IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received June 15, 2020, accepted June 21, 2020, date of publication June 25, 2020, date of current version July 6, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3004772

Towards QoS-Aware Load Balancing for High Density Software Defined Wi-Fi Networks

SOHAIB MANZOOR, ZE CHEN, YAYU GAO, (Member, IEEE), XIAOJUN HEI, (Member, IEEE), AND WENQING CHENG, (Member, IEEE)

School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan 430074, China

Corresponding author: Xiaojun Hei (heixj@hust.edu.cn)

This work was supported in part by the National Natural Science Foundation of China (No. 61972172).

ABSTRACT High density Wi-Fi networks require load balancing in order to ensure quality of service (QoS). In the traditional Wi-Fi networks, the wireless stations learn the access points (APs) load and make the association decisions themselves leading to uneven load distribution among the APs. The new paradigm, software defined networking (SDN), has a centralized architecture, which allows to manage, measure and control high density Wi-Fi networks easily. In this paper we propose a QoS-aware load balancing strategy (QALB) for software defined Wi-Fi networks (SD-Wi-Fi), as a solution to address the problem of Wi-Fi congestion among the OpenFlow enabled APs (OAPs). The SDN controller selects a load level up to which the association decisions are made by the OAPs without consulting the controller. The wireless stations from an overloaded OAP are handed to an underloaded OAP by considering multi-metrics such as the packet loss rate, received signal strength indicator (RSSI) and throughput. An emulation platform and a large-scale-low-cost testbed with the same settings are constructed to evaluate the performance of our load balancing strategy. The results show that in comparison to four non-static schemes such as, channel measurement based access selection scheme (CMAS), (DL-SINR) downlink-signal to interference plus noise ratio AP selection scheme (DASA), mean probe delay scheme (MPD) and RSSI scheme, the proposed QALB, optimizes the throughput up to 16%, reduces the average frame delay up to 19%, minimizes the number of re-transmissions by 49%, reduces the number of handoffs by 15%, improves the degree of load balancing by 22% and minimizes the re-association times by 38%, in high density SD-Wi-Fi.

INDEX TERMS Emulation, high density, load balancing, QoS, Wi-Fi, SDN, testbed.

I. INTRODUCTION

The popularity of the Wi-Fi networks has increased rapidly in the past decade due to their facile connectivity anywhere and anytime. Wi-Fi networks are commonly found in smart learning, smart health care, smart homes and shopping malls. The APs that provide access to the network are deployed in a spontaneous manner resulting in high variable AP densities and uneven load distribution among the APs [1], [2]. The smallest block in the Wi-Fi network is the basic service set (BSS), in which an AP acts as a bridge for wireless stations to provide access to the network. The wireless stations roam between one BSS to another sending re-association request to new APs in the vicinity. IEEE 802.11 standard supports the seamless roaming among the interconnected BSSs. The BSSs are separated in a single extended service set (ESS). The wireless stations without changing the network configuration can roam freely inside an ESS. The roaming and association decisions are carried at the client end by considering the RSSI values. This method is called the client driven method, which tends to cause an uneven distribution of load among the APs [3]–[5]. The ESS has many APs that share an overlapping region, but still the overloaded APs degrade the network performance in terms of QoS and the APs in the neighbor remain underloaded. To balance the load among the APs in an ESS for high density SD-Wi-Fi remains an open issue as no standardized method has addressed this problem.

The existing load balancing methods for Wi-Fi can be classified into either centralized or client-driven, depending on which entity is responsible for an AP association decisions. Most of the research in the past has focused on client driven methods, where the wireless station learns the AP load

The associate editor coordinating the review of this manuscript and approving it for publication was Byung-Seo Kim.

and make association decisions themselves. The client driven methods cannot make precise decisions for load balancing as it lacks the global view of the entire network. To reach the ideal load balancing condition a number of repeated associations are required which adds to the latency factor [6], [7]. The centralized methods on the other hand contribute well in the load balancing decisions by either controlling the association control mechanism [8], [9] or by adjusting the coverage area of a specific AP [10], [11]. Some of the centralized methods are based on a distributed approach, where an extensive message exchange is required between the custom hardware and the APs. The other centralized methods rely on a centralized controller, which incorporates too many unnecessary association/de-association decisions for the APs, hence overloading the controller. In heterogeneous networks, devices from different vendors and non-standardized protocols/hardware make it almost impossible to carry out load balancing operations in dense Wi-Fi networks.

Recently, a new paradigm, SDN is used to over lift the problems in the traditional Wi-Fi networks [12], [13]. The separated control and data plane in SDN simplify the network management tasks, bring programmability over wireless stations and introduce new network functions/applications. The abstraction of the control plane from wireless stations has motivated many vendors such as Microsoft, Google, HP and Cisco to support OpenFlow standards. By studying the existing literature on load balancing methods in Wi-Fi and the SDN, we believe that the hybrid software defined Wi-Fi networks (SD-Wi-Fi) architecture can incorporate better load balancing strategies instead of relying on custom protocols/hardware design.

In this paper we propose a QoS-aware load balancing strategy for high density Wi-Fi networks by taking advantages of the SDN to improve the overall network performance. The SDN controller harvesting an overall view of the network helps in collecting information from the OAPs and decide up to which load level the OAPs can carry out the association decisions by themselves, without consulting the centralized controller. There exists some literature that make use of the dedicated centralized controller to carry out the load balancing functions in the Wi-Fi networks. Our solution brings novelty in three ways. The existing load balancing methods relying on a centralized approach, renounce all the AP, RSSI based association and de-association decisions to the centralized controller. Our emulation and testbed results show that 100 percent dependency on the centralized controller increases the turn around time for AP association and de-association process and further overloads the controller. The proposed strategy allows the OAPs to make the association decisions themselves by dynamically adjusting the load level according to the network conditions. The second aspect of the proposed study is that, the destination OAP is selected through a multi-metric criteria (packet loss rate, RSSI and throughput), satisfying the least loaded conditions instead of just relying on one metric i.e., RSSI. Finally, our solution does not require any hardware changes and is applicable to any wireless device which supports the OpenFlow standards.

The core contributions of the proposed study are mentioned below:

- We propose a QoS-aware load balancing strategy (QALB) for a high density SD-Wi-Fi by designing two modules, one for the OAPs and one for the controller. The OAP module pushes the OAPs capacity and load information to the controller and manages the OAPs associations and de-associations with a minimal involvement of the controller by setting up a load level dynamically, according to the network conditions. The destination OAP is chosen based on a multi-metric criteria such as packet loss rate, RSSI and throughput. The controller module performs the load balancing operations by instructing the OAPs.
- We construct an emulation and a prototype testbed to evaluate the performance of a high density SD-Wi-Fi for monotonically increasing load variation.

The remainder of this paper is organized as follows. Section II explores the related work on load balancing in Wi-Fi, SDN and SD-Wi-Fi. Network model and problem formulation is dealt within Section III. Section IV explains the proposed QoS-aware load balancing solution for SD-Wi-Fi. Section V describes the emulation and the testbed platforms used for SD-Wi-Fi. Section VI discusses the performance evaluation of our proposed strategy against the previous load balancing schemes. Finally the paper is concluded in section VII. Tables 1 and 2 show the list of abbreviations and symbols, respectively, used in this study.

II. RELATED WORK

A. LOAD BALANCING IN Wi-Fi

The load balancing in Wi-Fi involves associating or re-associating the wireless stations to the least loaded APs regardless of which ever approach is used. In order to perform the load balancing operations, the wireless stations first perform the active or passive scanning [14]. In the active scan the mobile radio transmits a probe request message and then listens for the probe response message from the AP [15]. In the passive scan, in order for the wireless stations to get the APs information, they listen to the periodically transmitted beacon frames from the AP [16]. The load balancing in Wi-Fi can be classified into two methods, client-driven or centralized, depending upon which part of the network is in charge of AP association decisions [17].

1) CLIENT-DRIVEN METHODS

Most of the previous research has used the client-driven methods, where the wireless stations learn the AP load and make association decisions themselves. The AP with the highest RSSI value is chosen to be the best and most suitable candidate for association. The use of RSSI as the only parameter for AP selection may result in throughput degradation and

TABLE 1. List of abbreviations.

Abbreviations	Description
ACH	Adaptive connection hand-off
AHP	Analytical hierarchical process
AP	Access point
BSS	Basic service set
CLI	Command line interface
CMAS	Channel measurement based access selection
DCF	Distributed coordination function
DL-SINR	Downlink-signal to interference plus noise ratio
DPID	Daemon process ID
EASM	Efficiency-aware switch migration
EMA	Exponential moving average
ESS	Extended service set
FIFO	First in first out
hostapd	Host access point daemon
HP	High priority
IoT	Internet of things
iw	Configuration tool for wireless devices
LAP	Logical access point
LBBSRT	Load balancing based on server response time
LC	Local controller
LLF	Least loaded first
LP	Low priority
LVAP	Light virtual access point
MAC	Medium access control
MPD	Mean probe delay
OAP	OpenFlow enabled AP
OVS	Open virtual switch
QALB	QoS-aware load balancing strategy
QoS	Quality of service
RSSI	Received signal strength indicator
SDN	Software defined networking
SD-Wi-Fi	Software defined Wi-Fi network
SSF	Strongest signal first
SSID	Service set ID
SVM	Support vector machine
TCP/IP	Transmission control protocol/Internet protocol
Wi-Fi	Wireless fidelity
WLAN	Wireless local area network
XML	Extensible markup language

TABLE 2. List of symbols.

Symbols	Description
C	Capacity information
C_z	Capacity information of an OAP_z
cw	Current wireless stations
cw_z	Current wireless stations associated to an OAP_z
mw_z	Maximum number of wireless stations associated to an OAP_z
sw	Suggested number of wireless stations
sw_z	Suggested number of wireless stations to be associated to an OAP_z
snr_z	Average signal to noise ratio of an OAP_z associated connections
L	Load information
L_z	Load information of an OAP_z
P	Packet loss rate
$\hat{P}$	Packet loss rate with retransmissions
P_z	Packet loss rate at OAP_z
$RSSI_z$	Received signal strength indicator of an OAP_z
$\hat{T}_z$	Normalized throughput at OAP_z
J	Jain's fairness index of an ESS
nb	Number of BSS in an ESS
L_{avg}	Average load level of an ESS
L_L	Load level of an ESS
MAC	Medium access control
P_ϵ	Packet error rate
M_{cp}	MAC collision probability
$\hat{R}$	Maximum number of retransmissions
$T_{current}$	Current throughput
$T_{maximum}$	Maximum throughput
ϕ	Communication quality
$RSSI_{zx}$	RSSI strength of an OAP_z received by the wireless station x
$\phi_{z,x}$	ϕ of an OAP_z with respect to the wireless station x

crowded hotspots [18]. AP selection can be based on metrics other than RSSI. Wireless stations estimate the available bandwidth for AP selection [19]. The co-channel interference can be another parameter to select the most suitable AP for selection [20]. WiFi-seeker proposed a method to select the AP with least interference with other APs for different number of communication devices [21]. The client-driven methods lack in the global view of the entire network to make precise load balancing decisions. More time is required to achieve a load balanced equilibrium state as continuous and repeated AP associations and de-associations are involved. More over a single metric, i.e., RSSI, for AP selection may not be sufficient for the QoS satisfaction.

2) CENTRALIZED METHODS

The centralized methods help in achieving load balancing either by controlling the AP association/de-association or by adjusting the APs coverage area. The association control methods help the wireless stations to de-associate from the overloaded APs and re-associate to the underloaded APs. An adaptive admission control method based on Markov regenerative process coupled with fluid model is adopted for AP associations [22]. The scheme helps in balancing the packets, from different Internet of things (IoT) devices, being buffered in the APs. Network admission control is studied through a hierarchy of APs using an AP matrix, where a slave AP, a master AP and a central controller is used. The slave APs are outdated and hardly guarantee better performance. Master APs are powerful devices that take in charge of the slave APs. The AP matrix consisting of master and slave APs is controlled by a centralized controller for admission control [23]. An admission control algorithm is introduced for Wi-MAX networks which takes into account the bandwidth utilization [24]. The AP selection is performed based on the total size of the files from the mobile stations rather than the total number of mobile stations.

A WiFi coverage, development and management strategy is proposed, based on the user spatio-temporal association analytics [25]. The association history records are used to predict the AP load through machine learning algorithms and the APs which remain idle for predefined times are switched off. Another AP coverage adjustment method makes use of the transition-aware energy saving mechanism for dense WiFi networks [26]. Switching frequency is used as the control knob to switch the APs states to meet the user requirements. The scheme works on recent user behaviors and relevant network parameters such as energy consumption.

The distributed centralized methods require a massive exchange of information between the APs and the custom hardware support. The other methods perform AP associations through a controller, which overload the controller with unnecessary AP association and de-association control decisions. Due to large number of vendors and a lot of dependency among the non-standardized protocol/hardware designs, it is difficult to perform load balancing in high-density WiFi networks.

B. LOAD BALANCING IN SDN

SDN is used in several studies that involve load balancing. Higher scalability is focused, based on the server response time (LBBSRT) [27], which helps in achieving load balancing. A single server with minimum response time is selected. The Packet_In messages are processed by the controllers via the OpenFlow enabled switches and returned for processing. The user requests are then forwarded to the servers through controllers. The performance metric such as response time is calculated based on the time elapsed between the data packets transmitted and the data packets which are received successfully. The load balancing among the OpenFlow enabled switches/APs is ignored, as they handle the packets from the user devices and the controllers at the same time. A flow table design is proposed to handle the incoming traffic dynamically [28]. The proposed flow table is made of group of IP addresses. The status of the clients is responsible to make decisions for processing. If the status for most of the clients is the same then the processing becomes complex.

Load among the multiple controllers is balanced using load information strategy [29]. Average message arrival rate is used to determine the load at the controllers. Depending on the message arrival rate, the switches are placed in a descending order. The overloaded controller issues an instruction to make switch migration, which helps in mitigation of unnecessary resource utilization. The switch migration is a lengthier process as many unnecessary messages are exchanged between the controller and the switches. The idea of game model is used to perform game-switch migration [30]. The utilization ratio of switches is chosen as the criteria for switch migration. The game theoretic approach presents the best approach for switch migration but at the same time it involves complex multiple mathematical computations and allows numerous malicious requests to arrive into the network. Online controller load balancing scheme (OCLB) is designed to minimize the controller response time [31]. The OCLB based on the switch migration targets to enhance the scalability in the distributed control plane. The response diversity of flow requests at the data plane is neglected.

The load imbalance issue is resolved by a greedy method using switch migration [32]. The switch migration is initiated after acquiring the aggregate load information. Factors such as minimal path cost (between the switches and the controllers) and number of Packet_In messages (from switches to controllers) are used to calculate the load of each controller. A predefined threshold value of load diversity between the two controllers is used to trigger migration. The load increases due to the dynamic arrival of the packets at the data plane. SDN is used for load balancing using efficiency-aware switch migration (EASM) [33]. The sole purpose of EASM is to provide fairness in load balancing. Data interaction overhead, state synchronization and routing overhead are used to calculate the load at the controller. The switch to be migrated is determined through probability distribution. Multiple mathematical computations increase the complexity of the system. SDN is used to enhance the load balancing through solving controller placement problem [34]. The issues of clustering performance are resolved by using particle swarm optimization. The technique ignores the queuing model of flow request at the APs.

Load on the controllers is balanced through efficient switch migration load balancing (ESMLB) [35]. The loaded controller shifts the load to the underloaded controller by switch migration. The target underloaded controller is chosen through technique for order preference by similarity to an ideal solution (TOPSIS). Load balancing among the APs is not taken into account. The traffic flow arrivals are modeled using an undirected graph to balance the load [36]. Tie-sets along with the edge load factor balancing method are used to redirect the traffic flows as per the given routes. Due to the increased number of packets arrival, the maintenance of the graph becomes a tedious task. In order to enhance scalability DeepFlow framework is introduced [37]. The flows are measured in size and the aggregated sum is used to identify the least active switches. Without measuring the size of the flows it is not possible to monitor the flows and match them to flow rules. To balance the load in the control plane, controller adaption and migration decision (CAMD) is used [38]. A migration indexing factor is formulated which invokes the switch migration when a certain controller gets overloaded. Load balancing at data plane is neglected.

C. LOAD BALANCING IN SD-Wi-Fi

SD-Wi-Fi requires the mobility support to address the load imbalance issues among the IEEE 802.11 users. Logical access points (LAPs) are used to provide continuous connectivity [39]. RSSI is the only parameter considered for making handover decisions and balancing the load among the APs. A single metric for making handover decisions is not sufficient. A centralized controller is used to balance the load among the APs [40]. An adaptive load balancing algorithm based on RSSI, helps in identifying new users in the network and their connectivity to the APs. Load balancing in SD-WiFi is achieved by clustering of controllers and flow prioritization [41]. Analytical hierarchical process (AHP) is used to prioritize the flows. Clusters of local controllers update their load values to global controller periodically. The load imbalance among the APs is ignored. The load in the software defined WiFi networks (SDWN) is balanced by handoff-delay based association control [42]. The AP load is treated in term of queue backlogs and load balancing is modeled as utility maximum problem. The costs due to handoff delays are considered. The role of OpenFlow, OpenFlow message extensions and QoS guarantees for end users are neglected.

Two methods, strongest signal first (SSF) and the least loaded first (LLF) are used to perform handovers in SD-WiFi [43]. SSF provides connectivity to those devices which show highest RSSI values and ignores the load at the APs. LLF takes into the account, the load at the APs but ignores the signal strength values. The load is defined as the number of clients associated to the AP. The load is not fully characterized by the number of associations kept by an AP as the traffic conditions may vary for each wireless station [44]. In order to address the load imbalance issue at the APs, both load and RSSI values should be considered together. Load among the multi-controllers in SD-WiFi is balanced through type-2 fuzzy particle swarm optimization and support vector machine (SVM) [45]. The SVM acts as an efficient binary classifier at the data plane, where the traffic flows are classified into high priority (HP) and low priority (LP) classes. The Markov chain model predicts the future states of the local controllers (LCs) and in case of any overloaded LC the flows are transferred to a destination controller chosen by fuzzy and particle swarm optimization. The scheme neglects the load imbalance issue among the OAPs. A deterministic load balancing algorithm (Det-LB) is used to address the load imbalance issues in soft real time control systems [46]. Deadline miss ratio is chosen as a metric for AP association and load distribution is performed using first-price sealed-bid auction in game theory. The scheme only targets the deadline miss and packet loss as performance metrics.

An adaptive connection and hand-off (ACH) scheme is used to balance the load among the APs in SD-WiFi [47]. The load of an AP is determined by the number of associations it carries and the load balancing algorithm is triggered when the load difference between the two APs exceeded a certain threshold. The maximum load limit is not discussed and the number of wireless stations associated to a certain AP are not sufficient to reveal the exact load values. Load in SD-WiFi is defined and wireless traffic is prioritized for an healthcare scenario in [48], [49]. Packet generation times and packet payloads are used to tune load for minimum contention in wireless transmission. The role of an OpenFlow protocol between the controller and the APs is neglected. A testbed is designed to study the trade-off between programmability and performance in SD-WiFi [50]. Light virtual access points (LVAPs) are created to shift the load from the overloaded APs to the underloaded APs. With the increase in number of wireless stations, the LVAPs also increase, causing an excess burden on the centralized controller hence degrading the latency performance.

III. NETWORK MODEL

The overview of the QALB architecture is depicted in Fig. 1. The architecture is made up of 4 parts, which include applications, controller, OAPs and the wireless stations. The wireless stations get access to the network through the OAPs, which support the OpenFlow protocol and forward the data according to the flow rules defined by the SDN controller. In addition to just forwarding the data, the OAPs have the capability of recording the flow entries and pushing these entries to the controller whenever needed. The controller sits in the control plane and has the overall view of the network. Being a centralized entity, the controller is responsible for distributing the flow rules and controlling the forwarding behavior of the OAPs. The controller extracts the information from the OAPs and makes the load balancing decisions defined by the application plane.

The proposed network model considers a public Wi-Fi network such as a smart home, smart healthcare, campus Wi-Fi, etc. A single ESS, having multiple interconnected BSSs, i.e, $\{bss_z | z = 1, 2, \ldots, nb\}$, is considered. As a result, a dense Wi-Fi network, having many overlapped OAPs, controlled by a centralized SDN controller, is formed. The key mathematical notations used throughout in this study are listed in Table 2.

In order to accomplish the load balancing task, the SDN controller needs the load and capacity information from the OAPs. The capacity information (C_z) of a specific OAP say, OAP_z, depends upon the number of wireless stations currently associated to an OAP_z (cw_z), maximum number of wireless stations that an OAP_z can support (mw_z) and average signal to noise ratio of the OAP_z associated connections (snr_z). The load information (L_z) of an OAP_z depends upon the packet loss rate (P_z), RSSI ($RSSI_z$) and the normalized throughput ($\hat{T}_z$). By using the aforementioned information, we compute the Jain's fairness index (J) of the ESS from [51], in Eq.1.

$$J = \frac{\left(\sum_{z=1}^{nb} \frac{cw_z}{mw_z} maximum \ (L_z)\right)^2}{nb \sum_{z=1}^{nb} \left(\frac{cw_z}{mw_z} \ maximum \ (L_z)\right)^2} \tag{1}$$

The numerator in Eq.1, defines the square of overall loading of the ESS and the denominator defines the sum of square of each OAP times the number of OAPs. The J should be 1 if all OAPs are equally balanced. For the worst degree of load balancing, $J = \frac{1}{nb}$. The load levels for the OAPs are different, so it is important for the SDN controller to compute the

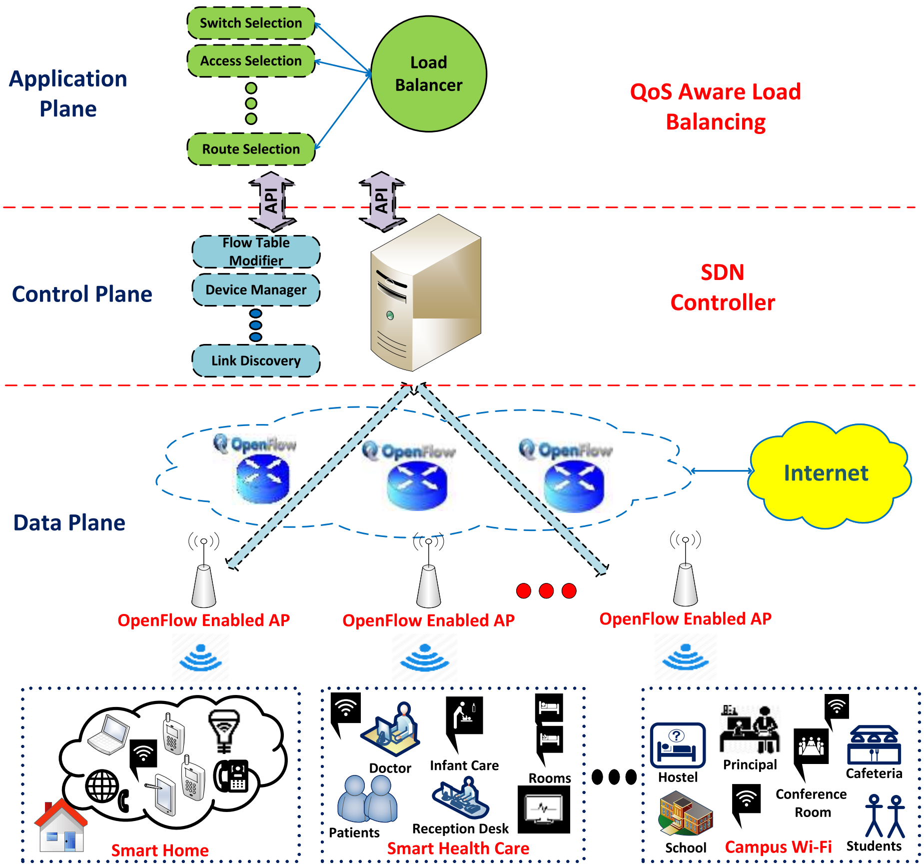

FIGURE 1. The proposed QoS-aware load balancing architecture for SD-Wi-Fi.

average load level (L_{avg}), given in Eq. 2.

$$L_{avg} = \frac{(\sum_{z=1}^{nb} \frac{cw_z}{mw_z} maximum\ (L_z))}{nb} \tag{2}$$

The L_{avg} ranges from 0 to 100, where 100 is the heaviest load and 0 represents the lightest load. The average load level metric is designed, so that it is convenient for the SDN controller to divide the OAPs into overloaded and underloaded classes. Once the overloaded and underloaded OAPs are classified, it is easier to shift the load from an overloaded OAP to an underloaded OAP.

Instead of relying on the SDN controller for all the OAP association and de-association decisions, our proposed method uses a parameter called load level (L_L), for the SDN controller to decide beyond which load threshold, itself shall make the load balancing decisions for an overloaded OAP. The L_L is calculated in Eq.3.

$$L_L = 100.(1 - \frac{sw_z}{mw_z}) \tag{3}$$

Unless the cw_z exceeds the suggested number of wireless stations that can be associated with an OAP$_z$ (sw_z), the SDN controllers will not issue the load balancing instructions to an OAP$_z$. The higher the value of L_L, the higher the number of wireless stations, an overloaded OAP, will hand over to an underloaded OAP.

Which underloaded OAP to be chosen, is another important decision, the SDN controller needs to take. The destination OAP is chosen based on multi-metrics such as packet loss rate, RSSI, and throughput. Packet loss rate (P) also influenced by packet error rate (P_ϵ) is defined as the ratio of the failed decoded packets to the number of packets sent [52]. Packets are loss due to the hidden terminal problems in wireless local area networks (WLANs). In order to enhance the communication quality at the medium access control (MAC) layer, P should be minimized. The relationship of P to P_ϵ and MAC collision probability (M_{cp}) is given in Eq.4.

$$P = 1 - (1 - P_\epsilon) * (1 - M_{cp}) \tag{4}$$

The P is the loss ratio of sent packets without considering the retransmissions. The packet loss rate considering the retransmissions is different. Let $\hat{P}$ represent the packet loss rate with retransmissions, and let $\hat{R}$ denote the maximum number of retransmissions. Given the same P_ϵ and M_{cp} the relationship of P to $\hat{P}$ is given in Eq.5. The OAP is capable to find out the P for down-link data by counting the retransmitted packets by itself. For the uplink data the retransmitted packets are calculated by inspecting the frame control field in the packet header. The retry indicator in the frame control field, when set to 1, means that the packet is retransmitted. The addition of the retry bit in the OpenFlow supervised flow table allows to count the retransmitted uplink packets.

$$\hat{P} = P^{\hat{R}+1} \tag{5}$$

Considering RSSI as the only metric for load determination of a certain OAP is not sufficient. Neglecting RSSI is also not appropriate as a number of factors such as data rate, bandwidth utilization, power efficiency and packet error rate relate to RSSI performance. RSSI values can be extracted from the packets exchanged between the wireless stations and the OAPs. The threshold defined for the RSSI in QALB is -75 dbm. The OAPs having the RSSI value above the threshold on the wireless stations side are considered as the suitable candidates for association.

Like RSSI, throughput alone is not sufficient to indicate the load on a certain OAP. To be able to compare the results obtained from the emulation platform and the testbed platform for monotonically increasing load variations the normalized throughput is preferred over the absolute throughput in the proposed method. The normalized throughput ($\hat{T}$) is calculated in Eq. 6.

$$\hat{T} = \frac{T_{current}}{T_{maximum}} \tag{6}$$

The current throughput is the amount of data passing through the target OAP in a certain amount of time and the maximum throughput is the maximum limit of throughput, that all the OAPs can support, which is set to 54 Mbps. The current throughput can be measured for any OAP among the 8 OAPs by calculating the packets received successfully in a given time.

The P_z and the $\hat{T}_z$ alone are not sufficient to calculate the load of a specific OAPz. We combine the impact of P_z and $\hat{T}_z$ to calculate the communication quality (ϕ) of an OAP_z in Eq. 7.

$$\phi\left(P_z, \hat{T}_z\right) = (1 - P_z) * (1 - \hat{T}_z) \tag{7}$$

where P_z and $\hat{T}_z$ denote the packet loss rate and normalized throughput, respectively, for an OAP_z. RSSI_{zx} represents the RSSI strength of an OAP_z received by the wireless station x. With the addition of RSSI, the communication quality of an OAP_z with respect to wireless station x is redefined in Eq.8.

$$\phi_{z,x} = (1 - P_z) * \left(1 - \hat{T}_z\right) * RSSI_{zx} \tag{8}$$

The larger the $\phi_{z,x}$, the best the communication quality between the OAP_z and the wireless station x. Higher values of ϕ satisfy the least loaded condition for an OAP, $\phi = \frac{1}{L}$. The OAP with the largest ϕ is chosen as the best candidate for re-association.

A. PROBLEM FORMULATION

A Wi-Fi network with a single ESS, having multiple interconnected BSS, i.e, $\{bss_z | z = 1, 2, \ldots, nb\}$, is considered

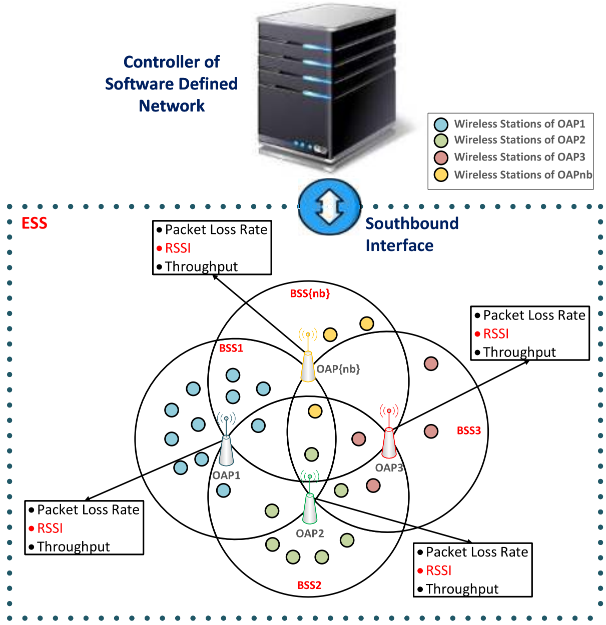

FIGURE 2. Illustration of OAP load balancing in dense SD-Wi-Fi.

in this study, as shown in Fig. 2. A BSS_z is related to a specific OAP_z. The C_z of an OAP_z is defined by cw_z, mw_z and snr_z whereas the L_z is defined by $RSSI_z$, P_z and $\hat{T}_z$. For a specific loading of the network the objective of the SDN controller is to find the minimum L_L for the ESS, while optimizing the throughput, minimizing the average frame delay, minimizing the average number of re-transmissions, minimizing the average number of handoffs, ensuring fair load distribution among the OAPs and reducing the re-association times. The minimum L_L corresponds to J value equal to 1. As seen in Fig. 2, the wireless stations near to OAP1 choose to associate with it rather than the other OAPs in the vicinity, due to higher RSSI, ignoring the load. This OAP selection based only on RSSI, increases the energy consumption of wireless stations, reduces the overall network performance and degrades the fairness among the network resources [53]. The second objective of the SDN controller is to handover the wireless stations from an overloaded OAP to an underloaded OAP which has a comparatively best ϕ.

IV. QoS-AWARE LOAD BALANCING

The section discusses the overview of the QALB for the SD-Wi-Fi. The OAP module is presented first, which manages the OAP capacity and load information reports for the SDN controller and also manages the OAPs associations and de-associations as instructed by the SDN controller. In the second half, the controller module, which is responsible for the load balancing among the OAPs, is presented. Finally, the modified OpenFlow message formats and the other implementation details are elaborated.

A. OVERVIEW

QALB distinguishes itself from the existing methods of load balancing that used a centralized approach, in three aspects.

The overloading of the centralized controller is alleviated in QALB, by a load level, that is used to offload the OAPs associations and de-associations from the controller to the OAPs. Secondly the minimum value of the load level is calculated automatically, which helps in maintaining the load balancing among the OAPs. The third aspect involves the OAP selection, based not only on RSSI but other important QoS factors such as packet loss rate and throughput.

The OpenFlow allows the OAPs to report the capacity and load information to the SDN controller, when ever they boot\re-boot or their status change. The status change reflects the occurrence of association\de-association or the load variation of an OAP. After receiving the OAPs information, the SDN controller computes the fairness index from Eq.1. If the fairness index is not close to 1, the SDN controller issues the load balancing instructions to the OAPs. The overloaded OAPs are instructed to de-associate the wireless stations with the weakest RSSI connection. The wireless stations that get de-associated will try to associate with another OAPs. Only the specific underloaded OAPs chosen by the SDN controller will respond to the association requests.

Algorithm 1 OAP Module Algorithm

OAPs boot\re-boot
OAPs connect to the SDN controller through the OpenFlow protocol
Check the OAP status and detect any new association\de-association
if *statusnotchanged* **then**
    Stable, check the status again in 10 seconds
else
    OAPs forward capacity (C) and load information (L) to the SDN controller
    Wait to receive J, L_{avg} and L_L from Eq.1, Eq.2 and Eq.3 respectively, from the SDN controller
    if $cw_z > sw_z$ **then**
        Overloaded OAP receives the de-association list
        Wireless stations with the weakest RSSI get de-associated from the overloaded OAP
        Underloaded OAP receives the re-association list
        Wireless stations re-associate to the underloaded OAPs having the best ϕ calculated from Eq.8
    else
        Stable, check the status again in 10 seconds
    end if
end if

B. OAP AND CONTROLLER MODULES

The operations of an OAP module are shown in Algorithm 1. The OAPs boot or reboot subject to the condition when the OAPs are operational. The OAPs communicate their load and capacity information to the SDN controller via OpenFlow session, whenever a status change occurs at the OAPs. If the status of the OAP is not changed, the status is rechecked after 10 seconds. If the status of the OAP changes means the OAP receives a new association request or the load of the OAP varies to the previous calculated load, the OAP reports the load and capacity information to the SDN controller. The controller on receiving the load and capacity information, calculates the J, L_{avg} and the L_L from Eqs. 1, 2 and 3 respectively. If current wireless stations (cw) associated to a certain OAP exceeds its suggested number of wireless stations (sw), the controller issues a de-association list of wireless stations to the overloaded OAP. The OAP de-associates the wireless stations with the weakest RSSI connections. The underloaded OAPs with the best ϕ are issued a re-association list by the controller to re-associate the wireless stations which were de-associated earlier on. In this way the load is optimized among the OAPs and wireless stations connect to only those OAPs having the best ϕ, maintaining the QoS at the users end. The process is repeated periodically after 10 seconds to ensure the fairness of load among the OAPs. 10 seconds provide enough time for the mobile wireless stations to associate, de-associate or re-associate to the neighborhood OAPs in the emulation environment.

Algorithm 2 SDN Controller Module Algorithm

Receive the OAPs information
if $nb == 1$ **then**
    Load is balanced
else
    Compute J from Eq.1
    Compute L_{avg} from Eq.2
    Compute L_L from Eq.3
    if $J == 1$ **then**
        Load is balanced
    else
        Prepare the de-association\re-association list
        Send the de-association list to the OAPs having load greater then L_{avg}
        Send the re-association list to the OAPs having load lesser then L_{avg}
    end if
end if

The design of the SDN controller module is shown Algorithm 2. The SDN controller collects the load and capacity information from the OAPs. On receiving the information, the SDN controller computes the number of BSSs in the ESS (nb). If $nb = 1$ then the load balancing condition is met. If there are more than one BSSs in the ESS the controller computes the J, L_{avg} and the L_L from Eqs. 1, 2 and 3 respectively. If J is computed to be 1, this means the load among the OAPs is balanced. On the other hand, if J is not equal to 1 then the SDN controller prepares a de-association\re-association list of the wireless stations for the OAPs. The OAPs load is computed by using the packet loss rate, RSSI and the normalized throughput values. The OAPs having load levels greater than the L_{avg} receive the de-association list of wireless stations having the weakest RSSI connections and the OAPs having load less than the L_{avg} receive the re-association list to re-associate the de-associated wireless stations.

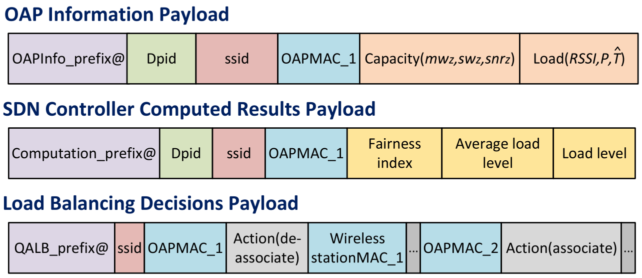

FIGURE 3. Payloads used in QALB for the OpenFlow messages.

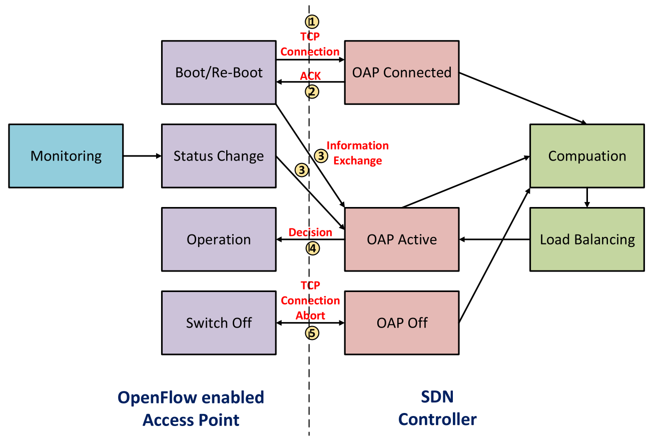

FIGURE 4. The state transition diagram for the OAPs and the SDN controller.

C. OpenFlow MESSAGE FORMATS

In order to support the load balancing between the SDN controller and the OAPs, three OpenFlow extension message formats are specified in Fig. 3. These message formats correspond to the payload, for the exchange of OAPs capacity\load information and the SDN controller computation results, between the SDN controller and the OAPs.

The payload *OAPInfo_prefix@* is used to forward OAP information to the controller. The fields cache the daemon process ID (*Dpid*), the service set ID (*ssid*), the MAC address of the source OAP ($OAPMAC_1$), the capacity information and the load information. The payload format *Computation_prefix@* is similar to the *OAPInfo_prefix@* with the exception that it carries the fairness index, average load level and the load level from the SDN controller to all the OAPs. The last payload *QALB_prefix@* is used to forward the association\de-association results from the SDN controller to the designated OAPs. The fields cache the information of the destination OAP MAC address, the actions to be taken by the specified OAP, and the MAC addresses of all the wireless stations that are chosen to be de-associated or re-associated.

D. IMPLEMENTATION DETAILS

The implementation details take into account the communication protocol between the OAPs and the SDN controller, the computations that the SDN controller performs and the actions that the OAPs need to take, to acquire load balancing. A state transition diagram as shown in Fig. 4, is used to depict the communication that takes place between the OAPs and the SDN controller, in the proposed QALB. *OAPConnected*, signifies the boot\re-boot state of the OAP, *OAPActive* state indicates, that the OAP is forwarding the information to the SDN controller, and the *OAPOff* state symbolizes, that now the OAP has terminated the communication with the SDN controller. Whenever any of the aforementioned states are triggered, the SDN controller computes the fairness index, average load level and the load level, to make the decisions for load balancing.

The OpenFlow extension messages depicted in Fig. 3, are used to exchange the information between the OAPs and the SDN controller. Fairness index, average load level and the load level are computed through Eq.1, Eq.2 and Eq.3 respectively. The de-association list of the wireless stations is prepared according to the First-In-First-Out (FIFO) principle. The load information of the OAPs is monitored through an exponential moving average (EMA) [54], as calculated in Eq. 9.

$$EMA_t = \gamma * L_t + (1 - \gamma) * EMA_0. \tag{9}$$

To detect the load values at OAPs considering all load variations, the EMA takes into account the smoothing constant (γ) which serves as the weighting factor, the load value (L_t) at time (t), and the previous calculated load EMA_0. Usually the range of γ is between [0-1]. In the proposed QALB the γ is set to 0.7 to make sure more weight is put to recent load values. If the previous forecasted load value $EMA_0 \equiv L_{EMA_0} = 15$ and the actual load value $L_1 = 25$, then by using Eq. 9, the new forecasted load value $EMA_1 \equiv L_{EMA_1}$ is calculated to be $= 0.7 * 25 + (1 - 0.7) * 15 = 22$. The difference between EMA_0 and EMA_1 reflects the status change in OAPs. To perform the associations and de-associations, the OAPs use the *hostapd_cli* utility [55].

V. PLATFORMS FOR SD-Wi-Fi

A number of evaluation platforms exist to perform experimentation on wired networks instead of the wireless networks. High density SD-Wi-Fi require an urgent need for such platforms, that can support the performance evaluation. We first emulate and then design a prototype testbed to evaluate the performance of the proposed QoS-aware load balancing strategy in high density SD-Wi-Fi.

A. EMULATION

Software defined network experiments are commonly performed using the Mininet emulator, which use a Linux based platform. Mininet has an inbuilt support for OpenFlow switches, but limited support for wireless links. Mininet and NS-3 together address this limitation by integrating the real time IEEE 802.11 channel emulation feature of NS-3 with Mininet.

Mininet-NS3-Wi-Fi [56] is used for the emulation of the SD-Wi-Fi. A real time emulation mode in NS-3 allows to integrate the real time devices either virtual or real with the simulation code. The emulation mode synchronizes the real time clock with the emulation clock. The connection of two virtual Mininet nodes, using an emulated NS-3 Wi-Fi channel is shown in Fig. 5. NetDevice and TapBridge interfaces act as the bridge for these connections. Each node in the Mininet has a unique network protocol stack and a Linux name. The Mininet node makes use of the Linux Tap NetDevice to get connected to the NS-3 channel. TapBridge allows the TAP Device of the Mininet nodes to get connected to the NS-3 channel. The interconnection of the TAP device and the TapBridge is seen as a NS-3 NetDevice by the NS-3 simulator. NS-3 connects to an external, real time network interface through the NetDevice, by performing simulation of the layer 2 network interface. By integrating the Mininet and NS-3 platforms, we are now able to evaluate the performance of real time high density SD-Wi-Fi.

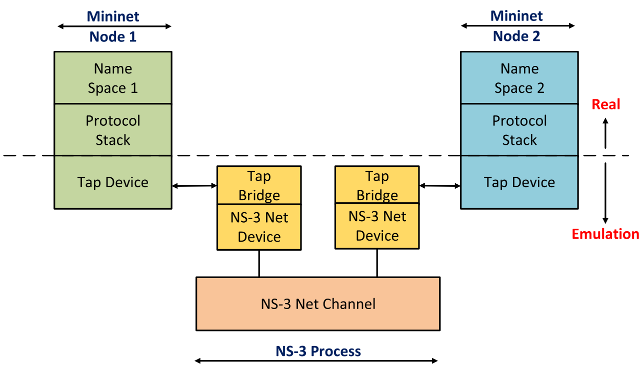

FIGURE 5. Mininet and NS-3 integration.

In the proposed QALB, the OAPs are the learning switches that have the ability to broadcast the packets through the wireless medium. The OAPs connect to the SDN controller through a TCP connection. After the TCP connection is initiated, the SDN controller modifies the flow rules at OAPs through the OpenFlow protocol. The Packet_In messages are directed from the OAPs to the controller and the Packet_Out messages are directed from the controller to the OAPs. The Flow_Mod messages from the controller to the OAPs contribute towards the change of the flow rules. The matching structure for the extended OpenFlow message formats, Packet_In messages, Packet_out messages and Flow_mod messages in Mininet-NS3-Wi-Fi are depicted in Fig. 6.

When the frame is lost in the IEEE 802.11 network, it is re-transmitted. By counting the re-transmitted frames to the total number of frames, the packet loss rate is calculated. The OAPs count the re-transmitted frames themselves for the down-link data. In the up-link data scenario, the frame control field in the packet header allows to calculate the number of re-transmitted frames. The frame control domain of the MAC

```
/ *Flow match field types for OpenFlow basic class extended version. */
enumoxm_ofb_match_fields
{ OFPXMT_OFB_IN_PORT = 0, / *Switch input port. */
OFPXMT_OFB_IN_PHY_PORT = 1, / *Switch physical input port. */
OFPXMT_OFB_METADATA = 2, / *Metadata passed between tables. */
OFPXMT_OFB_DST = 3, / *Destination address. */
OFPXMT_OFB_SRC = 4, / *Source address. */
OFPXMT_OFB_TYPE = 5, / *Frame type. */
OFPXMT_OFB_WLAN_VID = 6, / *WLAN id. */
OFPXMT_OFB_WLAN_PCP = 7, / *WLAN priority. */ ......}

action = of.ofp_action_output(port = out_port)
msg.actions.append(action)

# Send message to switch
 self.connection.send(msg)


fm = of.ofp_flow_mod()
fm.match.in_port = port number
fm.actions.append(of.ofp_action_output(port = port number))
```

FIGURE 6. Extending the basic OpenFlow message format.

header is shown in Fig. 7, where bit 11 is the Retry indicator. Whenever the frame is re-transmitted the Retry indicator is set to 1. The OAPs can determine whether the frame has been re-transmitted by checking the Retry bit of the received frames. The item indicating the Retry bit is added to the OpenFlow extended message formats to count the number of frames re-transmitted in the up-link scenario. The RSSI is extracted from any frame exchanged between the OAP and the wireless station. The current throughput of a specific OAP is calculated by counting the packets received successfully per unit time. The aforementioned information along with the number of associations maintained in the OAPs association table and the signal to noise ratio of the OAPs associated connections are forwarded to the controller. The load balancing application in the controller computes the J, L_{avg} and L_L from Eqs. 1, 2 and 3 respectively, to implement network wide load balancing.

B. TESTBED

There are a number of differences that exist between an emulation and a real time testbed. The differences arise due to the signal fading and channel interference which are usually ignored in the emulation. A prototype testbed may serve as another effective experimentation approach with more development efforts. The cost for a testbed is usually much higher than the emulation platforms, especially for the high density SD-Wi-Fi where a large number of wireless stations exist. There were a number of challenges while making the SD-Wi-Fi testbed, which are as follows:

- The OpenFlow has been used for wired networks in the past. In order for the SDN controller to achieve wireless management, the OpenFlow had to be extended for the wireless networks.
- The cost that incurs on making a testbed for high density SD-Wi-Fi is huge, as large number of wireless stations need to be used and connected to the OAPs.

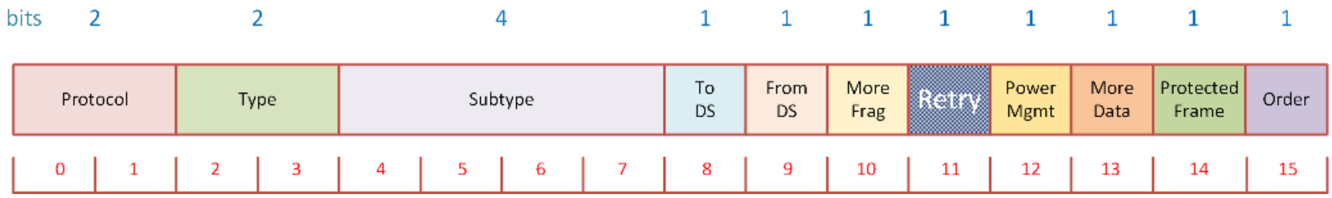

FIGURE 7. **The frame control domain in the MAC header.**

- The wireless stations do not always transmit the packets. To do so, an additional wireless station sensing functions are devised.

The prototype testbed as shown in Fig. 9, is built with the Ubuntu PCs, which supports the USB wireless cards and the Open Virtual Switch (OVS). Linux platform of Ubuntu 14.04 is installed on the Giada mini-PCs, which serves as the main host. Net Gear WNDA 3200 wireless USB cards are used. These USB cards have the softmac wireless driver support which is compatible with the Ubuntu systems. The technical details regarding the testbed setup are provided below:

Step 1: The USB wireless cards act as both, the wireless stations and the OAPs. In addition to the wireless USB cards the Ubuntu PCs are also used as wireless stations. The wireless configuration tools, *hostapd* and the *iw*, are used to configure the wireless USB cards to act as OAPs and wireless stations separately. Thus a low cost testbed with a large number of OAPs and wireless stations is implemented, where the wireless USB cards are configured with unique IP addresses.

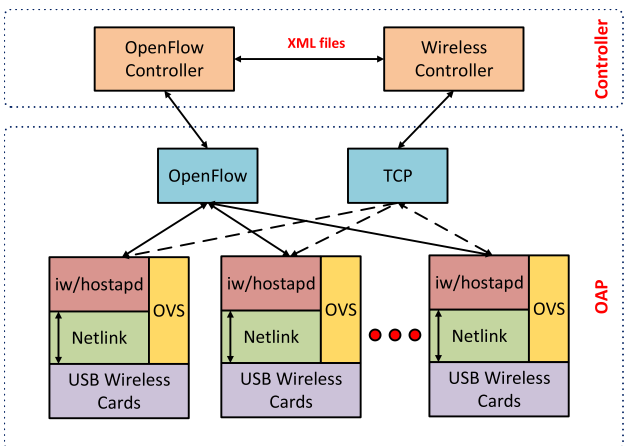

FIGURE 8. **The testbed architecture.**

Step 2: The architecture for the high density SD-Wi-Fi is shown in Fig. 8. The OAPs connect to the SDN switch, to support the management provided by the SDN controller. The wireless packets are manipulated by using the $ovs - vsctl$ command that helps in adding the USB wireless port to the OVS switch. This bridge operation is only used for the IP layer management and does not allow the controller to configure the wireless parameters such as the transmission rate, power control and the channel utilization. The wireless parameters are thus modified by using the *iw* and *hostapd*. A TCP connection is used to connect the OAPs to the controller. The information between the OpenFlow controller and the wireless controller, at the control plane, is shared through the parameter files in extensible markup language (*XML*).

Step 3: The *iw OAP_Namestation* dump command line interface (CLI) is used in the OAPs to collect the wireless stations information, which are in the transmission coverage of the OAP. Similarly the *iw STA_Name* scan CLI is used by the wireless stations to gather the OAPS information, which are discoverable by the wireless stations. The sensed information is forwarded to the SDN controller through a TCP connection.

Step 4: The OVS machines are managed by a RYU controller. The testbed consists of 100 wireless stations and 8 OAPs. The prototype testbed design is depicted in Fig. 9.

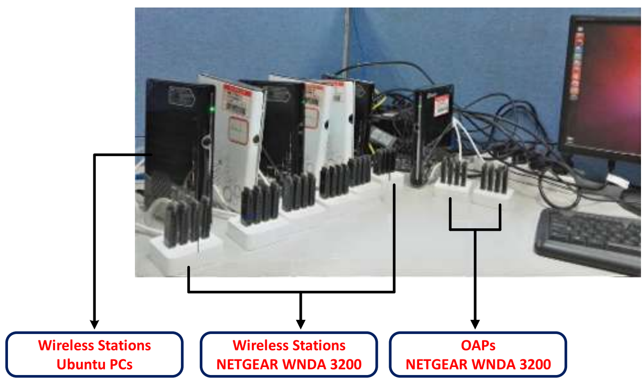

FIGURE 9. **The testbed demonstration with the wireless stations and the OAPs.**

VI. PERFORMANCE EVALUATION

A Mininet-NS3-Wi-Fi emulation platform and a Linux based testbed platform are used to perform comprehensive experiments in order to evaluate the performance of the proposed QALB. The comparison is made to four non-static schemes such as channel measurement based access selection scheme (CMAS) [57], (DL-SINR) downlink-signal to interference plus noise ratio AP selection scheme (DASA) [6], mean probe delay scheme (MPD) [58] and the traditional RSSI scheme.

The Wi-Fi network emulated in Mininet-NS3-Wi-Fi is deployed in a two-dimensional area. 8 OAPs and 50-100 wireless stations are distributed uniformly in 100*100 m^2 area with mobility set to 5 meters/s. The transmission range for each OAP is 1-33 m and every OAP have the same signal attenuation. The same parameter settings are used for the testbed. All the wireless stations and the OAPs are clustered in the testbed, so that each wireless station is within the OAP transmission range. However, the transmission power of each OAP is different from the other OAPs. The parameter settings are shown in the Table 3.

TABLE 3. Parameter settings.

Parameters	Values
Number of OAPs	8
Number of wireless stations	100
Transmission range of each OAP	1-33 m
Maximum permissible throughput of each OAP	54 Mbps
Wireless stations transmission rate	100 Kbps - 500 Kbps
Test time	30 minutes

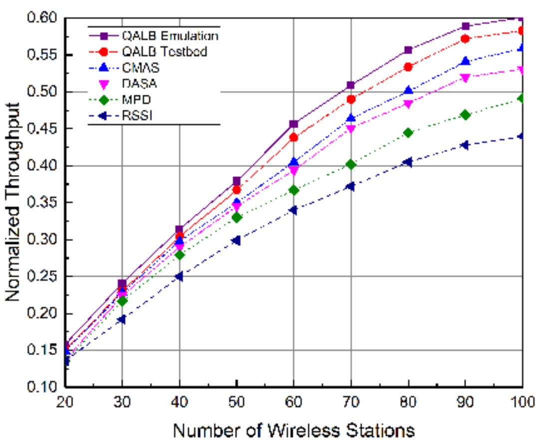

FIGURE 10. Performance of normalized throughput.

A. NORMALIZED THROUGHPUT

The normalized throughput performance is depicted in Fig. 10. It is observed that the normalized throughput increases by increasing the number of wireless stations. As the number of wireless stations increase the increase in the throughput slows down gradually. The result satisfies the analysis that, in a Wi-Fi network with the distributed coordination function (DCF), the throughput reaches a saturation limit or eventually decreases as the traffic load increases. The traditional RSSI based schemes shows the least throughput performance because only one parameter is chosen to select the destination OAP and the load among the OAPs is not balanced. Due to uneven load balancing scenarios, too many wireless stations get associated to a single AP, which result in packet collisions and degraded throughput performance. The worse fact in the traditional RSSI scheme is that the wireless stations keep their association to the AP unless they move away or the AP RSSI value become less due to some AP failure. The total contribution of throughput in RSSI based scheme is just 40% of the total throughput. In the proposed QALB, the load among the OAPs is balanced through a centralized controller and the destination OAP is chosen based on a multi-metric criterion fulfilling the QoS provisioning. If the OAP gets overloaded, the SDN controller prepares the de-association list for the overloaded OAPs and chooses an underloaded destination OAP for the de-associated wireless stations. In this way the load in the network is balanced dynamically and the contention is reduced, hence improving the throughput performance. In addition, the SDN controller de-associates the wireless stations with the lowest RSSI values, and re-associates the de-associated wireless stations to the least loaded OAP keeping the packet loss rate, RSSI and throughput in consideration. Due to this the packet delivery rate is also enhanced at the same time. A slight performance gap between the emulation and testbed results is observed due to uncontrolled environment for testbed where the physical medium has interference from external sources and the distance between wireless stations and AP cause fading.

CMAS shows better throughput performance in comparison to DASA and MPD. MPD shows the least throughput performance due to some unnecessary probing frames, which are used to estimate the load values. The delay between transmitting and receiving the probing frames is computed to estimate the load of the AP. In this way, the wireless stations choose the AP with the least load, resulting in less congestion in the network. The end-to-end aggregate throughput in DASA is improved by choosing an AP that has the best DL-SINR. In this way the PHY transmission rate between the AP and the wireless station is improved. In the CMAS scheme the load among the APs is balanced by estimating the bandwidth which is calculated using the DCF mechanism. Frame aggregation and hidden terminal problems are considered while evaluating the bandwidth in the DCF. While considering the frame aggregation, collision avoidance and load balancing, CMAS performs better than the DASA scheme, where only the SINR of the target AP is taken into the account. The SINR calculation introduces probing overheads, hence reducing the throughput performance. In the MPD scheme only round trip time is considered and larger probing delays can occur, even for the light loaded networks, when the packet collisions takes place by chance. The normalized throughput in QALB is optimized by 4%, 7%, 11% and 16% when compared to CMAS, DASA, MPD and RSSI schemes respectively.

B. AVERAGE FRAME DELAY

The average transmission delays of frames for 5 schemes including the proposed QALB is depicted in the Fig. 11. The average frame delay is the average of the time, when the wireless stations contend for the channel to the time when the OAP receives the frames successfully. It can be observed that when the number of wireless stations increase the average frame delay increases. It is concluded that the average frame delay has a close relationship to the throughput, where a lower throughput will result in a larger average delay for the same packet generation rate. The average frame delay in QALB is

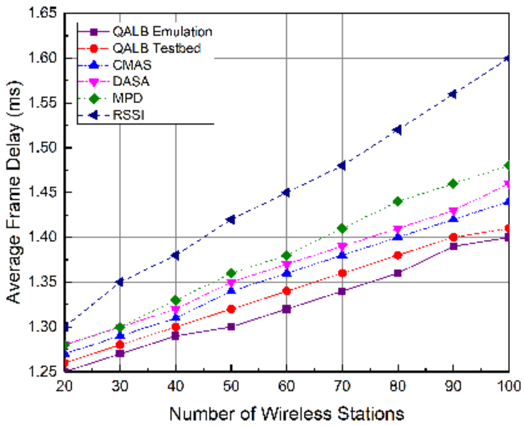

FIGURE 11. Performance of average frame delay.

reduced by 2%, 5%, 7% and 19% when compared to CMAS, DASA, MPD and RSSI schemes respectively.

C. AVERAGE NUMBER OF RE-TRANSMISSIONS

The performance of the average number of re-transmissions is depicted in Fig. 12. It can be observed that as the number of wireless stations increase the average number of re-transmissions for the proposed QALB, CMAS and DASA schemes increase slightly, while for MPD-5 (5 represents 5 probe-request frames) and RSSI schemes, the re-transmissions increase quickly. The average number of re-transmissions in QALB is less than that in RSSI, MPD, DASA and CMAS schemes by approximately 49%, 27%, 7% and 4% respectively. Increased throughput and fairness of load among the OAPs in QALB lead to fewer number of re-transmissions. These results show the efficiency of the proposed QALB in terms of load balancing among multiple OAPS in a software defined approach.

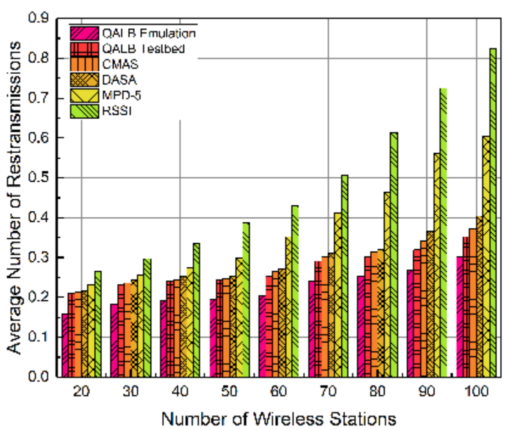

FIGURE 12. Performance of average number of re-transmissions.

D. AVERAGE NUMBER OF HANDOFFS

The relationship between the number of wireless stations and the average number of handoffs is shown in Fig. 13. Average number of handoffs is the ratio of the total number of handoffs executed by all the wireless stations to the total number of wireless stations. MPD-10 corresponds to the MPD scheme using 10 probe-request frames. For fewer number of wireless stations, the average number of handoffs for all 5 schemes is more or less the same. The performance gap among all the 5 schemes increases as the load and the number of wireless stations increase. Among 100 wireless stations, MPD-5 changes 33 APs. In comparison to MPD-5, the MPD-10 scheme has a smaller number of handoffs as more probe-request frames help in estimating the load of each AP more accurately. The proposed QALB has a least number of average handoffs. The reason for this is that the QALB, based on the wireless station's activity, can determine the sequence of OAPs handoffs. The load level for each OAP will decide the number of associations it can maintain. When the wireless stations connected to a certain OAP exceed the suggested number of wireless stations, the handoff will trigger. The wireless stations having the weakest RSSI connections will be chosen for the handoffs first. As a result, the load among the OAPs remain balanced resulting in fewer number of handoffs.

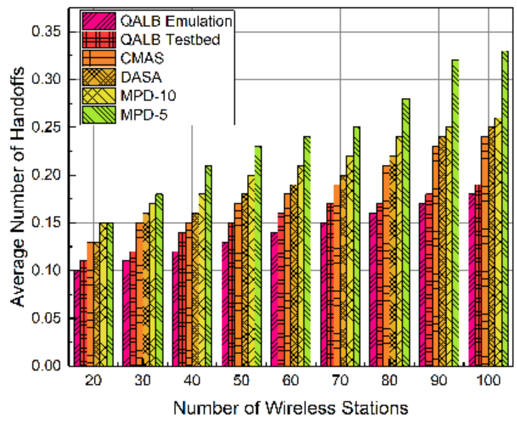

FIGURE 13. Performance of average number of handoffs.

In the MPD algorithm the handoffs are decided by the wireless stations independently. Inactive wireless stations play a greater role in the handoffs, hence increasing the average number of handoffs. In the proposed QALB the handoffs are monitored by the centralized controller, which determines the sequence of the handoffs. Therefore, when a certain OAP gets overloaded the wireless stations are shifted to the least loaded OAPs in an ordered manner, which avoids repetition of handoffs. In comparison to MPD scheme, DASA scheme relies on the SINR values of the AP to select the best destination AP. Higher time for SINR calculation, results in higher handoff times. The SINR measurement brings accuracy, which leads to lesser handoffs but higher probing overheads. In DASA scheme, the SINR measurement relies only on the probe-request and probe-response between the APs and the wireless stations. The additional probing frames that are required for the accurate calculation of the SINR, are not considered. The CMAS scheme, makes use of the available bandwidth, which serves as the metric for determining the handoffs. This metric is only helpful in calculating the contention in the Wi-Fi network, and is not fully capable to find the number of handoffs for a high density Wi-Fi networks. The proposed QALB reduces the average number of handoffs by 6%, 7%, 8% and 15% in comparison to CMAS, DASA, MPD-10 and MPD-5 schemes respectively.

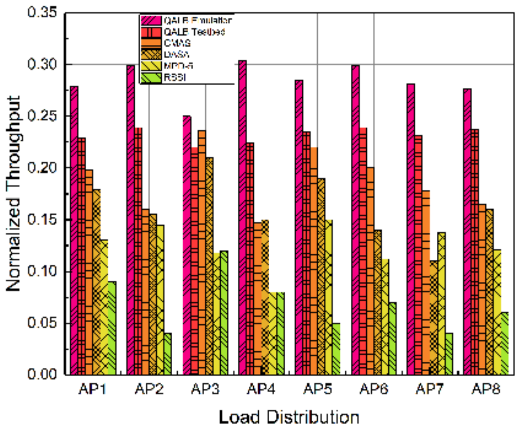

FIGURE 14. Load distribution among all APs.

E. LOAD DISTRIBUTION

In order to show the effectiveness of the load balancing algorithms, the normalized throughput for each AP is shown in Fig. 14. It can be observed that the MPD-5 and RSSI based schemes exhibit higher differences of load among the APs and show lower normalized throughput values. The same is for DASA, as load distribution among the APs is uneven. This is because in the SINR based AP selection, the underloaded AP can also have worse SINR values along with collisions and transient channel fading. The CMAS scheme shows higher throughputs but at the same time the loading difference among the APs is significant. The proposed QALB shows better load distribution among all OAPs, due to the centralized SDN architecture. Moreover the network performance in QALB is improved by less computations as the OAPs decide the associations and de-associations till a certain load level, hence reducing the communication complexity on the wireless stations. The load distribution in QALB is enhanced by 11%, 12%, 16% and 22% when compared to CMAS, DASA, MPD-5 and RSSI schemes respectively.

F. RE-ASSOCIATION TIME

In order to observe the performance evaluation of the re-association time, we conduct two experiments each on the emulation and the testbed platform. In both the experiments, each OAP can associate 50 wireless stations to it at most, and each OAP is maintained with association tables of different sizes {49, 34, 1, 23, 30, 46, 15, 38} respectively. In the first experiment all the OAPs are initialized with the same load values of 60, where as in the second experiment the OAPs are initialized with different load values {100, 80, 50, 40, 80, 70, 60, 30} respectively. Both the experiments are intended to study that how the re-association time is affected by tuning the load level for increasing load diversity and how the associations maintained by the OAPs effect the load balancing.

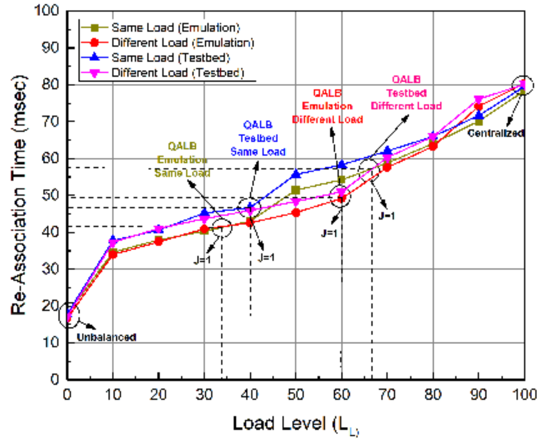

FIGURE 15. Performance of re-association time.

The re-association times for both the experiments for different load levels is depicted in Fig. 15. The results reveal that: 1) the Jain's fairness index converges to 1 before the load level reaches to 100; 2) as the load level increases, the wireless stations that needed to be handed from one OAP to another increase, hence increasing the re-association time; 3) the larger the re-association table size, the more the time the SDN controller takes for the computation process; 4) the SDN controller, adaptively finds the minimum load level, at which the re-association time is reduced and the Jain's fairness index for the Wi-Fi network becomes equal to 1. The convergence of fairness index to 1 in emulation is quicker than the real time testbed due to less fading and low channel interference present in the emulation environment. The load level = 0 means that the Wi-Fi network is unbalanced, whereas the load level = 100 represents the conventional centralized scheme where all APs association\de-association decisions are made by the centralized controller. In the proposed QALB, the adaptive tuning of the load level helps in reducing the re-association times by 27% to 38% as compared to the traditional centralized load balancing methods.

VII. CONCLUSION AND FUTURE WORK

In this paper, we proposed a QoS-aware load balancing scheme for high density software defined Wi-Fi networks. Our scheme is independent of the hardware, and is pertinent to all the devices that support the OpenFlow standards. The network management is accomplished as the SDN-based centralized control divides the OAPs into the overloaded and underloaded classes, manage the load level to balance the load among all the OAPs and choose the best underloaded destination OAP for the wireless stations to re-associate. The destination OAP is chosen based on multi-metrics such as packet loss rate, RSSI and throughput. The communication between the SDN controller and the OAPs is achieved through the extended OpenFlow messages; hence, network administrators may have a unified interface to add new applications and functions to our scheme without changing the original hardware and software framework. The proposed QALB outperforms in terms of the normalized throughput by 4% to 16%, the average frame delay by 2% to 19%, the average number of retransmissions by 4% to 49%, the average number of handoffs by 6% to 15%, the fairness among all the OAPs by 11% to 22% and the re-association times by 27% to 38% when compared to previous non-static load balancing schemes such as CMAS, DASA, MPD and the traditional RSSI schemes, respectively. We plan to extend the two evaluation platforms including the emulation and the testbed, used in this study, to gain a broader control over the wireless parameters and use the Zynq-based programmable Wi-Fi systems to design and implement an intelligent load balancing scheme in a software/hardware co-design approach [59]. We are also interested to evaluate the efficiency of OAPs and the controller module in terms of the iteration times.

REFERENCES

[1] C. Zhang, D. Qiu, S. Mao, X. Hei, and W. Cheng, "Characterizing interference in a campus WiFi network via mobile crowd sensing," in *Proc. Int. Conf. Collaborative Comput., Netw., Appl. Worksharing*, Springer, 2015, pp. 173–182.

[2] Y. Gao, L. Dai, and X. Hei, "Throughput optimization of multi-BSS IEEE 802.11 networks with universal frequency reuse," *IEEE Trans. Commun.*, vol. 65, no. 8, pp. 3399–3414, Aug. 2017.

[3] Y. He, M. Chen, B. Ge, and M. Guizani, "On WiFi offloading in heterogeneous networks: Various incentives and trade-off strategies," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 4, pp. 2345–2385, 4th Quart., 2016.

[4] D.-J. Deng, S.-Y. Lien, J. Lee, and K.-C. Chen, "On quality-of-service provisioning in IEEE 802.11ax WLANs," *IEEE Access*, vol. 4, pp. 6086–6104, 2016.

[5] M. Maity, B. Raman, and M. Vutukuru, "TCP download performance in dense WiFi scenarios: Analysis and solution," *IEEE Trans. Mobile Comput.*, vol. 16, no. 1, pp. 213–227, Jan. 2017.

[6] P. B. Oni and S. D. Blostein, "Decentralized AP selection in large-scale wireless LANs considering multi-AP interference," in *Proc. Int. Conf. Comput., Netw. Commun. (ICNC)*, Jan. 2017, pp. 13–18.

[7] U. Shafi, M. Zeeshan, N. Iqbal, N. Kalsoom, and R. Mumtaz, "An optimal distributed algorithm for best AP selection and load balancing in WiFi," in *Proc. 15th Int. Conf. Smart Cities, Improving Qual. Life Using ICT IoT (HONET-ICT)*, Oct. 2018, pp. 65–69.

[8] T. Lei, X. Wen, Z. Lu, and Y. Li, "A semi-matching based load balancing scheme for dense IEEE 802.11 WLANs," *IEEE Access*, vol. 5, pp. 15332–15339, 2017.

[9] M. Derakhshani, X. Wang, D. Tweed, T. Le-Ngoc, and A. Leon-Garcia, "AP-STA association control for throughput maximization in virtualized WiFi networks," *IEEE Access*, vol. 6, pp. 45034–45050, 2018.

[10] X. Li, J. Luo, W. Wu, and M. Yang, "Joint AP coverage adjustment and user association optimization for load balancing in multi-rate WLANs," in *Proc. IEEE 21st Int. Conf. Comput. Supported Cooperat. Work Design (CSCWD)*, Apr. 2017, pp. 390–395.

[11] M.-C. Tsai, F. Y.-S. Lin, and Y.-F. Wen, "Lagrangian-relaxation-based self-repairing mechanism for Wi-Fi networks," *IEEE Access*, vol. 7, pp. 15868–15883, 2019.

[12] C. Bernardos, A. de la Oliva, P. Serrano, A. Banchs, L. Contreras, H. Jin, and J. Zuniga, "An architecture for software defined wireless networking," *IEEE Wireless Commun.*, vol. 21, no. 3, pp. 52–61, Jun. 2014.

[13] B. Cao, F. He, Y. Li, C. Wang, and W. Lang, "Software defined virtual wireless network: Framework and challenges," *IEEE Netw.*, vol. 29, no. 4, pp. 6–12, Jul. 2015.

[14] Y. Gao, X. Sun, and L. Dai, "Sum rate optimization of multi-standard IEEE 802.11 WLANs," *IEEE Trans. Commun.*, vol. 67, no. 4, pp. 3055–3068, Apr. 2019.

[15] H. T. Dao and S. Kim, "Effective channel gain-based access point selection in cell-free massive MIMO systems," *IEEE Access*, vol. 8, pp. 108127–108132, 2020.

[16] G. Liang, H. Yu, X. Guo, and Y. Qin, "Joint access selection and bandwidth allocation algorithm supporting user requirements and preferences in heterogeneous wireless networks," *IEEE Access*, vol. 7, pp. 23914–23929, 2019.

[17] J. G. Andrews, S. Singh, Q. Ye, X. Lin, and H. S. 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.

[18] A. Raschella, F. Bouhafs, M. Seyedebrahimi, M. Mackay, and Q. Shi, "Quality of service oriented access point selection framework for large Wi-Fi networks," *IEEE Trans. Netw. Service Manage.*, vol. 14, no. 2, pp. 441–455, Jun. 2017.

[19] Y. Dai, D. Xu, and Y. Zhan, "Towards optimal access point selection with available bandwidth estimation," in *Proc. IEEE Int. Conf. Internet Things (iThings) IEEE Green Comput. Commun. (GreenCom) IEEE Cyber, Phys. Social Comput. (CPSCom) IEEE Smart Data (SmartData)*, Jun. 2017, pp. 762–769.

[20] Y. Kim, M.-S. Kim, S. Lee, D. Griffith, and N. Golmie, "AP selection algorithm with adaptive CCAT for dense wireless networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2017, pp. 1–6.

[21] H. Kim, W. Lee, M. Bae, and H. Kim, "Wi-Fi seeker: A link and load aware AP selection algorithm," *IEEE Trans. Mobile Comput.*, vol. 16, no. 8, pp. 2366–2378, Aug. 2017.

[22] S. R. Pokhrel, H. L. Vu, and A. L. Cricenti, "Adaptive admission control for IoT applications in home WiFi networks," *IEEE Trans. Mobile Comput.*, early access, Aug. 19, 2019, doi: 10.1109/TMC.2019.2935719.

[23] Z. Fu, X. Hu, X. Wang, C. Chen, and J. Li, "AP MATRIX: A new access point architecture for reliable public Wi-Fi services," in *Proc. Int. Conf. Comput., Netw. Commun. (ICNC)*, Feb. 2016, pp. 1–5.

[24] C. J. Modani, M. Lakshmanan, N. M. V, and M. Palanivelan, "Analysis of call admission control algorithm in WiMAX network," in *Proc. Online Int. Conf. Green Eng. Technol. (IC-GET)*, Nov. 2016, pp. 1–5.

[25] F. Lyu, L. Fang, G. Xue, H. Xue, and M. Li, "Large-scale full WiFi coverage: Deployment and management strategy based on user spatio-temporal association analytics," *IEEE Internet Things J.*, vol. 6, no. 6, pp. 9386–9398, Dec. 2019.

[26] F. Li, X. Wang, J. Cao, R. Wang, and Y. Bi, "A state transition-aware energy-saving mechanism for dense WLANs in buildings," *IEEE Access*, vol. 5, pp. 25671–25681, 2017.

[27] H. Zhong, Y. Fang, and J. Cui, "LBBSRT: An efficient SDN load balancing scheme based on server response time," *Future Gener. Comput. Syst.*, vol. 68, pp. 183–190, Mar. 2017.

[28] M. Qilin and S. Weikang, "A load balancing method based on SDN," in *Proc. 7th Int. Conf. Measuring Technol. Mechatronics Autom.*, Jun. 2015, pp. 18–21.

[29] J. Yu, Y. Wang, K. Pei, S. Zhang, and J. Li, "A load balancing mechanism for multiple SDN controllers based on load informing strategy," in *Proc. 18th Asia–Pacific Netw. Oper. Manage. Symp. (APNOMS)*, Oct. 2016, pp. 1–4.

[30] G. Cheng, H. Chen, H. Hu, and J. Lan, "Dynamic switch migration towards a scalable SDN control plane," *Int. J. Commun. Syst.*, vol. 29, no. 9, pp. 1482–1499, Jun. 2016.

[31] S. Zhang, J. Lan, P. Sun, and Y. Jiang, "Online load balancing for distributed control plane in software-defined data center network," *IEEE Access*, vol. 6, pp. 18184–18191, 2018.

[32] C. Wang, B. Hu, S. Chen, D. Li, and B. Liu, "A switch migration-based decision-making scheme for balancing load in SDN," *IEEE Access*, vol. 5, pp. 4537–4544, 2017.

[33] T. Hu, J. Lan, J. Zhang, and W. Zhao, "EASM: Efficiency-aware switch migration for balancing controller loads in software-defined networking," 2017, *arXiv:1711.08659*. [Online]. Available: http://arxiv.org/abs/1711.08659

[34] G. Li, X. Wang, and Z. Zhang, "SDN-based load balancing scheme for multi-controller deployment," *IEEE Access*, vol. 7, pp. 39612–39622, 2019.

[35] K. S. Sahoo, D. Puthal, M. Tiwary, M. Usman, B. Sahoo, Z. Wen, B. P. S. Sahoo, and R. Ranjan, "ESMLB: Efficient switch migration-based load balancing for multi-controller SDN in IoT," *IEEE Internet Things J.*, early access, Nov. 8, 2019, doi: 10.1109/JIOT.2019.2952527.

[36] M. Yamada, G. Ishigaki, and N. Shinomiya, "A tie-set based approach of software-defined networking for traffic load balancing," in *Proc. 7th Int. Conf. Comput. Commun. Netw. Technol. ICCCNT*, 2016, p. 20.

[37] A. Lazaris and V. K. Prasanna, "DeepFlow: A deep learning framework for software-defined measurement," in *Proc. 2nd Workshop Cloud-Assist. Netw. CAN*, 2017, pp. 43–48.

[38] K. S. Sahoo and B. Sahoo, "CAMD: A switch migration based load balancing framework for software defined networks," *IET Netw.*, vol. 8, no. 4, pp. 264–271, Jul. 2019.

[39] S. M. M. Gilani, T. Hong, W. Jin, G. Zhao, H. M. Heang, and C. Xu, "Mobility management in IEEE 802.11 WLAN using SDN/NFV technologies," *EURASIP J. Wireless Commun. Netw.*, vol. 2017, no. 1, p. 67, Dec. 2017.
[40] C.-Y. Lin, W.-P. Tsai, M.-H. Tsai, and Y.-Z. Cai, "Adaptive load-balancing scheme through wireless SDN-based association control," in *Proc. IEEE 31st Int. Conf. Adv. Inf. Netw. Appl. (AINA)*, Mar. 2017, pp. 546–553.
[41] S. Manzoor, X. Hei, and W. Cheng, "Towards dynamic two-tier load balancing for software defined WiFi networks," in *Proc. 2nd Int. Conf. Commun. Inf. Syst. ICCIS*, 2017, pp. 63–67.
[42] S. Lin, N. Che, F. Yu, and S. Jiang, "Fairness and load balancing in SDWN using handoff-delay-based association control and load monitoring," *IEEE Access*, vol. 7, pp. 136934–136950, 2019.
[43] H. T. Larasati, F. H. Ilma, B. Nuhamara, A. Mustafa, R. Hakimi, and E. Mulyana, "Performance evaluation of handover association mechanisms in SDN-based wireless network," in *Proc. 3rd Int. Conf. Wireless Telematics (ICWT)*, Jul. 2017, pp. 103–108.
[44] 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.
[45] S. Manzoor, X. Hei, and W. Cheng, "A multi-controller load balancing strategy for software defined Wifi networks," in *Proc. 4th Int. Conf. Cloud Comput. Secur.*, 2018, pp. 622–633.
[46] Y. Cheng, D. Yang, and H. Zhou, "Det-LB: A load balancing approach in 802.11 wireless networks for industrial soft real-time applications," *IEEE Access*, vol. 6, pp. 32054–32063, 2018.
[47] Z. Chen, S. Manzoor, Y. Gao, and X. Hei, "Achieving load balancing in high-density software defined WiFi networks," in *Proc. Int. Conf. Frontiers Inf. Technol. (FIT)*, Dec. 2017, pp. 206–211.
[48] S. Manzoor, C. Zhang, X. Hei, and W. Cheng, "Understanding traffic load in software defined WiFi networks for healthcare," in *Proc. IEEE Int. Conf. Consum. Electron. Taiwan (ICCE-TW)*, May 2019, pp. 1–2.
[49] S. Manzoor, P. Karmon, X. Hei, and W. Cheng, "Traffic aware load balancing in software defined WiFi networks for healthcare," in *Proc. Inf. Commun. Technol. Conf. (ICTC)*, May 2020.
[50] T. Zahid, X. Hei, W. Cheng, A. Ahmad, and P. Maruf, "On the tradeoff between performance and programmability for software defined WiFi networks," *Wireless Commun. Mobile Comput.*, vol. 2018, pp. 1–12, Apr. 2018.
[51] 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, Jun. 1989.
[52] Y. Gao and L. Dai, "Random access: Packet-based or connection-based?" *IEEE Trans. Wireless Commun.*, vol. 18, no. 5, pp. 2664–2678, May 2019.
[53] C. Chen, C. Wang, H. Liu, M. Hu, and Z. Ren, "A novel AP selection scheme in software defined networking enabled WLAN," *Comput. Electr. Eng.*, vol. 66, pp. 288–304, Feb. 2018.
[54] A. S. D. Alfoudi, S. H. S. Newaz, A. Otebolaku, G. M. Lee, and R. Pereira, "An efficient resource management mechanism for network slicing in a LTE network," *IEEE Access*, vol. 7, pp. 89441–89457, 2019.
[55] *hostapd*. [Online]. Available: http://w1.fi/hostapd/
[56] F. Pakzad, S. Layeghy, and M. Portmann, "Evaluation of mininet-WiFi integration via ns-3," in *Proc. 26th Int. Telecommun. Netw. Appl. Conf. (ITNAC)*, Dec. 2016, pp. 243–248.
[57] K. Hong, J. P. Kim, M.-S. Kim, and S. Lee, "Channel measurement-based access point selection in IEEE 802.11 WLANs," *Pervas. Mobile Comput.*, vol. 30, pp. 58–70, Aug. 2016.
[58] J.-C. Chen, T.-C. Chen, T. Zhang, and E. van den Berg, "Effective AP selection and load balancing in IEEE 802.11 wireless LANs," in *Proc. IEEE Globecom*, Dec. 2006, pp. 1–6.
[59] J. Kang, X. Hei, and J. Song, "A comparative study of Zynq-based OpenFlow switches in a software/hardware co-design," in *Proc. Int. Conf. Secur., Privacy Anonymity Comput., Commun. Storage*. Springer, 2017, pp. 369–378.

SOHAIB MANZOOR received the B.S. degree in electrical engineering from the Mirpur University of Science and Technology, Pakistan, in 2011, and the M.S. degree in electrical and electronics engineering from Coventry University, U.K., in 2014. He is currently pursuing the Ph.D. degree with the Huazhong University of Science and Technology, Wuhan, China. Since May 2012, he has been a Permanent Faculty Member with the Department of Electrical Engineering, Mirpur University of Science and Technology. His current research interests include software-defined networks, wireless LANs, and programming.

ZE CHEN received the B.S. and M.S. degrees from the Huazhong University of Science and Technology, Hubei, China, in 2015 and 2017, respectively. Since July 2017, he has been a Senior Engineer with the Department of Deep Learning, Huawei, Shanghai. His current research interests include neural computing, deep learning, knowledge graphs, and natural language processing.

YAYU GAO (Member, IEEE) received the B.Eng. degree in electronic engineering from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 2009, and the Ph.D. degree in electronic engineering from the City University of Hong Kong, Kowloon, Hong Kong, in 2014. Since 2014, she has been with the School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan, China, where she is currently an Assistant Professor. From October 2017 to September 2018, she visited FuNLab, University of Washington, Seattle, USA, where she collaborated with Prof. S. Roy on heterogeneous networks coexistence. Her current research interests include random access networks, wireless LANs, and heterogeneous coexisting networks in 5G NR unlicensed.

XIAOJUN HEI (Member, IEEE) received the B.Eng. degree in information engineering from the Huazhong University of Science and Technology, Wuhan, China, in 1998, the M.Phil. degree in electrical and electronic engineering from The Hong Kong University of Science and Technology, Hong Kong, in 2000, and the Ph.D. degree from the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, in 2008. From 2005 to 2007, he conducted a research visit on P2P networking with the Department of Computer Science and Engineering, NYU School of Engineering, Brooklyn, NY, USA. Since 2008, he has been with the Internet Technology and Engineering Research and Development Center, School of Electronic Information and Communications, Huazhong University of Science and Technology. He is currently an Associate Professor with the School of Electronic Information and Communications, Huazhong University of Science and Technology. His current research interests include edge networking, intelligent healthcare, and embedded network systems. He is a coauthor, together with Y. Liu and K. W. Ross, of the Best Paper in Multimedia Communications, in 2008, awarded by the Multimedia Communications Technical Committee of the IEEE Communications Society.

WENQING CHENG (Member, IEEE) received the B.S. degree in telecommunication engineering and the Ph.D. degree in electronics and information engineering from the Huazhong University of Science and Technology, Wuhan, China, in 1985 and 2005, respectively. She is currently a Professor with the School of Electronic Information and Communications, Huazhong University of Science and Technology. Her current research interests include information systems and e-learning applications.