset of clients requires a brute-force exhaustive search on all possible client sets in downlink MU-MIMO LANs, which is not practical also due to high complexity. Hence, many recent works has focused on suboptimal greedy client selection scheme, and some of them were based on ZFBBF [13], [15], [22], [23]. In [15], swap (GUSS) algorithm adopts 'add', 'delete' and 'swap' operations in the client selection algorithm to avoid trapping into local optimum and shows that an efficient effective-channel-gain updating strategy can achieve most of the sum rate achieved by exhaustive search. In [16], MIMOMate is proposed in a leader-contention-based uplink MU-MIMO MAC protocol that can reduce the contention overhead for concurrent streams because it only selects the leader of the matched client to content for transmission opportunities. In [3], the authors propose a rate adaptation protocol for uplink MU-MIMO LANs in that each client can adapt their bit rate for concurrent transmission to improve the throughput.

Nevertheless, client selection schemes for downlink MU-MIMO LANs [13], [15], [22], [23] requires the correct channel state information (CSI) from the feedbacks of clients to select the optimal subset of clients for transmission with the aim of maximizing the sum rate. In [4], a low-complexity client selection scheme based on ZFBBF is proposed in a limited feedback system in which each client feeds back the channel direction information (CDI) and the channel quality information (CQI) based on SINR. They also analyze the sum rate performance with the tradeoff between the number of clients and the number of feedback bits, and show that having more clients can reduce the feedback load.

The above-mentioned studies have investigated theoretical capacity analyse; little research on the implementation of MU-MIMO beaforming [2], [24] has been done. In [2], the authors show that smaller separation distance between two receivers can achieve the maximum sum rate and outdated channel information will affect the performance because of mobility and

environment variation. In [24], the authors propose a novel scheme to achieve the time and phase synchronization between coordinated APs in distributed MU-MIMO, and implement it in the FPGA of the WARP radio platform.

Association control for clients has been an important subject in recent research, which mostly focuses on optimizing different network parameters (i.e., fairness [6], throughput, interference [7], energy saving [25], handoff frequency, power adaptation [26], load balancing [6], etc) for maximizing client throughput.

In [6], the authors introduced a centralized algorithm for client association which takes into consideration the factor of fairness and load balancing. Their goal was to guarantee max-min fair bandwidth allocation, and the simulation results showed that their algorithm significantly outperforms popular heuristics. Other client association problems are discussed in a wireless mesh network(WMN) [27], [28]. A WMN can easily exchange information on backhauls between APs and has the advantage of dynamically self-configuring and self-organizing. In [27], a novel client association is proposed to achieve fair bandwidth allocation and considers the load balancing, signal strength, and WMN's multi-hop characteristic. Due to the overhead of gathering information in a centralized approach, the authors also designed a distributed approaches for convenient deployment.

As mentioned above, the centralized approach is known to have problem of overhead in collecting information and inconvenient deployment [7], so research has been done on the distributed approaches for client association [25], [26], [29]–[32]. These works usually require information provided by AP which leads to the modifications of APs themselves. However, in [29], [32], only client devices needs to be modified for making client association. In [29], the authors proposed a practical online client association with reasonable computational overhead, and that can be implemented on commercial hardware without any modification. In addition to receiving information from AP for making client association, the clients can mutually exchange information to determine which AP is the optimal one that can provide best throughput performance [25], [30]. For saving energy power on device, the client devices use Bluetooth to transmit information when meeting other clients in the same coverage area of AP [25]. Taking the impact of heterogeneous 802.11a/b/g clients in 802.11n WLANs into consideration, [33] models the AP association problem as bi-dimensional Markov chain.

In [34], the authors propose a SAC protocol for hybrid nature of wireless network to improve

client association, where multiple APs are connected to a central control. They consider the factors of bandwidth and fairness, and formulate a max-flow problem to decide optimal client association. In [35], the authors provide guidelines for optimizing the performance of wireless network and explore the interrelation between channel allocation, user association and power control influences the performance of network. However, handoff cost is a non-negligible overhead for dynamic client association that the throughput performance of client maybe is affected. In [36], the authors use virtualized wireless network interface(WNIC) on each station that can be connected to multiple APs concurrently, so the handoff overhead can be negligible. The authors also adopted an association topology to achieve load balancing of AP in different traffic conditions.

Our work focused on making proper client association for downlink multi-cell MU-MIMO LANs, and adopts ZFBF scheme with low complexity on client selection to achieve maximum throughput.

III. SYSTEM MODEL

In this section, we first give the background of MU-MIMO networks and then describe our system model.

A. Zero-forcing Beamforming (ZFBF) for MU-MIMO

We consider a downlink MU-MIMO network that consists of an AP equipped with N antennas and a set of clients \mathcal{U} . Say the AP serves a group of clients $\mathcal{G} \subseteq \mathcal{U}$ simultaneously, and transmits a symbol s_u to client u , for all $u \in \mathcal{G}$. To eliminate interference, the AP pre-codes (multiplies) each transmitted symbol s_u by a beamforming weight unit vector $\mathbf{w}_u \in \mathbb{C}^{N \times 1}$. The signal received by a client $u \in \mathcal{G}$ can then be expressed by

$$y_u = \sqrt{P_u} \mathbf{h}_u \mathbf{w}_u s_u + \sum_{u' \in \mathcal{G}, u' \neq u} \sqrt{P_{u'}} \mathbf{h}_u \mathbf{w}_{u'} s_{u'} + n_u, \quad (1)$$

where P_u is the transmit power allocated to client u , $\mathbf{h}_u \in \mathbb{C}^{1 \times N}$ is the channels between the AP and client u , which is independently and identically distributed (i.i.d.) complex Gaussian random variables with zero mean and unit variance, and n_u is the Additive Gaussian White Noise (AWGN) at client u with variance σ_u^2 . Note that the sum transmit power should be restricted by

the maximum power, i.e., $\sum_u P_u \leq P_{max}$. In Eq. (1), the first term is the target signal for client u , and the second term is the interference from the streams of other clients.

Prior work [37] has shown that finding an optimal beamforming weight vector \mathbf{w}_u for all $u \in \mathcal{G}$ that maximizes the system capacity is a difficult non-convex optimization. Practical systems [2] hence instead exploit a simpler technique, called zero-forcing beamforming (ZFBF), to find the beamforming vectors \mathbf{w}_u . In particular, to cancel (null) out the interfering signals at client u , the AP can pick the beamforming vectors $\mathbf{w}_{u'}$ such that $\mathbf{h}_u \mathbf{w}_{u'} = 0$, for all $u' \in \mathcal{G}$ and $u' \neq u$. Such ZFBF ensures that each client u does not see the second interference term in Eq. (1), and hence can exploit the standard decoder to recover the desired signal in the first term.

Let $\mathbf{H}(\mathcal{G}) = [\mathbf{h}_1 \mathbf{h}_2 \cdots \mathbf{h}_{|\mathcal{G}|}]^T$ be the matrix form of the channel vectors of all clients in the group \mathcal{G} , and, similarly, let $\mathbf{W}(\mathcal{G}) = [\mathbf{w}_1 \mathbf{w}_2 \cdots \mathbf{w}_{|\mathcal{G}|}]$ be the matrix form of the beamforming vectors of all clients in \mathcal{G} . According to [38], the ZFBF beamforming matrix can be obtained by the pseudo inverse of $\mathbf{H}(\mathcal{G})$ as follows.

$$\mathbf{W}(\mathcal{G}) = \mathbf{H}(\mathcal{G})^\dagger = \mathbf{H}(\mathcal{G})^* (\mathbf{H}(\mathcal{G}) \mathbf{H}(\mathcal{G})^*)^{-1} \quad (2)$$

The sum-rate can be computed through the equation represented as

$$R(\mathcal{G}) = \max_{P_u: \sum_{u \in \mathcal{G}} \gamma_u^{-1} P_u \leq P} \sum_{u \in \mathcal{G}} \log(1 + P_u), \quad (3)$$

where

$$\gamma_u = \frac{1}{\|\mathbf{w}_u\|^2} = \frac{1}{\left[(\mathbf{H}(\mathcal{G}) \mathbf{H}(\mathcal{G})^*)^{-1} \right]_{u,u}} \quad (4)$$

can be regarded as the effective channel gain of client u [39]. The optimal power allocation P_u can be determined by the well-known water-filling algorithm. Since different groups of clients result in different system capacities, the AP should explicitly select a proper group of clients for concurrent transmissions. Note that an N -antenna AP can serve at most N clients at the same time. In addition, serving more concurrent clients does not mean improving the sum-rate because channel correlation between clients might increase when more clients are served simultaneously. Therefore, the problem of optimal client grouping for sum-rate maximization can be formulated as

$$\mathcal{G}^* = \arg \max_{\mathcal{G} \subseteq \mathcal{U}, |\mathcal{G}| \leq N} R(\mathcal{G}), \quad (5)$$

which can be solved by exhausted search or approximated by greedy algorithms [13], [22], [40]. For example, the basic idea of exhausted search is to search the best beamforming group \mathcal{G}^* that produces the maximal sum-rate from \mathcal{U} , then exclude those clients (i.e., $\mathcal{U} = \mathcal{U} \setminus \mathcal{G}^*$), and repetitively select the best group from the remaining clients in \mathcal{U} until every client belongs to one beamforming group. The average achievable sum-rate of beamforming groups associated with the same AP can then be estimated by

$$R(\mathcal{U}) \approx \sum_{\mathcal{G} \in \mathcal{S}} R(\mathcal{G})/|\mathcal{S}|, \quad (6)$$

where \mathcal{S} is the collection of all the selected groups \mathcal{G}^* .

B. Client-AP Association Problem in MU-MIMO

We consider a multi-cell scenario with a set of APs \mathcal{A} and a set of clients \mathcal{U} . Each AP $a \in \mathcal{A}$ is equipped with N_a antennas, and can serve at most N_a clients simultaneously. Assume that neighboring APs can leverage spacial reuse, and operate over non-overlapping channels to avoid inter-cell interference. We leave jointly considering the client-AP association and channel assignment problems as our future work. Let \mathcal{C}_a denote the subset of clients that are covered by AP $a \in \mathcal{A}$. Then, each client $u \in \mathcal{U}$ can associate with any of its neighboring APs a if $u \in \mathcal{C}_a$. The clients that finally associate with AP a are collected as a set \mathcal{U}_a , and hence $\mathcal{U}_a \subseteq \mathcal{C}_a$.

The achievable sum-rate of an association set \mathcal{U}_a depends on how clients are classified into beamforming groups, as mentioned in Eqs. (5) and (6). To simplify explanation, we use exhausted search to group clients in this work. However, our association algorithm can be combined with any client grouping algorithms. Without loss of generality, we assume that each AP knows the channel state information (CSI) of the clients within its coverage range to perform client grouping. In practice, the CSI of a client can either be estimated by its neighboring APs using reciprocity, or be estimated by clients and reported to the APs. Therefore, our objective is to associate each client with an AP, i.e., determining the proper association set \mathcal{U}_a for all $a \in \mathcal{A}$, with consideration of beamforming grouping in a MU-MIMO downlink transmission scenario such that the sum-rate of all clients $\sum_{a \in \mathcal{A}} R(\mathcal{U}_a)$ can be maximized. Table I summarizes the notations used in this paper, and the above MU-MIMO client-AP association problem can be formulated as follows.

$$\max \sum_{u \in \mathcal{U}} \sum_{a \in \mathcal{A}} r_{u,a} \quad (7)$$

subject to

$$\sum_{a \in \mathcal{A}} I_{u,a} \leq 1, \forall u \in \mathcal{U} \quad (8)$$

$$J_{\mathcal{G}_a} \leq I_{u,a}, \forall u \in \mathcal{G}_a \in G_a, a \in \mathcal{A} \quad (9)$$

$$\sum_{u \in \mathcal{G}_a \in G_a} J_{\mathcal{G}_a} \leq 1, \forall a \in \mathcal{A} \quad (10)$$

$$r_{u,a} = \frac{\sum_{u \in \mathcal{G}_a \in G_a} r_{u,\mathcal{G}_a} J_{\mathcal{G}_a}}{\sum_{\mathcal{G}_a \in G_a} J_{\mathcal{G}_a}}, \forall u \in \mathcal{U}, a \in \mathcal{A} \quad (11)$$

$$I_{u,a} \in \{0, 1\}, \forall u \in \mathcal{U}, a \in \mathcal{A} \quad (12)$$

$$J_{\mathcal{G}_a} \in \{0, 1\}, \forall \mathcal{G}_a \in G_a, a \in \mathcal{A} \quad (13)$$

We introduce a binary variable $I_{u,a}$ indicating whether client u associates with AP a for all $u \in \mathcal{U}$ and $a \in \mathcal{A}$. Specifically, $I_{u,a}$ equals 1 if client u is served by AP a , and equals 0, otherwise. Hence, Eq. (8) ensures that each client u associates with at most one AP. We note that, in a MU-MIMO network, the achievable rate of a client depends on channel orthogonality among concurrent clients. That is, a client might achieve a positive throughput when it communicates alone, but get zero throughput when it joins an improper beamforming group, i.e., clients with high channel correlation. Thus, for fair bandwidth sharing, we deem a subset of clients \mathcal{G} as a *candidate beamforming group* only if all the members $u \in \mathcal{G}$ can obtain a non-zero throughput when they communicate concurrently, i.e., $r_{u,\mathcal{G}} > 0$. For ease of representation, we collect all *candidate beamforming groups* of AP a as a set G_a , which is defined as the set of candidate beamforming groups covered by AP a , i.e.,

$$G_a \triangleq \{\mathcal{G} : \mathcal{G} \subseteq \mathcal{C}_a, |\mathcal{G}| \leq N_a, r_{u,\mathcal{G}} > 0, \forall u \in \mathcal{G}\}. \quad (14)$$

To model client grouping, we also introduce another binary variable $J_{\mathcal{G}_a}$ to represent whether a candidate beamforming group \mathcal{G}_a is selected. Eq. (9) restricts that a group \mathcal{G}_a can only be selected, i.e., $J_{\mathcal{G}_a} = 1$, if all its members $u \in \mathcal{G}_a$ are associated with AP a , i.e., $I_{u,a} = 1, \forall u \in \mathcal{G}_a$. In addition, each client u can at most join one group, which is constrained by Eq. (10). Eq. (11) indicates that the expected achievable rate $r_{u,a}$ of client u from AP a can be estimated by the rate it obtains from joining group \mathcal{G}_a divided by the number of groups served by AP a , because

Parameters	Definition
\mathcal{A}	the set of APs
N_a	the number of antennas equipped at AP $a \in \mathcal{A}$
\mathcal{U}	the set of clients
\mathcal{C}_a	the subset of clients that are covered by AP $a \in \mathcal{A}$
\mathcal{G}_a	a beamforming group of clients covered by AP $a \in \mathcal{A}$, in which all the member can communicate concurrently with AP a and achieve a non-zero rate
G_a	the set of all candidate beamforming groups covered by AP a
$r_{u,\mathcal{G}}$	the achievable rate of client u if it communicates with AP a when joining group \mathcal{G}
Variables	Definition
$I_{u,a}$	binary variable indicating whether client u associates with AP a
$J_{\mathcal{G}_a}$	binary variable indicating whether the beamforming group $\mathcal{G}_a \in G_a$ is selected by AP a
$r_{u,a}$	the expected achievable throughput of client u when it associates with AP a
\mathcal{U}_a	the set of clients that associate with AP a

TABLE I: Definition of notations

the capacity of AP a should be shared by all the beamforming groups. Finally, the objective is to maximize the sum-rate of all clients.

C. Hardness result

We show that our client-AP association problem in multi-cell MU-MIMO network can not be solved in polynomial time unless $P = NP$, i.e. the problem is NP-hard. Our reduction is via 3-dimensional matching problem, which is NP-complete.

Definition1 : Given an instance of the following problem: Disjoint sets $B = \{b_1, \dots, b_n\}$, $C = \{c_1, \dots, c_n\}$, $D = \{d_1, \dots, d_n\}$, and a family $F = T_1, \dots, T_m$ of triples with $|T_i \cap B| = |T_i \cap C| = |T_i \cap D| = 1$ for $i=1, \dots, m$. Question: Does F contain a matching, i.e., a subfamily F' for which $|F'| = n$ and $\bigcup_{T_i \in F'} T_i = B \cup C \cup D$?

Theorem1: client-AP association problem for maximizing sum-rate in MU-MIMO network is NP-hard, where multiple APs and multiple clients are equipped with multiple antennas and

one antenna, respectively.

Proof : We slightly adapting the reduction in [41] [42], and we assume that $m > n$ in our reduction. We call the triples that contain b_j be the number of triples of type j for $j=1,...,n$. AP i corresponds to the triple T_i for $i=1,...,m$. There are $2n$ element clients, corresponding to the $2n$ elements of $C \cup D$. There are t_j-1 dummy client of type j for $j=1,...,n$, and the total number of clients is $m - n$. AP i corresponding to a triple of type j , say $T_i=(b_j, c_k, d_l)$, the corresponding element clients c_k and d_l have an original rate of R , respectively. The ZFBF gain when scheduling two element clients together is S , and each obtains a rate of SR . Each of the dummy clients of type j has a rate of R . However, the ZFBF gain S' when an element client and a dummy client of type j are assigned to one of the t_j APs, is much smaller than S . All other clients obtain a rate from APs are 0. We assume that a matching exists if our objective function $F^* \geq F$ where $F = 2nGR + (m - n)R$ (The GR denotes the rate $r_{u,a}$ after client selection in our objective function).

Suppose there is a matching. For each $T_i = (b_j, c_k, d_l)$ in the matching, associate element client c_k and d_l with AP i . For each j , this leaves t_j-1 idle APs corresponds to triples of type j that are not in the matching; associate the $t_j - 1$ dummy client of type j with these t_j-1 APs. This association strategy comply with our objective function value of $F = 2nGR + (m - n)R$. Conversely, suppose that there is such an association strategy with objective function $F^* \geq F$. Each dummy client of type j is associated with an AP of type j . Therefore, APs should not be idle. If possible, with our objective function for obtaining a better solution, there is exactly one AP corresponding to a triple of type j that is associated with two element clients. Each AP is equipped with two antennas, therefore, no AP should get more than 3 clients. For those APs with two clients, it is better to associate two element clients corresponding to AP of type j , because the ZFBF gain is better. So in any of these assignment, there is exactly one AP for each type j equipped with two antennas that be associated with two element clients corresponding to type j . There are all $2n$ element clients which implies a 3-dimensional matching in the original problem. This problem can be extend to the scenario that each AP is equipped with three antennas, it is also implies a 3-dimensional matching.

IV. MU-MIMO AP ASSOCIATION ALGORITHM

In this section, we first introduce our centralized greedy algorithm (Algorithm 1) to make an optimal client-AP association, and then we propose another adaptation algorithm (Algorithm 2) to execute periodic client reassociation based on the current network condition (i.e., the variance in channel characteristic from client mobility, a client arrives/departs in/from the network). We also show that the adaptation algorithm can be extended to a distributed scheme.

A. Maximization sum-rate greedy algorithm

The key idea of the proposed greedy algorithm is to iteratively associate a candidate beamforming group with its best AP that contributes the maximum incremental throughput. The iterative association procedure terminates until each of the clients $u \in \mathcal{U}$ is connected to an AP and included in one beamforming group. Algorithm 1 summarizes our greedy algorithm. Recall that G_a is the set of all candidate beamforming groups within the coverage of AP a . We create a set \mathcal{S}_a to collect all the beamforming groups selected to associate with AP a , which is initially set as an empty set $\mathcal{S}_a = \phi$. To pick the group that produces maximum incremental throughput, we estimate the achievable throughput $T(\mathcal{G}_a)$ of all candidate groups $\mathcal{G}_a \in G_a$, for all APs $a \in \mathcal{A}$,

$$T(\mathcal{G}_a) = R(\mathcal{G}_a) / |\mathcal{S}_a \cup \mathcal{G}_a|, \quad (15)$$

where \mathcal{S}_a is the set of beamforming groups that have already associated with AP a . The rationale of the above equation is that each beamforming group should share the medium, and hence the available bandwidth, with other groups that associate with the same AP. Therefore, the actual achievable throughput of a group is not only related to its instantaneous sum-rate, but also depends on the load of its associated AP. In other words, a group could achieve a relatively low throughput if it connects to a heavy loaded AP that already serves many groups, even if the instant sum-rate of the group, i.e., $R(\mathcal{G})$ is high. Here, we assume that all the associated groups share an equal proportion of channel time to communicate with the AP, and hence we can estimate the achievable throughput of a group by its sum-rate divided the number of associated groups. However, some protocols, e.g., 802.11, support packet fairness, instead of time fairness. We can use more sophisticated method, e.g., [43], to estimate the achievable throughput based on the design of the MAC protocol, without affecting the operation of our algorithm.

Algorithm 1 Greedy algorithm for MU-MIMO AP association

Input: set of APs \mathcal{A} ; set of clients \mathcal{U} ; set of clients \mathcal{C}_a covered by AP $a \in \mathcal{A}$; set of candidate beamforming groups G_a covered by AP $a \in \mathcal{A}$

Output: set of clients \mathcal{U}_a associated with AP $a \in \mathcal{A}$

- 1: $\mathcal{U}_a \leftarrow \{\}, \mathcal{S}_a \leftarrow \{\}$ for all $a \in \mathcal{A}$
 - 2: **while** $\mathcal{U} \neq \{\}$ **do**
 - 3: $T(\mathcal{G}_a) = R(\mathcal{G}_a)/|\mathcal{S}_a \cup \mathcal{G}_a|, \forall \mathcal{G}_a \in G_a, a \in \mathcal{A}$
 - 4: $(a^*, \mathcal{G}^*) = \arg \max_{a \in \mathcal{A}, \mathcal{G}_a \in G_a} T(\mathcal{G}_a)$
 - 5: $\mathcal{U}_{a^*} \leftarrow \mathcal{U}_{a^*} \cup \mathcal{G}^*, \mathcal{S}_{a^*} \leftarrow \mathcal{S}_{a^*} \cup \{\mathcal{G}^*\}$
 - 6: $\mathcal{U} \leftarrow \mathcal{U} \setminus \mathcal{G}^*$
 - 7: $G_a \leftarrow G_a \setminus \{\mathcal{G} : \mathcal{G} \in G_a, \mathcal{G} \cap \mathcal{G}^* \neq \emptyset\}$ for all $a \in \mathcal{A}$
 - 8: **end while**
 - 9: **return** \mathcal{U}_a for all $a \in \mathcal{A}$
-

With the estimated achievable throughput, we can then pick the group along with its associated AP that achieves the maximum throughput gain as follows.

$$\begin{aligned}
 (a^*, \mathcal{G}^*) &= \arg \max_{a \in \mathcal{A}, \mathcal{G}_a \in G_a} T(\mathcal{G}_a) \\
 &= \arg \max_{a \in \mathcal{A}, \mathcal{G}_a \in G_a} R(\mathcal{G}_a)/|\mathcal{S}_a \cup \mathcal{G}_a|.
 \end{aligned} \tag{16}$$

Once we find the best beamforming group \mathcal{G}^* along with its proper AP a^* , we can associate all the members $u \in \mathcal{G}^*$ with AP a^* , i.e., $\mathcal{U}_{a^*} = \mathcal{U}_{a^*} \cup \mathcal{G}^*$. After each selection iteration, we discard all the other candidate beamforming groups that contain the clients $u \in \mathcal{G}^*$, as shown in line 7 of Algorithm 1, and also remove the clients in \mathcal{G}^* from the original client set \mathcal{U} . The selection procedure is repeatedly executed until \mathcal{U} becomes empty, meaning that all the clients in \mathcal{U} associate with an AP. Therefore, different from conventional client-AP association approaches, our algorithm not only associates clients with proper APs, but also, at the same time, forms beamforming groups that can produce high throughput.

Since most of APs in an enterprise/campus WLAN are connected via backhaul wired networks, our proposed algorithm can be practically realized by delegating a lead AP to perform coordination. One way is to let the lead AP collect the information about the achievable rates of

the candidate groups G_a from all APs and solve Algorithm 1. Alternatively, we can reduce the amount of information exchange by iteratively asking each AP to report the achievable sum-rate of its best beamforming group, i.e., $T(\mathcal{G}^*)$, to the lead AP. The lead AP can pick the best client-AP association tuple once it collects all the reports, and announces the results to all the other APs and waits for the next iteration of best group selection. On the other hand, association needs to be updated when the channels of clients change. However, frequent re-association will also cause expensive overhead and oscillation. We hence ask the APs to execute the association algorithm periodically. Deciding the optimal update duration that best balances the tradeoff between the overhead and performance is a potential future research direction.

An example is shown in Table II, where there are two APs equipped with two antennas in the MU-MIMO system and three clients with single antenna. Initially, the network conditions are presented in Table II(a). Each AP selects a subset of clients with the maximum sum rate from their set G_a , then the maximum subset of clients (2,3) are selected from all APs (Table II(b)). The selected subset of clients (2,3) with the maximum sum rate are associated with the AP A, then we subtract them from the set \mathcal{U} and G_a to avoid selecting the duplicate users (2,3) at next iteration (Table II(c)).

However, with the considerations of channel variety and client mobility, we propose another adaptation algorithm to execute client reassociation, which will be introduced in the next subsection. The adaptation algorithm uses a client reassociation strategy only for those clients, whose data rates decline due to client mobility or those who arrives/departs in/from the network after client reselection.

B. Adaptation Algorithm

In this section, we describe the adaptation algorithm that we proposed for executing client reassociations for maximizing the sum rate. The condition of a network is constantly changing. For example, when a new client enters into the network or an existing leaves the network, the condition of the network changes. Therefore, we also consider the variability of network in the design of our method. Thus, we propose an adaptation algorithm (Algorithm 2) for adapting network variability while preserving maximum sum capacity in MU-MIMO LANs.

First, we consider a scenario that a new client enters into the network or an existing client leaves from the coverage of an AP in the network. In order to achieve the maximum sum rate

Set	Clients
\mathcal{U}	1 , 2 , 3
AP	C_a
A	1 , 2
B	1 , 2 , 3
AP	G_a
A	(1, 2) (1) (2)
B	(1, 2) (1, 3) (2, 3) (1) (2) (3)

(a) The network condition at first iteration

AP	Sum-rates(units)
A	(1, 2) = 10
B	(2, 3) = 15

(b) Each AP selects a subset of clients with the maximum sum rate

Set	Clients
\mathcal{U}	1
AP	\mathcal{U}_a
A	
B	2 , 3
AP	G_a
A	(1)
B	(1)

(c) The maximum subset of clients (2,3) are selected from all APs

TABLE II: An example of associating clients with APs

of MU-MIMO system in this scenario, Algorithm 2 makes this new client to be associated with an optimal AP. The equation for the selection of the optimal AP to be associated with a new client is:

$$a^* = \arg \max_{a \in A_u} \left(\sum_{u \in \mathcal{U}'_a} r'_{u,a} / |\mathcal{U}'_a| - t_a \right) \quad (17)$$

The new client u may be covered by many APs. We let \mathcal{A}_u denote APs whose coverages cover

the new client u . First we need to compute each data rate $r'_{u,a}$ given the client u is associated with each AP $a \in \mathcal{A}_u$. Let t_a denote the original throughput of AP a before the association with client u , and \mathcal{U}'_a denote the set of clients that AP a currently serves, including client u . In Eq.(17), the new throughput can be estimated by $r'_{u,a}/|\mathcal{U}'_a|$, and to subtract the original throughput t_a from it indicates the gain for the new AP a from serving the new client u . We search all APs in the set \mathcal{A}_u to find the AP a^* that can achieve the maximal gain for our MU-MIMO system, and then we associate client u with AP a^* and update the set \mathcal{U}_{a^*} with the new set \mathcal{U}'_{a^*} that includes new client u .

We execute client reselection to match the maximal sum rate of the subset of clients on an AP to achieve maximal multiplexing gain when a client connects/disconnects to an AP or the channel characteristic changes due to client mobility. The Algorithm 2 focuses on adopting reassociation strategy for the client based on the current network condition at per period. We first compute the load of all APs $a \in \mathcal{A}$, denoted by L_a , and the load definition is interpreted in [5]. The load should be inversely proportional to the bandwidth that the clients obtain from an AP. Assuming that each AP a provides the same traffic volume of size 1 to all its associated clients $u \in \mathcal{U}_a$, and $r'_{u,a}$ denotes the current latest data rate(i.e, $L_a = \sum_{u \in \mathcal{U}_a} \frac{1}{r'_{u,a}}$). And we sort all APs a in the set M_a by L_a in an descending order, then we start to switch clients $u \in \mathcal{U}_a$ to other APs from the maximum load of AP to the minimum load of AP for balancing the load of all APs $a \in M_a$.

We further divide all clients $u \in \mathcal{U}_a$ into two categories sets: (i) The first set U_d includes the clients with reduction on its data rate. Let $r'_{u,a}$ denote the latest data rate after executing client reselection due to client mobility on the AP a , and $r_{u,a}$ is the last data rate obtained from the AP a . (ii) The second set U_o includes the clients, whose data rates remain unchanged. Then sorting all clients u in these two sets U_d and U_o by the weight W_{declne} and $W_{original}$ in a descending order, respectively. These two weights are represented as

$$W_{original} = [(SNR_{alone} - SNR_{new})/SNR_{alone}] \quad (18)$$

$$W_{declne} = [(SNR_{old} - SNR_{new})/SNR_{old}] * W_{original} \quad (19)$$

,where SNR_{alone} denotes the client's SNR that an AP communicates with only this client at one time, SNR_{old} denotes the last client's SNR that concurrently communicates with other clients

with an AP, and SNR_{new} indicates the latest client's SNR after client reselection. Thus, $W_{original}$ indicates the percentage of SNR reduction for the client as it concurrently communicates with other clients with an AP, that can be channel orthogonality between other clients for the client u . The amount of SNR reduction is inversely proportional to channel orthogonality. Let $W_{decline}$ indicates the percentage of SNR reduction for the client after client reselection that be multiplied by the weight $W_{original}$.

In Algorithm 2, at per period, we only switch the clients to other APs, which in the set U_o or U_d in each APs $a \in M_a$. Therefore, if the set U_d is empty, we will do client reassociation for clients in set U_o . However, in order to guarantee that the client can obtain higher rate and produce maximal total incremental throughput for our MU-MIMO network, we formulate two qualifications that must be satisfied by any client u in set U_o or U_d before reassociation. The qualifications are represented as

$$r_{u,j} - r'_{u,a} > 0 \quad (20)$$

$$(t''_a + t'_{u,j}) - (t'_{u,a} + t_j) > 0 \quad (21)$$

Let t_j and $t'_{u,a}$ indicate the throughput of AP $j \in \mathcal{A}_u$ that has not served the client u and the latest throughput of the AP $a \in \mathcal{A}_u$ that currently serve the client u respectively. And we denote $t'_{u,j}$ and t''_a as the renew throughput of the AP j that associates to client u after client reassociation and the renewed throughput of the AP a that be disconnected from client u . In order to produce incremental gain on rate of client u and the total throughput for our downlink system, the Eq.(20) and Eq.(21) also must be satisfied before switching client u to AP j from AP a . Then, If the any AP $j \in \mathcal{A}_u$ satisfies the above restrictions, we put this AP j into the set M_c . Because all coordinated APs are connected to a centralized controller that constantly monitors the whole condition of network, the above situation can be detected and computed to determine whether the client reassociation is executed.

For achieving the maximal total incremental throughput, the Algorithm 2 determine that a client u switches to which one AP j^* providing the highest benefit from an original AP a . If there is no AP j complies with above qualifications, the client u still keep connecting with

original AP a and adopting original $r'_{u,a}$ in downlink. The main equation for finding the optimal AP $j^* \in \mathcal{A}_u, a \neq j^*$ for client u is as follows

$$j^* = \arg \max_{a \neq j, j \in M_c} \left[(t''_a + t'_{u,j}) - (t'_{u \rightarrow a} + t_j) \right]. \quad (22)$$

Through the Eq.(22), thus, the client u is switched to another AP j^* to obtain the best performance. Then, the new set \mathcal{U}'_{j^*} and \mathcal{U}''_a , the former includes the client u and the latter denotes the remainder of clients after client u disconnects to AP a , both of them update the original the set \mathcal{U}_{j^*} and \mathcal{U}_a , respectively.

In our design, each AP $a \in \mathcal{A}$ always check that whether the channel characteristics between its clients $u \in \mathcal{U}_a$ have be changed due to client mobility or the client is connected/disconnected to it, then it executes client reselection. Our Algorithm 2 is executed to adopt an optimal client reassociation for maximizing the sum rate based on the current network condition at per period. It adopts the client reassociation on the maximum load of AP in the set M_a until the all APs in set M_a also adopt this operation once for theirs clients. Thus, each AP can switch theirs clients to other optimal APs based on the current condition of clients.

C. Distributed Algorithm

The above two algorithms assume that each AP a are connected to a centralized controller so that it has global information of whole network and can be efficient to dynamically change the network topology for client reassociation. However, it might not be the practice solution for a network, where APs have be set up in the environment and are not connected to a centralized controller. Hence, we extend our Algorithm 2 in a distributed scheme, where APs use local information to decide optimal client association strategy at per period.

We first let each AP a can periodically broadcast the information: (i) the set U_d and U_o (ii) all clients' data rate in set U_d and U_o , (iii) its load, and collect these information from its neighboring APs. Then, each AP creates its *neighbor_list* that includes these informations, and checks that whether it has the maximum load in its *neighbor_list*. If AP a has maximum load in its *neighbor_list*, it starts to decide whether to switch its clients in the set U_d or U_o to its neighboring APs for maximizing sum rate. If other APs $j \in \mathcal{A}, j \neq a$ also include the AP a with the maximum load in theirs *neighbor_list*, they compute $t'_{u,j}$ and $r_{u,j}$ for the client u in

the AP a 's set U_d or U_o and also be covered by these APs j . Then these APs j broadcast $t'_{u,j}$ and $r_{u,j}$, the AP a can check that whether switching the client u to AP j can achieve a positive gain, i.e., satisfy the Eq.(20) and Eq.(21). In order to obtaining the maximal gain for our MU-MIMO system, AP a switches client u to the AP j^* in Eq.(22). When the AP a adopt the client reassociation strategy for all clients in set U_d or U_o , it sends a termination information to its neighboring APs. Then the second maximum load of AP in AP a 's *neighbor_list* will execute the same procedure, and the remainder of APs in AP a 's *neighbor_list* also do it with descending order of load. The Algorithm 2 in a distributed scheme will be periodically executed, and each AP a adopts a client reassociation for their clients based on the descending order of load in their *neighbor_list* until all AP $a \in \mathcal{A}$ execute this procedure once. The biggest difference between this distributed scheme and the centralized scheme with the Algorithm 2 is that the APs must exploit local information exchange to adjust the client configuration. Therefore, a distributed scheme can also cope with the network dynamics such that the clients are mobile, or join and leave dynamically.

D. Computational complexity

Each iteration of Algorithm 1 includes two steps: (i) each AP selects its best beamforming group, and (ii) each AP removes the unnecessary candidate groups from G_a . We note that the complexity of Algorithm 1 highly depends on the complexity of the client grouping algorithm we apply. If we apply exhausted search, we need to search the best beamforming group from the set of candidates G_a . Since the size of G_a is $\sum_{k=1}^{N_a} C(|\mathcal{U}|, k)$ for each AP $a \in \mathcal{A}$ and we need at most $|\mathcal{U}|$ iterations, the complexity of Algorithm 1 is $O(|\mathcal{U}|^{N_a} |\mathcal{A}| |\mathcal{U}|)$. However, we can combine our algorithm with some lightweight client grouping algorithms, e.g., [13] [22], with the complexity of $O(N_a^3 |\mathcal{U}|)$ for each iteration of client grouping. Then, the overall complexity can be reduced to $O(N_a^3 |\mathcal{A}| |\mathcal{U}|)$. The complexity of Algorithm 2 is $O(N_a^3 |\mathcal{A}| |\mathcal{U}|^2)$. To balance the tradeoff between complexity and performance, we keep our algorithm flexible, and let the system deployer determine the client grouping algorithm. One can apply exhausted search when the number of antennas at each AP is small, while applying greedy algorithms, otherwise.

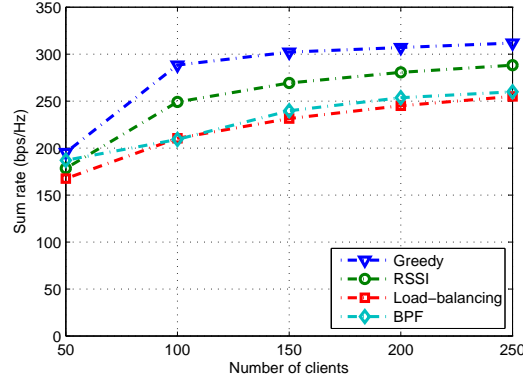
E. Practical issue

To obtain the channel information of all users at APs for selecting a better subset of users, the users can leverage *channel reciprocity* [44] to learn their channels by exploiting the beacons and report this channel information to the AP. Therefore, the CSI is available to both AP and users. To improve the network performance for users whose data rates decline, the Algorithm 2 is executed to reassociate them to other better APs. However, the reassociation mechanism can cause the latency of transmission, and a user may experience a connectivity gap. In order to address the seamless connections in WLANs, smooth reassociation mechanism that use a second wireless interface has been identified as an interesting way to reduce reassociation latency and improve user experience [45], [46]. Therefore, before disconnecting from an original AP, the second interface is triggered to connect to a candidate AP to provide continuous connectivity. In our MU-MIMO system, our cooperated AP also provides the information of candidate AP (e.g., BSSID, its channel), the user can directly associate the candidate AP to avoid the overhead in scanning phase. Therefore, the reassociation mechanism can be executed practically to improve the network performance and maintain continuous connectivity.

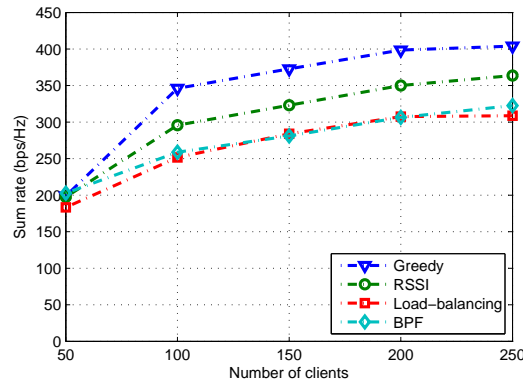
V. SIMULATION RESULTS

We conduct simulations to evaluate our algorithm in a multi-cell MU-MIMO network. Clients and APs are uniformly randomly distributed in a $500 \times 500m$ square area. Each AP is equipped with 3-4 antennas, while each client is equipped with a single antenna. The communication range of each AP is $150m$. Each node applies the free-space path loss model with the exponent 3. The transmit power is set to 15 dBm, while the noise level is set to -95 dBm. Each adjacent AP uses different channels to enable spatial reuse. We assume that all the clients have continues downlink traffic. Each AP hence transmits concurrent packets to the members of each beamforming group in a round-robin manner.

We compare our greedy algorithm with three association schemes. One is the RSSI-based association scheme, which allows each client to select the AP with the strongest signal strength. The second scheme is the load-balancing association scheme, where the client selects the AP with consideration of the load, i.e., number of served clients, of the APs. To be specific, each client predicts its achievable throughput from all the APs within its coverage range, and selects the one that provides the highest single-link throughput, without considering channel correlation



(a) Sum-rate in the 3-antenna AP scenarios

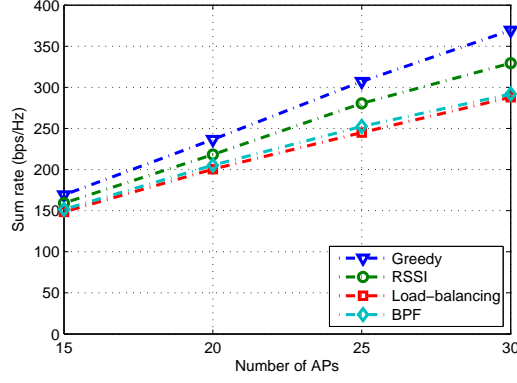


(b) Sum-rate in the 4-antenna AP scenarios

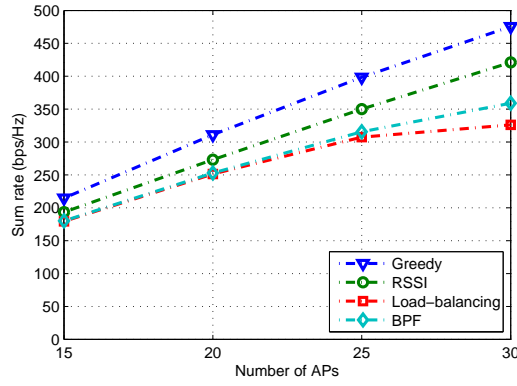
Fig. 2: The performance of varying numbers of clients (25 APs)

among clients. However, since the throughput estimation depends on the number of clients that have associated with an AP, we let clients sequentially select their suitable APs in a random order. The third scheme is BPF [8] based on a novel performance revenue function, which jointly considers AP association and fair bandwidth allocation. Thus, It can achieve proportional fairness. The All the comparison schemes use exhausted search to find beamforming groups. We check the performance of the comparison schemes for varying numbers of APs, numbers of clients and numbers of antennas equipped at each AP. For each simulation setting, we report the average result of 50 simulation runs.

We first consider the impact of traffic load on the throughput performance by changing the number of clients from 50 to 250. The number of APs is fixed to 25. Figs. 2(a) and 2(b) plot the average throughput of each beamforming group for 3-antenna and 4-antenna APs scenarios,



(a) Sum-rate in the 3-antenna AP scenarios



(b) Sum-rate in the 4-antenna AP scenarios

Fig. 3: The performance of varying numbers of APs (200 clients)

respectively. The results show that the performance of the load-based and BPF schemes are even worse than the simple RSSI-based scheme. This illustrates that estimating the throughput without considering channel correlation among clients might overestimate the available bandwidth from an AP, and hence performs worse in a MU-MIMO environment. By jointly considering client grouping, our greedy algorithm outperforms the other three association schemes. The average gain of our greedy algorithm over the RSSI-based, loading-based and BPF schemes is 11% (12%), 26% (28%) and 22% (30%), respectively, in the 3-antenna (4-antenna) AP scenarios. The maximum gain is up to 15% (17%), 37% (38%) and 34% (38%), respectively, when the number of clients is 100. The performance of the BPF scheme is similar to the performance of the Load-based scheme. When the number of clients is 50, each AP may cover only a small number of clients, as a result reducing the benefit of diversity. On the other hand, when the

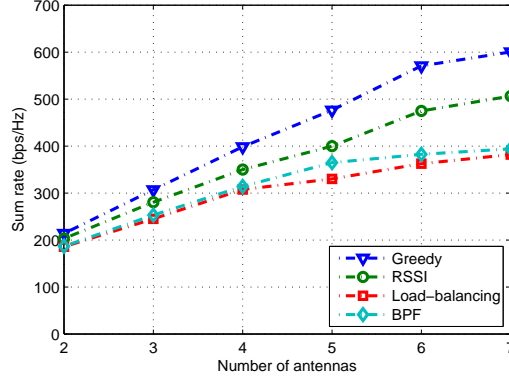


Fig. 4: Varying number of antennas at each AP (25 APs, 200 clients)

number of clients increases, even without proper client-AP association, the AP might still be able to select acceptably good beamforming groups. This explains why the throughput gain for the 250-client case is less significant than the 100-client case. In the 4-antenna AP scenarios, the clients achieve a higher throughput than that obtained in the 3-antenna AP scenarios because the APs can support more concurrent clients. However, the general performance trend of two scenarios are similar.

We then check the performance when the number of APs varies from 15 to 30 and the number of clients is fixed to 200. Figs. 3(a) and 3(b) plot the average aggregate throughput in the 3-antenna and 4-antenna AP scenarios, respectively. The results also show that our proposed greedy algorithm outperforms the three association schemes. In general, the performance of all the schemes increases when there are more APs sharing the traffic load of clients and providing their clients a higher bandwidth. However, without considering channel correlation among clients, the load-based scheme and BPF scheme cannot fully utilize the bandwidth of each AP and, hence, its throughput improvement is less significant. In other words, the gain of our greedy algorithm increases as the number of APs increases because each client has more association choices. The maximum gain over the load balancing scheme is 28% (45%) for 3-antenna (4-antenna) AP scenarios when the number of APs is 30.

In Fig. 4, we examine the performance when the number of APs' antennas changes from 2 to 7. In this simulation, we fix the number of APs and clients to 25 and 200, respectively. The figure shows that our algorithm outperforms the other schemes and the gain over the RSSI-based, load-

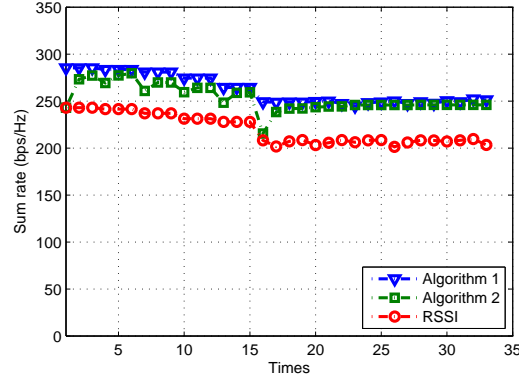


Fig. 5: Comparison of sum rate performance with mobility of clients and variability of channel characteristic (25 APs ; 200 clients).

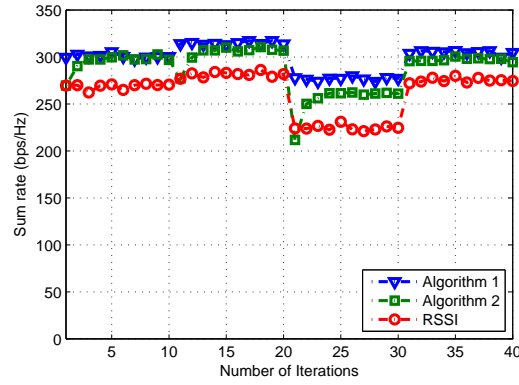


Fig. 6: Iterations required to converge (25 APs ; 150,200,100,180 clients).

based and BPF schemes increases as the number of AP's antennas increases. This is because, when the number of concurrent clients increases, the probability that any two concurrent clients have correlated channels also increases, as a result decreasing the average rate of a beamforming group. In this case, careful client grouping and AP association become more important. This explains why our algorithm can benefit more from increasing the number of antennas at each AP.

We finally compare the performance of our Algorithm 2 with our proposed Algorithm 1 and RSSI association scheme. In this simulation, we assume that initially all clients associate with the strongest signal strength of APs. In Fig.5, we evaluate the performance of Algorithm 2 when there are mobility of clients and variability of channel characteristic. The result shows that the

Algorithm 2 can quickly converge to the performance of Algorithm 1 that just only reconfigures some client-AP associations based on current network condition without recalculation of whole network configuration from Algorithm 1. And in fig.6, we divide the simulation to four phases and change the number of clients in each phase (the four phase are set to 150, 200, 100, 180 clients, respectively.). The result also shows that the Algorithm 2 can converge within less than five iterations, and it's performance can approach the performance of Algorithm 1.

VI. CONCLUSIONS

In this paper, we consider the problem of client-AP association in a MU-MIMO WLAN. We have formulated the joint problems of client-AP association and client grouping as a mathematical model. Due to the NP-hardness of the joint problem, we propose two greedy algorithm that computes the achievable sum-rate of each beamforming group based on channel orthogonality among clients and then iteratively associates the best beamforming group with its optimal AP that produces the maximal sum-rate. We also consider the practical deployment, thus, we extend our Algorithm 2 in a distributed scheme. Our evaluation shows that the proposed algorithm achieves a throughput gain of 11–28%, 26–45% and 22–38% over the RSSI-based , load-based and BPF schemes, respectively. The result also showed that Algorithm 2 has similar performance with that of Algorithm 1, and converges quickly in a few iterations.

REFERENCES

- [1] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2658–2668, 2003.
- [2] E. Aryafar, N. Anand, T. Salonidis, and E. W. Knightly, "Design and Experimental Evaluation of Multi-user Beamforming in Wireless LANs," in *ACM MobiCom*, 2010.
- [3] W.-L. Shen, Y.-C. Tung, K.-C. Lee, K. C.-J. Lin, S. Gollakota, D. Katabi, and M.-S. Chen, "Rate Adaptation for 802.11 Multiuser MIMO Networks," in *ACM Mobicom*, 2012.
- [4] T. Yoo, N. Jindal, and A. Goldsmith, "Multi-Antenna Downlink Channels with Limited Feedback and User Selection," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 7, pp. 1478–1491, 2007.
- [5] F. Xu, X. Zhu, C. C. Tan, Q. Li, G. Yan, and J. Wu, "SmartAssoc: Decentralized Access Point Selection Algorithm to Improve Throughput," *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 12, pp. 2482–2491, 2013.
- [6] Y. Bejerano, S.-j. Han, and L. Li, "Fairness and Load Balancing in Wireless LANs Using Association Control," *IEEE/ACM Transactions on Networking*, vol. 15, no. 3, pp. 560–573, 2007.
- [7] A. Baid, M. Schapira, I. Seskar, J. Rexford, and D. Raychaudhuri, "Network Cooperation for Client-AP Association Optimization," in *International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt)*, 2012.

- [8] 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 Transactions on Networking*, vol. 22, no. 1, pp. 191–202, Feb 2014.
- [9] E. Telatar, "Capacity of Multi-antenna Gaussian Channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, 1999. [Online]. Available: <http://dx.doi.org/10.1002/ett.4460100604>
- [10] A. El Gamal and T. M. Cover, "Multiple user information theory," *Proceedings of the IEEE*, vol. 68, no. 12, pp. 1466–1483, 1980.
- [11] D. Gesbert, M. Kountouris, R. W. Heath, C.-B. Chae, and T. Salzer, "Shifting the MIMO Paradigm," *IEEE Signal Processing Magazine*, vol. 24, no. 5, pp. 36–46, 2007.
- [12] P. Viswanath and D. N. C. Tse, "Sum Capacity of the Vector Gaussian Broadcast Channel and Uplink-downlink Duality," *IEEE Transactions on Information Theory*, vol. 49, no. 8, pp. 1912–1921, 2006. [Online]. Available: <http://dx.doi.org/10.1109/TIT.2003.814483>
- [13] G. Dimic and N. D. Sidiropoulos, "On Downlink Beamforming with Greedy User Selection: Performance Analysis and a Simple New Algorithm," *IEEE Transactions on Signal Processing*, vol. 53, no. 10, pp. 3857–3868, 2005.
- [14] G. Gupta and A. K. Chaturvedi, "Conditional Entropy Based User Selection for Multiuser MIMO Systems," *IEEE Communications Letters*, vol. 17, no. 8, pp. 1628–1631, 2013.
- [15] S. Huang, H. Yin, J. Wu, and V. C. M. Leung, "User Selection for Multi-user MIMO Downlink with Zero-Forcing Beamforming," *CoRR*, vol. abs/1206.6, 2012.
- [16] T.-W. Kuo, K.-C. Lee, K.-J. Lin, and M.-J. Tsai, "Leader-Contention-Based User Matching for 802.11 Multiuser MIMO Networks," *IEEE Transactions on Wireless Communications*, vol. 13, no. 8, pp. 4389–4400, Aug 2014.
- [17] Z. Shen, R. Chen, J. G. Andrews, R. W. Heath, and B. L. Evans, "Sum Capacity of Multiuser MIMO Broadcast Channels with Block Diagonalization," *IEEE Transactions on Wireless Communications*, vol. 6, no. 6, pp. 2040–2045, 2007.
- [18] —, "Low complexity user selection algorithms for multiuser MIMO systems with block diagonalization," *IEEE Transactions on Signal Processing*, vol. 54, no. 9, pp. 3658–3663, 2006.
- [19] C. Shepard, H. Yu, N. Anand, E. Li, T. Marzetta, R. Yang, and L. Zhong, "Argos: Practical Many-antenna Base Stations," in *ACM Mobicom*, 2012.
- [20] K. Tan, H. Liu, J. Fang, W. Wang, J. Zhang, M. Chen, and G. M. Voelker, "SAM: Enabling Practical Spatial Multiple Access in Wireless LAN," in *ACM MobiCom*, 2009.
- [21] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian MIMO broadcast channel," in *IEEE Transactions on Information Theory*, 2004, pp. 174–.
- [22] T. Yoo and A. Goldsmith, "On the Optimality of Multiantenna Broadcast Scheduling Using Zero-Forcing Beamforming," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 3, pp. 528–541, 2006.
- [23] J. Wang, D. J. Love, and M. D. Zoltowski, "User Selection With Zero-Forcing Beamforming Achieves the Asymptotically Optimal Sum Rate," *IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp. 3713–3726, 2008.
- [24] H. V. Balan, R. Rogalin, A. Michaloliakos, K. Psounis, and G. Caire, "AirSync: Enabling Distributed Multiuser MIMO With Full Spatial Multiplexing," *IEEE/ACM Transactions on Networking*, vol. 21, no. 6, pp. 1681–1695, 2013.
- [25] J. R. Pedrasa, M. A. Pedrasa, and A. P. Seneviratne, "Information Exchange for Enhanced Network Selection," in *IEEE WCNC*, 2010.
- [26] C.-Y. Huang and P. Ramanathan, "Joint Association and Power Adaption in WiFi Community Networks," in *Annual Conference on Information Sciences and Systems (CISS)*, 2011.

- [27] Y. Cui, T. Ma, J. Liu, and S. Das, "Load-balanced AP association in multi-hop wireless mesh networks," *The Journal of Supercomputing*, vol. 65, no. 1, pp. 383–409, 2013. [Online]. Available: <http://dx.doi.org/10.1007/s11227-010-0519-7>
- [28] P. Lv, X. Wang, M. Xu, and Y. Chen, "Network-Leading Association Scheme in IEEE 802.11 Wireless Mesh Networks," in *IEEE Communications (ICC)*, 2011, pp. 1–5.
- [29] F. Xu, C. C. Tan, Q. Li, G. Yan, and J. Wu, "Designing a Practical Access Point Association Protocol," in *IEEE INFOCOM*, 2010.
- [30] S. Keranidis, T. Korakis, I. Koutsopoulos, and L. Tassiulas, "Contention and Traffic Load-Aware Association in IEEE 802.11 WLANs: Algorithms and Implementation," in *International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt)*, 2011.
- [31] K. Sundaresan and K. Papagiannaki, "The Need for Cross-layer Information in Access Point Selection Algorithms," in *Proceedings of the 6th ACM SIGCOMM Conference on Internet Measurement*, ser. IMC '06, 2006, pp. 257–262. [Online]. Available: <http://doi.acm.org/10.1145/1177080.1177113>
- [32] T. Sun, Y. Zhang, and W. Trappe, "Improving Access Point Association Protocols through Channel Utilization and Adaptive Switching," in *IEEE Mobile Adhoc and Sensor Systems (MASS)*, 2011, pp. 155–157.
- [33] D. Gong and Y. Yang, "AP association in 802.11n WLANs with heterogeneous clients," in *IEEE INFOCOM*, 2012, pp. 1440–1448.
- [34] S. K. Dandapat, B. Mitra, R. R. Choudhury, and N. Ganguly, "Smart Association Control in Wireless Mobile Environment Using Max-Flow," *IEEE Transactions on Network and Service Management*, vol. 9, no. 1, pp. 73–86, 2012.
- [35] I. Broustis, K. Papagiannaki, S. V. Krishnamurthy, M. Faloutsos, and V. P. Mhatre, "Measurement-Driven Guidelines for 802.11 WLAN Design," *IEEE/ACM Transactions on Networking*, vol. 18, no. 3, pp. 722–735, 2010.
- [36] M. Kawada, M. Tamai, and K. Yasumoto, "A trigger-based dynamic load balancing method for WLANs using virtualized network interfaces," in *IEEE Wireless Communications and Networking Conference (WCNC)*, 2013, pp. 1091–1096.
- [37] S. Ye and R. S. Blum, "Optimized Signaling for MIMO Interference Systems with Feedback," *IEEE Transactions on Signal Processing*, vol. 51, no. 11, pp. 2839–2848, 2003.
- [38] A. Wiesel, Y. C. Eldar, and S. Shamai, "Zero-Forcing Precoding and Generalized Inverses," *IEEE Transactions on Signal Processing*, vol. 56, no. 9, pp. 4409–4418, 2008.
- [39] T. Yoo and A. Goldsmith, "Optimality of Zero-Forcing Beamforming with Multiuser Diversity," in *IEEE ICC*, 2005.
- [40] S. Huang, H. Yin, J. Wu, and V. Leung, "User Selection for Multiuser MIMO Downlink With Zero-Forcing Beamforming," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 7, pp. 3084–3097, Sept 2013.
- [41] T. Bu, L. Li, and R. Ramjee, "Generalized Proportional Fair Scheduling in Third Generation Wireless Data Networks," in *IEEE INFOCOM*, 2006, pp. 1–12.
- [42] J. K. Lenstra, D. B. Shmoys, and E. Tardos, "Approximation Algorithms for Scheduling Unrelated Parallel Machines," *Math. Program.*, vol. 46, no. 3, pp. 259–271, 1990. [Online]. Available: <http://dx.doi.org/10.1007/BF01585745>
- [43] K.-J. Lin and C.-F. Chou, "Exploiting multiple rates to maximize the throughput of wireless mesh networks," *IEEE Transactions on Wireless Communications*, vol. 8, no. 12, pp. 6038–6049, Dec. 2009.
- [44] M. Guillaud, D. T. M. Slock, and R. Knopp, "A practical method for wireless channel reciprocity exploitation through relative calibration," in *Signal Processing and Its Applications*, vol. 1, August 2005, pp. 403–406.
- [45] V. Brik, A. Mishra, and S. Banerjee, "Eliminating handoff latencies in 802.11 w lans using multiple radios: Applications, experience, and evaluation," in *Proceedings of the 5th ACM SIGCOMM Conference on Internet Measurement*, ser. IMC '05. USENIX Association, 2005, pp. 27–27. [Online]. Available: <http://dl.acm.org/citation.cfm?id=1251086.1251113>

- [46] K. Ramachandran, S. Rangarajan, and J. Lin, “Make-before-break mac layer handoff in 802.11 wireless networks,” in *IEEE ICC*, vol. 10, June 2006, pp. 4818–4823.

Algorithm 2 Adaptation algorithm for maximization sum-rate

Input: set of all APs \mathcal{A} ; the sets of clients \mathcal{U}_a associated with AP $a \in \mathcal{A}$; the set of $A_u \subset \mathcal{A}$ that cover the same client u

Output: the sets of clients \mathcal{U}_a associated with AP $a \in \mathcal{A}$

```

1: if a new client  $u$  arrivals then
2:    $a^* = \arg \max_{a \in \mathcal{A}_u} (\sum_{u \in \mathcal{U}'_a} r'_{u,a} / |\mathcal{U}'_a| - t_a)$ 
3:    $\mathcal{U}_{a^*} \leftarrow \mathcal{U}'_{a^*}$ 
4: end if
5: while TIME = T do
6:   for all  $a \in \mathcal{A}$  do
7:      $L_a = \sum \frac{1}{r_{u,a}}, u \in \mathcal{U}_a$ 
8:   end for
9:    $M_a = \{a \mid L_a, a \in \mathcal{A}\}$ 
10:  sort AP  $a$  in  $M_a$  by  $L_a$  in an descending order
11:  for all  $a \in M_a$  in order do
12:    if  $r'_{u,a} - r_{u \rightarrow a} < 0, u \in \mathcal{U}_a$  then
13:       $U_d \leftarrow u$ 
14:      sort client  $u$  in  $U_d$  by  $W_{decline}$  in an descending order
15:    else
16:       $U_o \leftarrow u$  for all  $u \in \mathcal{U}_a$ 
17:      sort client  $u$  in  $U_o$  by  $W_{original}$  in an descending order
18:    end if
19:    for all  $u \in (U_d \text{ or } U_o)$  in order do
20:      (if  $U_d$  equals empty set, we execute the operation in set  $U_o$ )
21:      for all  $j \in \mathcal{A}_u, a \neq j$  do
22:        if  $r_{u,j} - r_{u,a} > 0 \parallel [(t''_a + t'_{u,j}) - (t'_{u,a} + t_j)] > 0$  then
23:           $M_c \leftarrow j$ 
24:        end if
25:      end for
26:       $j^* = \arg \max_{j \in M_c} [(t''_a + t'_{u,j}) - (t'_{u,a} + t_j)]$ 
27:       $\mathcal{U}_{j^*} \leftarrow \mathcal{U}'_{j^*}$ 
28:       $\mathcal{U}'_a \leftarrow \mathcal{U}''_a$ 
29:    end for
30:  end for
31: end while

```