IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 23, 2017, accepted July 19, 2017, date of publication July 28, 2017, date of current version August 22, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2733083

A Semi-Matching Based Load Balancing Scheme for Dense IEEE 802.11 WLANs

TAO LEI[1,2], XIANGMING WEN[1,2,3], ZHAOMING LU[1,2,3], AND YUE LI[4]

[1]Beijing Key Laboratory of Network System Architecture and Convergence, Beijing University of Posts and Telecommunications, Beijing 100876, China
[2]Beijing Laboratory of Advanced Information Networks, Beijing 100876, China
[3]Beijing Advanced Innovation Center for Future Internet Technology, Beijing 100876, China
[4]Xi'an Datang Telecom Co., Ltd, Xi'an 710075, China

Corresponding author: Tao Lei (leitao3120@bupt.edu.cn)

This work was supported by the WLAN achievement transformation based on SDN project of Beijing Municipal Commission of Education under Grant 201501001.

ABSTRACT A load balancing mechanism can adjust the load distribution among access points (APs) and improve resource utilization for dense wireless local area networks (WLANs). In this paper, we propose a semi-matching-based load balancing scheme for the IEEE 802.11 dense WLANs. The proposed scheme runs in a centralized controller. The controller judges whether the load is unevenly distributed according to the collected channel busy time ratio information of the entire network, and triggers the load balancing mechanism accordingly. In order to realize load balancing among APs and maximize the overall network throughput, we model the station to AP association problem as a weighted bipartite graph matching problem and find the optimal semi-matching using the Kuhn–Munkres (K-M) algorithm. Simulation results show that the proposed scheme achieves performance improvement comparing with traditional schemes.

INDEX TERMS IEEE 802.11, dense WLANs, load balancing, semi-matching, channel busy time ratio (CBTR).

I. INTRODUCTION

To meet the sustained growth of mobile traffic demands, IEEE 802.11 based wireless local area networks (WLANs) have been widely deployed. In some hot spots, WLANs have become more and more dense. In order to cope with the expected growth in the number and density of WLANs, since May 2014, the IEEE 802.11ax task group has started the development of a new standard for high efficiency WLANs (HEWs). The 802.11ax task group aims at improving the network efficiency and the network throughput in dense network deployment with hundreds of access points (APs) and stations (STAs) [1].

In current implementations of the 802.11 devices, AP selection decision is based on the Received Signal Strength Indication (RSSI). A STA simply selects the AP from which it has received the strongest signal during the scanning process [2]. In dense WLAN scenarios, since users tend to concentrate in some areas and the signal strength strongly depends on the distance, the overall network load is usually unevenly distributed. This can degrade the fairness in STA's throughput and harm an efficient use of the WLAN resource. This problem motivates a load balancing mechanism for dense WLANs. The goal is to make the overall load evenly distributed in the network thus to increase the whole system throughput.

A sound load balancing scheme should address two primary issues. The first is how to define and measure the load of a WLAN. The second concerns how to distribute overall traffic load among all available APs [3]. For the first issue, so far there are no widely accepted metrics. Different algorithms may use different criteria. The most straightforward metric to measure an AP's load is the number of STAs associated with it [4]. However, the number of associated STAs cannot accurately represent the load condition, because some STAs may not send data packets all the time or just send some few data packets. It is possible that the load of APs with fewer STAs is heavier than that of APs with more STAs. Another commonly used metric is the throughput per AP, including uplink and downlink traffic [5]. When all the associated STAs work at the same rate, this metric can correctly evaluate the AP's load. But today's 802.11 devices generally run different rate control algorithms. STAs adjust their transmission rate adaptively according to specific rate control algorithms. In infrastructure mode, a single AP together with all associated STAs is called a basic service set (BSS) [6]. The BSS provides the basic building-block of an 802.11 based WLAN.

If the channel of a BSS is occupied by a large number of low rate STAs for a long time, the throughput is low but the load is very heavy. Therefore, AP's throughput is also a debatable metric.

For the second issue, the most straightforward way is to adjust the load distribution according to the selected load metric. A load level reaching some predefined threshold or exceeding the current average by a certain hysteresis can be an indication of overload. According to this method, some STAs associated with the overloaded AP will be switched to a light load AP proactively or passively. But which STA should be switched and how to choose the target AP remain problems. Therefore, it is difficult to obtain the optimal result only by using the load metric as load balancing adjustment indicator. RSSI is a metric widely used by STAs to select target APs. But the result is not necessarily the optimal one, because neither the load at the APs nor the cross-BSS interference is accounted for [7]. In recent years, there has been growing interest in analyzing and optimizing the BSS association to improve the overall network performance using signal-to-interference-plus-noise ratio (SINR). It is the first-order predictor of the user experience, or at least of the link reliability, because it is directly related to the real-time modulation and coding scheme. But a general utility maximization of SINR, subject to a resource or/and power constraint, is NP hard and not computable even for modest-sized networks [8].

It is a good idea to solve the second issue using graph theory. In [9], the authors analyzed a task assignment problem, in which each task requires one unit of processing time, and must be assigned to some machine that can process it. The tasks are to be assigned to machines in a manner that minimizes two optimization objectives. One objective is to minimize the average completion time and the other objective is to maximize the fairness of the assignment from the machines' point of view, i.e., to minimize the variance of the loads on the machines. They referred to this problem as the optimal semi-matching problem and presented two efficient algorithms for computing optimal semi-matchings.

Inspired by the task assignment problem, we model the WLAN load balancing problem using graph theory and propose a semi-matching based load balancing scheme for IEEE 802.11 based dense WLANs. The proposed scheme runs in a centralized controller, it considers both the load balancing among APs and the optimization of the long-term network throughput. The main contribution can be summarized as follows:

- We take both load status and interference effects into account in the load balancing problem, the load status determines the number of STAs that each AP can serve and the SINR is utilized to determine the final STA-AP association set.
- We use the channel busy time ratio (CBTR) as load metric, it is estimated by the average network allocation vector (NAV) in a specific time window.
- We model the STA-AP association problem as weighted bipartite graph matching problem and formulate the problem of finding a balanced semi-matching in which the total weight is maximized for the weighted bipartite graph.
- We use the SINR between a STA and its associated AP as the weight. By pre-assigning the number of STAs per BSS according to the CBTR, we solve the weighted bipartite graph matching problem using the Kuhn-Munkres (K-M) algorithm [10].

II. RELATED WORK

There have been a lot of studies on load balancing over WLAN. All existing load balancing schemes in IEEE 802.11 networks distribute traffic load by managing associations between APs and STAs. According to the part that is in charge of such management, these schemes can be classified into two types: STA-based and network-based. STA-based schemes run distributed algorithms, selecting APs simply from the STA's view, are difficult to achieve system-wide load balance. Besides, how to acquire the AP's load condition accurately is also a problem. STAs usually need the assistance from APs or any network-side entity. For example, in [11], a stand-alone server is used to count the load status of each BSS and provides it to the STAs. Cao *et al.* [12] proposed a dynamic STA association method wherein a STA determines which AP to connect to, taking into consideration of multiple factors such as RSSI, link quality, and location of the users. In [13], the authors presented a quality of service (QoS) aware AP selection scheme for pre-load-balancing in multi-BSSs WLANs. They proposed an access category station count element to specify the number of STAs including both active and inactive ones. By inserting such information into frames advertised from APs, STAs are able to reassociate with the APs that are not the strongest one but with less potential load, and hence reduces the probability of being rejected when initiation a call.

For the network-based schemes, it is a network-side entity that controls the distribution of AP's load. Han *et al.* [14] presented an adaptive load balancing scheme for Software-Defined enterprise WLANs which takes into account of AP load condition in addition to RSSI value as load metrics. The proposed adaptive load balancing scheme implementing on top of SDN controller maintains a central view of the network and periodically checks the load condition of each AP. In case that unbalanced traffic load of AP is detected, it proactively re-assigns STAs to less loaded APs with relatively good RSSI value. In [15], the authors proposed a joint user association and load balancing scheme for WLAN with a centralized coordinator. The first objective is to balance the load among the APs. The second objective is to associate each user with its best AP for maximizing the network throughput. The last objective is to decide whether to turn on/off each AP given the current load condition. They formulated the problem into mixed integer non-linear fractional programming which is

NP-hard, and transformed the original problem into an equivalent mixed integer linear programming (MILP) problem which can be optimally solved using the branch and bound technique. Harada *et al.* [16] formulated the min-weight load balancing problem of finding a balanced semi-matching that minimizes the total weight for weighted bipartite, and proposed a variational Hungarian algorithm to solve the load balancing problem. Feng *et al.* [17] presented a concept of virtual resource chain for software defined campus WLANs (SD-WLANs). The virtual resource chain binds all the resources in WLAN, and can be developed to support various network applications. Using the virtual resource chain, they implemented a user-oriented load balancing application on the SD-WLAN controller to improve the resource utilization.

III. SYSTEM MODEL AND PROBLEM STATEMENT

A. SYSTEM MODEL

In densely deployed WLAN scenarios, a lot of APs' coverage areas overlap with each other. STAs located in overlapping coverage areas can associate with multiple candidate APs. In order to model densely deployed scenarios, we don't consider any channel assignment scheme in this work, that is, all the APs work on the same channel. According to the distributed coordination function (DCF) mechanism of the 802.11 medium access control (MAC) layer specification [6], APs need to compete for wireless channel with the STAs equally [18]. In this paper, we assume that APs' transmission queues are empty and all the STAs' transmission queues are never empty (*i.e.*, we study the load balancing problem in saturated uplink traffic condition). We consider a WLAN which is comprised of X BSSs and Y STAs. Fig. 1 shows the system architecture of the proposed scheme, all the APs are connected to a single distribution system (DS), and a centralized controller which connects to the same DS runs the proposed load balancing algorithm. The centralized controller has a global view of the whole network, and can use southbound protocol to obtain network status information and user information (such as CBTR, SINR, RSSI, etc.) from the underlaying APs. It can also realize a variety of network management operations through the southbound protocol, such as switching a STA from one AP to another AP etc.. Each AP runs an agent. The agent can communicate with the controller and handle the control logic issued by the controller. These mechanisms can be implemented on the software-defined wireless networks (SDWN) framework [19], and the load balancing controller can be a network management application on top of the SDWN controller [20].

Each STA is spatially separated from APs, the channels are assumed to be log-normal shadowing radio propagation. By considering the shadow fading and path loss, the received signal strength at a distance d_{ij} from an AP is

$$P_{d_{ij}} = P_{tx} - P_{r0} + 10\delta\log_{10}(d_{ij}/d_0) - X_0(dbm), \quad (1)$$

where P_{tx} is the transmission power (all STAs are assumed to transmit at the same power), P_{r0} is the path loss at reference

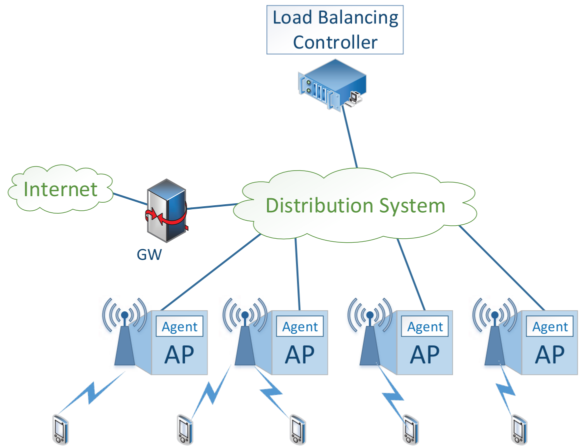


FIGURE 1. The proposed load balancing algorithm runs on a centralized controller.

distance $d_0 = 1m$, δ is the path loss exponent, X_0 is a Gaussian random variable with zero mean and a variance in db. For a given uplink from STA i to AP j, the received SINR of the AP can be calculated as

$$SINR_{ij} = \frac{P_{d_{ij}}}{I_r + \sigma^2}, \quad (2)$$

where

$$I_r = \sum_{k \in Y, k \neq i} P_{d_{kj}}, (1 \leq k \leq Y, 1 \leq j \leq X) \quad (3)$$

is the cumulative interference produced by all other transmitting STAs (except the tagged STA i for this AP j), σ^2 denotes the value of additive white Gaussian noise power.

There are two conditions for an AP j to successfully receive a frame from STA i:

$$\begin{cases} P_{d_{ij}} > P_0 \\ SINR_{ij} > \beta. \end{cases} \quad (4)$$

The first condition imposes that the received signal power should be larger than the carrier sense threshold P_0. IEEE 802.11 standard dictates several minimum requirements for the operation of a receiver. For instance, the sensitivity for the minimum transmission mode should be −82 dBm for a 20 MHz channel and the carrier sensing function should declare the medium busy for any signal 20 dBm above the minimum sensitivity requirement (hence, −62 dBm for a 20 MHz channel).

The second condition imposes that the SINR should be larger than the threshold β. The value of β determines the modulation scheme and the instantaneous rate [21]. For 802.11n, a SINR level of $\beta_1 = 2$ dBm for a 20 MHz channel is the minimum required for the 802.11 node (STA or AP) to be covered; it then gets an instantaneous bit rate of (at least) $r_1 = 6.5$ Mb/s. Similarly $\beta_2 = 5$ dBm, $\beta_3 = 9$ dBm and $\beta_4 = 11$ dBm are required for the higher bit rates of $r_2 = 13$ Mb/s, $r_3 = 19.5$ Mb/s and $r_4 = 26$ Mb/s respectively. As indicated, if the SINR at a 802.11 node is less than $\beta_1 = 2$ dBm, the node is not covered. In this paper, we don't consider any adaptive rate adjustment mechanism, that is, all nodes send data packets at the same rate. Therefore, the value of β can be set to a constant. In this paper, STAs are assumed to send data packets at the rate of 13 Mb/s, and β is set to 5 dBm accordingly.

B. PROBLEM STATEMENT

The way to achieve load balancing in dense WLANs is to redistribute the unevenly distributed load of the whole network by finding an optimal STA-AP association scheme, thus maximizing the total system throughput available to users. This is a combinatorial optimization problem. Since the achievable rate is computed as the logarithm of SINR, the higher the SINR is, the higher rate the STA can achieve. Therefore, the load balancing problem in dense WLANs in the uplink can be formalized as

$$\textit{maximize} \quad \sum_{j=1}^{X}\sum_{i=1}^{Y} SINR_{ij}x_{ij} \tag{5a}$$

$$\textit{subject to} \quad \sum_{j=1}^{X} x_{ij} = 1, \quad \forall i \in Y \tag{5b}$$

$$\sum_{j=1}^{X} SINR_{ij}x_{ij} > \beta, \quad \forall i \in Y \tag{5c}$$

$$\sum_{j=1}^{X} P_{d_{ij}}x_{ij} > P_0, \quad \forall i \in Y \tag{5d}$$

$$x_{ij} \in \{0, 1\}, \quad 1 \le i \le Y,\ 1 \le j \le X, \tag{5e}$$

where the constraint in 5(b) and 5(e) ensure that each STA associates with exactly one AP, the constraint in 5(c) and 5(d) ensure that AP j can successfully receive a frame from its associated STA i. The setting of the values of P_0 and β has been discussed above.

The associations from STAs to APs in a WLAN can be modeled using a bipartite graph. For example, Fig. 2 (a) shows a simple WLAN scenario which consists of three APs and six STAs. The coverage areas of the three APs overlap with each other. STAs located in the coverage area of an AP have one candidate AP and STAs located in the overlapping coverage areas have multiple candidate APs. Since we consider the saturated uplink traffic condition, the dashed arrows indicate the candidate direction of the uplink data flow. Let $G = (U \cup V, E)$ be a bipartite graph, where U and V denote a set of vertices and E denotes a set of edges between U and V with weights. We denote the edge with ends in u and v by $\{u, v\}$ for $u \in U$ and $v \in V$. As Fig. 2 (b) shows, U stands for the set of STAs, V stands for the set of APs, and E is the set of potential wireless links. For example, vertex $2 \in U$ is matched to vertices $\{1, 3\} \in V$ via two edges, which indicates that STA_2 has two candidate APs (AP_1 and AP_3). To relate the edges E with the quality of wireless links, the edge from STA i to AP j can be weighted by w_{ij} representing the uplink SINR of STA i to AP j.

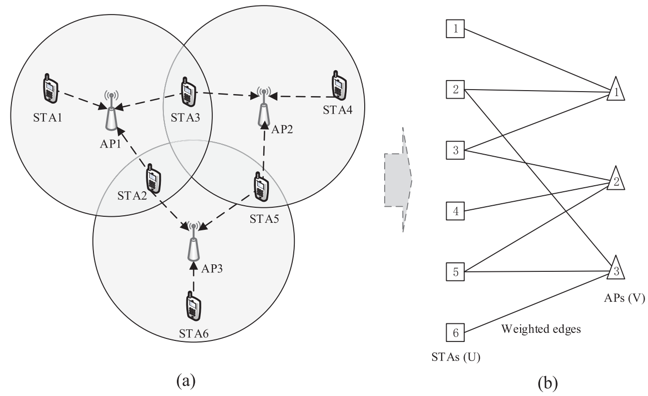


FIGURE 2. Modeling the STA-AP association problem as weighted bipartite graph matching problem. (a) Simple WLAN scenario which consists of three APs and six STAs. (b) Modeled weighted bipartite graph of scenario (a).

In the actual network, each STA can connect to exactly one AP. The STA-AP association scheme is a subset of edges in E. A semi-matching is defined as a set of edges $M \subseteq E$ such that each vertex in U is an end-point of exactly one edge in M. Edge $e \in M$ is called matching edge and edge $e \notin M$ is called non-matching edge respectively [16]. Therefore, the aforementioned optimization problem can be reformulated as finding a semi-matching $M \subseteq E$ which has the maximal total weight from a bipartite graph $G = (U \cup V, E)$.

IV. THE PROPOSED SCHEME

The K-M algorithm [10] can be used to solve the minimum or maximum weight bipartite graph matching problem. But it requires that the number of vertices in U be equal to that in V, that is, the number of STAs be the same as the number of APs, and each AP can only serve one STA. In this situation, there is no need to execute load balancing.

In this paper, we consider two aspects in solving the aforesaid reformulated bipartite graph matching problem. Firstly, from the perspective of the network, the assignment of traffic load to APs should be appropriate according to it's real-time load condition, otherwise users connected to an overloaded AP cannot get effective and reliable communication. Secondly, from the user's perspective, the total system throughput availabla to users should be maximized by finding an optimal STA-AP association scheme. Accordingly, the proposed scheme consists of two phases: the pre-process phase and the semi-matching phase. The pre-process phase allocates the affordable number of STAs for each AP according to the load distribution of the whole network. In the semi-matching phase, we re-construct the weight matrix Q, which will be introduced below, into a square matrix, then use the K-M algorithm to find the optimal semi-matching with the maximal total weight.

A. PRE-PROCESS PHASE

Before the pre-process phase, we should determine the metric used to measure the WLAN load. In the proposed scheme, we use the CBTR as load metric. It is estimated by the average

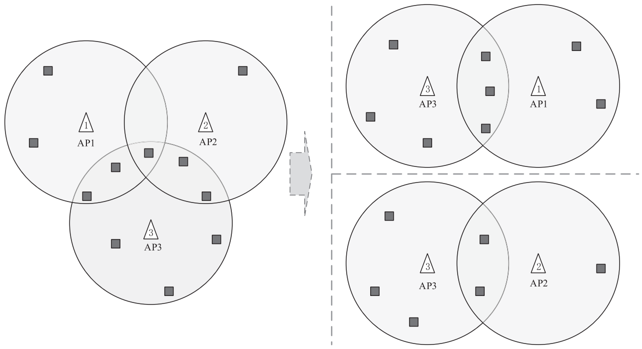


FIGURE 3. Allocating the affordable numbers of STAs for the overloaded AP and its neighboring APs.

$$\begin{bmatrix} 18.7 & 0 & 0 \\ 0 & 20.6 & 13.9 \\ 16.2 & 0 & 19.1 \\ 19.6 & 12.0 & 0 \\ 15.4 & 17.9 & 21.8 \\ 14.3 & 22.1 & 0 \\ 0 & 18.3 & 16.5 \end{bmatrix} \quad \begin{bmatrix} 18.7 & 18.7 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 20.6 & 20.6 & 13.9 & 13.9 & 13.9 \\ 16.2 & 16.2 & 0 & 0 & 19.1 & 19.1 & 19.1 \\ 19.6 & 19.6 & 12.0 & 12.0 & 0 & 0 & 0 \\ 15.4 & 15.4 & 17.9 & 17.9 & 21.8 & 21.8 & 21.8 \\ 14.3 & 14.3 & 22.1 & 22.1 & 0 & 0 & 0 \\ 0 & 0 & 18.3 & 18.3 & 16.5 & 16.5 & 16.5 \end{bmatrix}$$

(a) (b)

FIGURE 4. Original weight matrix Q and its corresponding square matrix Q'. (a) Weight matrix Q of a bipartite graph which contains 7 STAs and 3 APs. (b) Obtained square weight matrix Q'.

NAV in a specific time window. NAV is commonly used in IEEE 802.11 based WLANs [6] to avoid collision by setting a busy duration on hearing frame transmissions from other mobile hosts. Thus, the NAV can well reflect the channel busy status. Besides, the NAV information can be easily obtained by APs. In the proposed scheme, the APs (agents) periodically send the calculated CBTR information of their BSS to the load balancing controller via the southbound protocol. The controller then judges whether the load condition is unevenly distributed according to the CBTR information of the entire network. If the network is indeed load imbalanced, that is, some APs are overloaded while others are light loaded, the load balancing algorithm starts execution.

In the pre-process phase, the proposed algorithm calculates the ideal number of STAs for the overloaded APs and their neighboring APs according to APs' CBTR and the distribution of STAs. The principle is that the heavier the AP's load is, the fewer STAs the AP should be assigned to. And only the STAs located in the overlapping coverage area can be adjusted, that is, switching from the overloaded AP to the lightly loaded AP. Therefore, we assign the STAs in the overlapping coverage area in inverse proportion of the CBTR value. For example, on the left part of Fig. 3 (small squares represent the STAs), we assume that AP_3 is overloaded, then the load balancing algorithm starts execution. The load balancing controller converts the network scenario into two parts (which depends on the number of the overloaded AP's neighboring APs) and allocates the number of STAs for each AP as

$$\begin{aligned} d[AP_1] &= L_1 + \frac{CBTR_3 \times O_{13}}{CBTR_1 + CBTR_3} \\ d[AP_2] &= L_2 + \frac{CBTR_3 \times O_{23}}{CBTR_2 + CBTR_3} \\ d[AP_3] &= L_3 + \frac{CBTR_1 \times O_{13}}{CBTR_1 + CBTR_3} + \frac{CBTR_2 \times O_{23}}{CBTR_2 + CBTR_3}, \end{aligned} \tag{6}$$

where L_1, L_2 and L_3 denote the numbers of STAs which are only covered by AP_1, AP_2 and AP_3 respectively, O_{13} and O_{23} represent the numbers of STAs which are located in the overlapping coverage areas of AP_1, AP_3 and AP_2, AP_3. $CBTR_1$, $CBTR_2$ and $CBTR_3$ represent the channel busy time ratio of AP_1, AP_2 and AP_3 respectively. More generally, the allocated number of STAs of overloaded AP_i can be expressed as

$$d[AP_i] = L_i + \sum_{j=1}^{h} \frac{CBTR_j}{CBTR_i + CBTR_j} O_{ij}, \tag{7}$$

where L_i denotes the number of STAs which are only covered by AP_i and h is the total number of neighboring APs of AP_i. After the pre-process phase, we get the STA allocation vector D. It holds the allocated number of STAs for each AP, and will be used in the semi-matching phase.

B. SEMI-MATCHING PHASE

The pre-process phase allocates the affordable number of STAs for each AP from the perspective of network when the load balancing mechanism is triggered. Then the load balancing controller informs the APs (agents) to send SINR information from all the STAs to the controller. In the semi-matching phase, the proposed scheme calculates the final STA-AP association set which has the maximal total SINR. We model the STA-AP association problem as weighted bipartite graph matching problem, the edge from STA i to AP j is weighted by w_{ij} representing the uplink SINR of STA i to AP j and constituting the weight matrix Q. For example, Fig. 4 (a) shows a weight matrix of a bipartite graph which contains 7 STAs and 3 APs (the rows represent the STAs, and the columns represent the APs). According to the previous introduction, the K-M algorithm can be used to solve the weighted bipartite graph matching problem, but it can only get the optimized matching when the weight matrix is a square matrix. For non-square weight matrices, using the K-M algorithm directly is invalid. Considering that all APs have been assigned corresponding numbers of STAs in the pre-process phase, and it is obvious that $Y = \sum_{j=1}^{X} d[AP_j]$. Therefore, we can transfer the original weight matrix to a square matrix Q'. To form the square matrix Q' used in the K-M algorithm, duplicate columns should be added based on the STA allocation vector D. We should first construct STA index vector D' from D (vector D has X elements and vector D' has Y elements), where

$$D'[k] = \begin{cases} 0 & \text{if } k < D[0], \\ x & \text{if } \sum_{p=0}^{x-1} D[p] \le k < \sum_{p=0}^{x} D[p], 0 < x < X, \end{cases} \tag{8}$$

$$\begin{bmatrix} 1&0&0&0&0&0&0\\ 0&0&1&0&0&0&0\\ 0&0&0&0&0&0&1\\ 0&1&0&0&0&0&0\\ 0&0&0&0&0&1&0\\ 0&0&0&1&0&0&0\\ 0&0&0&0&1&0&0 \end{bmatrix} \quad \begin{bmatrix} 1&0&0\\ 0&1&0\\ 0&0&1\\ 1&0&0\\ 0&0&1\\ 0&1&0\\ 0&0&1 \end{bmatrix}$$

(a) (b)

FIGURE 5. Output matrix *M* of the K-M algorithm and its corresponding final matching matrix *M'*. (a) Output matrix M of Q shown in Fig. 4(a) and *D* = [2, 2, 3]. (b) Final matching matrix M'.

where k $(0 \le k < Y)$ is the index of vector D'. Then the transferred $Y \times Y$ square matrix Q' can be expressed as

$$Q'[i,j] = Q[i, D'[j]],\ (0 \le i < Y; 0 \le j < X)\,. \qquad (9)$$

It duplicates some columns of the original weight matrix Q according to the STA allocation vector D.

For example, $D = [2, 2, 3]$ and Q shown in Fig. 4 (a) can create $D' = [0, 0, 1, 1, 2, 2, 2]$ and Q' shown in Fig. 4 (b) according to equation (8) and equation (9) respectively. After obtaining the square weight matrix Q', it is used as the input of the K-M algorithm, and the square matching matrix M can be obtained. Fig. 5 (a) shows the output matrix M of Q shown in Fig. 4 (a) and $D = [2, 2, 3]$. Finally, the final matching matrix M' can be acquired by merging the previously duplicated columns according to equation (10)

$$M'[i,x] = \begin{cases} \sum\limits_{j=0}^{D[0]-1} M[i,j] & if\ x = 0, \\ \sum\limits_{j \in k} M[i,j] & if\ \sum\limits_{p=0}^{x-1} D[p] \le k < \sum\limits_{p=0}^{x} D[p], \\ & \quad 0 < x < X, \end{cases} \qquad (10)$$

where k $(0 \le k < Y)$ is the index of vector D'. Fig. 5 (b) shows the final matching matrix M' of the above example.

Algorithm 1 Maximum weight semi-matching algorithm.

Input: X, Y, weight matrix Q, STA allocation vector D;
Output: matching matrix M'.
Step 1: Calculate STA index vector D' according to D;
Step 2: Transfer matrix Q to square matrix Q';
Step 3: Call the K-M algorithm, calculate K-M(Q');
Step 4: Calculate the matching matrix M' based on the result of K-M(Q') and D'.

Algorithm 1 shows the proposed maximum weight semi-matching algorithm. It runs in the semi-matching phase, generating the final matching matrix M'. The load balancing controller then adjusts the STA-AP association scheme according to the matching matrix M', thus balancing the load distribution of the whole network and maximizing the total capacity available to users.

The APs periodically send the CBTR information of their BSS to the load balancing controller before the pre-process phase and send the SINR information from all the STAs to the load balancing controller before the semi-matching phase. Assuming that one controller manages 1024 APs, each AP serves 50 STAs, the information uploading period is 1 second, and the length of the uploaded data unit is 15 bytes. Then the controller receives 120 Kb per second before the pre-process phase and 5.8 Mb per second before the semi-matching process phase. The controller must be configured to support the corresponding computational overhead.

As for the algorithm complexity, steps 1, 2 and 4 have the complexity of $O(m^2)$, which is less than the complexity of the K-M algorithm, *i.e.*, $O(m^3)$ [22]. Therefore, the complexity of the aforementioned algorithm is $O(m^3)$.

TABLE 1. Simulation parameters.

Parameter	Value
CCA threshold	-62 dBm
Transmition power	100 mW
Pathloss exponent, δ	3.2
Minimum CW	31
Maximum CW	1023
SIFS	10 us
DIFS	50 us
Slot time	20 us
Packet size	1000 bytes
Background noise	-90 dBm
MAC header	272 bits
Bit rate for RTS/CTS/ACK	1 Mbps
Bit rate for data packets	13 Mbps
PHY header	192 bits
RTS	160 bits + PHY header
CTS, ACK	112bits + PHY header

V. SIMULATION RESULTS

This section investigates the performance of the proposed semi-matching based load balancing scheme. The entire framework for the simulation is built in MATLAB. The simulation scenario comprises 4 APs uniformly deployed in a $70m \times 70m$ area (i.e., the area is subdivided into four parts, and the APs are positioned in the center of each of these subdivisions). All the APs work on the same channel (channel 11). STAs are randomly distributed across the APs' coverage areas and generate saturated uplink traffic. The parameters used in the simulation correspond to the IEEE 802.11n standard. The detailed simulation parameters are provided in Table 1.

The following STA-AP association schemes are used for comparison purposes.

- Random association scheme: STAs randomly select the target APs to associate with. The effect of this scheme is used as baseline for comparison.
- RSSI based scheme: STAs select the target APs which provide the strongest signal when the load balancing mechanism is triggered. Similar to the proposed scheme, this scheme can be implemented in the centrally controlled WLAN architecture.

At the beginning of the simulations, STAs in the simulation network use the random association scheme to select the target APs firstly. The RSSI based scheme or the proposed scheme is triggered accordingly if the load distribution of the entire network is detected to be imbalanced. The triggered load balancing scheme then redistributes the load in the network by adjusting the STA-AP association schemes.

In addition, we use the well known Jain's fairness index introduced in [23] as balance index to reflect the degree of load balance among APs. The balance index ω is defined as

$$\omega = \frac{\left(\sum B_i\right)^2}{\left(X \cdot \sum B_i^2\right)}, \tag{11}$$

where B_i is the average throughput of AP i, and X is the total number of APs. The value of ω becomes 1 when all APs share equal load, and it approaches $1/X$ when load of APs are extremely imbalanced.

The first scenario considers the 4 APs with different numbers of STAs. The purpose is to investigate the load balancing effects of the three STA-AP association schemes under different numbers of STAs.

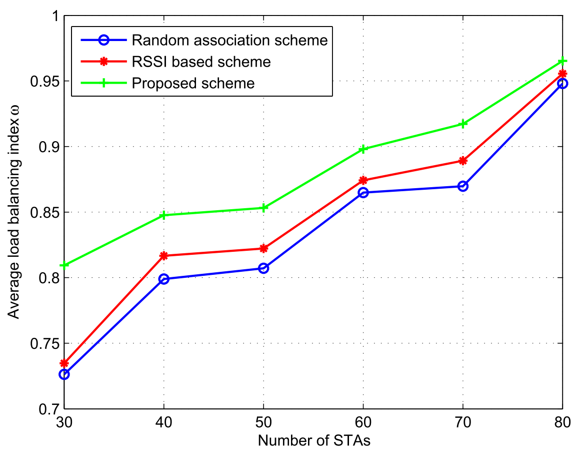


FIGURE 6. Load balancing index versus different numbers of STAs.

Fig. 6 shows the load balancing index ω versus different numbers of STAs, for each simulated scheme. In each case, the plotted ω is the average value from 100 simulation runs. It is shown that the load balancing index ω of all the three schemes increase with the number of STAs. This is because the competition between the STAs becomes more and more intense as the number of STAs grows and the uplink throughput of the four BSS gradually becomes saturated. The differences in throughput between the four APs becomes smaller, so the load balancing index becomes bigger (i.e., closer to 1). The proposed scheme and the RSSI based scheme significantly outperform the random association scheme. The reason is that when the load balancing controller perceives the imbalanced load distribution of the entire network, it immediately switches the STAs located in the overlapping coverage areas from the APs which are overloaded or saturated to the lightly loaded or non-saturated APs, thus balancing the throughput among APs. Besides, the proposed scheme outperforms the RSSI based scheme. This is because the proposed scheme can find the STA-AP association set which has the maximal total uplink SINR (i.e, the total capacity available to users is correspondingly maximized). The differences in throughput between the four APs are further reduced. Therefore, the proposed scheme has better load balancing effect.

The second scenario consists of the 4 APs and 50 STAs. Under this scenario, we can compare the effects of the three STA-AP association schemes by investigating the cumulative distribution function of the total network throughput.

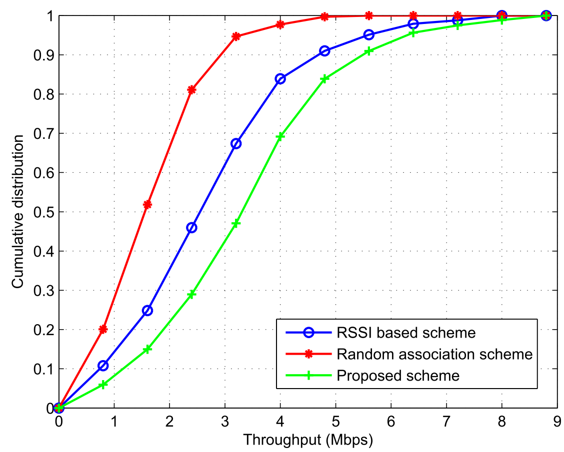


FIGURE 7. Cumulative distribution function of overall network throughput.

Fig. 7 shows the cumulative distribution function of the total network throughput under different STA-AP association schemes when the number of STAs is 50. Again, the presented data is the average from 100 simulation runs. It can be observed that the percentage gain of the proposed scheme against the baseline scheme is quite large. For instance, the throughput of the proposed scheme is about two times that of the random association scheme when the percentage is 50%, and also higher than that of the RSSI based scheme. This is because the proposed maximum weight semi-matching algorithm can find the STA-AP association set which has the maximal total capacity available to users. Therefore, the proposed scheme can not only balance the load among APs but also optimize the long-term network throughput.

VI. CONCLUSION

This paper proposes a semi-matching based load balancing scheme for allocating the load and maximizing the total network throughput in dense WLANs. The proposed scheme runs in a centralized controller and uses the CBTR information as load metric. We model the STA-AP association problem as a weighted bipartite graph matching problem and find the optimal semi-matching using the K-M algorithm. Simulation results show that the proposed scheme can effectively improve the load balancing index and overall network throughput. One of our future work is to validate the performance of the proposed scheme in multi-rate WLANs. Furthermore, we are working on realizing the proposed load balancing scheme on our SDN based WLAN experimental network.

REFERENCES

[1] H. A. Omar, K. Abboud, N. Cheng, K. R. Malekshan, A. T. Gamage, and W. Zhuang, ''A survey on high efficiency wireless local area networks: Next generation WiFi,'' *IEEE Commun. Surveys Tuts.*, vol. 18, no. 4, pp. 2315–2344, 4th Quart., 2016.

[2] T. Lei, X. Wen, Z. Lu, W. Jing, B. Zhang, and G. Cao, ''Handoff management scheme based on frame loss rate and RSSI prediction for IEEE 802.11 networks,'' in *Proc. Int. Symp. Wireless Commun. Syst. (ISWCS)*, Sep. 2016, pp. 555–559.

[3] L. H. Yen, T. T. Yeh, and K. H. Chi, ''Load balancing in IEEE 802.11 networks,'' *IEEE Internet Comput.*, vol. 13, no. 1, pp. 56–64, Jan. 2009.

[4] Y. Yukuda and Y. Oie, ''Decentralized access point selection architecture for wireless LANs deployability and robustness,'' in *Proc. IEEE 60th Veh. Technol. Conf. (VTC-Fall)*, vol. 2. Sep. 2004, pp. 1103–1107.

[5] H. Velayos, V. Aleo, and G. Karlsson, ''Load balancing in overlapping wireless LAN cells,'' in *Proc. IEEE Int. Conf. Commun.*, vol. 7. Jun. 2004, pp. 3833–3836.

[6] *IEEE Standard for Information Technology–Telecommunications and Information Exchange Between Systems Local and Metropolitan Area Networks–Specific Requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, IEEE Standard 802.11-2012, I. S. Association, 2012.

[7] K. Shen and W. Yu, ''Load and interference aware joint cell association and user scheduling in uplink cellular networks,'' in *Proc. IEEE 17th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jul. 2016, pp. 1–5.

[8] J. Andrews, S. Singh, Q. Ye, X. Lin, and H. Dhillon, ''An overview of load balancing in HetNets: Old myths and open problems,'' *IEEE Wireless Commun.*, vol. 21, no. 2, pp. 18–25, Apr. 2014.

[9] N. J. Harvey, R. E. Ladner, L. Loväsz, and T. Tamir, ''Semi-matchings for bipartite graphs and load balancing,'' *J. Algorithms*, vol. 59, no. 1, pp. 53–78, 2006.

[10] H. W. Kuhn, ''The Hungarian method for the assignment problem,'' *Naval Res. Logistics*, vol. 52, no. 1, pp. 7–21, 2005.

[11] L.-H. Yen and T.-T. Yeh, ''SNMP-based approach to load distribution in IEEE 802.11 networks,'' in *Proc. IEEE 63rd Veh. Technol. Conf.*, vol. 3. May 2006, pp. 1196–1200.

[12] F. Cao, Z. Zhong, Z. Fan, M. Sooriyabandara, S. Armour, and A. Ganesh, ''User association for load balancing with uneven user distribution in IEEE 802.11ax networks,'' in *Proc. 13th IEEE Annu. Consum. Commun. Netw. Conf. (CCNC)*, Jan. 2016, pp. 487–490.

[13] L. Du, M. R. Jeong, A. Yamada, Y. Bai, and L. Chen, ''QoS aware access point selection for pre-load-balancing in multi-BSSs WLAN,'' in *Proc. IEEE Wireless Commun. Netw. Conf.*, Mar. 2008, pp. 1634–1638.

[14] Y. Han, K. Yang, X. Lu, and D. Zhou, ''An adaptive load balancing application for software-defined enterprise WLANs,'' in *Proc. Int. Conf. Inf. Commun. Technol. Converg. (ICTC)*, Oct. 2016, pp. 281–286.

[15] M. H. Dwijaksara, W. S. Jeon, and D. G. Jeong, ''A joint user association and load balancing scheme for wireless lans supporting multicast transmission,'' in *Proc. 31st Annu. ACM Symp. Appl. Comput.*, New York, NY, USA, 2016, pp. 688–695. [Online]. Available: http://doi.acm.org/10.1145/2851613.2851852

[16] Y. Harada, H. Ono, K. Sadakane, and M. Yamashita, ''Optimal balanced semi-matchings for weighted bipartite graphs,'' *Inf. Media Technol.*, vol. 2, no. 4, pp. 1103–1112, 2007.

[17] J. Feng, L. Zhao, C. Chen, Z. Ren, and J. Du, ''User-oriented load balance in software-defined campus wlans,'' in *Proc. IEEE 83rd Veh. Technol. Conf. (VTC Spring)*, May 2016, pp. 1–5.

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

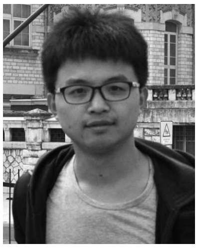
[19] C. Xu, W. Jin, G. Zhao, H. Tianfield, S. Yu, and Y. Qu, ''A novel multipath-transmission supported software defined wireless network architecture,'' *IEEE Access*, vol. 5, pp. 2111–2125, 2017.

[20] L. Tao, W. Xiangming, L. Zhaoming, Z. Xing, L. Yangchun, and Z. Biao, ''SWN: An SDN based framework for carrier grade Wi-Fi networks,'' *China Commun.*, vol. 13, no. 3, pp. 12–26, 2016.

[21] H. Q. Nguyen, F. Baccelli, and D. Kofman, ''A stochastic geometry analysis of dense IEEE 802.11 networks,'' in *Proc. IEEE INFOCOM*, May 2007, pp. 1199–1207.

[22] H. Zhu, M. Zhou, and R. Alkins, ''Group role assignment via a Kuhn–Munkres algorithm-based solution,'' *IEEE Trans. Syst., Man, A, Syst. Humans*, vol. 42, no. 3, pp. 739–750, May 2012.

[23] D.-M. Chiu and R. Jain, ''Analysis of the increase and decrease algorithms for congestion avoidance in computer networks,'' *Comput. Netw. ISDN Syst.*, vol. 17, no. 1, pp. 1–14, 1989.



TAO LEI received the B.Eng. degree in electronics and information engineering from the China University of Geosciences, Beijing, China. He is currently pursuing the Ph.D. degree in communications engineering with the Beijing University of Posts and Telecommunications. His current research interests include software defined wireless local access network, WLAN management strategies, next generation Wi-Fi technologies, and wireless network performance analysis.

XIANGMING WEN received the M.Sc. and Ph.D. degrees in information and communication engineering from the Beijing University of Posts and Telecommunications. He is currently the Director of the Beijing Key Laboratory of Network System Architecture and Convergence, where he has managed several projects related to open wireless networking. He is also the Vice President of the Beijing University of Posts and Telecommunications. His current research interests focus on radio resource and mobility management, software defined wireless networks, and broadband multimedia transmission technology.

ZHAOMING LU received the Ph.D. degree from the Beijing University of Posts and Telecommunications in 2012. He joined the School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, in 2012. His research includes open wireless networks, QoE management in wireless networks, software defined wireless networks, and cross-layer design for mobile video applications.

YUE LI received the B.Eng. degree in computer science and engineering from Northwestern Polytechnical University, Xi'an, China, in 1999, and the M.Eng. degree in applied computer from Xidian University, Xi'an, China, in 2008. He is currently with the Datang Telecom Technology Co., Ltd., as the GM of the Wireless Products Division. His research interests include cellular networks and wireless communication theory.