Optimal Mapping of Stations to Access Points in Enterprise Wireless Local Area Networks

Suzan Bayhan and Anatolij Zubow
Telecommunication Networks Group, Technische Universität Berlin, Germany
{bayhan,zubow}@tkn.tu-berlin.de

ABSTRACT

Efficient resource allocation in enterprise wireless local area networks (WLAN) has become more paramount with the shift of traffic toward WLANs and increasing share of the video traffic. Unfortunately, current practise of client-driven association to APs has several shortcomings, e.g., *sticky client problem*. As a remedy, we propose to move the AP association decision to a periodically-running central controller which aims to maximize the proportionally-fair network throughput. After formulating the optimal mapping problem, we devise several heuristics requiring various degrees of knowledge, e.g., pairwise user-AP link rates, throughput demand of each user. Our analysis via simulations on realistic scenarios (conference, office, and shopping mall) shows the superior performance of our proposals in terms of aggregate logarithmic throughput. While the utility gain over the conventional client-driven approach is modest, up to 18%, the resulting increase in the weakest user's throughput is significant (71-120%) as well as that of AP load balance and fairness of user throughputs. Moreover, our evaluations reveal a very small optimality gap (between 0.1-5%). The highest gain is observed in the conference setting where the users are unevenly distributed in the network and hence there is a huge load imbalance among the APs. While schemes requiring more knowledge, i.e., on handover-cost and traffic demands, perform the best, a naive approach which runs periodically and assigns each user to the AP providing the highest signal level to that user maintains up to 41% gain in the weakest user's throughput over the client-driven handover approach.

KEYWORDS

AP association, enterprise WLAN, handover

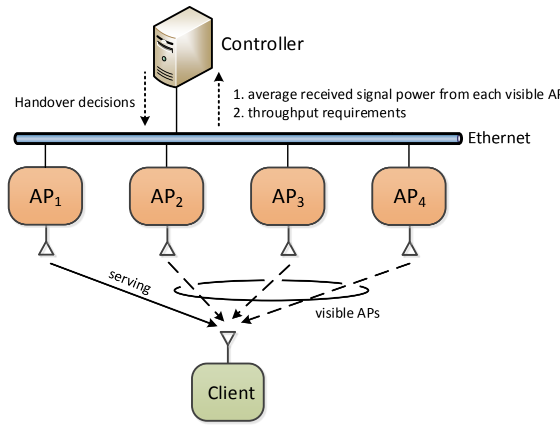


Figure 1: Dense enterprise WLAN setting. The controller collects channel state information and client throughput requirements from the APs for the optimal user-AP association.

1 INTRODUCTION

IEEE 802.11 wireless local area networks (WLANs) have recently become the predominant access networks due to a variety of reasons: ease of deployment, the maturity of the WiFi standard, operating on the licence-exempt spectrum compared to the costly cellular bands, offloading strategy for network operators, to name possibly the most important ones. With this increasing shift to WLANs, managing the WiFi resources has gained more importance to provide a pleasant user experience. Enterprise WLANs, in contrast to its residential counterparts, are amenable to centralized resource management owing to the infrastructure with common management and control authority. For example, in an enterprise network, a controller can decide user-access point (AP) associations and trigger user handovers to improve the network performance, e.g., balancing the network loads on APs [4, 10, 16], managing user mobility [17] or interference by hidden nodes.

Conventionally, a WiFi station[1] connects to the AP providing the highest signal strength and sticks to this AP until the received signal level is below some threshold received signal strength indicator (RSSI), e.g., -80 dBm [11]. When the signal level is low, the user scans for other APs and selects the AP having the strongest signal towards the user without considering the AP's load or the interference situation at the APs. However, as user's knowledge of the network state is much limited and local compared to that of an AP or the network, infrastructure-driven handover proposes to control or assist the handovers using a broader knowledge or even a global view of the network. Different than client-driven handovers which are triggered mostly as a result of decreasing RSSI of a mobile user, infrastructure-driven handovers can be triggered also to improve the network performance. On the other hand, since users experience some outage during a handover—around 4 s with today's hardware [17], frequent handovers will degrade the user experience. Therefore, it is not straightforward to decide how often a controller should manage user-AP associations and how much such a control would improve network performance over the conventional client-driven handover scheme.

accepted for presentation at MSWiM '17, November 2017, Miami, FL, USA

ACM ISBN xxxx...$15.00
https://doi.org/AUTHORCOPY

[1]We use the terms *user*, *client*, and *station* interchangeably.

An efficient user-AP association scheme must ensure high user throughput as well as fairness among users. Client-driven approach obviously falls short of meeting these requirements as it is not responsive to changes in the link quality unless the received signal drops below some threshold RSSI value. Moreover, selecting the link with the highest SNR does not guarantee a pleasant user experience as WiFi is a shared access medium and airtime must be shared with other users. Given that video traffic is becoming more predominant, ensuring high SNR and high airtime may not still provide user satisfaction unless minimum throughput requirement of an application is fulfilled. Hence, an efficient scheme should be aware of the application's throughput requirements. Finally, as aforementioned, handover-cost awareness is another desirable aspect.

Contributions: Our contributions in this paper are threefold:

- Based on the above-listed requirements, we first formulate optimal mapping of users to APs as a proportionally-fair network throughput maximization problem. Due to the hardness of this problem, we design efficient algorithms that are only slightly sub-optimal. Different than the existing solutions which are executed only when a user joins the network or leaves the coverage of an AP, we investigate the benefits of periodic management of user-AP associations.
- Different than existing solutions, we consider the entailed overhead of handover in terms of time lost for handover during which the user cannot get service from its AP. Moreover, with the increasing shift to video traffic, user-AP association without considering the throughput requirement of applications may fall short of providing a pleasant user experience. To represent this fact, we consider minimum rate requirements of users while deciding on user-AP associations.
- Our simulations of realistic scenarios—conference, office, and shopping mall settings, show the superior performance of our proposals under increasing controller period, number of users, and handover latency. We observe the highest gain (e.g., up to 120% increase in the weakest user's throughput) in the conference setting where there is a significant density imbalance in the network.

The rest of the paper is organized as follows. Section 2 describes the considered enterprise WLAN setting while Section 3 introduces the optimal user-AP association problem by modelling the airtime and expected throughput of each user under a particular AP association. Next, Section 4 presents lower-complexity user-AP association schemes as the formulated optimization problem is computationally hard. Section 5 assesses the performance of our proposals along with the traditional client-driven handover approach. Section 6 overviews the related work while Section 7 concludes the paper.

2 SYSTEM MODEL

We consider an enterprise WLAN with a central controller operating in a time-slotted manner. The controller collects statistics from APs at the beginning of each time slot after which it may trigger the APs to handover their users to the selected APs. Fig.1 shows the considered setting and Table 1 lists the used notations.

Let $\mathcal{A} = \{AP_j, \cdots, AP_K\}$ represent the set of APs with K APs and $C = \{1, \cdots, F\}$ the set of channels these APs can operate on. As enterprise networks are planned with the goal of ensuring high service continuity and high capacity for mobile users, AP coverage areas are mostly overlapping [6]. Each AP covers a circular region of radius r meters. We can represent the network topology by a connectivity graph $G = (\mathcal{A}, \mathcal{E})$ where APs are abstracted as the vertices and an edge between two APs, e.g., AP_j and AP_k, means that the two APs have some overlapping coverage region. We represent the channel allocation with $\mathcal{F} = [a_j^f]$ where a_j^f yields 1 if AP_j is assigned channel f for its operation. Denote the set of APs which are assigned to f by $\mathcal{A}_f$ and we refer to the APs in this set as *co-channel APs*. Note that an efficient channel assignment scheme should guarantee minimal interference among co-channel APs [5]. We assume that two APs which are in the interference range are assigned orthogonal channels.

Let $\mathcal{U} = \{u_i, \cdots u_n\}$ denote the set of n users in this network. As users might have different traffic types, e.g., voice vs. video, their throughput requirements may also vary. We denote by r_i^{min} the minimum required rate for user u_i. We assume that uplink traffic is negligible and focus only on the saturated downlink traffic.

As a result of overlapping cells, a user u_i may be receiving signals from a number of APs with varying signal strength. Let $d_{i,j}$ denote the distance between u_i and AP_j. We call the set of APs a user overhears as *visible APs* for this user, i.e., received signal power from the AP is above the receiver sensitivity of the station. We denote the visibility of AP_j at u_i by a binary variable $v_{i,j}$, which yields value 1 if AP_j is a visible AP for u_i. We consider mobile users which are moving with speed values uniformly distributed in interval $[v_{min}, v_{max}]$ m/s. We assume a Random-Waypoint mobility with some probability of pausing. After a user pauses or hits the borders, it changes its direction of movement with an angle ~U(0, 2π).

The controller maps users to APs periodically —every T time units, and we call the number of time slots elapsed between the two consecutive mapping as *controller period*. The collected statistics may include average SNR to each visible AP for each user and user's traffic requirements. Based on the collected statistics, the controller may trigger one or a subset of APs for switching their users from one to another. Note that in IEEE 802.11ac, users have already such SNR information for each primary and secondary channels. More specifically, users measure their links in each channel to APs which operate on the whole spectrum using multiple channels as secondary channels in addition to the primary channel. We assume such a system in our paper.

We call the users whose visible AP set has more than one AP as *handover candidates* denoted by $\mathcal{U}^{ho}$. The rest of the users, i.e., $u_k \in \mathcal{U} - \mathcal{U}^{ho}$, are either under outage or have only one visible AP. For the former, there is nothing a controller could do, and in fact, this case should rarely occur under a careful coverage planning, i.e., sufficiently dense AP deployment. For the latter, the controller assigns the user to its only option.

In the next section, we present several ways the controller can decide on these handover events and discuss how to set the controller period T. The controller sends the new user-AP association decision to the APs after which users with change in their AP associations are switched to the new APs. During handover, a user experiences some outage period which may disrupt its communications. We represent the handover cost t_{sw} in terms of total time to complete association to a new AP. When controller is not active,

Table 1: Summary of Key Parameters

Symbol	Description
$\mathcal{A} = \{AP_j\}, K$	Set of access points, number of APs
$\mathcal{U} = \{u_i\}, n$	Set of users, number of users
$\mathcal{U}_j, n_j$	Set and number of users associated to AP_j
B	Bandwidth available at each AP
a_j^f	Binary variable yielding 1 if AP_j uses channel f
P_j	Transmission power of AP_j
C, F	Set and number of channels
$x_{i,j}$	Binary decision variable showing if user i is assigned to AP j
$v_{i,j}$	Binary state variable showing whether u_i is in AP_j's service region
σ	Path loss coefficient of the environment
$\gamma_{i,j}$	Received signal strength of AP_j at u_i
$d_{i,j}$	Distance of u_i from AP_j
$r_{i,j}$	Capacity of the channel between u_i and AP_j
$R_{i,j}$	Throughput of u_i if associated to AP_j
$\mathcal{R}_{i,j}$	Utility of u_i if associated to AP_j
r_i^{min}	Min. rate required for u_i's application
$\alpha_{i,j}^{min}$	Needed min. airtime for u_i if served by AP_j
$\beta_{i,j}$	Acquired airtime for u_i if served by AP_j
$\alpha_{i,j}^{sw}$	Airtime for a switching u_i if served by AP_j
$\alpha_{i,j}$	Airtime for u_i who is already associated to AP_j
T	Controller period
t_{sw}	Handover cost in terms of outage time
ϕ_i	Binary variable showing if u_i switches from one AP to another AP at the current association period

conventional *client-driven handover scheme* (CD) is in effect, i.e., when the user-AP link becomes weak such that the signal level is below the handover threshold SNR, the user associates to the AP with the highest SNR regardless of the AP load.

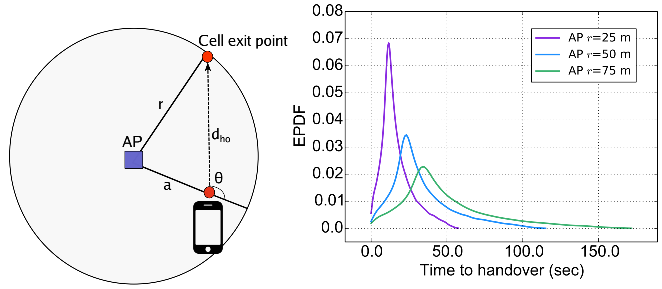
3 OPTIMAL MAPPING OF STATIONS TO APS

3.1 Airtime share under handover latency

Assume that u_i is connected to AP_s. If the user or the controller has decided that the user connects to a destination AP denoted by AP_j, the user needs to perform association steps for $AP_j \neq AP_s$. To denote the handover status of u_i, we define a binary variable ϕ_i which yields 1 if $AP_j \neq AP_s$. We assume that handover operations are performed at the beginning of a time slot.

During handover period, a user which is being switched to AP_j is in outage and hence cannot be served by AP_j. As a result, the channel (medium) will be shared only among the users who are already associated to AP_j. Let $\mathcal{U}_j$ be the set of users including the switching users served by AP_j and n_j be the number of these users, i.e., $|\mathcal{U}_j| = n_j$. Moreover, denote the number of users who are being switched to AP_j by n_j^{sw}. We assume each AP targets fairness in airtime share in the downlink, e.g., IEEE 802.11e. As a result, each user gets an equal share of the medium for the downlink.

Assuming that T is set considering the expected time the population of an AP service set would not change drastically, we calculate



(a) Expected distance d_{ho} from the cell exit point.

(b) EPDF of time to handover calculated from Monte Carlo simulations, $v \sim U(1, 5)$ m/sec.

Figure 2: Cell geometry to model expected time to handover and calculate controller period using Monte Carlo runs.

the expected channel airtime a switching user gets from AP_j for its downlink traffic as:

$$\alpha_{i,j}^{sw} = \frac{T - t_{sw}}{T} \cdot \frac{1}{n_j}. \tag{1}$$

On the other hand, airtime a user which is already associated to AP_j gets is:

$$\alpha_{i,j} = \frac{t_{sw}}{T} \cdot \frac{1}{n_j - n_j^{sw}} + \frac{T - t_{sw}}{T} \cdot \frac{1}{n_j}. \tag{2}$$

In (2), the first term of the summation shows the airtime share of each user in the handover period and the second term represents the airtime each user gets after all switching users are associated to AP_j. Notice that the second term in (2) equals to $\alpha_{i,j}^{sw}$ in (1) showing that users which are already connected to this AP gets more airtime.

3.2 Controller period T

Since users are mobile, their signal qualities evolve with time. A controller needs to track such changes and to trigger handovers promptly to maintain high user satisfaction in the network. That means, controller period T should be short enough to react to degrading user-AP links. On the other hand, we need to set T sufficiently long to avoid a large fraction of airtime, i.e., t_{sw}/T, be lost to handover overhead.

Assume that a user's distance from its AP is $a \sim U(0, r)$ and it moves with a speed $v \sim U(v_{min}, v_{max})$ toward the cell edge with an angle $\theta \sim U(0, 2\pi)$ (cf. to Fig.2a). We need to find the expected time for a randomly-picked user to reach the cell edge where the user-to-AP link quality is poor and hence a handover needs to be performed. A similar problem is investigated in the context of handovers from WLANs to 3G networks, e.g., in [15]. We can derive the expected distance from the cell exit point, the point on the cell edge where this user leaves the AP coverage, denoted by d_{ho}, using the cosine theorem: $a^2 + d_{ho}^2 - 2 \cdot a \cdot d_{ho} \cdot \cos(\pi - \theta) = r^2$. We reorganize the above equation as a quadratic univariate equation where the only unknown is d_{ho}. Then, we can find the root of the above equation as: $d_{ho} = -a \cdot \cos\theta \pm \sqrt{(a \cdot \cos\theta)^2 + (r^2 - a^2)}$. Given this distance d_{ho}, we can derive $T_{ho} = d_{ho}/v$ which denotes the expected time to span d_{ho} with speed v. We can set T using T_{ho}. In Fig.2b, we plot experimental pdf of T_{ho} values driven from Monte Carlo simulations for a cell under various coverage radius values. The user speed is uniformly distributed with (1, 5) m/sec.

Unsurprisingly, we observe in the figure that the time to handover is longer for APs with larger footprints. Consequently, we can set controller period longer for such settings. Moreover, Fig.2b shows that there can be very short duration to next handover for users who are already close to the edge of an AP's coverage. As a result, we must set controller period small to detect these handovers. However, this analysis does not provide us how much performance loss a network will experience with longer controller periods. In Section 5, we address this question by the help of simulations.

3.3 Throughput for a given user-AP mapping

We can calculate the capacity of the downlink channel between AP_j and u_i, denoted by $r_{i,j}$, as a function of the signal-to-noise-plus-interference ratio (SINR) of AP_j's signal received by u_i. More formally, SINR of an AP signal operating at channel f with bandwidth B equals to:

$$\gamma_{i,j} = \frac{P_j d_{i,j}^{-\sigma}}{B\eta_0 + \sum_{k \in \mathcal{A}_f} P_k d_{i,k}^{-\sigma}}, \tag{3}$$

where P_j denotes the transmission power of AP_j, σ is the path loss coefficient, and η_0 is noise power per unitary bandwidth. SNIR of a link is a function of the channel assignment decision which is reflected in $\mathcal{A}_f$—the set of APs assigned to channel f. Corresponding capacity of the channel with bandwidth B units is then:

$$r_{i,j} = B \log(1 + \gamma_{i,j}) \text{ bits per second.} \tag{4}$$

If u_i is associated to AP_j, its downlink throughput is a function of $\gamma_{i,j}$ and the airtime it will get from AP_j.[2] As we assumed orthogonal channel assignment, there is only one AP operating at a particular WiFi channel in a collision domain. In other words, co-channel APs are outside their carrier sensing range and resulting interference is insignificant, i.e. below noise floor. Hence, each AP utilizes all the airtime itself without sharing it with other APs.

We set ϕ_i to 1 if u_i handovers. Then, we can calculate the expected throughput of u_i from AP_j, denoted by $R_{i,j}$, as:

$$R_{i,j} = r_{i,j}(\alpha_{i,j}^{sw}\phi_i + \alpha_{i,j}(1-\phi_i)) \text{ bits,} \tag{5}$$

where the term $(\alpha_{i,j}^{sw}\phi_i + \alpha_{i,j}(1-\phi_i))$ represents the airtime a user gets depending on whether it is switched to AP_j or not.

Although an AP allocates its airtime equally among its users, the actual airtime a user needs may differ across users. For instance, a user browsing the web would need less airtime compared to another having a video conference. We denote the airtime need of a user as $\alpha_{i,j}^{min}$ and airtime acquired by it as $\beta_{i,j}$. We calculate the minimum needed airtime for u_i from AP_j with rate requirement r_i^{min} as:

$$\alpha_{i,j}^{min} = \frac{r_i^{min}}{r_{i,j}}. \tag{6}$$

[2]We user Shannon's capacity formula to calculate the rate of this user-AP link. However, in reality the actual rate depends also on the selected modulation and coding scheme (MCS) according to the channel quality.

Note that satisfying $\beta_{i,j} \geqslant \alpha_{i,j}^{min}$ inequality is essential for some applications such as video communications whereas for best-effort traffic it is not a stringent requirement. In fact, we set $r^{min} = 0$ for best-effort traffic. If the throughput a user gets from its AP is at least equal to the requested minimum throughput, we call such user a *satisfied user* and define its utility as a function of its throughput. For unsatisfied users, the utility is zero as the user cannot get the bare minimum for a pleasant user experience. To reflect the two goals of our controller, i.e., high throughput efficiency and fairness among users, we define the utility of a user as its logarithmic throughput. More formally, utility of a user u_i connected to AP_j is defined as:

$$\mathcal{R}_{i,j} = \begin{cases} \log(1 + R_{i,j}), & \text{if } \beta_{i,j} \geqslant \alpha_{i,j}^{min} \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

3.4 Problem formulation

Let $x_{i,j}$ denote the binary decision variable yielding value 1 if the controller assigns user u_i to AP_j. We formulate centralized optimal user-AP assignment problem as follows:

$$\textbf{P1}: \max_{\mathrm{X}=[x_{i,j}]} \sum_{AP_j \in \mathcal{A}} \sum_{u_i \in \mathcal{U}} \log\left(1 + x_{i,j} r_{i,j}(\alpha_{i,j}^{sw}\phi_i + \alpha_{i,j}(1-\phi_i))\right) \tag{8}$$

$$\sum_{AP_j \in \mathcal{A}} x_{i,j} \leqslant 1 \qquad \forall u_i \in \mathcal{U} \tag{9}$$

$$x_{i,j} \leqslant v_{i,j} \qquad \forall u_i \in \mathcal{U}, \forall \mathbf{AP}_j \in \mathcal{A} \tag{10}$$

$$\sum_{u_i \in \mathcal{U}} x_{i,j}\alpha_{i,j}^{min} \leqslant 1 \qquad \forall \mathbf{AP}_j \in \mathcal{A} \tag{11}$$

$$x_{i,j}\alpha_{i,j}^{min} \leqslant \alpha_{i,j}^{sw}\phi_i + \alpha_{i,j}(1-\phi_i), \forall u_i \in \mathcal{U}, \forall \mathbf{AP}_j \in \mathcal{A} \tag{12}$$

$$\alpha_{i,j}^{sw} = \frac{T - t_{sw}}{T \cdot \sum_{u_i \in \mathcal{U}} x_{i,j}} \tag{13}$$

$$\alpha_{i,j} = \frac{t_{sw}}{T \cdot \sum_{u_i \in \mathcal{U}} x_{i,j}(1-\phi_i)} + \frac{T - t_{sw}}{T \cdot \sum_{u_i \in \mathcal{U}} x_{i,j}} \tag{14}$$

$$x_{i,j} \in \{0, 1\} \qquad \forall u_i \in \mathcal{U}, \forall \mathbf{AP}_j \in \mathcal{A}. \tag{15}$$

The objective function in (8) states that users must be associated to the APs that result in the highest network utility which is a function of logarithmic throughput maintained by each user. Const. (9) signifies that each user can be mapped to at most one AP, whereas Const. (10) is necessary for a feasible assignment, i.e., a user can only be associated to a visible AP. Const. (11) states that the minimum airtime demand of associated users cannot exceed the capacity of an AP, i.e., 100% airtime, whereas Const. (12) ensures that the airtime share this user will get from an AP is higher than its minimum bandwidth requirement. Eqns. (13) and (14) formally state the airtime a user gets if it is a switching user or otherwise, respectively. Finally, Const. (15) denotes that each assignment variable is a binary variable.

Note that the controller solves **P1** every T time units and sends to all APs only the changes in user-AP associations; entries of $X = [x_{i,j}]$ where $x_{i,j} = 1$ and different from the previous assignment. Complexity of **P1** depends on the number of users in the cell edges with overlapping AP coverages and the number of APs visible to each such user. More particularly, it increases exponentially with number of such users: given that the number of users in these cell regions is n in the worst case and the number of APs each user can get service is K in the worst case, complexity is $O(K^n)$. The high computational complexity of **P1** renders it impractical for practical operation. Hence, we design lower complexity heuristics in the next section.

4 LOW-COMPLEXITY USER-AP ASSOCIATION SCHEMES

Two design goals for our heuristics are as follows: (i) minimum throughput requirements of the users must be satisfied, and if not, the utility of an unsatisfied user is zero, and (ii) a heuristic should have polynomial complexity in number of users and number of APs. For all schemes, the controller first identifies handover candidates ($\mathcal{U}^{ho}$) and assigns the rest to their only visible AP, if any.

4.1 Highest-SNR AP association (h-SNR)

A simple yet efficient handover scheme a controller can implement is to assign each user to the AP providing the highest signal strength. Note that the conventional *client-driven handover* (CD) has the same approach but with a difference that handover is triggered only after the user's AP can not provide the minimum required signal level for a reliable link. In h-SNR, the user does not stick to its AP but instead switches to the AP with the highest signal level. With h-SNR, we can assess the benefit of periodic handover management in comparison to client-driven handovers. Note that h-SNR is handover-cost oblivious and does not consider minimum rate requirements.

4.2 Airtime-aware AP association (AIR)

While a high SNR value ensures high link rate, it overlooks the time-sharing nature of WiFi. A user's throughput is a multiplication of its link rate and how much airtime it receives. Hence, to account for both parameters, we design airtime-aware AP association (AIR). More specifically, this scheme first calculates rates for each user-AP link as in (4). To have some notion of fairness, AIR starts with a randomly-picked user and assigns this user the AP which promises highest (air-time × rate). After each assignment, we update the airtime a user can get from each AP by considering the new number of associated users for each AP. Note that AIR is handover-cost aware, but it does not consider the minimum rate requirements.

4.3 Demand-aware AP association (DAW)

While airtime-aware AP association calculates expected throughput, it does not consider how a new association affects the performance of existing users of the AP. Given that some users require minimum rate to have a satisfactory quality of experience, demand-aware AP association (DAW) avoids violating the minimum rate requirements.

Similar to airtime-aware scheme, DAW first calculates (air-time× rate) for each user-AP pair as in (4). Different than the previous scheme, DAW checks if an AP has *spare* airtime. Given that an AP allocates its airtime equally among its users, the number of users it can serve is limited by the minimum rate requirements. Let $\mathcal{U}_j$ denote the set of users served by AP_j and total airtime demand from users of this AP as $\alpha_j^{min} = \sum_{u_i \in \mathcal{U}_j} \alpha_{i,j}^{min}$. While an AP with $\alpha_j^{min} = 0$ can in theory serve unlimited number of users, an AP with $\alpha_j^{min} > 0$ can serve at most n_j^{max} users where n_j^{max} is:

$$n_j^{max} = \lfloor \frac{1}{\max\limits_{u_i \in \mathcal{U}_j} \{\alpha_{i,j}^{min}\}} \rfloor. \tag{16}$$

Then, spare airtime s_j of this AP that can be allocated to a newly-joining user equals to:

$$s_j = \begin{cases} \frac{T - t_{sw}}{T(n_j+1)}, & \text{if } n_j < n_j^{max} \\ 0, & \text{otherwise.} \end{cases} \tag{17}$$

Let $\mathcal{A}^+$ be the set of APs that can accommodate newly joining users without violating the minimum rate requirements of the existing users, i.e., $\mathcal{A}^+ := \bigcup AP_j$ where $s_j > 0$ for $AP_j \in \mathcal{A}$. Our aim is to switch handover candidates to such APs in $\mathcal{A}^+$.

For a candidate AP, e.g., AP_j, we need to also consider the decrease in aggregate throughput of users in $\mathcal{U}_j$ after a new user joins. Since the number of users increases by 1 in AP_j's service region, users of AP_j will have less airtime and therefore will sustain lower throughput. The decrease in throughput results in a decrease in utility. Since our aim is to maximize the utility in (8), we consider the decrease in utility as follows:

$$\Delta u_{i,+j}^{-} = \sum_{u_k \in \mathcal{U}_j} \log\left(1 + \frac{r_{k,j}}{n_j}\right) - \sum_{u_k \in \mathcal{U}_j \setminus u_i} \log\left(1 + r_{k,j}\left(\frac{t_{sw}}{T n_j} + \frac{T - t_{sw}}{T(n_j+1)}\right)\right). \tag{18}$$

In (18), the first term represents the aggregate utility of AP_j before u_i associates to this AP whereas the second term is the updated utility (excluding u_i) after u_i joins. Then, considering u_i's expected utility, we calculate the net utility of assigning user i to AP_j as:

$$\Delta u_{i,j} = \log\left(1 + \frac{r_{i,j}(T - t_{sw})}{T(n_j+1)}\right) - \Delta u_{i,+j}^{-}$$

Then, we take the pair user i^* and AP j^* achieving the highest net utility: $i^*, j^* = \arg\max \Delta u_{i,j}$. We update $\mathcal{U}^{ho}$ by removing the assigned user u_{i*} from the set of handover candidates. To avoid violation of minimum rate requirements, we update the spare capacity of each AP in $\mathcal{A}^+$ using (16) and (17). Next, we remove those APs with zero spare capacity from $\mathcal{A}^+$. DAW terminates when $\mathcal{A}^+$ or $\mathcal{U}^{ho}$ equals to empty set.

4.4 Comparison of heuristics

In Table 2, we compare the proposed heuristics and CD, according to their awareness in terms of AP load, handover cost (`HO cost` column), and traffic demands (`demand` column). We also denote if these schemes can be implemented in a distributed manner (`distributed` column). For CD and h-SNR, each client can decide by itself as these algorithms need only the user-received SNR from each AP which is already available at each user. On the contrary, more information on the whole network is required for AIR and DAW. For AIR, a controller needs to know the number of users at each AP and their handover status to calculate the airtime the user will get. In addition to this information, DAW requires the user-AP link rates for all users to make mapping decisions. The complexity of h-SNR, CD, and AIR equals to $O(nK)$ as for each client these algorithms iterate over all APs to find the AP with the highest utility, i.e., signal level or airtime×rate. DAW is of complexity $O(n^2 K)$ as it iterates over all user-AP pairs to find the best mapping in each step.

Table 2: Comparison of Heuristics

Heuristic	AP load	HO cost	Demand	Distributed
CD	-	-	-	✓
h-SNR	-	-	-	✓
AIR	✓	✓	-	-
DAW	✓	✓	✓	-

5 PERFORMANCE EVALUATION

5.1 Performance metrics

We use the following performance metrics in our analysis.

- *Utility* is the objective function defined in (8).
- *Fraction of satisfied users* is the ratio of the users whose minimum required rate is satisfied over all users.
- *Fairness of user throughputs* is the Jain's fairness index considering the throughput distribution over all users.
- *Load balance across APs* is the Jain's fairness index considering the number of users served by each AP.
- *Probability of handover* is calculated as the ratio of total number of handovers over all connected users at a time slot.
- *Gain in weakest user's throughput* is the improvement achieved by a heuristic in the minimum user throughput over CD.

5.2 Parameters and scenarios

Our scenarios are similar to indoor scenario defined in IMT guidelines [12], e.g., indoor environments isolated from external interference and consisting of stationary or low-mobility pedestrians. More specifically, we define the following three scenarios (cf. to Table 3). For all scenarios, we set $n = 80$ and $K = 10$.

- **Conference scenario:** In this scenario, we model a conference environment in which a room with three APs hosts a large number of users. For the initial placement of users, 90% of the users are located in this room. Outside the conference hall, there are 7 APs deployed in a grid-like topology to serve the remaining 10% of the users.
- **Office scenario:** In this scenario, we model an office setting with a grid-like topology, e.g., we deploy APs on a grid and change the AP locations in a small radius to account for building imperfections. Users are deployed based on a skewed Pareto distribution which diverges from uniform deployment depending on the skewness parameter. Only a small fraction of users are mobile.
- **Shopping mall scenario**: In this scenario, we model WiFi usage in a shopping mall. APs are deployed in a grid topology and users are uniformly placed in the area. A large fraction of users are mobile while only a small fraction has minimum bandwidth requirements.

Notice that although all scenarios have the same number of APs and users, AP deployment and user deployment differ across

Table 3: Scenarios

Scenario	Fraction of mobile users	Fraction of users with throughput demand
Conference	0.5	0.3
Office	0.3	0.5
Shopping mall	0.9	0.3

scenarios. To quantify the resulting imbalance in user distribution in the considered area, we define a metric called *density balance* which reflects the homogeneity in user distribution. More specifically, on each time slot, for each AP, we record the number of users for whom this AP is the nearest AP and is expected to provide the highest SNR. Then, we calculate the Jain's fairness index considering the number of users of each AP. Intuitively, for a grid AP topology and uniform deployment of users, the density balance is close to 1. Unless otherwise stated, we will use the settings listed in Table 3.

For a fair comparison across different settings, we assume that the total available bandwidth is the same in all settings, i.e., B_{tot} = 100 MHz. Based on the network topology, we first find the minimum number of channels needed for an orthogonal frequency assignment by solving a graph coloring problem on our AP topology. Then, we find the bandwidth per AP by dividing the available bandwidth to the required number of channels, i.e., $B = B_{tot}/\chi$ where χ is the chromatic number of the AP graph.

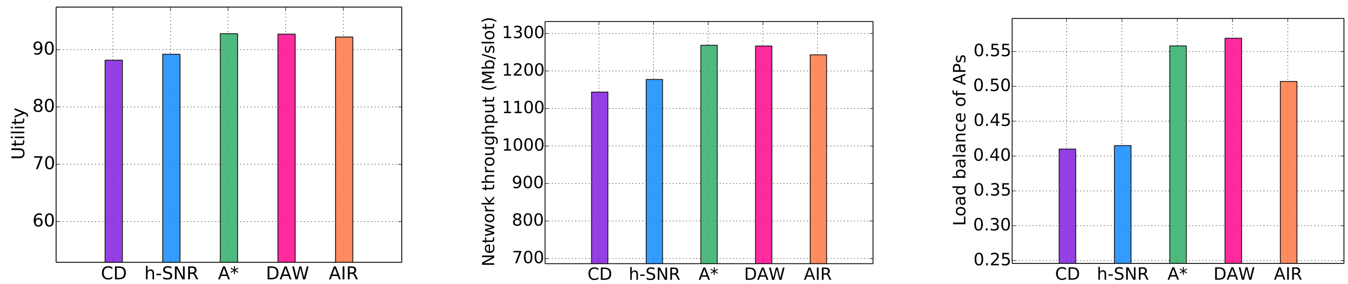
We set $t_{sw} = 0.2$ s while a time slot = 1 s, minimum rate requirements $\approx [5, 15]$ Mbps considering the video rates reported by Skype and Netflix.[3] All scenarios are assumed to cover an area of [150 m, 100 m]. For the conference case, the conference hall is located at the centre of the area and its size is [50 m, 30 m]. Mobility model is random-waypoint mobility with 0.2 pausing probability at each time slot and user speed $\approx [1, 5]$ m/s. [4] We model the user-AP links using Keenan-Motley model. While we calculate fairness of user throughputs, we consider the throughput accumulated in a time window of 5 slots, rather than instantaneous throughput value in that particular time slot. We report the average results of 100 repetitions for each scenario and a simulation time of 100 time slots.

5.3 Optimality gap

To understand the performance gap between our heuristics and the optimal solution, we find the optimal user-AP association by running A* search [13] considering utility in (8). Due to the high computation time, we show the performance for a smaller conference setting; 6 APs and 15 users in an area of 120m x 80m. Note that we expect a wider optimality gap for a larger scenario with many more users and APs. Fig. 3 depicts the utility, aggregate throughput, and load balance of the network. Regarding utility in Fig. 3a, the optimality gap of DAW is 0.1%, i.e., it achieves 0.1 percent less utility than the optimal solution. To give a perspective, CD has optimality gap of 5% while h-SNR and AIR has 3.9% and 0.6% lower utility than

[3]https://support.skype.com/en/faq/FA1417/how-much-bandwidth-does-skype-need, and https://help.netflix.com/en/node/306.

[4]Note that this interval covers also higher speeds than typical walking speeds. However, we chose these values for practical purpose of having more handover events in the given simulation period.



(a) Utility.

(b) Aggregate throughput.

(c) Load balance.

Figure 3: Comparison of heuristics with the optimal solution A^* for the conference setting with $n = 15$ and $K = 6$.

the optimal scheme, respectively. Similarly, Fig. 3b compares the network throughput. CD maintains 9.9% lower throughput than A* while those of h-SNR and AIR are 7.2% and 2% lower, respectively. DAW's performance approaches to that of optimal (0.2% lower) in terms of achieved aggregate throughput. Finally, as Fig. 3c shows, DAW outperforms A* solution only slightly (0.2%) while CD and h-SNR has around 26% and AIR 9% worse performance than A*.

5.4 Impact of controller period

First, we analyze the impact of controller period across different scenarios by setting $T = \{1, 2, 4, 10, 20\}$ time slots. Figs. 4, 5, and 6 show the impact of increasing controller period for conference, office, and mall scenarios, respectively.

Let us first focus on the case when $T = 1$. In this case, the highest gain in utility over CD is achieved in the conference scenario where we have the lowest density balance. The average density balance for each scenario is 0.33 (conference), 0.76 (office), and 0.95 (mall) while the corresponding gain in utility is up to 18%, 5%, and 2%. Note that while utility improvement values are moderate to low, the resulting gain on the minimum rate of the stations is up to 120%, 73%, and 71%. Regarding h-SNR, while it achieves the highest total network throughput (not plotted), its lack of fairness notion results in a much lower utility. Thus, DAW and AIR both outperform h-SNR in terms of utility, throughput fairness, and thereby gain in the weakest user's throughput. In addition, h-SNR is oblivious to handover cost and to traffic demands, which results in lower fraction of satisfied users around 0.92 (figure not depicted) compared to DAW and h-SNR around 0.98. However, since conference setting has only a small number of users with minimum rate requirements, we do not observe a significant difference in terms of satisfied users, e.g., satisfaction ratio is 0.91 for CD. DAW and AIR perform very similarly despite the fact that AIR neglects the traffic requirements. We attribute this behaviour to the available network resources, e.g., the capacity is not tight. However, with increasing user density with video traffic, we expect to see the superior performance of DAW over AIR.

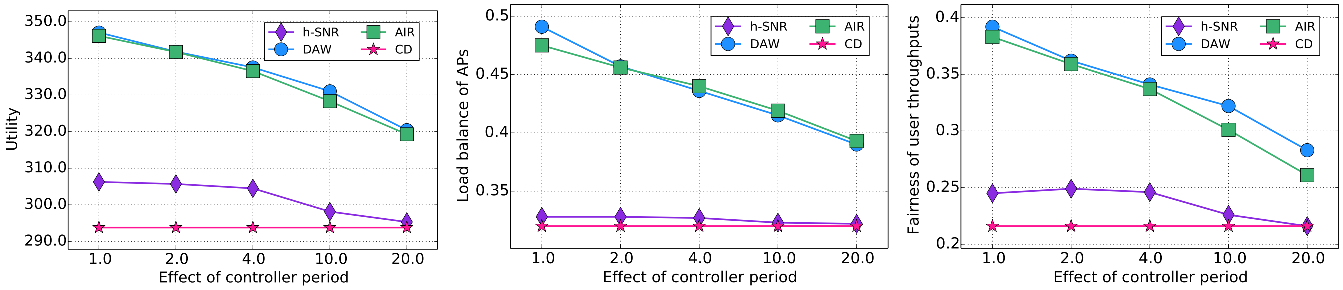
With increasing T, the impact of controller decisions become less significant and eventually performance of all heuristics are expected to approach that of CD. Please recall that when controller is not active, the legacy CD approach at each client is in charge of AP association. Moreover, with longer T, the expected throughput calculated using (1) and (2) deviates from the actual user throughput due to variations in user location and channel states. We observe a steeper decrease in performance for the conference scenario comparing Figs.4a, 5a, and 6a. Primary reason for this trend is again the imbalance of AP loads. Note that although mall scenario has the highest mobility, the observed decrease in utility with increasing T is less pronounced due to the high density balance of the network. The average number of visible APs of a user is 3.47 APs for conference, 4.20 APs for office, and 3.89 APs for the mall setting. When controller period is longer, opportunities of more efficient user-AP association are lost. On the other hand, for the office and mall scenarios, almost all APs have the same number of users and they are deployed on a grid, there is less room for optimizations. Notice almost perfect AP load balance and throughput fairness in these scenarios depicted in Figs.5b, 6b, 5c, and 6c.

To understand whether our heuristics result in frequent handovers, we plot the probability of handover for each scenario in Fig.7. While our heuristics result in higher switching probability compared to CD and h-SNR, the handovers are yet very low —less than 0.10 probability. We attribute this behaviour to the low mobility of the considered settings. After a closer look to CD curves, we can observe the sticky user problem, e.g., very low probability of handover if clients decide on AP association.

In short, the controller should consider the expected density balance to set the period appropriately without sacrificing from the performance significantly. Another observation is that even a naïve heuristic like h-SNR which does not require any global knowledge can improve utility significantly if performed periodically rather than only when the station is about to lose its connectivity. In fact, h-SNR could be a good option as it can run in a distributed manner as opposed to AIR and DAW (as shown in Table 2).

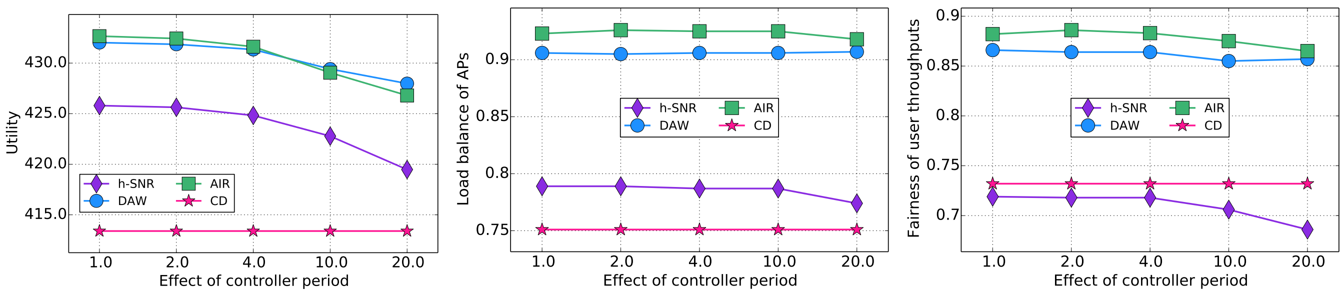
5.5 Impact of handover cost

To see the impact of handover cost more clearly, we set the fraction of mobile users to 1 for the conference setting. Fig. 8 depicts the impact of increasing handover cost on each user-AP association scheme. As we consider scenarios with low mobility, the impact of handover on the overall utility is only marginal for all schemes. Moreover, this slight impact is due to the increased airtime for users who are already associated to their APs. Recall that as shown in (2), airtime lost by switching stations are used for stations that



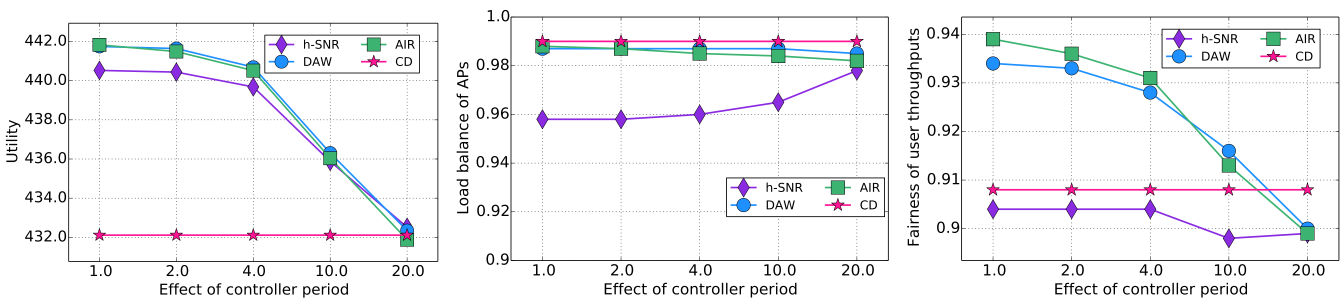
(a) Utility. (b) Load balance across APs. (c) Fairness of user throughputs.

Figure 4: Impact of controller period for conference scenario where the mean density balance is 0.33.



(a) Utility. (b) Load balance across APs. (c) Fairness of user throughputs.

Figure 5: Impact of controller period for office scenario where the mean density balance is 0.76.



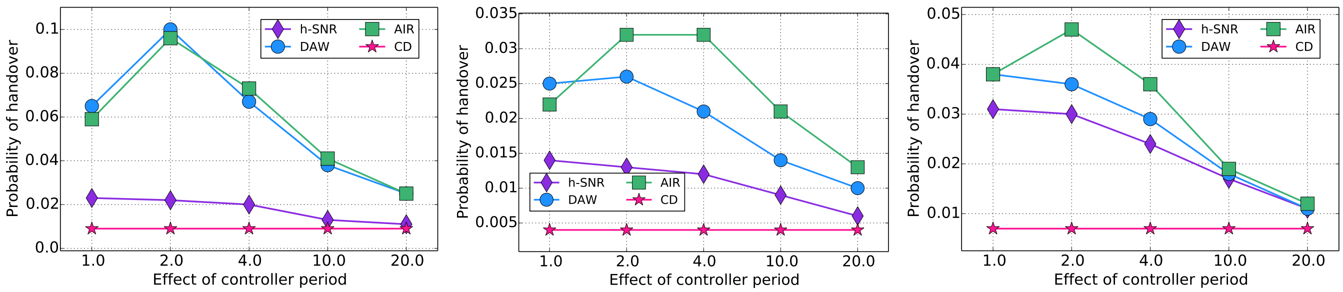
(a) Utility. (b) Load balance across APs (c) Fairness of user throughputs.

Figure 6: Impact of controller period for mall scenario where the mean density balance is 0.95.

are already connected. In that respect, airtime of the AP is conserved. On the other hand, decrease in the gain of the weakest user's throughput in Fig.8b becomes more visible with increasing handover latency. For example, for h-SNR, the gain drops from 48% to 26% while that of AIR is from 140% to 114% in the considered cost range. Additionally, switching users may experience low satisfaction if their allocated airtime is insufficient to provide the required throughput for their applications. Fig. 8c shows that handover-cost aware schemes avoid changes in user-AP mappings with increasing handover cost.

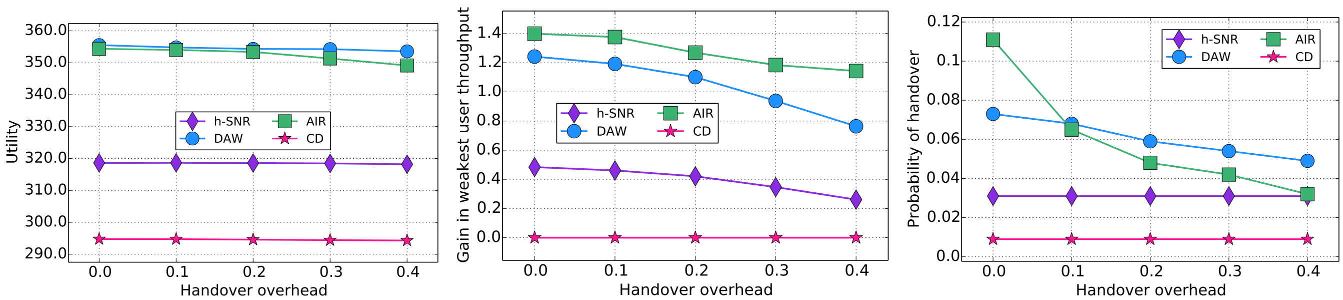
5.6 Impact of user density

Fig. 9 shows the impact of increasing user density for the conference scenario. For low user density, all schemes including CD can meet the users' expectations as reflected by 100% satisfaction in Fig.9b. However, with increasing number of users, we observe how naive approaches fall short of providing user satisfaction while DAW and AIR can still maintain a high fraction of satisfied users. For example, for $n = 130$, satisfaction ratio is 0.78 for CD and 0.80 for h-SNR whereas it is 0.86 and 0.85 for DAW and AIR, respectively. Hence, we argue that for dense networks with many users implementing smart controller schemes becomes more paramount compared to



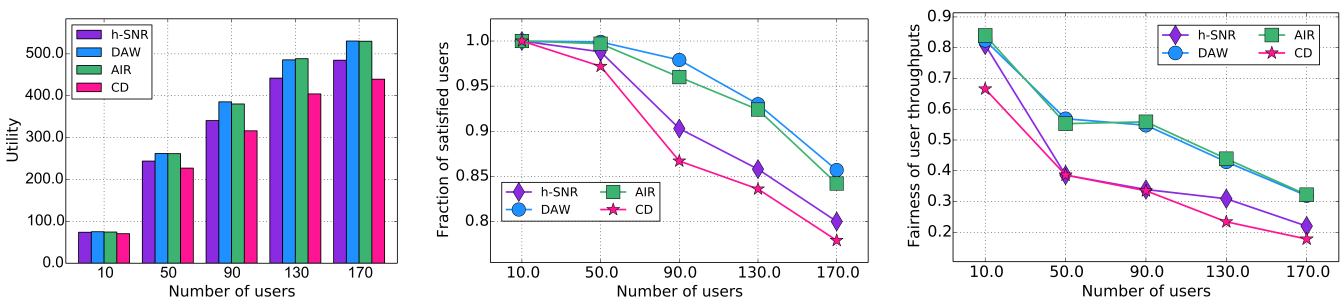
(a) Conference. (b) Office. (c) Shopping mall.

Figure 7: Probability of user handover with increasing controller period for $n = 80$.



(a) Utility. (b) Gain in the weakest user throughput. (c) Probability of handover.

Figure 8: Impact of increasing handover latency for conference scenario.



(a) Utility. (b) Fraction of satisfied users. (c) AP load balance.

Figure 9: Impact of increasing number of users for conference scenario.

sparse networks. Recall that only a small fraction, i.e., 0.3, has minimum rate requirements. For scenarios where there are more users with minimum rate requirements, we expect the performance gap between smart and naive schemes as well as between DAW and AIR to be more visible. For scenarios with only a few high-demand users, AIR provides a good balance between complexity and performance as it requires less knowledge compared to DAW.

6 RELATED WORK

Several works [2, 7, 9] have designed various optimal user-AP association schemes using different objectives and under different assumptions. For example, Amer et al. [2] assume that all users associate to an AP will get the same throughput due to the packet-level fairness of 802.11 WLAN and they all have the same traffic demands, whereas Karimi et al. [9] consider a scenario where a user might have explicit time limitations for using an AP's resources. Both [2] and [9] take the logarithmic network-wide downlink throughput as maximization objective to achieve a tradeoff between aggregated throughput and fairness among users. Authors of [8] define a MAC efficiency metric to account for both uplink and downlink throughput of a user and focus on the heterogeneity of the clients, i.e., IEEE 802.11a/b/g/n, while deciding on the optimal users-AP associations in an IEEE 802.11n network. For multiple WiFi deployments, [3] argues that cooperation among different WiFi networks can increase the throughput of the weakest user by alleviating the inter-network interference.

Although optimal user-AP association has been widely studied in the literature [1, 8, 9], when to trigger association control other than upon a new user joining the network is yet to be understood. A user-AP association decision which is optimal at the time of a user joining the network might later become suboptimal due to several reasons, e.g., mobility of the user or new traffic arrival to the user's AP. Amer et al. [2] propose to solve the formulated optimization problem periodically without discussing how long this period should be, whether there is a need for such periodic changes in the user-AP association, and most importantly, how a handover as a result of new association decision affects ongoing communications. Similarly, Karimi et al. [9] envisage that users may change deployment parameters, which then triggers the controller, i.e., the upstream WLAN provider in [9], to run its optimal association algorithm. Different than these works, we provide a thorough analysis of how a range of WLAN scenarios can be affected by the increase in the period of user-AP mapping algorithm.

The key reason for avoiding the periodic or frequent changes in user-AP association is the long duration of switching period due to the set of control messages between the user and the old and new APs for connection breakdown and re-association. To decrease the overhead, [10] virtualizes the wireless NIC as if the user is connected to multiple APs simultaneously. BigAP proposed in [17] exploits the Dynamic Frequency Selection functionality of IEEE 802.11n/ac cards, as if a radar signal is detected, to trigger the handover of some clients on congested/highly-loaded APs to less-loaded APs. Handover is seamless in BigAP as all APs are assigned the same BSSID. Our work can be considered as a complementary solution to [17] in that user-AP mapping decisions are applied by such a mechanism in [17] with very smooth handover.

Another line of related research is on load balancing among APs [4, 10, 14, 16]. In cases where a group of WLAN clients are compactly packed in the coverage of an AP, clients all connect to the same nearby AP. To tackle with the a potential load imbalance, [4] proposes to dynamically adjust the signal power of AP beacons in a way similar to cell breathing in cellular networks. While our proposals do not directly optimize load balance, schemes with AP-load awareness achieve a drastic improvement in load balance compared with the client-driven AP association scheme.

As a WLAN's load balance changes with incoming and leaving users, understanding the nature of these events is crucial. Analysis in [14] on the traces of a campus WLAN shows a high correlation between some users in their AP association history and uplink/downlink traffic volume. Exploiting the social relation among these users, [14] proposes to map such users to different APs to avoid sudden changes in AP loads and overloading of some APs. We plan to extend our analysis with more realistic group mobility models considering the results of [14].

7 CONCLUSION

Motivated by the need for more efficient enterprise WLANs, we have formulated optimal user-AP association as proportionally-fair sum throughput maximization problem solved by a central controller periodically. Different than the existing user-AP association proposals, our proposal is handover-cost aware and considers the minimum throughput requirements (e.g., video or best-effort traffic) of each user while assigning users to APs. Due to the hardness of the optimal solution, we provide several sub-optimal yet efficient heuristics whose achieved utility is slightly lower than the optimal solution. Via simulations, we assess the performance of our proposals for realistic scenarios, e.g., a conference environment. Our results prove the superiority of periodic user-AP association control over the conventional client-driven AP association which is only triggered when the user joins a network or is about to lose connectivity to its serving AP due to user mobility. Generally speaking, our proposals achieve the highest gain for scenarios where users are unevenly distributed in the network.

As future work, we plan to consider also the uplink traffic which has been increasing with wide usage of cloud storage as well as video streaming applications. Another possible direction is to explore how mobility affects the performance of each scheme.

REFERENCES

[1] Murad Abusubaih and Adam Wolisz. 2007. An optimal station association policy for multi-rate IEEE 802.11 wireless LANs. In ACM MsWiM. ACM, 117–123.

[2] Mohammed Amer, Anthony Busson, and Isabelle Guérin Lassous. 2016. Association Optimization in WiFi Networks: Use of an Access-based Fairness. In ACM MsWiM. 119–126.

[3] Akash Baid, Michael Schapira, Ivan Seskar, Jennifer Rexford, and Dipankar Raychaudhuri. 2012. Network cooperation for client-AP association optimization. In IEEE WiOpt. 431–436.

[4] Yigal Bejerano and Seung-Jae Han. 2009. Cell breathing techniques for load balancing in wireless LANs. IEEE Trans. on Mobile Comp. 8, 6 (2009), 735–49.

[5] Surachai Chieochan, Ekram Hossain, and Jeffrey Diamond. 2010. Channel assignment schemes for infrastructure-based 802.11 WLANs: A survey. IEEE Communications Surveys & Tutorials 12, 1 (2010), 124–136.

[6] Cisco. 2017. Channel Planning Best Practices. https://documentation.meraki.com/MR/WiFi_Basics_and_Best_Practices/Channel_Planning_Best_Practices. (2017).

[7] Sourav Kumar Dandapat, Bivas Mitra, Romit Roy Choudhury, and Niloy Ganguly. 2012. Smart association control in wireless mobile environment using max-flow. IEEE Trans. on Network and Service Management 9, 1 (2012), 73–86.

[8] Dawei Gong and Yuanyuan Yang. 2014. On-line AP association algorithms for 802.11n WLANs with heterogeneous clients. *IEEE Trans. Comput.* 63, 11 (2014), 2772–86.

[9] O. B. Karimi, J. Liu, and J. Rexford. 2014. Optimal collaborative AP association in wireless networks. In IEEE INFOCOM.

[10] Masahiro Kawada, Morihiko Tamai, and Keiichi Yasumoto. 2013. A trigger-based dynamic load balancing method for WLANs using virtualized network interfaces. In IEEE Wireless Comms. and Netw. Conference (WCNC). 1091–96.

[11] Nicolas Montavont, Alberto Blanc, Renzo Navas, Tanguy Kerdoncuff, and German Castignani. 2015. Handover triggering in IEEE 802.11 Networks. In IEEE Symp. on a World of Wireless, Mobile and Multimedia Netws. (WoWMoM).

[12] ITUR Resolution. 2009. Guidelines for evaluation of radio interface technologies for IMT-Advanced. (2009).

[13] Thomas Weise. 2009. Global Optimization Algorithms– Theory and Application.

[14] Guangtao Xue, Qi He, Hongzi Zhu, Tian He, and Yunhuai Liu. 2013. Sociality-aware access point selection in enterprise wireless LANs. IEEE Trans. on Parallel and Distributed Systems 24, 10 (2013), 2069–78.

[15] Xiaohuan Yan, Nallasamy Mani, and Y Ahmet Sekercioglu. 2008. A traveling distance prediction based method to minimize unnecessary handovers from cellular networks to WLANs. IEEE Communications Letters 12, 1 (2008), 14–16.

[16] Li-Hsing Yen, Tse-Tsung Yeh, and Kuang-Hui Chi. 2009. Load balancing in IEEE 802.11 networks. IEEE Internet Computing 13, 1 (2009), 56–64.

[17] Anatolij Zubow, Sven Zehl, and Adam Wolisz. 2016. BigAP–Seamless Handover in High Performance Enterprise IEEE 802.11 Networks. In IEEE/IFIP Network Operations and Management Symposium (IEEE NOMS).