Demand-aware Load Balancing in Wireless LANs Using Association Control

Liu Yang* †, Yong Cui*, Heyi Tang* Shihan Xiao*
*Department of Computer Science and Technology, Tsinghua University, Beijing, PR China
Email: 09232022@bjtu.edu.cn, cuiyong@tsinghua.edu.cn, tangheyi.09@gmail.com, xiaoshihan.xsh@gmail.com
†State Key Laboratory of Networking and Switching Technology, BUPT, Beijing, PR China

***Abstract*—The densely deployed Access Points (APs) have overlapping coverage areas. In the hotspot area, a user can usually receive signals of more than ten APs. The signal-based association in IEEE 802.11 may result in significant unbalanced loads among APs. Moreover, diverse user demands on bandwidth further exacerbate the load unbalance. Some APs are too overloaded when multiple high-demand users gather on them due to the strongest signal, while the others are light-loaded with a few low-demand users associated to them. The severe load unbalance degrades the user performance. In this paper, by adding different user bandwidth demands as new constraints, we formulate the joint AP association and bandwidth allocation problem. We comprehensively analyze the solution space of optimal bandwidth allocation while satisfying the time-based fairness among users. We develop a 1/2-approximation algorithm to solve the problem. Our extensive trace-driven evaluations show that our algorithm achieves better load balance. As a result, it greatly improves the aggregated throughput and provides better user fairness than conventional association schemes.**

I. Introduction

Recently, to meet the growing demands on Internet access of users, the widely deployed Wireless Local Area Networks (WLANs) based on 802.11 protocol are constantly developing. The users usually connect to the AP with the best Received Signal Strength Indicator (RSSI). Due to the nonuniform distribution of the users in an area, this strategy will lead to load unbalance.

To address this problem, many researchers focus on designing the optimized association schemes based on some specific objectives. Some researchers introduce local-view strategies [1]–[7], while some others consider parameters of the whole network and give global-view association strategies [8]–[13]. Another important issur is the fair bandwidth allocation of AP to users. Many bandwidth allocation strategies are proposed to achieve proportional fairness based on the users'priority [14]–[17]. Some others [18]–[20] try to achieve the max-min fairness among the users. [1]–[7]. Xiaohui et al. [7] investigate the AP selection problem using game theory with local information.

Existing AP association and bandwidth allocation strategies are proposed based on the assumption that every user has a huge bandwidth demand. It means the user's demand cannot be satisfied, no matter how to allocate the bandwidth of AP to the users. Then these strategies allocate bandwidth equal or proportional to the transmission rate of users and assume the users can run out their allocated bandwidths. However, it is common that the bandwidth demand of each user differs vastly due to their different preferences on network applications. The allocated bandwidth may be wasted if users just enjoy surfing the web-sites; while it is not enough, if users watch videos. Therefore, existing allocation strategies without consideration of the user bandwidth demand are not optimal. This bandwidth demand is mentioned in [21], but not analyzed in detail.

To address this challenge, we introduce the bandwidth demand of users as a new constraint to formulate the Joint AP Association and bandwidth Allocation (JAA) problem. The consideration of bandwidth demand fundamentally changes the nature of AP association and bandwidth allocation problems, and essentially calls for a separate study. Inspired by existing work [14], [15], [17], in which the number of associated users is used to measure the AP load, we consider the joint effect of user number and each user's demand. In the single AP case, the conventional schemes that allocate bandwidth equal or proportional to the transmission rate may result in bandwidth waste. Further, for the multi-AP case, we formulate the JAA problem as a mixed-integer nonlinear programming problem. The maximum aggregated bandwidth utility is used as an object to balance the throughput gain and user fairness.

To solve JAA, we give an approximate Maximum Aggregated Bandwidth Utility (MABU) algorithm to achieve load balance among APs with the estimation of AP utilization. With a detailed analysis on the properties of the optimal bandwidth allocation, we design a fair bandwidth allocation strategy for APs. We prove that the approximation factor of MABU is 1/2 under the imported settings. Our performance evaluations based on the real data traces demonstrate that MABU increases the user's throughput by up to 23.1% and achieves better user fairness compared with conventional schemes.

The rest of the paper is organized as follows. We formulate the JAA problem in section II and propose algorithm MABU to solve it in section III. Finally, we present the performance evaluations in section IV and conclude the paper in section V.

II. Problem Formulation

A. Application Scenario

We first give an example to illustrate the application of our schedule on AP association and bandwidth allocation. We consider a typical AC (Access Controller) - AP architecture [22]. AC, the centralized scheduler, manages multiple APs. The AC accepts the data from the Internet and sends it to the APs. We assume that each user is covered by at least one AP. As in [8], we further assume that there are sufficient channels and frequency planning has been done ahead, thus each AP

Supported by NSFC (no. 61422206, 61120106008) and National 863 project (no. 2015AA015701).

TABLE I. NOTATIONS

M	The number of users
N	The number of APs
T	The allocable transmission time of an AP
S_i	The size of data arriving for user i
B_i	The bandwidth demand of user i
x_{ij}	The association coefficient of user i and AP j
r_{ij}	The transmission rate of user i and AP j
T_{ij}	The transmission time demand of user i on AP j
t_{ij}	The allocated transmission time of user i by AP j
b_{ij}	The allocated bandwidth of user i by AP j

is assigned an interference-free channel. We consider a period of time T, in which the network is stable. A user is allowed to associate to one AP within one period and the AP assigns fractional transmission time to each associated user.

We then illustrate the workflow of the architecture. AC caches the arriving data in one period and delivers them to users in the next period. At the same time, it accepts data from Internet. In terms of an arbitrary period, AC first calculates the optimal schedule; then it achieves fast association switching with the virtual AP technique presented in [23] and sends data to APs. After that, APs send data to users. To calculate the optimal schedule, the demand and traffic rate of users are necessary. We use the size of down link data arriving in the AC to represent the user demand, since the down link data produces the dominated traffic for the real applications [24], [25]. The traffic rate is added to the periodic control message exchanged between AC and AP [22].

With low overhead, the virtual AP technique makes the short switching period, about 100ms [23], which leads to small period T and thus small caching delay. Further, the calculation delay can be also ignored due to the low complexity of our algorithm as shown in section III. Consequently, the delay of our application scenario is competitive to the traditional AC-AP architecture. In this paper, we focus on designing the efficient schedule during one period T, i.e., the allocable transmission time of AP is T.

B. Network Model

We use the coefficient x_{ij} to indicate the association relationship between user i and AP j. The coefficient is defined as follows:

$$x_{ij} = \begin{cases} 1, & \text{if user } i \text{ is associated to AP } j, \\ 0, & \text{otherwise,} \end{cases}$$

where $i = 1, 2, \cdots, M$, $j = 1, 2, \cdots, N$. M is the number of users and N is the number of APs.. The allocated transmission time of user i from AP j is denoted as t_{ij}. After the association is decided, the allocated transmission time of user i is given by $\widetilde{t}_i = \sum_{j=1}^{N} x_{ij} t_{ij}$. A list of symbols used in this paper are summarized in Table I.

Definition 1. *The bandwidth demand of user* i is the size of data that arrives for user i during per unit of time, denoted by B_i. It is formulated by

$$B_i = \frac{S_i}{T}, \tag{1}$$

where S_i is the size of data that arrives for user i.

Definition 2. *The transmission time demand of user i associated to AP* j is the length of time that the user needs to download the arriving data from AP j, which can be formulated as

$$T_{ij} = \frac{S_i}{r_{ij}} = \frac{B_i}{r_{ij}} T, \tag{2}$$

where r_{ij} is the transmission rate of user i associated to AP j, determined by the experienced Signal to Interference plus Noise Ratio (SINR) of user i from AP j [14]. After the association is decided, the transmission rate of user i is given by $\widetilde{r}_i = \sum_{j=1}^{N} x_{ij} r_{ij}$, and the transmission time demand of user i is given by $\widetilde{T}_i = \sum_{j=1}^{N} x_{ij} T_{ij}$.

Definition 3. *The allocated bandwidth of user i from AP* j is the average number of bits that the user downloads from AP j during per unit of time, as the AP is only partly occupied by user i. It is given by

$$b_{ij} = \frac{r_{ij} t_{ij}}{T}. \tag{3}$$

After the association is decided, the allocated bandwidth of user i is given by $\widetilde{b}_i = \sum_{j=1}^{N} x_{ij} b_{ij}$.

C. Bandwidth Allocation of Single AP

In a single AP, the allocated bandwidth is decided by the transmission rate and the allocated transmission time. Since the transmission rate is fixed after the association is decided, the problem of bandwidth allocation transforms to the problem of time scheduling. Note that in the multi-rate network, taking the user fairness into account is necessary. Our goal is to allocate the transmission time to the users as equally as possible under the constraint of users' transmission time demand, i.e., the variance of allocated transmission time is minimized. We illustrate our goal is equivalent to the goal of maximizing the aggregated bandwidth utility proposed by [26] as follows.

The objective of maximizing the aggregated bandwidth utility is presented as

$$\max \quad \textstyle\sum_{i=1}^{m} \log(\widetilde{b}_i) = \sum_{i=1}^{m} \log \frac{\widetilde{r}_i \widetilde{t}_i}{T}.$$

$\log(\widetilde{b}_i)$ is the bandwidth utility of user i. m is the number of users associated to the AP. Since $\widetilde{r}_i$ and T are constant, the objective can be transformed into $\max \sum_{i=1}^{m} \log \widetilde{t}_i$. If the objective is achieved, the variance of allocated transmission time is minimized. We prove the conclusion with theorem 1.

Theorem 1. *Suppose that* $\{\widetilde{t}_i^*\}$ *is the optimal solution of problem (4), then* $\{\widetilde{t}_i^*\}$ *is the optimal solution of problem (5).*

$$\begin{array}{ll} \max & \sum_{i=1}^{m} \log \widetilde{t}_i \\ \text{subject to:} & \sum_{i=1}^{m} \widetilde{t}_i = T \\ & 0 \le \widetilde{t}_i \le \widetilde{T}_i, 1 \le i \le m. \end{array} \tag{4}$$

$$\begin{array}{ll} \min & \sum_{i=1}^{m} (\widetilde{t}_i - \bar{\widetilde{t}})^2 / m \\ \text{subject to:} & \sum_{i=1}^{m} \widetilde{t}_i = T \\ & 0 \le \widetilde{t}_i \le \widetilde{T}_i, 1 \le i \le m. \end{array} \tag{5}$$

Proof. Since $\widetilde{t}_i^* \le \widetilde{T}_i$, the indexes of $\{\widetilde{t}_i^*\}$ can be divided into two subsets: set $A : \{i | \widetilde{t}_i^* = \widetilde{T}_i\}$, set $B : \{i | \widetilde{t}_i^* < \widetilde{T}_i\}$. To prove the theorem, we first give the two following lemmas:

Lemma 1. $\{\widetilde{t}_i^*\}$ *has the following two properties:*

(1) The value of each $\widetilde{t}_i^*$ *with index in set* B *equals to each other, i.e.,* $\exists t_B, \widetilde{t}_i^* = t_B, \forall i \in B$.

(2) Based on (1), the value of each $\widetilde{t}_i^*$ *is no greater than* t_B*, i.e.,* $\widetilde{t}_i^* \le t_B, 1 \le i \le m$.

Proof. We prove the lemma by solving (4) with the KKT method. Based on the analysis on the solving process, we obtain the properties of $\{\widetilde{t}_i^*\}$. This completes the proof. The details will be in the extended version of this paper. □

Lemma 2. *If some solution of (5), $\{\widetilde{t}_i\}$, satisfies the properties presented in lemma 1, $\{\widetilde{t}_i\}$ is the optimal solution of (5).*

Proof. We prove the lemma by contradiction. Assume there is another solution $\{\widetilde{t}'_i\}$, which satisfies the properties in lemma 1, but different from $\{\widetilde{t}_i\}$. If it is also the optimal solution of (5), we have the contradiction that the variance of $\{\widetilde{t}'_i\}$ is larger than that of $\{\widetilde{t}_i\}$, i.e., $\{\widetilde{t}'_i\}$ is not the optimal solution. The details will be in the extended version of this paper. □

According to lemma 1, $\{\widetilde{t}_i^*\}$ has the two properties. According to lemma 2, the solution with the two properties is the optimal solution of (5). Thus $\{\widetilde{t}_i^*\}$ is the optimal solution of (5). This completes the proof. □

We then give theorem 2 to illustrate that the optimal solution $\{\widetilde{t}_i^*\}$ can be calculated based on the properties in lemma 1.

Theorem 2. *Suppose $\{\widetilde{t}'_i\}$ is one solution of (4). If it satisfies the two properties in lemma 1, $\{\widetilde{t}'_i\}$ must be $\{\widetilde{t}_i^*\}$, i.e., the optimal solution of (4).*

Proof. We prove the theorem by contradiction. We assume $\{\widetilde{t}'_i\}$ is different from $\{\widetilde{t}_i^*\}$, the variance of them are respectively V' and V^*. According to lemma 1, $\{\widetilde{t}_i^*\}$ satisfies the two properties. Since the properties are satisfied, $V^* < V'$ on the basis of lemma 2's proof. Besides, $\{\widetilde{t}'_i\}$ satisfies the properties, thus $V' < V^*$. The two conclusions contradict each other. Therefore, if some solution satisfies the two properties of the optimal solution, it is equivalent to the optimal solution $\{\widetilde{t}_i^*\}$. This completes the proof. □

D. JAA Problem

In this part, we focus on the JAA problem when multiple APs are considered. Taking the objective of maximizing the aggregated throughput based on the fairness criterion, we need to decide $\{x_{ij}\}$ and $\{t_{ij}\}$. The JAA problem's formulation is presented as follows.

$$\max \quad \sum_{i=1}^{M} \log \sum_{j=1}^{N} \frac{x_{ij} r_{ij} t_{ij}}{T} \tag{6}$$

subject to:

$$x_{ij} \in \{0,1\}, 1 \le i \le M, 1 \le j \le N \tag{7}$$

$$\sum_{j=1}^{N} x_{ij} = 1, 1 \le i \le M \tag{8}$$

$$0 \le t_{ij} \le T_{ij}, 1 \le i \le M, 1 \le j \le N \tag{9}$$

$$0 \le \sum_{i=1}^{M} x_{ij} t_{ij} \le T, 1 \le j \le N. \tag{10}$$

Our objective (6) is to maximize the sum of all the users' bandwidth utility. Constraint (7) indicates that the association coefficient is a binary variable and constraint (8) limits that the user can be associated to one AP with in one period. Constraint (9) shows that the user's allocated transmission time must be no more than its transmission time demand and constraint (10) means that the sum of transmission time allocated from the AP must be no more than its allocable transmission time.

III. MABU ALGORITHM

In this section, we first propose the Fairness-based Bandwidth Allocation (FBA) algorithm to solve the bandwidth allocation problem in section II-C. Then we propose the 1/2-approximation MABU algorithm to solve the JAA problem in section II-D. FBA is called in MABU.

A. Bandwidth Allocation

We propose FBA to allocate the bandwidth of single AP to its associated m users. As we illustrate in section II, we can construct the optimal time allocation $\{\widetilde{t}_i^*\}$ according to the two properties of the optimal solution in lemma 1.

The details of Algorithm 1 are as follows. We first sort $\widetilde{T}_i$s in ascending order and calculate the current average allocable transmission time as t_B. We guess k from 1 to m by comparing $\widetilde{T}_k$ with t_B. We allocate each user k with $\widetilde{T}_k$ when $\widetilde{T}_k \le t_B$. Then we update t_B with the rest allocable time and go into the next loop. We repeat the operation in lines 5-13 till $\widetilde{T}_k > t_B$. Then t_B is allocated to each of the rest users.

Algorithm 1 FBA

Input: T, $\{\widetilde{T}_1, ..., \widetilde{T}_m\}$
Output: $\{\widetilde{t}_1^*, ..., \widetilde{t}_m^*\}$
1: Sort $\widetilde{T}_i$s in ascending order
2: Initialize $T_{total} = T, count = m, k = 1$
3: Guess the value $t_B = T_{total}/m$
4: $A = \emptyset$
5: **while** $k \le m$ **do**
6:   **if** $\widetilde{T}_k > t_B$ **then**
7:     break
8:   **else**
9:     add k into set A
10:     $\widetilde{t}_k^* = \widetilde{T}_i$, $T_{total} -= \widetilde{T}_i$, $count -= 1$, $k += 1$
11:     guess $t_B = T_{total}/count$
12:   **end if**
13: **end while**
14: **for** each $i \notin A$ **do**
15:   $\widetilde{t}_i^* = t_B$, add i into set B
16: **end for**

B. AP Association and Bandwidth Allocation

We then propose MABU to solve JAA. In MABU, we first decide the association relationship between the APs and users and then allocate the bandwidth of each AP by calling FBA.

As Algorithm 2 illustrates, it first sorts the users by the data size S_i in descending order. In terms of an arbitrary user i, it tries to associate to each AP j and estimates the allocated transmission time of AP j. The allocated transmission time of AP equals to the aggregated transmission time demand of the users associated to the AP, including user i. The user chooses the AP with the least allocated transmission time, as presented in line 4. After the association decision, we use FBA to allocate the bandwidth of each AP.

Algorithm 2 MABU

Input: $\{S_1, ..., S_M\}, \{T_{11}, ..., T_{MN}\}$
Output: $\{x_{11}, ..., x_{MN}\}$
1: Initialize $x_{ij} = 0, \forall i \in 1..M, j \in 1..N$
2: Sort the users by the size of data S_i in descending order
3: **for** each user i in the sorted users **do**
4: $\quad j_{min} = \underset{j}{\operatorname{argmin}}(\sum_{k=1}^{i-1} x_{kj}T_{kj} + T_{ij})$
5: $\quad x_{ij_{min}} = 1$
6: **end for**
7: **for** each AP j **do**
8: $\quad$ Use Algorithm 1 to schedule the transmission time
9: **end for**

In Algorithm 2, we first take $\mathcal{O}(M \log M)$ operations to sort the users (line 2). The time complexity of association decision for M users is $\mathcal{O}(MN)$ (line 3-6). In Algorithm 1, we first sort $\widetilde{T}_i$s with $\mathcal{O}(M \log M)$ operations (line 1) and then find k with $\mathcal{O}(M)$ operations (line 2-13). Then the time complexity of bandwidth allocation for N APs is $\mathcal{O}(MN \log M)$. Therefore, the time complexity of MABU is $\mathcal{O}(MN \log M)$.

C. Bound Analysis of MABU

In this part, we analyze the lower bound of the solution given by MABU. We assume the transmission rate of all the users associated to different APs are the same. Therefore, r_{ij} equals to a constant, denoted as r. Then T_{ij} is simplified as T_i, with $T_i = S_i/r$. $\mathbb{T} = \sum_{k=1}^{M} T_k$, i.e., $\mathbb{T}$ is the sum of the transmission time demand.

In order to analyze the lower bound of MABU, we first give the following lemma to estimate the upper bound of AP's load when the MABU algorithm is adopted.

Lemma 3. *Suppose for N APs and M users, each user's transmission time demand is no more than $\frac{2\mathbb{T}}{N+1}$. With MABU algorithm, the maximum load of one AP will be no greater than $\frac{2\mathbb{T}}{N+1}$, i.e.,*

$$max \sum_{k=1}^{M} x_{kj}T_k \leq \frac{2\mathbb{T}}{N+1}, j = 1, 2, ..., N. \tag{11}$$

Proof. We prove the lemma by contradiction. Suppose that when the i_{th} user is associated to AP j, the aggregated transmission time demand of the users associated to AP j is more than $\frac{2\mathbb{T}}{N+1}$, i.e. $\sum_{k=1}^{i-1} x_{kj}T_k + T_i > \frac{2\mathbb{T}}{N+1}$. We discuss the contradiction under the two cases:

Case 1: $T_i > \frac{\mathbb{T}}{N+1}$. Since the users are sorted by the size of data in descending order and the transmission rate between the APs and users are the same, the transmission time demand is also in descending order according to (2). Thus $\sum_{k=1}^{i} T_k \geq iT_i$. As we assume that each T_i is no more than $\frac{2\mathbb{T}}{N+1}$, it satisfies $i \geq N+1$. We have $iT_i \geq (N+1)T_i > (N+1)\frac{\mathbb{T}}{N+1} = \mathbb{T}$. Combining the two expressions above, we have $\sum_{k=1}^{i} T_k > \mathbb{T}$. It means the aggregated transmission time demand of the i users is more than $\mathbb{T}$, which is contradictory against the assumption that the aggregated transmission time demand of the M users equals to $\mathbb{T}$.

Case 2: $T_i \leq \frac{\mathbb{T}}{N+1}$. Adopting the similar analysis, we obtain the contradiction same to that in case 1. This completes the proof. Please see the details in the technical report [27]. $\square$

Then we compare the actual load with the upper bound to calculate the gap between MABU's solution and optimal solution with the following theorem.

Theorem 3. *Let U' be the aggregated bandwidth utility given by MABU and U^* be the aggregated bandwidth utility given by the optimal solution. We have*

$$U^* - U' \leq M \log \frac{2\mathbb{T}/(N+1)}{T}. \tag{12}$$

Proof. We assume $T \leq \frac{2\mathbb{T}}{N+1}$. If user i is allocated to transmission time $\frac{T}{2\mathbb{T}/(N+1)}T_i$ from its associated AP, the aggregated allocated transmission time of each AP is no more than T by applying lemma 3. Since the time scheduling proposed by Algorithm 1 is optimal with the same AP association, we have $U' \geq \sum_{i=1}^{M} \log \frac{r\frac{T}{2\mathbb{T}/(N+1)}T_i}{T} = \sum_{i=1}^{M} \log \frac{rT_i}{T} - M \log \frac{2\mathbb{T}/(N+1)}{T}$.

Since the allocated transmission time of user i is no more than its transmission time demand, we have $\sum_{i=1}^{M} \log \frac{rT_i}{T} \geq U^*$. Then we have the conclusion that $U^* - U' \leq M \log \frac{2\mathbb{T}/(N+1)}{T}$.

The above proof is based on the assumption that each T_i is no more than $\frac{2\mathbb{T}}{N+1}$. When there is some T_i that is larger than $\frac{2\mathbb{T}}{N+1}$, we can think it equals to $\frac{2\mathbb{T}}{N+1}$. This is because thinking it equals to $\frac{2\mathbb{T}}{N+1}$ will not influence the value of U^* and U'. When we think it equals to $\frac{2\mathbb{T}}{N+1}$, the assumption that each T_i is no more than $\frac{2\mathbb{T}}{N+1}$ is satisfied and we can use the proof above. This completes the proof. $\square$

At last, we prove that the approximation factor of MABU is 1/2 with corollary 1.

Corollary 1. *If $\overline{b}'$ is the geometric mean of $\{\widetilde{b}_i'\}$, $\overline{b}^*$ is the geometric mean of $\{\widetilde{b}_i^*\}$ and $\mathbb{T} \leq NT$, then we have $\overline{b}' > 1/2\overline{b}^*$. $\{\widetilde{b}_i'\}$ is the bandwidth allocation given by MABU and $\{\widetilde{b}_i^*\}$ is the bandwidth allocation given by the optimal solution.*

Proof. According to the definition of U^* and U', we have $U^* - U' = \log \frac{\prod_{i=1}^{M} \widetilde{b}_i^*}{\prod_{i=1}^{M} \widetilde{b}_i'} = \log(\frac{\overline{b}^*}{\overline{b}'})^M = M \log \frac{\overline{b}^*}{\overline{b}'}$. Combing (12) with the equation above, we have $\frac{\overline{b}^*}{\overline{b}'} \leq \frac{2\mathbb{T}}{(N+1)T}$. When $\mathbb{T} \leq NT$, i.e., the aggregated transmission time demand is no larger than the allocable transmission time of all the APs, we have $\frac{2\mathbb{T}}{(N+1)T} \leq \frac{2NT}{(N+1)T} < 2$. Then we get $\overline{b}^*/\overline{b}' < 2$, i.e., $\overline{b}' > 1/2\overline{b}^*$. This completes the proof. $\square$

IV. Performance Evaluation

A. Evaluation Setup and Methodologies

In this part, we report our simulation results for the scenario where the network contains static users. We compare the performance of MABU with those of the following ones:

- Strong Signal First (SSF): the default AP selection scheme in the 802.11 standard.
- NLAO-PF: a time-based fairness algorithm in [14].

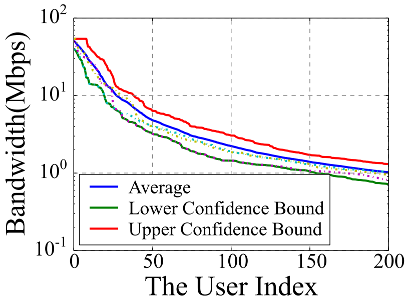


Fig. 1. User bandwidth demand in real traffic scenario

- Norm Load-Based AP selection (NLB): a decentralized AP selection algorithm with the objective of minimizing the norm load of AP proposed in [19].

All the algorithms are examined carefully according to the following performance metrics:

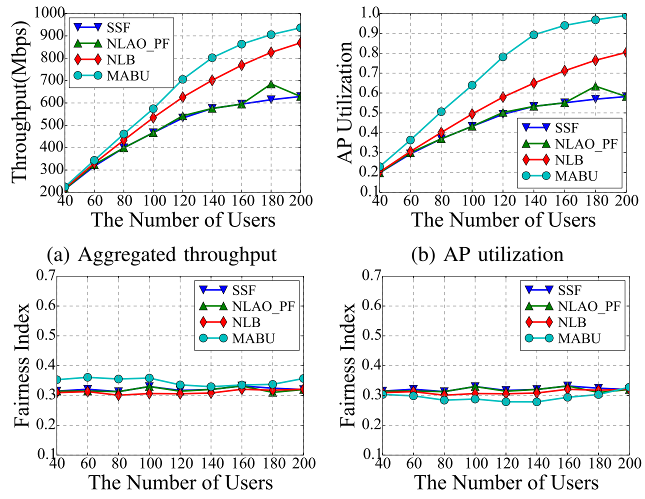
- Normalized transmission time demand on AP, is computed by $\frac{\sum_{i=1}^{M} x_{ij} T_{ij}}{T}$;
- Average AP utilization, is computed by $\frac{\sum_{j=1}^{N} (\sum_{i=1}^{M} x_{ij} t_{ij})/T}{N}$;
- Per-user throughput in Mbps and the corresponding statistical information;
- Jain's Fairness Index [28], is defined as $J = \frac{(\sum_{i=1}^{M} y_i)^2}{M(\sum_{i=1}^{M} {y_i}^2)}$, where y_i is the allocated transmission time, the allocated bandwidth of user, or the demand on AP. Note that a larger value of $J \in [0, 1]$ indicates a better fairness.

For ease of comparison, we employ the same settings as those in [14]. The network contains 20 APs placed on a 5×4 grid, with each on a grid point. The distance between two adjacent APs is 100m and the coverage area of AP is about 150m. There are users residing in the network with different bandwidth demands, resulting in different levels of network loads. Two types of user distribution are considered: (1) Uniform: the users are uniformly positioned within the coverage area of the network, with each on a grid point; (2) Hotspot: the users are randomly positioned in a circle-shaped hotspot area with a radius of 100m near the center of the 20-AP network. Since the transmission rate of a user is difficult to collect, we derive it based on the distance between the AP and user as in [14].

We capture the down link data arriving for users within 120 seconds in a campus network to simulate the bandwidth demand of users. We plot the bandwidth demand of the most active 200 users in Fig.1. The dash line represents the demand of each user in one second. The average value, lower and upper confidence bounds are highlighted in the figure. We can find that the bandwidth demand differs vastly between different users.

B. Evaluation Results

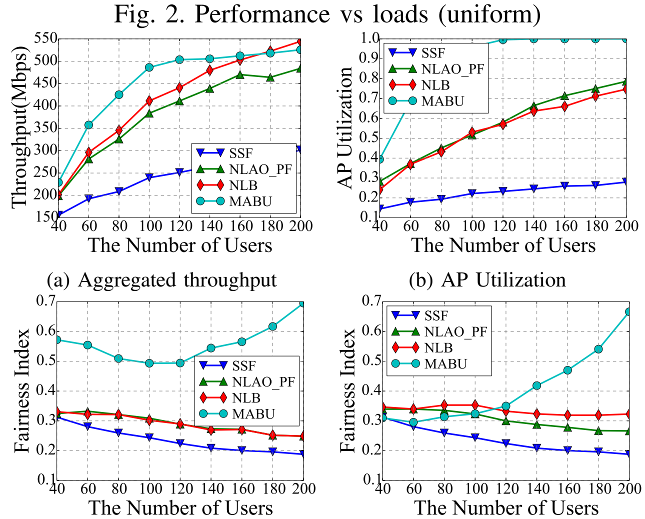
In this part, we evaluate the performance of MABU with comparison to other algorithms. We mainly focus on the two aspects of performance: throughout increase and fairness guarantee. We first analyze the performance of MABU under different loads. Then we investigate why MABU can increase throughput while ensuring fairness. We find that MABU redistributes the load carefully and increases average AP utilization, leading to high throughput and fairness guarantee.



(a) Aggregated throughput (b) AP utilization

(c) Time fairness index (d) Bandwidth fairness index

Fig. 2. Performance vs loads (uniform)



(a) Aggregated throughput (b) AP Utilization

(c) Time fairness index (d) Bandwidth fairness index

Fig. 3. Performance vs loads (hotspot)

1) Performance under Different Loads: In this part, we evaluate the performance of MABU under different loads in the two distributions. We fix the number of AP as 20 and change the scale of users to change the load. For each scale, the algorithm runs for 50 times to obtain the average effect. As Fig.2a and Fig.3a show, when the load is light, the throughput of all the algorithms is very close. At this point, the AP utilization is lower than 50%, as Fig.2b and Fig.3b show.

With the growth of user scale, the AP utilization increases. The AP utilization of MABU increases faster than other algorithms, as Fig.2b and Fig.3b show. As a result, MABU gets a significant throughput increase, i.e., 14.4% in the uniform and 23.1% in hotspot distribution, compared with NLB. Since the fairness we use in our paper is time equality, the time fairness index of MABU is higher than other algorithms, as Fig.2c and Fig.3c show. The time equality leads to well bandwidth equality. Hence, the bandwidth fairness index of MABU is close to that of the other algorithms, as shown in Fig.2d and Fig.3d.

When the network is overloaded, the AP utilization of MABU reaches 1, as Fig.2b and Fig.3b show. The throughput

TABLE II. STATISTICS OF THE FAIRNESS RESULTS

		Fairness Of User's Allocated Time		Fairness Of User's Allocated Bandwidth		Fairness Of User's Demand On APs	
Type	Algorithm	Mean	Jain's Fairness Index	Mean	Jain's Fairness Index	Mean	Jain's Fairness Index
Uniform	MABU	0.132	0.273	5.776	0.216	0.969	0.967
	NLAO_PF	0.076	0.232	4.078	0.232	0.794	0.484
	NLB	0.089	0.230	4.819	0.230	0.794	0.620
	SSF	0.076	0.232	4.078	0.232	0.794	0.484
Hotspot	MABU	0.199	0.448	5.277	0.217	1.155	0.956
	NLAO_PF	0.094	0.213	3.996	0.192	0.735	0.439
	NLB	0.087	0.222	3.703	0.241	0.724	0.336
	SSF	0.045	0.172	2.427	0.172	0.623	0.148

can not be increased with load balancing, thus the advantage of MABU becomes not obvious. The throughput of all the algorithms is close to each other.

2) Fairness Guarantee: In this part, the user fairness is investigated. In Table II, the time fairness index of MABU is higher than that of the other algorithms in each distribution. That is because MABU focuses more on the fair time allocation. Furthermore, the mean of the allocated transmission time of MABU is higher than that of other algorithms. It results from the increase of the AP utilization. As to the bandwidth fairness index, the value of MABU is close to that of the other algorithms. It illustrates that the time fairness brings the bandwidth fairness. The mean of the bandwidth of MABU is higher than that of other algorithms. The reason is that the increase of AP utilization leads to the increase of throughput.

V. Conclusion

The wide spread of IEEE 802.11 based WLAN applications makes the network management more complex and challenging. Load balancing and fair bandwidth allocation are two of the most critical issues. Different from the existing work, we formulate the JAA problem with consideration of the differences of the user's bandwidth demand. Then the MABU algorithm is proposed to solve the problem. As shown in the real-trace based evaluations, the conventional AP association strategies lead to an obvious load unbalance, both in the uniform and hotspot cases. Considering the user's bandwidth demand, our association strategy effectively balances the loads and brings a significant throughput increase. We have also shown that the bandwidth allocation algorithm provides better user fairness.

References

[1] H. Gong and J. Kim, "Dynamic load balancing through association control of mobile users in wifi networks," *Consumer Electronics, IEEE Transactions on*, vol. 54, no. 2, pp. 342–348, 2008.

[2] W. Ge, H. Ji, V. C. Leung, and P. Si, "Access point selection for wlans with cognitive radio: a restless bandit approach," in *ICC*. IEEE, 2011, pp. 1–5.

[3] L.-H. Yen, J.-J. Li, and C.-M. Lin, "Stability and fairness of ap selection games in ieee 802.11 access networks," *Vehicular Technology, IEEE Transactions on*, vol. 60, no. 3, pp. 1150–1160, 2011.

[4] Z. Chen, Q. Xiong, Y. Liu, and C. Huang, "A strategy for differentiated access service selection based on application in wlans," in *INFOCOM WKSHPS*. IEEE, 2014, pp. 317–322.

[5] M. Lu and J. Wu, "Localized access point selection in infrastructure wireless lans with performance guarantee," *Wireless Communications and Mobile Computing*, vol. 11, no. 8, pp. 994–1010, 2011.

[6] L. Chen, "A distributed access point selection algorithm based on no-regret learning for wireless access networks," in *VTC*. IEEE, 2010, pp. 1–5.

[7] X. Chen, W. Yuan, W. Cheng, W. Liu, and H. Leung, "Access point selection under qos requirements in variable channel-width wlans," *Wireless Communications Letters, IEEE*, vol. 2, no. 1, pp. 114–117, 2013.

[8] D. Gong and Y. Yang, "Ap association in 802.11 n wlans with heterogeneous clients," in *INFOCOM*. IEEE, 2012, pp. 1440–1448.

[9] X. Chen, Y. Zhao, B. Peck, and D. Qiao, "Sap: Smart access point with seamless load balancing multiple interfaces," in *INFOCOM*. IEEE, 2012, pp. 1458–1466.

[10] M. Abusubaih and A. Wolisz, "An optimal station association policy for multi-rate ieee 802.11 wireless lans," in *MSWiM*. ACM, 2007, pp. 117–123.

[11] G. Xue, Q. He, H. Zhu, T. He, and Y. Liu, "Sociality-aware access point selection in enterprise wireless lans," *IEEE Transactions on Parallel and Distributed Systems (TPDS)*, vol. 24, no. 10, pp. 2069–2078, 2013.

[12] L. Du, Y. Bai, and L. Chen, "Access point selection strategy for large-scale wireless local area networks," in *WCNC*, 2007, pp. 2161–2166.

[13] S. K. Dandapat, B. Mitra, R. R. Choudhury, and N. Ganguly, "Smart association control in wireless mobile environment using max-flow," *Network and Service Management, IEEE Transactions on (TNSM)*, vol. 9, no. 1, pp. 73–86, 2012.

[14] 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," *Networking, IEEE/ACM Transactions on (TON)*, vol. 22, no. 1, pp. 191–202, 2014.

[15] S. Wang, F. Huang, and C. Wang, "Adaptive proportional fairness resource allocation for ofdm-based cognitive radio networks," *Wireless networks*, vol. 19, no. 3, pp. 273–284, 2013.

[16] T. Bu, L. Li, and R. Ramjee, "Generalized proportional fair scheduling in third generation wireless data networks," in *INFOCOM*. IEEE, 2006, pp. 1–12.

[17] L. Li, M. Pal, and Y. R. Yang, "Proportional fairness in multi-rate wireless lans," in *INFOCOM*. IEEE, 2008.

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

[19] 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 (TPDS)*, p. 1, 2013.

[20] S. Keranidis, T. Korakis, I. Koutsopoulos, and L. Tassiulas, "Contention and traffic load-aware association in ieee 802.11 wlans: Algorithms and implementation," in *WiOpt*. IEEE, 2011, pp. 334–341.

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

[22] G. Conradi, "Current status and overview of the capwap protocol," *Dostopno na: http://www. cse. wustl. edu/jain/cse574-10/ftp/capwap/index. html*, 2010.

[23] L. Suresh, J. Schulz-Zander, R. Merz, A. Feldmann, and T. Vazao, "Towards programmable enterprise wlans with odin," in *SIGCOMM WKSHPS*. ACM, 2012, pp. 115–120.

[24] A. Akella, G. Judd, S. Seshan, and P. Steenkiste, "Self-management in chaotic wireless deployments," *Wireless Networks*, vol. 13, no. 6, pp. 737–755, 2007.

[25] Y. Cao, T. Jiang, C. Wang, and L. Zhang, "Crac: cognitive radio assisted cooperation for downlink transmissions in ofdma-based cellular networks," *Selected Areas in Communications, IEEE Journal on*, vol. 30, no. 9, pp. 1614–1622, 2012.

[26] F. Kelly, "Charging and rate control for elastic traffic," *European Transactions on Telecommunications*, vol. 8, pp. 33–37, 1997.

[27] "Technical report," http://www.4over6.edu.cn/others/association%20technical%20report.pdf.

[28] R. Jain, D.-M. Chiu, and W. R. Hawe, *A quantitative measure of fairness and discrimination for resource allocation in shared computer system*. Eastern Research Laboratory, Digital Equipment Corporation Hudson, MA, 1984.