This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2017.2668066, IEEE Transactions on Vehicular Technology

QoS-Driven Efficient Client Association in High-Density Software Defined WLAN

Jiacheng Chen, Bo Liu, *Member, IEEE,* Haibo Zhou, *Member, IEEE,* Quan Yu, *Senior Member, IEEE,* Lin Gui, *Member, IEEE,* and Xuemin (Sherman) Shen, *Fellow, IEEE*

***Abstract*—Software defined wireless local area network (SDWLAN) has gained significant interest recently from both academic and industrial communities, and initiated a paradigmatic reconsideration on the stereotyped management and control of current WLAN due to its flexibility and programmability. However, with the proliferation of mobile devices, efficient client association with quality of service (QoS) guarantees in high-density SDWLAN is a very challenging issue. In this paper, we study the client association problem in SDWLAN with new features including centralized association, global network state awareness, seamless handoff and flow-level association. We formulate the client association in high-density scenario as an optimization problem aiming to minimize the inter packet delay of individual flows, based on an unsaturated and heterogeneous Markovian analytical model. Further, we interpret the optimization problem as an NP-hard supermodular set function minimization problem, which is solved by two low-complexity heuristic methods, namely the greedy algorithm and the bounded local search algorithm. Through simulations, we demonstrate the effectiveness of our proposed client association solution.**

***Index Terms*—Software defined networking (SDN), Software defined WLAN (SDWLAN), client association, quality of service (QoS), high-density.**

I. Introduction

FUTURE demand on high-quality wireless provisioning in small-scale hotspot areas is envisioned to take a big jump from what current wireless local area network (WLAN) can supply, especially with the emergence of cutting-edge applications such as high-definition video streaming and virtual reality. To fill the gap, evolvement on transmission technologies has been in steady progress and prospective solutions have been brought forth and standardized. However, the management and control of WLAN still remains in the original paradigm, which is inefficient in terms of resource utilization and has become the impedance for realizing advanced functionalities.

As an important part of management and control, client association in current WLAN is conducted in a spontaneous and arbitrary manner, and therefore is not globally optimal. It is often the case that some of the access points (APs) are congested while others are underutilized, which leads to inefficiency, network performance degradation, and quality of service (QoS) dissatisfaction. What is more, only very limited information is exposed to clients, which is insufficient for clients to make appropriate association decisions. The information can even be misleading in certain cases. For example, in the enterprise scenario, an AP usually shares the backhaul with many other APs, femto base stations and desktops. Therefore, the effective backhaul capacity of the AP can become limited and bottleneck the wireless transmission. However, clients are unaware of the backhaul of APs and may connect with an AP with high signal strength but limited backhaul capacity. Moreover, current association mechanism cannot react to the network dynamics swiftly and efficiently due to the lack of flexibility, which is reflected in the following aspects: i) only client-level association is supported; ii) one client can only be associated with one AP at the same time; and iii) re-association can cause service interruption. On the contrary, a carefully designed client association scheme within a more flexible and powerful management paradigm is able to take advantage of the densely deployed APs and enhance WLAN in an easy-to-implement way that requires no change at PHY and MAC layers. Towards efficient client association in WLAN, software defined networking (SDN) [1] promises to deliver more flexible and manageable networks and provide diverse QoS guarantees through the separation of control plane and data plane. Such an integration of SDN and WLAN is referred to as software defined WLAN (SDWLAN).

As a newly emerging technology, SDWLAN still faces many challenges with respect to client association. Fig. 1 illustrates the basic idea of client association in SDWLAN. As shown in the figure by solid arrows, the controller is responsible for the client association through interactions with the APs and clients. Also, the dotted arrows describe the general process for client association. Since the above process is quite different from the current WLAN, the first challenge is how to exploit the advanced features of SDWLAN so as to enhance client association. Realistic and useful functionalities need to be first identified from the existing SDWLAN prototypes, and then integrated with the client association workflow and utilized during the decision-making process. The second challenge is how to efficiently conduct client association in the high-density scenario, which is common in future wireless networks [2] and has appealed to both academia and industries [3][4][5]. With limited orthogonal channels in the unlicensed

This work is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) and Huawei Technologies Canada.

J. Chen, Q. Yu and L. Gui are with the Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: jiacheng1989@gmail.com; yuquan61@qq.com; guilin@sjtu.edu.cn).

B. Liu is with the School of Information Technology, Deakin University, Burwood VIC 3125, Australia (e-mail: bo.liu@deakin.edu.au).

H. Zhou and X. Shen are with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (email: haibozhouuw@gmail.com; xshen@bbcr.uwaterloo.ca). H. Zhou is the corresponding author.

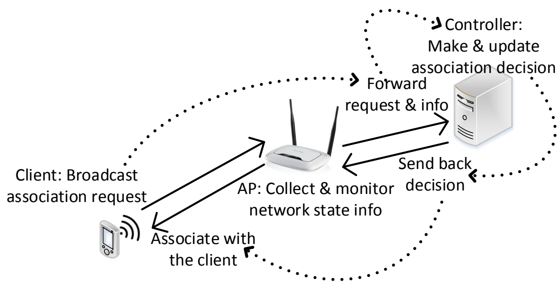


Fig. 1: Basic idea of client association in SDWLAN.

bands (e.g., only three orthogonal channels in the 2.4GHz band), some of the APs have to work on the same channel. Therefore, simply associating clients with less loaded APs working on different channels is not optimal enough and cannot take full advantage of the densely deployed APs. The last challenge is how to enhance the QoS provision capability of SDWLAN with client association. Future applications tend to have higher requirements on QoS. Also, a client may have several concurrent QoS-sensitive services. Each of them needs to be treated separately.

In order to address the aforementioned challenges, this paper studies the QoS-driven efficient client association in high-density SDWLAN scenario. To begin with, we identify the new features of SDWLAN, which differs the client association in SDWLAN from that in current WLAN. The feasibility of these features in practice has been demonstrated by prototypes of SDWLAN. However, how to utilize the new features to develop a novel client association process has not been studied by literatures so far. The first feature is *centralized client association*, which means association decisions for all the clients are made by the SDWLAN controller. This feature is due to the fact that data flows in SDWLAN are centrally controlled by the controller. The second feature is *global network state awareness* of the controller, which realizes global optimization of client association with the aid of network monitoring techniques. In this paper, the network state mainly refers to the data rates of links between clients and APs, the data rates of services requested by the clients and the backhaul capacity of APs. The third feature is *flow-level association*, owing to the flexible flow operations and control by SDWLAN controllers. Specifically, different flows of the same client can be associated with different APs simultaneously, thereby realizing the dedicated management of each individual flow. The last feature is *seamless handoff*, which means association decisions can be updated timely through re-association without significant performance degradation and service interruption. The seamless handoff property also reveals the potential for adopting SDWLAN in the high-mobility scenario, such as vehicular networks [6].

With the above features, we further consider client association in the high-density scenario. In this paper, "high-density" mainly implies the overlapped basic service set (OBSS) case, in which all the APs and clients are located within a certain coverage area and operate on the same channel. Therefore, OBSS can represent the high-density scenario in terms of both spatial and frequency perspectives. Then, client association is formulated as an optimization problem, with the objective of minimizing the inter packet delay of individual flows. Such an objective aims to improve the QoS provision by reducing the chance of service interruption and flow suspension. The problem formulation is based on the derivation of inter packet delay from an appropriate analytical model [7], which is an extension of Bianchi's Markov model [8] in the unsaturated and heterogenous scenario. The formulated problem is further interpreted as an NP-hard supermodular set function minimization problem, and two heuristic algorithms are developed as the solutions, namely the greedy algorithm and the bounded local search algorithm. In summary, the contributions of this paper are listed as below:

- We present the system model of SDWLAN and the workflow of client association in SDWLAN, which not only shed light on how advanced SDWLAN features can be utilized to improve client association, but also exemplify how traditional network management and control can be enhanced with the new SDN paradigm.
- Based on the derivation of inter packet delay from an unsaturated and heterogenous Markovian analytical model, we formulate a novel flow-level delay-minimized client association problem in high-density SDWLAN, so as to improve the QoS of each individual application in the resource-restricted scenario.
- In order to find an efficient and scalable way to solve the problem, we interpret the formulated problem into an NP-hard supermodular set function minimization problem, and thereby develop two sub-optimal low-complexity solutions, namely the greedy algorithm and the bounded local search algorithm.

The remainder of the paper is organized as follows. Section II introduces the literatures related to this paper. System model of SDWLAN and the workflow of client association are presented in Section III. Problem formulation of the client association in SDWLAN is given in Section IV, which is followed by the development of heuristic solutions in Section V. Simulation results are provided in Section VI and conclusion of the paper is drawn in Section VII.

II. Related Works

A. SDN

SDN [1] is an emerging networking paradigm originated from McKeown's seminal work [9]. The main concept of SDN is the separation of control plane and data plane in the network layer, compared with traditional networks where control and forwarding functionalities are coupled within the switches. Specifically, the network flows are controlled by the centralized controller, in which various network applications (e.g., routing) can be deployed. Therefore, the switches only need to implement packet processing functions (e.g., forwarding, dropping) based on the rules set by the controller. OpenFlow [9] is the *de facto* communication protocol between the controller and switches. It standardizes the flow table formats of the switches such that packets can be processed in a "match-action" manner. The benefits of SDN include but not limited to: network programmability and flexibility, robustness, vendor neutralization and global optimality. Due to these advantages, SDN has been successfully adopted in wired networks, and recently is shifted to the wireless networks such as wireless mesh network [10], vehicular network [11][12], and WLAN.

B. Implementation of SDWLAN

Since the emergence of SDN, researchers have been focusing on realizing the advanced SDN features in WLAN. Recently, several SDWLAN systems have been designed and prototyped to demonstrate the features. In [13], Yap *et al.* present OpenRoads and develop a wireless extension of OpenFlow. Odin [14] is another implementation of SDWLAN, which is built on the concept of light virtual AP (LVAP) abstraction. In CloudMAC [15], MAC frames are generated and processed by the virtual APs (VAPs) residing in datacenters. Hence, physical APs only need to forward these MAC frames. BeHop [16] utilizes SDN to build a testbed for dense Wi-Fi networks, in which APs only implement low-level managements and time-sensitive tasks. In COAP framework [17], Patro and Banerjee propose a set of SDN-based Open APIs for home WLAN management so as to achieve coordination among APs through centralized control. In the following, we give more details on how the new features of SDWLAN are implemented.

1) Centralized Client Association: In [18], Murty *et al.* configure the APs so that they only send beacons with hidden SSIDs. Each AP maintains a local access control list containing the MAC addresses of the clients that are allowed to associate with the AP and only responses to the probe requests from the clients in its list. Similarly, in [16], the white-listing mechanism of commercial off-the-shelf APs is utilized to implement the centralized client association. After the white list of each AP is set by the controller, a client can only associate with the AP whose list contains it.

2) Global Network State Awareness: Many existing open-source utilities can support the monitoring and gathering of network state information. For example, packet-level statistics can be obtained through Wireshark; in-frame information delivery can be realized by Radiotap headers; and flow type identification can be achieved via deep packet inspection (DPI). In [17], the authors use Click to implement link-level statistics gathering. [19] utilizes RFlow for client traffic monitoring and Wiviz for channel information collection. The collector module in [16] is able to collect data from three sources: Wi-Fi statistics, raw packets with radio metadata and traffic on the wired backhauls.

3) Seamless Handoff: Seamless handoff can be achieved in SDWLAN through AP virtualization. One of the representative works is Odin [14], which builds an LVAP for each client in the physical APs. In this way, each client is given the illusion that it always connects to the same AP. Then, handoff between APs can be realized through moving the client's LVAP from the old physical AP to the new one. A similar method is proposed in CloudMAC [15], where the state information of associations is kept in the VAPs residing in datacenters and physical APs are only responsible for packet transmissions on the wireless interface. Hence, handoff simply requires changing the forwarding rules of the OpenFlow switches.

4) Flow-level Association: Generally, the controller can control the forwarding paths of the flows through setting the flow tables of the switches. Also, APs can know the destinations of flows from the controller and then transmit flows to corresponding clients. In OpenRoads [13], a client can be associated with two or even more APs, forming the bi-casting or n-casting links. [20] utilizes a group of APs to communicate with a certain client, which is referred to as "AP diversity".

C. Analytical Model of WLAN

Performance analysis of IEEE 802.11 distributed coordination function (DCF) has been well studied in the past decades. Bianchi's model [8] is the most influential one, which adopts Markov chains to model the possible states of a node. However, several important assumptions are made in the model such that it cannot be used in many practical scenarios. Therefore, many extensions of Bianchi's model have been created. In [21], the authors analyze the delay in the IEEE 802.11e enhanced distributed channel access (EDCA) scenario. Malone *et al.* consider nonsaturated and heterogenous conditions in [22] by adding post-backoff states to the original Markov model. Felemban and Ekici add a channel state Markov chain in [23] so as to improve the accuracy of backoff process modeling. Aurzada *et al.* consider the frame aggregation feature in IEEE 802.11n and extend the analysis of Bianchi's model in [24]. In [7], Gong and Yang present another nonsaturated and heterogenous Markov model, which is the one we adopt in this paper. On the one hand, the clients in our scenario are also nonsaturated and have various sizes of flows. On the other hand, compared with the other extensions mentioned above, it is the most appropriate model for the client association problem formulation, in terms of complexity and accuracy. Based on the node transmission probability given by the model, we further derive the average inter packet delay of individual flows.

D. Client Association Optimization

Client association optimization becomes a hot topic recently in both WLAN and small cell networks, since it can improve the network performance in dense deployment scenarios under the current transmission technologies. In [18], Murty *et al.* present client association and load balancing mechanisms based on the "Available Capacity" metric. Bejerano *et al.* achieve max-min fairness client association in [25]. In [26], Broustis *et al.* propose measurement-driven guidelines for WLAN designs, which include channel assignment of APs, client association and power control of APs. In [7], an online client association algorithm is proposed with the purpose of maximizing MAC efficiency. An access link virtualization approach is developed in [27] in order to let the clients make use of the pooled Wi-Fi resources. Backhaul-awareness client association is considered in [28], [29] and [30]. In [28], a delay-based (including both wireless and backhaul delays) access control policy is proposed in the heterogeneous network

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2017.2668066, IEEE Transactions on Vehicular Technology

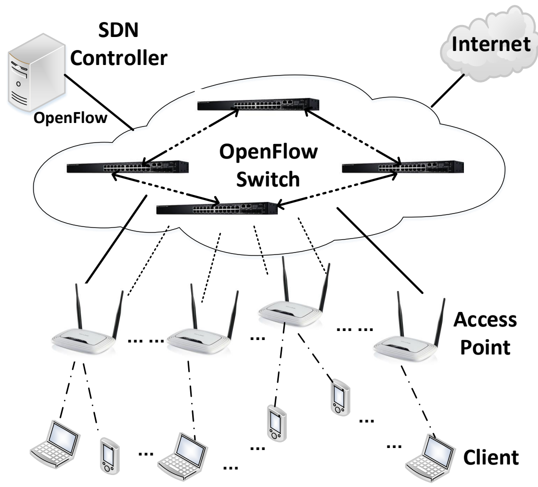


Fig. 2: An illustrative architecture of SDWLAN.

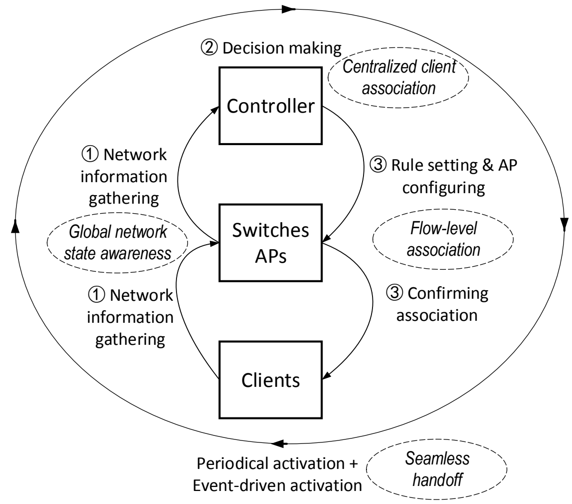


Fig. 3: Workflow of centralized client association in SDWLAN.

scenario based on a hierarchical network model. Beyranvand *et al.* study the client association in fiber-wireless (FiWi) enhanced LTE-A heterogeneous networks in [29] and take into account the delay and reliability of fiber backhauls. In [30], Chai and Shin propose to aggregate backhaul links of multiple APs by associating a client with different APs in different periods. Different from the above works, this paper studies client association in the SDWLAN scenario by considering the new SDN-enabled features. Also, we focus on client association in OBSS, which is a common case in high-density WLAN. Moreover, we aim to improve the QoS of individual flows in SDWLAN through the minimization of average inter packet delay.

III. System Model

Fig. 2 shows an illustrative architecture of the SDWLAN. The OpenFlow switches are controlled by the SDN controller such that traffic can be routed to the APs based on the rules. Specifically, a packet coming from the Internet first arrives at an OpenFlow switch, and is then forwarded it to other switches and finally to the destination AP based on the corresponding flow table entries set by the controller. If it is the first packet of a flow, then it is first sent to the SDN controller, which then forwards the packet as well as the flow table settings to all the related switches. Packet transmission between APs and clients is the same as that in current WLAN.

The workflow of the centralized client association process in SDWLAN is illustrated in Fig. 3. In the first step, due to the global network state awareness feature of SDWLAN, the controller gathers all the necessary information from other components in the system. In this paper, the information mainly refers to the wireless channel quality of each client-AP pair, the backhaul link capacity of each AP and the data rate of each flow. APs act as the gateway for the information gathering of clients, which cannot contact the controller directly. Then, with the global information, the controller makes the client association decision in a centralized manner. In the third step, the controller distributes the decision to the data plane, i.e., setting the forwarding rules of each switch and configuring each AP. Also, each client gets response from the AP that it is assigned to such that the association is confirmed. In order to adapt to the dynamics of the network, the above association process is activated periodically and can be triggered by certain events. With the seamless handoff feature of SDWLAN, the connection state of each client can be maintained all the time, i.e., changes in association decisions are transparent to the ongoing services.

In our scenario, due to the flow-level association feature of SDWLAN, a client can upload to one AP and download from another AP at the same time. Hence, it can be decomposed as an upload client and a download client. We also consider flow-level routing for the download traffic, so a download client can be associated with multiple APs, each of which transmits a fraction of the original traffic. Actually, since a download client does not really transmit, it can be equivalently regarded as several dummy download clients, each being associated with one AP and receiving corresponding flows. Specifically, we consider the case that each dummy client only has one flow (here we assume that a single flow cannot be split). Note that if flow-level association is disabled, then all the dummy clients should be associated with the same AP.

In this paper, we consider the OBSS scenario where there exist n co-channel transmitting nodes consisting of n_u upload clients and n_a APs, which deliver packets to n_d download clients, i.e., $n = n_u + n_a$. Let A, N_u and N_d represent the set of APs, upload clients and download clients respectively (the cardinalities are n_a, n_u and n_d). We also define N_u^a and N_d^a as the upload and download clients associated with AP a. According to the above analysis, we have $N_u \cap N_d = \emptyset$. Also, N_d and N_d^a represent the sets of dummy download clients instead of the real ones. Throughout the analysis, a can only refer to an AP while i, j can refer to either a general node or a client.

IV. Problem Formulation

In this section, we will formulate the centralized client association as an optimization problem based on a Markovian analytical model. Specifically, the objective is to derive the mean inter packet delay of download flows from the model and minimize it in the formulated problem, where the backhaul link capacity of each AP will be considered as the main constraint.

A. Markovian Analytical Model of WLAN

The IEEE 802.11 DCF is featured by the random backoff process and can be characterized by a Markovian analytical model. We adopt the two-dimension Markov model in [7] to help describe the behavior of each node in the network. A brief review of the model is given below. The maximum number of backoff stages is m, so a node i can be in the backoff stage $0, 1, 2, ..., m$. The contention window size of a backoff stage l can be derived as per the binary exponential backoff mechanism, i.e., $W_l = 2^l W_0$, where W_0 is the minimum contention window size CW_{min} and $W_m = 2^m CW_{min}$ is the maximum contention window size CW_{max}. A backoff state in the model is represented by (l, k), where the backoff counter $k \in \{0, 1, 2, ..., W_l - 1\}$. There is also a special state $(-1, 0)$, which represents the state that a node has no packet to transmit. After a successful transmission, each node returns to this state and waits until the arrival of the next packet that is to be transmitted. The stationary distribution of each state of node i is denoted by $b_{l,k}(i)$ and can be solved from the properties of the Markov chain. The transmission probability of node i at the beginning of each state is denoted by τ_i and can be known with $b_{l,k}(i)$.

We consider both packet collision and channel error, so the packet error probability of node is given by:

$$p_i = 1 - (1 - p_{ci})(1 - p_{ei}) \tag{1}$$

where p_{ci} is the conditional packet collision probability and p_{ei} is the packet error probability due to channel error. Then, the conditional successful transmission probability of node i is given by:

$$p_{si} = 1 - p_i = (1 - p_{ei}) \prod_{j \in A \cup N_u, j \neq i} (1 - \tau_j) \tag{2}$$

We use q_i to denote the packet arrival probability of node i, i.e., the probability that node i has a packet to transmit at state $(-1, 0)$. For $q_i, i \in N_u$, under the Poisson process model, we have:

$$q_i = 1 - e^{-\lambda_i E_s} \tag{3}$$

where E_s is the mean state length and λ_i is the packet arrival rate (in number of packets per unit time). For AP a, the corresponding packet arrival probability is:

$$q_a = 1 - \prod_{i \in N_d^a} (1 - q_i) \tag{4}$$

which also gives:

$$\lambda_a = \sum_{i \in N_d^a} \lambda_i \tag{5}$$

B. Calculation of Mean State Length

During a state transition, three events can happen: the channel being idle, only one node transmitting (i.e., successful transmission of a packet if there is no failure due to channel error) and more than one nodes transmitting (i.e., collision). In the first event, the state length is equal to the backoff slot length σ. Also, the probability of the event is:

$$P_{idle} = \prod_{i \in A \cup N_u} (1 - \tau_i) \tag{6}$$

Hence, the event length (i.e., the state length multiplied by event probability) is:

$$T_{idle} = P_{idle} \sigma \tag{7}$$

The IEEE 802.11 protocol has both time and bit overheads. It also has two access mechanisms, i.e., basic mode and RTS/CTS mode. We use $T_{o,s}$ and $L_{o,s}$ ($T_{o,c}$ and $L_{o,c}$) to respectively represent the time and bit overheads in a successful transmission (collision) for both access modes (mode is labeled on the superscript when necessary). Details of these overheads can be found in [8]. Then, in the second event, the mean length $T_{tr,s}$ is given as:

$$\begin{aligned} T_{tr,s} = & \sum_{i \in N_u} \tau_i \frac{p_{si}}{(1-p_{ei})} \left(T_{o,s} + \frac{\overline{L}_i + L_{o,s}}{r_i} \right) \\ & + \sum_{a \in A} \tau_a \frac{p_{sa}}{(1-p_{ea})} \sum_{i \in N_d^a} h_i^a \left(T_{o,s} + \frac{\overline{L}_i + L_{o,s}}{r_i} \right) \end{aligned} \tag{8}$$

where $\overline{L}_i$ is the average payload size in bit, r_i is the date rate between node i and its associated AP, and h_i^a denotes the probability that the packet arriving at AP a belongs to client i. Assuming that h_i^a is proportional to the packet arrival rate of i, i.e., λ_i, we have:

$$h_i^a = \frac{\lambda_i}{\sum_{j \in N_d^a} \lambda_j} \tag{9}$$

In the third event, the calculation of mean length $T_{tr,c}$ is slightly different for the two access methods, as shown below:

$$T_{tr,c}^{bas} = \sum_{N_c} P_{c,N_c} \cdot \left(T_{o,c}^{bas} + \max_{i \in N_c} \left\{ \frac{\overline{L}_i + L_{o,c}^{bas}}{r_i} \right\} \right) \tag{10}$$

$$T_{tr,c}^{rts} = \sum_{N_c} P_{c,N_c} \cdot \left(T_{o,c}^{rts} + \max_{i \in N_c} \left\{ \frac{L_{o,c}^{rts}}{r_i} \right\} \right) \tag{11}$$

where N_c is the set of nodes involved in a collision and P_{c,N_c} is the corresponding probability. N_c should satisfy:

$$\begin{aligned} & N_c \subseteq \left(\bigcup_{a \in A} N_d^a \right) \cup N_u \ , \ |N_c| \geq 2 \ , \\ & \forall i, j \in N_c \ , \ i \in N_d^a \ , \ j \in N_d^{a'} \ : \ a \neq a' \end{aligned} \tag{12}$$

Let a_i denote the AP serving client i and A_c denote the set of APs involved in the collision, i.e., $A_c = \{a_i : i \in N_c - N_u\}$. Then, P_{c,N_c} can be calculated:

$$\begin{aligned} P_{c,N_c} = & \prod_{i \in N_c \cap N_u} \tau_i \cdot \prod_{i \in N_u - N_c} (1 - \tau_i) \cdot \\ & \prod_{i \in N_c - N_u} \tau_{a_i} h_i^{a_i} \cdot \prod_{a \in A - A_c} (1 - \tau_a) \end{aligned} \tag{13}$$

With the three event lengths calculated above, the mean state length E_s can be given by:

$$E_s = T_{idle} + T_{tr,s} + T_{tr,c} \tag{14}$$

Also, the mean throughput of node i is the mean successfully transmitted bits during the mean state length, namely:

$$S_i = \frac{\tau_i p_{si} \overline{L}_i}{E_s} \tag{15}$$

C. Delay Minimization Problem Formulation

In this paper, we focus on the inter packet delay of download flows. First, we suppose that a packet of $i \in N_d^a$ has arrived at a and calculate the access delay of the packet. The duration of a successful transmission is (applicable for both basic access method and RTS/CTS method):

$$\Delta_{s,i,a} = T_{o,s} + \frac{\overline{L}_i + L_{o,s}}{r_i} \tag{16}$$

The duration of a collision involving i is denoted as $\Delta_{c,i,a}$ and it can be calculated from Eq. (10) and Eq. (11), with N_c being the set of colliding nodes containing i. Hence, the access delay for i is:

$$\begin{aligned} \Delta_{i,a}^{access} &= \sum_{l=0}^{\infty} {p_a}^l (1 - p_a) \\ &\cdot (\Delta_{s,i,a} + l\Delta_{c,i,a} + \sum_{b=0}^{l} \tfrac{2^{\min\{b,m\}} W_0 - 1}{2} {E_s}') \end{aligned} \tag{17}$$

where E_s' is the mean state length when a is not transmitting. The mean access delay for AP a is given by:

$$\Delta_a^{access} = \sum_{i \in N_d^a} h_i^a \Delta_{i,a}^{access} \tag{18}$$

Then, we can calculate the waiting delay, namely the duration between the last successful packet transmission and a new packet arrival, i.e.,

$$\Delta_a^{wait} = \frac{1}{q_a} {E_s}' \tag{19}$$

Therefore, the delay between two consecutive successful packet transmissions is:

$$\Delta_a = \Delta_a^{access} + \Delta_a^{wait} \tag{20}$$

For $i \in N_d^a$, it takes average $1/h_i^a$ times for the AP to transmit its packet, so the delay between two consecutive packets of i is $1/h_i \cdot \Delta_a$. Therefore, the mean inter packet delay for the clients associated with AP a is:

$$\Delta_{pkt,a} = \sum_{i \in N_d^a} h_i \frac{1}{h_i} \Delta_a = |N_d^a| \Delta_a \tag{21}$$

We use $X = \{x_{i,a}\}$ as the indicator for association between clients $i \in N_u \cup N_d$ and APs $a \in A$. Under a given X, the association between clients and APs can be known, and the delay can be calculated from the above analysis. The objective is to find an association X that minimizes the sum of all the APs' average inter packet delay. The formulation of the problem is given as follows:

objective:

$$X = \arg\min \left\{ \sum_{a \in A} \Delta_{pkt,a} \right\} \tag{22}$$

subject to:

$$x_{i,a} \in \{0, 1\} \tag{23}$$

$$\sum_{a \in A} x_{i,a} \le 1 \tag{24}$$

$$x_{i,a} r_{i,a} > 0,\ \forall\, x_{i,a} > 0 \tag{25}$$

$$\lambda_a \le B_a \tag{26}$$

where $r_{i,a}$ is the data rate between AP a and client i and B_a is the downlink backhaul capacity of AP a. Constraint (23) corresponds to the assumption that a single flow cannot be split. Constraint (24) indicates a client cannot be associated with multiple APs, since we have assumed that each client only has one flow. Constraint (25) means a client cannot be associated with an AP when the data rate between them is 0. Constraint (26) suggests the total download traffic of an AP should be less than its backhaul capacity.

V. Client Association Algorithms

In this section, we first introduce how to solve the Markov model. Then, the presented optimization problem is interpreted in another way and the NP-hardness property is shown. Accordingly, two heuristic algorithms are proposed, namely the greedy method and the local search method. At last, some discussions on the centralized client association algorithms are given.

A. Solving the Non-linear System

Basically, in order to calculate the delay, we should at least obtain the values of transmission probabilities τ_i for $i \in A \cup N_u$. This requires to solve at least n non-linear equations. Due to the complexity of these equations, we choose to solve the non-linear system containing $2n + 1$ equations, i.e., equations for τ_i for n nodes, equations for p_{si} for n nodes, and equation for E_s. For simplicity, all the equations are represented in the form of $f(\cdot) = 0$ and the system is represented as $\mathbf{F}(\cdot) = \mathbf{0}$. By solving the system, variables τ_i, p_{ci} for $i \in A \cup N_u$ and E_s can be obtained. In this paper, we use Newton's method to solve the above non-linear system since it is known to have fast convergence speed. As an iterative method, Newton's method ends when the $l2$-norm of $\mathbf{F}(\cdot)$ is less than the accuracy control variable ϵ. Also, we make sure that the obtained solution is the desired one (i.e., not other local optima which are usually invalid) by carefully selecting an appropriate starting point.

B. Problem Interpretation

In this subsection, we interpret the problem formulated in Sec. IV into supermodular set function optimization, which is a powerful method and has been used in similar scenarios [31][32]. The original objective (22) is minimizing the sum of average inter packet delay of all APs. For the convenience of problem interpretation, we slightly change it into minimizing

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2017.2668066, IEEE Transactions on Vehicular Technology

the sum of inter packet delay of all download clients. Also, we interpret the variables X in another way. Note that for the upload clients in our scenario, what they can do to minimize the total delay is only to reduce the channel occupation time when it has the transmission opportunity. Therefore, we adopt the maximum received signal strength indicator (RSSI) association method for the upload clients. In the following, we thus only consider the association of download clients. We define the universe set $\mathcal{U} = \{u_{i,a}\}$ for $i \in N_d$ and $a \in A$. Each element in $\mathcal{U}$ represents a feasible association between a download client and an AP. Also, the objective is defined as a set function on the power set of $\mathcal{U}$, denoted as $f : 2^{\mathcal{U}} \rightarrow R^+$. So our job becomes to find $U = \arg\min_{U \in 2^{\mathcal{U}}} f(U)$ satisfying certain constraints, which is described in the following proposition:

Proposition 1: The problem described above is a non-decreasing supermodular function minimization problem satisfying cardinality constraint, partition matroid constraint and knapsack constraint.

We first give some explanations on the terms mentioned in the above proposition [33][34]. A set function is a function that is defined on a family of sets (e.g., a subset $\mathcal{I}$ of power set $2^{\mathcal{U}}$). A non-decreasing set function (also called a monotone set function) satisfies $f(A) \geq f(B)$ for $A \supseteq B$. A supermodular function (i.e., negative of a submodular function) has the following property:

$$f(A) + f(B) \leq f(A \cap B) + f(A \cup B) \tag{27}$$

It can also be written equivalently as below:

$$f(A \cup \{e\}) - f(A) \geq f(B \cup \{e\}) - f(B) \tag{28}$$

for $A \supseteq B$ and $e \in 2^{\mathcal{U}} \backslash A$. Cardinality constraint requires the number of elements of a solution U be limited, i.e., $|U| \leq k$. Partition matroid constraint means U should belong to a partition matroid. For a matroid $\mathcal{M} = (\mathcal{U}, \mathcal{I})$, $\mathcal{U}$ is partitioned into ℓ non-intersecting subsets $U_1, U_2, ...U_\ell$ with associated integers $k_1, k_2, ..., k_\ell$. A set $U \subseteq \mathcal{U}$ belongs to the partition matroid (i.e., $U \in \mathcal{I}$) iff $|U \bigcap U_i| \leq k_i$. Knapsack constraint requires the value of $\sum_{u \in U} c(u)$ be less than a constant value, where $c(\cdot)$ is a function defined on $\mathcal{U}$.

Now we give proofs on Proposition 1. Here we define $f(\emptyset) = 0$. As more clients are associated with APs, the total delay gets larger, so the objective function is non-decreasing. If a new client is added, the delay of all the existing clients also increases due to contention, queuing and so on. Therefore, the more clients existing in the network, the larger the total delay becomes when a new client is added. From property (28), the objective function is supermodular. Since all the clients will be associated with APs at the end, we have $|U| \leq N_d$ and the cardinality constraint exists. We partition the universe $\mathcal{U}$ into $|N_d|$ subsets $\mathcal{U}_1, \mathcal{U}_2, ...\mathcal{U}_{|N_d|}$, where $\mathcal{U}_i$ contains $u_{i,a}$ for $a \in A$. Hence, a feasible association U satisfies $U \bigcap \mathcal{U}_i \leq 1$ because a client can only be associated with one AP, which gives the partition matroid constraint. Knapsack constraint corresponds to constraint (26), i.e., the sum of download flows should not exceed the backhaul capacity of the AP that they are associated with.

It is known that submodular function maximization problem and its counterpart supermodular function minimization problem are NP-hard [33][35]. Therefore, we adopt low-complexity heuristic algorithms for the download client association. In our algorithms, the knapsack constraint is relaxed through the following relationship:

$$\lambda_a = \min \left\{ \sum_{i \in N_d^a} \lambda_i, B_a \right\} \tag{29}$$

namely, the backhaul capacity constraint is represented by limiting the flow arrival rate of the AP. In this way, the inter packet delay of download flows will become larger due to the limited backhaul capacity, which serves as the penalty for violating the constraint.

C. Heuristic Algorithms

We first present a basic greedy algorithm, as shown in Alg. 1. In each iteration, a client is associated with an AP that minimizes the total delay. The loop ends until all the clients are associated. It can be seen that the algorithm has $|N_d|$ iterations, each of which contains at most $|\mathcal{U}| = |N_d||N_a|$ computations of $f(U_i \cup \{u_{j,a}\}) - f(U_i)$, so the complexity of greedy association algorithm is $O(|N_d||N_a|^2)$.

Algorithm 1: GREEDY ASSOCIATION ALGORITHM

Input: $\mathcal{U}$
Output: U
$i = 0$, $U_0 = \emptyset$
while $\mathcal{U} \neq \emptyset$ **do**
  $u_{j,a} \leftarrow \arg\min_{u_{j,a} \in \mathcal{U}} \{f(U_i \cup \{u_{j,a}\}) - f(U_i)\}$
  $U_{i+1} \leftarrow U_i \bigcup \{u_{j,a}\}$
  $\mathcal{U} \leftarrow \mathcal{U} - \{u_{k,a} : u_{k,a} \in \mathcal{U}, k = j\}$
  $i \leftarrow i + 1$
return U_i

Next, based on the algorithm given in [35], we present another polynomial-time and bounded local search method, as shown in Alg. 2. The algorithm requires an initial feasible association and an accuracy control variable $\varepsilon \in (0, 1)$ as inputs. We adapt the client association to the algorithm in [35] since it only considers cardinality constraint. Specifically, when calculating θ, we assume that for every download client, there exist $|N_a|$ identical clients, each of which is associated with a different AP respectively. In this way, the partition matroid constraint is not considered temporarily and $f(\mathcal{U})$ can be calculated. Afterwards, the local search is conducted—in each iteration, one element v in the original association U_i is replaced by another u. The selection of u and v aims at minimizing the total delay. Note that the partition matroid constraint should be complied with, i.e., u and v should belong to the same partition set $\mathcal{U}_j$. The local search terminates when the reduced total delay is small enough. In each iteration, there are $|N_d|(|N_a|-1)$ computations of $f(U_i + \{u\} - \{v\})$. According to [35], the algorithm ends within at most $O((|N_d||N_a|)^3)$ iterations. Hence, the complexity of the local search association

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2017.2668066, IEEE Transactions on Vehicular Technology

is $O((|N_d||N_a|)^4)$. Also, the original algorithm in [35] has a bounded approximation ratio $\frac{1}{1-\varepsilon}\left[1+\frac{\theta}{(1-\theta)^2}\right]$. Due to the partition matroid constraint, the solution space of our problem is a subset to that of the original problem. Therefore, our algorithm at least has the same bounded approximation ratio. In other words, for our solution $\widetilde{U}$ and the optimal one U^*, we have:

$$f(\widetilde{U}) \le \frac{1}{1-\varepsilon}\left[1+\frac{\theta}{(1-\theta)^2}\right] f(U^*) \tag{30}$$

Algorithm 2: LOCAL SEARCH ASSOCIATION ALGORITHM

Input: $\mathcal{U}, U_0, \varepsilon$
Output: U

$\theta \leftarrow \max_{u\in\mathcal{U}}\left\{1-\frac{f(\{u\})}{f(\mathcal{U})-f(\mathcal{U}-\{u\})}\right\}$, $i \leftarrow -1$
repeat
    $i \leftarrow i+1$
    $(u,v) \leftarrow \arg\min_{u,v\in\mathcal{U}_j,\forall j\in N_d,u\in\mathcal{U}\backslash U_i,v\in U_i}\{f(U_i+\{u\}-\{v\})\}$
    $U_{i+1} \leftarrow U_i+\{u\}-\{v\}$
until $f(U_i)-f(U_{i+1}) \le \frac{(1-\theta)\varepsilon}{n_a n_d} f(U_i)$
return U_{i+1}

D. Discussions

In this subsection, several issues related with the centralized client association algorithms are briefly discussed. The first issue is about scalability, which is a problem that all centralized methods will encounter. A feasible solution [27] is to conduct the centralized association only for the "elephant" flows (e.g., video streaming). In practice, one client usually does not have several concurrent "elephant" flows, and hence such a solution can greatly reduce the computation load of the centralized method. For the "mice" flows, they can have the same APs with the flows of the same client, or just use a simple maximum RSSI association. The involved complexity-performance compromise is acceptable since "mice" flows cannot significantly influence the network. A more robust method is to build a virtualized network for these "mice" flows and allocate dedicated resources for them by slicing the Wi-Fi network.

The second issue is about the network dynamics since the centralized client association is based on the network "snapshot", i.e., the current network state. In SDWLAN, the inter-AP handoff overheads are trivial due to the seamless handoff feature. Also, the controller has the global view of the realtime network state through monitoring. Therefore, two methods can be adopted to make client association adapt to the dynamic change of network topology. In the first method, the controller periodically updates the association, similar to [19]. The length of the period is dependent on the compromise between the timeliness of the association and the computation costs. In the second method, the update of client association can be triggered by certain events [18][7] that significantly change the current network state, such as the joining/leaving of clients and significant flow variations.

The last issue is about the configuration of APs, which influences the hidden/exposed terminal problems in Wi-Fi networks. Generally, the transmission power and clear channel assessment (CCA) threshold of APs can be adjusted. However, as shown in [36], it is impossible to eliminate the hidden/exposed terminal problems together, so they can only be balanced. What is more, [37] proves the existence of node starvation problem if these parameters are badly tuned. [18] conducts experiments to confirm the findings in [37] and indicates that the best policy is to simply use the maximum transmission power. With considerations on all these conclusions, in this paper, we let all the APs use the same transmission power and CCA threshold. Also, the CCA threshold is adjusted such that each node can detect the transmissions from other nodes. This is practical in the OBSS scenario, where APs and clients are supposed to be densely distributed within a certain area. With the above configurations, the network tends to have more toleration on the exposed terminal problem than the hidden terminal problem, since the latter can cause more collisions and spoil the transmissions.

VI. SIMULATION RESULTS

In this section, we conduct simulations to evaluate the outcomes presented in the paper. The density of APs operating on a single channel can be as high as 0.004 AP/m^2. Also, there exist at most 10 APs and 30 clients operating on the channel within the area similar to the coverage of a single AP. The above network density is close to other works on high-density WLAN such as [38]. Parameters used in the simulation are given in Tab. I. All the MAC protocol parameters are consistent with IEEE 802.11n and can be found in [24].

TABLE I: Values of the parameters used in the simulation.

Parameter	Value
simulation time	10^6 slots
maximum backoff stage m	6
minimum window size W_0	16
packet failure probability p_e	10^{-5}
slot length σ	9 us
propagation delay δ	1 us
MAC overheads	see [24]
packet length L	2304 Bytes
MCS	$\{65, 58.5, 52, 39, 26, 19.5, 13, 6.5\}$ Mbps
flow data rate	$\{100, 200, 300, 400, 500\}$ KB/s

A. Model Verification

In this subsection, we validate the accuracy of the Markov model. With different network sizes and in both basic and RTS/CTS modes, the system throughput and average inter packet delay derived from the model are compared with the results from the simulation, as shown in Fig. 4 and Fig. 5, respectively. From both figures, it can be found that the analytical results can match the simulation outcomes well, which validates the accuracy of the model. Hence, the results

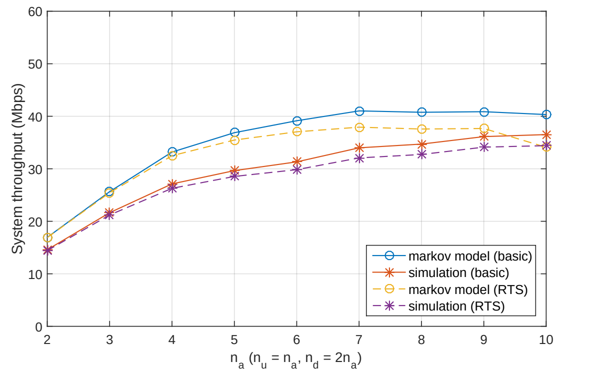


Fig. 4: System throughput from Markov model and simulation in basic mode and RTS/CTS mode.

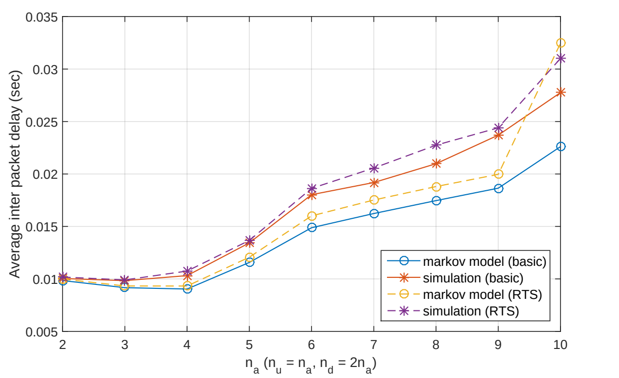


Fig. 5: Average inter packet delay from Markov model and simulation in basic mode and RTS/CTS mode.

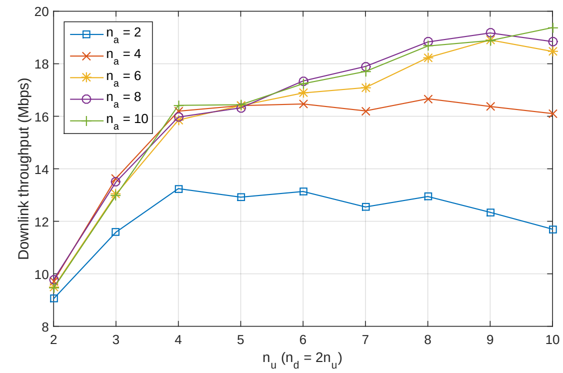


Fig. 6: Downlink throughput with different n_a.

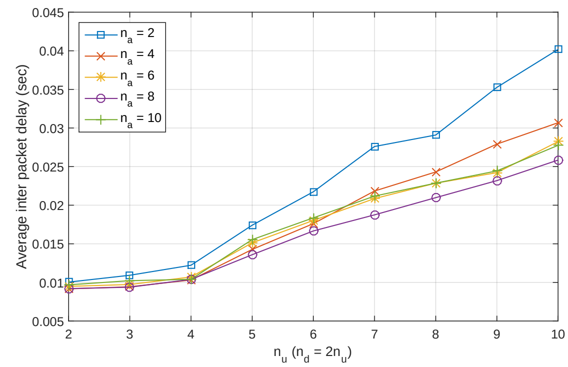


Fig. 7: Average inter packet delay with different n_a.

from the model can be further used for the formulation of client association problem. Note that in the following, we only consider the basic mode in order to focus on the primary purposes of the simulations.

B. Effect of High-Density APs and Backhual Capacity

In this subsection, we first show the effect of high-density co-channel APs to the network performance. We change the number of APs at each time (the backhaul capacity of each AP is set to be sufficient) and compare the downlink throughput and average inter packet delay with different numbers of clients, as given in Fig. 6 and Fig. 7, respectively. In Fig. 6, we can see the traditional WLAN characteristic, i.e., the throughput slightly drops after the network becomes saturated. Also, Fig. 6 shows that when the number of APs is small, the downlink throughput increases as the number of APs gets larger. There are two main reasons for this. On the one hand, download clients are more likely to have larger transmission rates. On the other hand, they also have more opportunities to receive packets, since the “uplink and downlink asymmetry” phenomenon is relieved due to the increased number of APs. This can also be justified by the observation that as the number of APs gets larger, the downlink throughput is much harder to become saturated. However, more APs can also cause more contentions and collisions, which may slightly decrease the throughput when the number of APs is relatively large, as shown in Fig. 6. Also, in Fig. 7, we can find similar results, i.e., the average inter packet delay becomes larger when the number of APs either gets larger or smaller. Therefore, it can be concluded that the high-density co-channel APs can be utilized to improve the performance if they are properly deployed.

Next, we show how the backhaul capacity of APs affects the throughput and inter packet delay of download clients. We set the number of APs to 1 and changes the backhaul capacity each time. As shown in Fig. 8, the downlink throughput increases when the backhaul capacity becomes larger. Also, as the number of clients gets larger, the downlink throughput begins to decrease due to the contention from upload clients. This happens when the backhaul capacity is saturated (marked by the red circles) or when the number of upload clients gets large enough. Similarly, from Fig. 9, it can be seen that the average inter packet delay becomes larger when the backhaul capacity is more limited.

C. Algorithm Comparison

In this subsection, we compare the algorithms presented in this paper with three existing algorithms, which are described as below:

1) RSSI: For the RSSI-based method, each client is associated with the AP that offers the largest transmission rate.

2) FAME: The FAir Mac Efficiency (FAME) [7] algorithm aims at maximizing the minimum weighted MAC efficiency of all the clients. We use the default weight 0 as indicated in

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2017.2668066, IEEE Transactions on Vehicular Technology

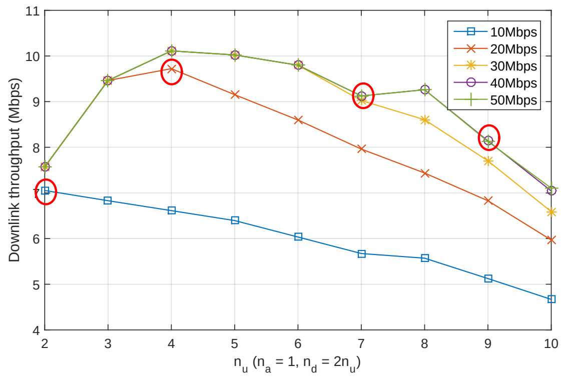


Fig. 8: Downlink throughput with different backhaul capacities.

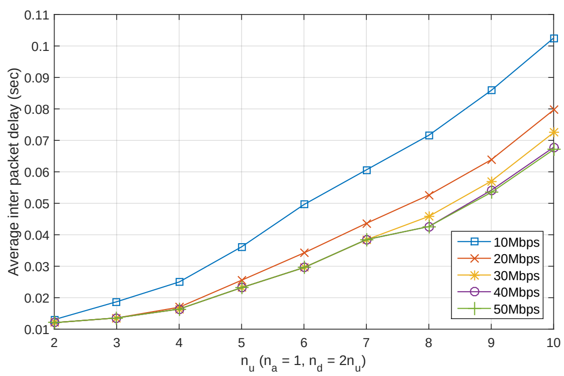


Fig. 9: Average inter packet delay with different backhaul capacities.

[7], so the MAC efficiency of client i is (Eq. (5) in [7]):

$$\alpha_i = \frac{S_i^{up} + S_i^{down}}{\min\{1, u_i + d_i\} \cdot r_{i,a_i}} \tag{31}$$

where r_{i,a_i} is the rate between i and its AP a_i, S_i^{up} and u_i (S_i^{down} and d_i) are the throughput and packet arrival probability of upload (download) flows of client i, respectively. In our scenario, since all the clients and APs share the same medium, the association mechanism is slightly changed from [7]. Specifically, when a new client comes, it is assumed to be associated with a different AP at each time and the minimum MAC efficiency of all the clients that have been associated with APs is calculated. Then, the client is finally associated with the AP that maximizes the minimum MAC efficiency.

3) Access Link Virtualization: In [27], peak load minimization of all APs is achieved through access link virtualization, Wi-Fi pooling and SDN-enabled traffic steering. The client association problem is reduced to the classical job shop scheduling problem and the best known longest processing time (LPT) heuristic is used. Specifically, at each time, the client with the most traffic demand is assigned to the AP with the largest residue bandwidth.

Comparisons among the algorithms are conducted in three different scenarios. In the first scenario, the clients and APs are uniformly distributed and the backhaul capacity is large enough. The average inter packet delay is given in Fig. 10a. It can be seen that the performance of greedy and local search methods is better than the FAME and Virtualization algorithms. This is because that FAME is designed under the assumption that each AP has a unique and isolated channel, while Virtualization only considers traffic demand and backhaul capacity but neglects the influence of transmission rates. Also, RSSI achieves nearly the same results as greedy and local search methods, since in this scenario, the transmission rates are the primary factor that affects the performance. In the second scenario, we fix the number of APs to 3 and set the backhaul capacity of the APs to 10Mbps, 20Mbps and sufficiently large, respectively. From Fig. 10b, we can see that FAME and Virtualization are still worse than the other three methods. However, in this scenario, local search and greedy methods perform better than RSSI, since the latter does not take the backhaul capacity into account. In the third scenario, the number of APs remains 3 but the clients are distributed in the hotspot manner. Specifically, we suppose that most clients are close to a certain AP and are relatively far away from the other two APs. Fig. 10c shows the comparison results, from which we can see similar outcomes as the second scenario. The reason is that RSSI does not balance the traffic load among APs, which leads to the situation that a large number of clients are associated with one AP. Hence, the inter packet delay becomes larger due to long queues at the congested AP.

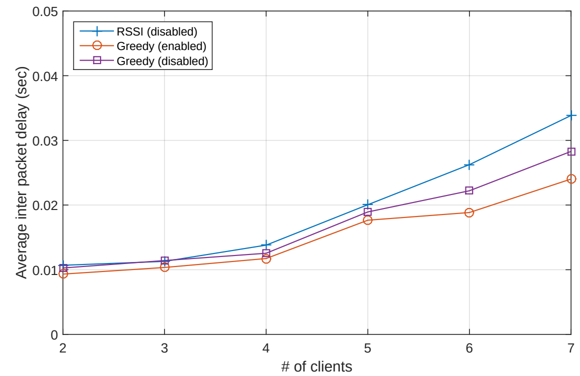


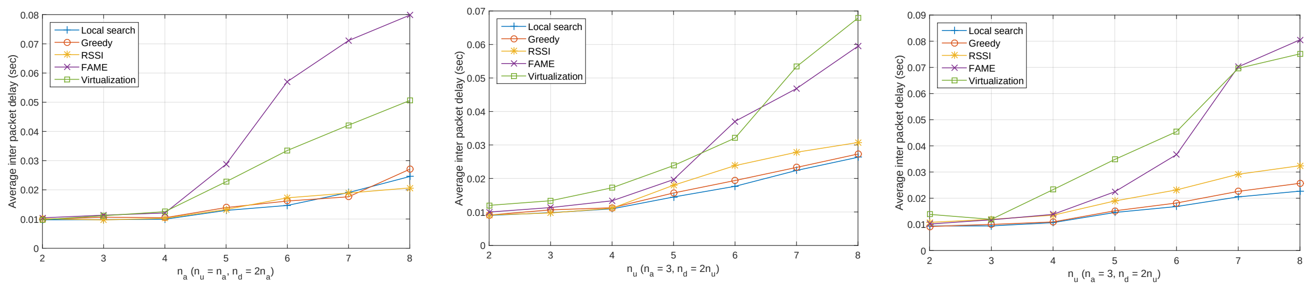
Fig. 11: Average inter packet delay of methods with different flow-level association capabilities (scenario: hotspot, insufficient backhaul capacity).

At last, we show the effect of flow-level association to the network performance. Enabling such a feature will endow the client association with more flexibility. The number of APs is set to 3 and the backhaul capacity of the APs is 20Mbps, 20Mbps and sufficiently large, respectively. Also, clients are supposed to behave in the hotspot manner and each client is supposed to have one upload flow and two download flows. There are three cases compared in this scenario, i.e., RSSI, which cannot realize flow-level association; and greedy method with and without flow-level association capabilities. The average inter packet delay of all the download flows is shown in Fig. 11. It can be seen that the performance is better when flow-level association capability is enabled, due to more flexibility for the association.

VII. Conclusion

In this paper, we have studied the client association in SDWLAN by considering several new SDN features, namely centralized client association, global network state awareness, seamless handoff and flow-level association. The workflow of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2017.2668066, IEEE Transactions on Vehicular Technology



(a) (b) (c)

Fig. 10: Average inter packet delay of different methods under different scenarios: (a) uniform, sufficient backhaul capacity; (b) uniform, limited backhaul capacity; (c) hotspot, sufficient backhaul capacity.

client association has been formulated as a delay-minimized optimization problem, where the inter packet delay of flows has been derived from a two-state Markov model. The formulated problem has been further transformed into an NP-hard supermodular set function minimization problem and two low-complexity heuristic algorithms have been proposed to solve the problem. For our future work, we will jointly consider channel allocation and client association problems in SDWLAN to further improve the network efficiency.

References

[1] D. Macedo, D. Guedes, L. Vieira, M. Vieira, and M. Nogueira, "Programmable networks—from software-defined radio to software-defined networking," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 2, pp. 1102–1125, 2015.

[2] N. Zhang, N. Cheng, A. T. Gamage, K. Zhang, J. Mark, and X. Shen, "Cloud assisted hetnets toward 5g wireless networks," *IEEE Communications Magazine*, vol. 53, no. 6, pp. 59–65, 2015.

[3] A. Baid and D. Raychaudhuri, "Understanding channel selection dynamics in dense wi-fi networks," *IEEE Communications Magazine*, vol. 53, no. 1, pp. 110–117, 2015.

[4] K. Shin, I. Park, J. Hong, D. Har, and D.-H. Cho, "Per-node throughput enhancement in wi-fi densenets," *IEEE Communications Magazine*, vol. 53, no. 1, pp. 118–125, 2015.

[5] V. Jones and H. Sampath, "Emerging technologies for wlan," *IEEE Communications Magazine*, vol. 53, no. 3, pp. 141–149, 2015.

[6] N. Lu, N. Cheng, N. Zhang, X. Shen, J. W. Mark, and F. Bai, "Wi-fi hotspot at signalized intersection: Cost-effectiveness for vehicular internet access," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 5, pp. 3506–3518, 2016.

[7] D. Gong and Y. Yang, "On-line ap association algorithms for 802.11 n wlans with heterogeneous clients," *IEEE Transactions on Computers*, vol. 63, no. 11, pp. 2772–2786, 2014.

[8] G. Bianchi, "Performance analysis of the ieee 802.11 distributed coordination function," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 3, pp. 535–547, 2000.

[9] N. McKeown, T. Anderson, H. Balakrishnan, G. Parulkar, L. Peterson, J. Rexford, S. Shenker, and J. Turner, "Openflow: enabling innovation in campus networks," *ACM SIGCOMM Computer Communication Review*, vol. 38, no. 2, pp. 69–74, 2008.

[10] H. Huang, P. Li, S. Guo, and W. Zhuang, "Software-defined wireless mesh networks: Architecture and traffic orchestration," *IEEE Network*, vol. 29, no. 4, pp. 24–30, 2015.

[11] Q. Zheng, K. Zheng, H. Zhang, and V. C. Leung, "Delay-optimal virtualized radio resource scheduling in software-defined vehicular networks via stochastic learning," *IEEE Transactions on Vehicular Technology*, to appear.

[12] H. Li, M. Dong, and K. Ota, "Control plane optimization in software defined vehicular ad-hoc networks," *IEEE Transactions on Vehicular Technology*, to appear.

[13] K.-K. Yap, M. Kobayashi, R. Sherwood, T.-Y. Huang, M. Chan, N. Handigol, and N. McKeown, "Openroads: Empowering research in mobile networks," *ACM SIGCOMM Computer Communication Review*, vol. 40, no. 1, pp. 125–126, 2010.

[14] L. Suresh, J. Schulz-Zander, R. Merz, A. Feldmann, and T. Vazao, "Towards programmable enterprise wlans with odin," in *Proceedings of the first workshop on Hot topics in software defined networks*. ACM, 2012, pp. 115–120.

[15] J. Vestin, P. Dely, A. Kassler, N. Bayer, H. Einsiedler, and C. Peylo, "Cloudmac: towards software defined wlans," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 16, no. 4, pp. 42–45, 2013.

[16] Y. Yiakoumis, M. Bansal, A. Covington, J. van Reijendam, S. Katti, and N. McKeown, "Behop: A testbed for dense wifi networks," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 18, no. 3, pp. 71–80, 2015.

[17] A. Patro and S. Banerjee, "Coap: a software-defined approach for home wlan management through an open api," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 18, no. 3, pp. 32–40, 2015.

[18] R. Murty, J. Padhye, R. Chandra, A. Wolman, and B. Zill, "Designing high performance enterprise wi-fi networks," in *Proceedings of the 5th USENIX Symposium on Networked Systems Design and Implementation*. USENIX Association, 2008, pp. 73–88.

[19] V. Sivaraman, J. Matthews, C. Russell, S. T. Ali, and A. Vishwanath, "Greening residential wi-fi networks under centralized control," *IEEE Transactions on Mobile Computing*, vol. 14, no. 3, pp. 552–564, 2015.

[20] P. Lv, X. Wang, X. Xue, and M. Xu, "Swimming: Seamless and efficient wifi-based internet access from moving vehicles," *IEEE Transactions on Mobile Computing*, vol. 14, no. 5, pp. 1085–1097, 2015.

[21] P. E. Engelstad and O. N. Osterbo, "Analysis of the total delay of ieee 802.11 e edca and 802.11 dcf," in *IEEE International Conference on Communications*. IEEE, 2006.

[22] D. Malone, K. Duffy, and D. Leith, "Modeling the 802.11 distributed coordination function in nonsaturated heterogeneous conditions," *IEEE/ACM Transactions on Networking*, vol. 15, no. 1, pp. 159–172, 2007.

[23] E. Felemban and E. Ekici, "Single hop ieee 802.11 dcf analysis revisited: Accurate modeling of channel access delay and throughput for saturated and unsaturated traffic cases," *IEEE Transactions on Wireless Communications*, vol. 10, no. 10, pp. 3256–3266, 2011.

[24] F. Aurzada, M. Lévesque, M. Maier, and M. Reisslein, "Fiwi access networks based on next-generation pon and gigabit-class wlan technologies: A capacity and delay analysis," *IEEE/ACM Transactions on Networking*, vol. 22, no. 4, pp. 1176–1189, 2014.

[25] Y. Bejerano, S.-J. Han, and L. Li, "Fairness and load balancing in wireless lans using association control," *IEEE/ACM Transactions on Networking*, vol. 15, no. 3, pp. 560–573, 2007.

[26] I. Broustis, K. Papagiannaki, S. Krishnamurthy, M. Faloutsos, and V. Mhatre, "Measurement-driven guidelines for 802.11 wlan design," *IEEE/ACM Transactions on Networking*, vol. 18, no. 3, pp. 722–735, 2010.

[27] V. Sivaraman, T. Moors, H. Habibi Gharakheili, D. Ong, J. Matthews, and C. Russell, "Virtualizing the access network via open apis," in *Proceedings of the ninth ACM conference on Emerging networking experiments and technologies*. ACM, 2013, pp. 31–42.

[28] D. Chen, T. Quek, and M. Kountouris, "Backhauling in heterogeneous cellular networks: Modeling and tradeoffs," *IEEE Transactions on Wireless Communications*, vol. 14, no. 6, pp. 3194–3206, 2015.

[29] H. Beyranvand, W. Lim, M. Maier, C. Verikoukis, and J. A. Salehi, "Backhaul-aware user association in fiwi enhanced lte-a heterogeneous networks," *IEEE Transactions on Wireless Communications*, vol. 14, no. 6, pp. 2992–3003, 2015.

[30] E. Chai and K. G. Shin, "Sidekick: Ap aggregation over partially over-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2017.2668066, IEEE Transactions on Vehicular Technology

lapping channels," in *19th IEEE International Conference on Network Protocols*. IEEE, 2011.

[31] Y. Xu, H. Yang, F. Ren, C. Lin, and X. Shen, "Frequency domain packet scheduling with mimo for 3gpp lte downlink," *IEEE Transactions on Wireless Communications*, vol. 12, no. 4, pp. 1752–1761, 2013.

[32] S. He, J. Chen, X. Li, X. Shen, and Y. Sun, "Leveraging prediction to improve the coverage of wireless sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 4, pp. 701–712, 2012.

[33] A. Krause and D. Golovin, "Submodular function maximization," *Tractability: Practical Approaches to Hard Problems*, vol. 3, p. 19, 2012.

[34] G. Calinescu, C. Chekuri, M. Pál, and J. Vondrák, "Maximizing a submodular set function subject to a matroid constraint," in *Integer programming and combinatorial optimization*. Springer, 2007, pp. 182–196.

[35] X. Zhe, F. Zhang, W. Wang, and S. He, "A polynomial time algorithm for minimizing a nondecreasing supermodular set function and its performance guarantee," *International Journal of Pure and Applied Mathematics*, vol. 52, no. 5, pp. 637–643, 2009.

[36] J. Zhu, B. Metzler, X. Guo, and Y. Liu, "Adaptive csma for scalable network capacity in high-density wlan: A hardware prototyping approach." in *IEEE International Conference on Computer Communications.* IEEE, 2006.

[37] V. P. Mhatre, K. Papagiannaki, and F. Baccelli, "Interference mitigation through power control in high density 802.11 wlans," in *26th IEEE International Conference on Computer Communications.* IEEE, 2007.

[38] F. G. Debele, M. Meo, D. Renga, M. Ricca, and Y. Zhang, "Designing resource-on-demand strategies for dense wlans," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 12, pp. 2494–2509, 2015.

Jiacheng Chen received his B.S. degree from the School of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian, China, in 2013. He is pursuing his PhD degree in the Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai, China, and was a visiting PhD student with BBCR Lab, the University of Waterloo, Canada in 2015 -16. His current research interests include resource management in wireless networks, vehicular networks and white space networks.

Bo Liu received the BS degree from Nanjing University of Posts and Telecommunications, Nanjing, China, in 2004 and the MS and PhD degrees, in 2007 and 2010, respectively, from Shanghai Jiao Tong University, Shanghai, China, all in electrical engineering. Since 2010, he has been an assistant researcher and then associate researcher at the Electrical Engineering Department of Shanghai Jiao Tong University. He is currently a postdoctoral research fellow in Deakin University, Australia, since November 2014.

Haibo Zhou received the Ph.D. degree in Information and Communication Engineering from Shanghai Jiao Tong University, Shanghai, China, in 2014. He is currently a Post-Doctoral Fellow with the Broadband Communications Research (BBCR) Group, University of Waterloo. His current research interests include resource management and performance analysis in cognitive radio networks and vehicular networks.

Quan Yu (SM'16) received the B.S. degree in information physics from Nanjing University in 1986, the M.S. degree in radio wave propagation from Xidian University in 1988, and the Ph.D. degree in fiber optics from the University of Limoges in 1992. Since 1992, Dr. Yu joined the faculty of the Institute of China Electronic System Engineering Corporation, where he has been a Senior Engineer, and currently a Research Fellow. His main areas of research interest are the architecture of wireless networks, optimization of protocols and cognitive radio. Dr. Yu was invited to serve as the Honorary Dean of School of Telecommunications Engineering, Xidian University in 2012. Currently, Dr. Yu serves as adjunct professor and Ph.D supervisor in Xidian University and Shanghai Jiao Tong University. Dr. Yu is a Fellow of Chinese Academy of Engineering (CAE).

Lin Gui (M'08) received the Ph.D. degree from Zhejiang University, Hangzhou, China, in 2002. Since 2002, she has been with the Institute of Wireless Communication Technology, Shanghai Jiao Tong University, Shanghai, China. Currently, She is a Professor in Shanghai Jiao Tong University. Her current research interests include HDTV and wireless communications.

Xuemin (Sherman) Shen (IEEE M'97-SM'02-F'09) received the B.Sc.(1982) degree from Dalian Maritime University (China) and the M.Sc. (1987) and Ph.D. degrees (1990) from Rutgers University, New Jersey (USA), all in electrical engineering. He is a Professor and University Research Chair, Department of Electrical and Computer Engineering, University of Waterloo, Canada. He is also the Associate Chair for Graduate Studies. Dr. Shens research focuses on resource management in interconnected wireless/wired networks, wireless network security, social networks, smart grid, and vehicular ad hoc and sensor networks. He is an elected member of IEEE ComSoc Board of Governor, and the Chair of Distinguished Lecturers Selection Committee. Dr. Shen served as the Technical Program Committee Chair/Co-Chair for IEEE Globecom16, IEEE Infocom14, IEEE VTC10 Fall, and IEEE Globecom07, the Symposia Chair for IEEE ICC10, the Tutorial Chair for IEEE VTC11 Spring and IEEE ICC08, the General Co-Chair for ACM Mobihoc15, Chinacom07 and QShine06, the Chair for IEEE Communications Society Technical Committee on Wireless Communications, and P2P Communications and Networking. He also serves/served as the Editor-in-Chief for IEEE Network, Peer-to-Peer Networking and Application, and IET Communications; a Founding Area Editor for IEEE Transactions on Wireless Communications; an Associate Editor for IEEE Transactions on Vehicular Technology, Computer Networks, and ACM/Wireless Networks, etc.; and the Guest Editor for IEEE JSAC, IEEE Wireless Communications, IEEE Communications Magazine, and ACM Mobile Networks and Applications, etc. Dr. Shen received the Excellent Graduate Supervision Award in 2006, and the Outstanding Performance Award in 2004, 2007, 2010, and 2014 from the University of Waterloo, the Premiers Research Excellence Award (PREA) in 2003 from the Province of Ontario, Canada, and the Distinguished Performance Award in 2002 and 2007 from the Faculty of Engineering, University of Waterloo. Dr. Shen is a registered Professional Engineer of Ontario, Canada, an IEEE Fellow, an Engineering Institute of Canada Fellow, a Canadian Academy of Engineering Fellow, a Royal Society of Canada Fellow, and a Distinguished Lecturer of IEEE Vehicular Technology Society and Communications Society.