Ph.D. DISSERTATION

Resource Management Schemes for IEEE 802.11-based Enterprise Wireless LANs

IEEE 802.11 기반 Enterprise 무선 LAN을 위한
자원 관리 기법

February 2019

DEPARTMENT OF ELECTRICAL ENGINEERING
AND COMPUTER SCIENCE
COLLEGE OF ENGINEERING
SEOUL NATIONAL UNIVERSITY

Made Harta Dwijaksara

Abstract

Resource Management Schemes for IEEE 802.11-based Enterprise Wireless LANs

Made Harta Dwijaksara
The School of Electrical Engineering and Computer Science
The Graduate School
Seoul National University

As the IEEE 802.11 (WiFi) becomes the defacto global standard for wireless local area network (WLAN), a huge number of WiFi access points (APs) are deployed. This condition leads to a densely deployed WLANs. In such environment, the conflicting channel allocation between the neighboring access points (APs) is unavoidable, which causes the channel sharing and interference between APs. Thus, the channel allocation (channelization) scheme has a critical role to tackle this issue. In addition, when densely-deployed APs covering a certain area are managed by a single organization, there can exist multiple candidate APs for serving a user. In this case, the user association (UA), i.e., the selection of serving AP, holds a key role in the network performance both in quasi-static and vehicular environments. To improve the performance of WiFi in a densely deployed WLANs environment, we propose a channelization scheme. The proposed channelization scheme utilizes the interference graph to assign the channel for each AP and considers channel bonding. Then, given the channel bonding assignment, the primary channel location for each AP is determined by observing whether the AP supports the static or dynamic channel bonding.

Meanwhile, the UA problem in the quasi-static and vehicular environments are slightly different. Thus, we devise UA schemes both for quasi-static and vehicular

environments. The UA schemes for quasi-static environment takes account the load balancing among APs and energy saving, considering various techniques for performance improvement, such as multicast transmission, multi-user MIMO, and AP sleeping, together. Then, we formulate the problem into a multi-objective optimization and get the solution as the UA scheme. On the other hand, the UA scheme in vehicular environment is realized through handover (HO) scheduling mechanism. Specifically, we propose a HO scheduling scheme running on a server, which determines the AP to which a user will be handed over, considering the road topology. Since a user only needs to decide when to initiate the connection to the next AP, a very fast and efficient HO in the vehicular environment can be realized. For this purpose, we utilize the graph modeling technique to map the relation between APs within the road. We consider a practical scenario where the structure of the road is complex, which includes straight, curve, intersection, and u-turn area. Then, the set of target APs for HO are selected for each user moving on a particular road based-on its moving path which is predicted considering the road topology. The design objective of the proposed HO scheduling is to maximize the connection time on WiFi while minimizing the total HO latency and reducing the number of users which contend for the channel within an AP. Finally, we develop a WLAN testbed to demonstrate the practicality and feasibility of the proposed channelization and UA scheme in a quasi-static environment. Furthermore, through extensive simulations, we compare the performance of the proposed schemes with the existing schemes both in quasi-static and vehicular environments.

Keywords: Enterprise WLAN, resource management, channel allocation, user association, handover scheduling

Student Number: 2013-30773

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background and Motivation

The proliferation of 802.11-enabled devices into daily human life has grown exponentially. This condition leads to the surge of the wireless Internet traffic demand. There have been more and more Wi-Fi access points (APs) installed to cope with this issue, which creates a densely deployed wireless local area networks (WLANs). Since a representative example is the WLAN installed in a company, this type of WLAN is called the enterprise WLAN. The enterprise WLAN is not limited only to WLANs in the companies, but also it can more generally refer to any large scale WLANs in the schools, public area, government building, etc. In such environment, the service area of many APs overlaps with each other. Considering the limited number of available channels, the channel sharing and interference between APs can significantly degrade the network performance. Thus, the first fundamental problem in the enterprise WLAN is channelization, which is the process of allocating a channel to a particular AP. Furthermore, when the service area of many APs overlaps with each other, the user may

have many candidate APs to be selected as its serving AP. If the serving AP is not selected appropriately, the performance of the network can be degraded. It leads to the second fundamental problem in the enterprise WLAN i.e., user association (UA), which is the process to determine the serving AP among one or more available APs for any particular user. Since the channels and APs are resources used to serve the users, these two problems can be generally referred to as resource allocation problems.

The channel allocation aims in reducing the interference between APs by minimizing the conflicting channel assignment between the APs. To improve performance, the channel allocation scheme may also exploit the latest feature of the 802.11 standard i.e., channel bonding [1]. The channel bonding allows the AP and station to combine several adjacent 20 MHz channels, referred to as the basic channels, for their transmission to potentially achieve higher throughput. However, the performance of the WLAN may be worse when the AP uses the channel bonding, particularly in the dense WLAN environment. The reason is that due to the limited number of channels, the channel bonding may increase the channel sharing and interference level between APs. Therefore, in order to get the advantage of channel bonding, the channel bonding should be carefully allocated to the AP. Moreover, to use the channel bonding, the AP should decide one basic channel as the primary channel, in which the ordinary backoff-based contention is performed, among the allocated basic channels. The location of the primary channel also affects the performance of WLAN [2]. Thus, the channelization scheme should not only assign the channel with appropriate width but also decide the primary channel location carefully.

On the other hand, the user association (UA) should be performed considering the WLAN environment. This is because the users do not only access WiFi from indoor (quasi-static environment[1]) but also outdoor even when they move in vehicular

[1]Note that the quasi-static environment refers to the typical mobility pattern of the user. In such environment, the users may move slowly within a short time duration. After they stop, they tend to stay on particular location for relatively long duration.

speed (vehicular environment). Since the characteristic of quasi-static and vehicular environment is different, the UA problem for those environments should be addressed differently. The most simple and traditional UA scheme is to choose the AP with the highest received signal strength (RSS) as the serving AP [3]. This policy, however, may lead to the load unbalancing among APs particularly in quasi-static environment. This is because most of the users around a certain location have a tendency to be associated with a particular AP. In this situation, the users associated with heavily-loaded APs suffer from long delay and low throughput. Moreover, there can be lightly or no loaded APs, of which existence implies the cost inefficiency and the waste of operating power from the network point of view. For solving this unbalance problem in quasi-static environment, a number of UA policies or schemes have been studied and developed.

Meanwhile, the requirement of mobile Internet access from a moving vehicle has been also increasing. Nowadays, many people enjoy various services from a simple news-feed to high rate video streaming, inside moving vehicles [4]. Cellular network, because of its reliability and wide availability, has become a top priority wireless access system for providing such Internet service in the vehicular environment. This has led to the drastic growth of data traffic in the cellular network, causing a traffic-overload problem. Hence, cellular network operators have been driven to offload the traffic to other alternative technology systems [5]. On the other hand, most of the users usually have the limited access to the cellular network due to the high-cost issue. Unless the users register for an expensive unlimited data plan, they cannot freely access the Internet through the cellular network. Thus, not only the cellular network operators but also the users need alternative technologies being much cheaper.

The IEEE 802.11 standard, also known as WiFi, has become a popular alternative technology of cellular network. Among the available WiFi versions, the IEEE 802.11p is particularly designed for vehicular communication. However, as an alternative of

the cellular network, the IEEE 802.11p is still less attractive compared to the IEEE 802.11b/g/n/ac, even in the vehicular environment. First, the IEEE 802.11p network infrastructure is not widely available yet. Hence, to provide Internet access using the IEEE 802.11p, a huge number of IEEE 802.11p roadside units should be initially installed, which requires a high cost. Second, the commercial end-user devices on the market do not typically support the IEEE 802.11p standard, which causes a compatibility issue. Third, even if it is available, the IEEE 802.11p is not likely able to support high-speed Internet access, due to its low transmission rates.

On the other hand, the IEEE 802.11b/g/n/ac has become a *de facto* global standard for wireless local area network (WLAN), and most of the recent mobile electronic devices are equipped with this WiFi chipset. Furthermore, its access points (APs) have been deployed rapidly, not only indoors but also outdoors so that the WiFi service area also covers the road[6]. Recently, such a deployment condition can be easily found in the campus area, company area, government complex, etc. In such environment, the users have been eager to utilize WiFi whenever possible as an alternative of cellular network. This trend is also supported by the fact that the default connection setting of two major mobile operating systems (i.e., Android and iOS) prefers WiFi to cellular network [7]. Based on these reasons, we focus on the IEEE 802.11b/g/n/ac-based WiFi as an alternative of the cellular network in the vehicular environment.

However, when the users access the Internet through WiFi in the vehicular environment, they may seriously suffer from the frequent occurrence of handover (HO) events with long latency. The frequent occurrence of HO events is caused by the small coverage area of an AP [9]. If a user moves in a vehicular speed, its sojourn time inside the service area of an AP is relatively short and thus, in order to maintain the WiFi connection, the user should often trigger the handover from an AP to another AP. On the other hand, since the WiFi was originally designed for static or semi-static users, it finds a target AP for HO by scanning all channels. Such full channel scanning usually

incurs the long HO latency [10].

It is obvious that the frequent HO occurrence and long HO latency may cause severe connection interruption to users. Accordingly, in providing satisfactory services to the users accessing the Internet through WiFi inside a moving vehicle, the fast handover with short latency between APs can be essential [8].

1.2 Related Works

There have been many channel allocation schemes in the literature [12]–[15]. These schemes can be classified into distributed and centralized ones [12]. In the distributed channel allocation scheme, the AP acts independently to select a channel. On the other hand, for the centralized scheme, the AP usually sends a report message to the centralized controller for the channel allocation process, which is done by the controller. Ref. [13] presents the Linux automatics channel selection (ACS) scheme. Each AP executes the ACS independently. It starts by scanning all the available channels, and selects one channel with the least interference for the AP. Furthermore, the authors in [14] used not only interference level but also the load of each AP as the channel allocation metrics. Therefore, the AP may switch its channel dynamically according to the offered load on its working channel. In [15], the authors proposed a centralized channel allocation scheme. In this scheme, each AP estimates the interference level of the other APs on each channel. Then, it builds a bipartite graph representation by using the measured interference level of each AP and assigns each edge with a weight calculated based on the number of AP shared the same channel. The channel allocation problem is modeled as a maximum weight matching problem, which is solved using Hungarian method. Moreover, There is a work in [16] which considers channel bonding during the channel allocation process. However, the solution is based on a simple heuristic algorithm which fails to produce an optimal result. In [2], the authors proposed only a primary channel selection scheme, given the channel allocation for

each AP. Furthermore, given the channel allocation, the authors in [17] proposed a selection mechanism of channel width for data transmission dynamically by allowing transmission with non-contiguous channels.

On the other hand, there have been also extensive works on UA problem for load balancing in WLANs. In some UA schemes, e.g., in [18] – [21], the load balancing is an explicit objective. In [18], a heuristic UA scheme for load balancing was proposed. This scheme marks the AP of which channel usage is greater than a certain level as the overloaded AP. Then, the overloaded APs are avoided to be associated with users. In [19], the UA was formulated into a min-max load balancing problem. In [20], Li *et al*. stated that the UA based on max-min fairness is not appropriate to be applied for multi-rate WLAN and proposed a UA scheme achieving a proportional fairness. Their UA scheme allocates bandwidth to users in proportion to their bit rates so that the sum of the throughput is maximized. The UA problem for load balancing in [21] was modeled as a weighted bipartite graph matching problem. The authors used the channel busy time ratio as a load metric and derived the UA solution using the Kuhn-Munkres (K-M) algorithm.

Some other UA schemes do not claim explicitly that their objective is the load balancing. Nevertheless, the UA objectives such as the delay minimization [3], the throughput maximization [22], [23], or the fairness [24], [25] may lead to the load balancing as a resulting effect. In [3], [23] and [24], the distributed UA schemes based on a game theoretic approach were proposed, respectively. However, since these schemes cannot exploit the advantage of the centralized management, they are less attractive for the enterprise WLANs. In [22], Gong *et al*. considered the 802.11n WLAN supporting the legacy 802.11a/b/g users with a lower transmission rate also for providing backward compatibility. For this type of WLAN, the authors developed a UA scheme for maximizing the minimum user MAC efficiency, where the MAC efficiency of a user is defined as the user throughput divided by the optimal data rate of the user. In [25], the

UA problem was formulated into maximizing a logarithmic utility function of users' throughputs. Then, a heuristic algorithm based on local search technique was proposed to solve the problem.

Besides considering load balancing, some other issues related to UA have been studied for performance improvement of enterprise WLANs. One of them is to minimize the negative effect caused by multicast transmission. In enterprise WLANs, there are many applications taking the advantages of multicast transmission [26]. Typically, the APs carry out multicast transmission with the rate being suitable to the user with lowest transmission rate. Thus, a frequent multicast transmission may cause performance degradation of other users. In the large-scale WLANs such as enterprise WLANs, since there are many users and APs deployed densely, the multicast users can be grouped according to the transmission rate. Then, by associating the multicast users which belong to the same group with as few APs as possible, the negative effect due to the multicast transmission can be minimized. In [27], Chen *et al*. proposed a UA scheme for multicast users. However, this work did not treat how the unicast users should be associated when they coexist with the multicast users.

Another UA-related issue is originated from a distinguished load characteristic of enterprise WLAN, which is a drastic traffic change with time in a certain area. For example, consider the traffic change in a class room (or in a conference room) between during-class (during-meeting) and after-class (after-meeting). Since the deployment density of APs in such an area is usually determined so as to accommodate the peak traffic rate, the APs are likely to be under very low load condition during most of time while wasting the energy. To reduce the energy consumption and thus the operating cost of an enterprise WLAN, it is necessary to switch some APs off during light load condition (e.g., after-class, after-meeting, at night time, etc.). The user association can be effectively exploited to select the APs which should be put into sleep. In [28], Rossi *et al*. proposed a UA scheme which switches on or off an AP according to its current

traffic load, where an AP with light load tries to relocate its users to other AP in light or normal load state so that it can go to sleep. An AP having heavy load also relocates its users to the neighbor AP in light or normal load condition for load balancing. Xu *et al*. [29] formulated the UA problem for energy saving into a mixed integer linear programming (MILP). The main goal is to put as many APs as possible to sleep while satisfying the traffic load requirement of all users.

The AP sleeping also can be considered together with the channel sharing. If there are a large number of users with light traffic, the number of working APs may not be small even when overall system load is low. In such a case, due to the limited available channels, the same channel may be assigned to multiple APs. In [30], Karimi *et al*. formulated a UA problem for WLAN system where the channel sharing between APs is allowed, but they pursued merely throughput maximization without considering the AP sleeping. However, by putting mainly the APs participating in the channel sharing into sleep under light load, the system can operate energy-efficiently while maintaining high channel utilization. Thus, in designing a UA scheme, channel sharing and AP sleeping need to be considered together.

On the other hand, in WLAN systems which adopt the multiuser MIMO (MU-MIMO) technology, multiple users can be served in parallel. Note that, while a traditional single-user MIMO system acts mainly as a physical (PHY) layer performance booster, the MU-MIMO technique can be regarded as a multiple access scheme with which the users share the channel spatially [31]. Recently, Hsu *et al*. [32] presented a UA algorithm customized for MU-MIMO WLANs, where they consider MU-MIMO for increasing the aggregate system throughput. Since many users separated spatially can be served at once by an AP using MU-MIMO, there is room for more sleeping APs. Accordingly, with MU-MIMO, the energy saving as well as the throughput enhancement can be simultaneously targeted.

Finally, the HO decision of wireless networks has been extensively studied. The

works in [33] and [34] present comprehensive survey regarding HO decision, in which most of HO decision schemes focused on selecting a single target AP for HO on the basis of various selection metrics. Regarding the HO in vehicular networks, the recent works in [35] – [37] deal with the vertical HO, whereas the works in [38] – [11] handle the horizontal HO of IEEE 802.11 networks.

In [35], a cloud-based vertical HO scheme for vehicular networks was proposed, where the problem of selecting a target network for HO was formulated as a coalition formation game having a Nash equilibrium. Each coalition is associated with a particular PoA and, thus, joining a coalition means selecting the corresponding PoA. The authors in [36] formulated a weighted multi-objective optimization problem for a vertical HO in vehicular environments. Four metrics of throughput, packet delay, packet loss, and price (cost) were used for selecting a target PoA. The scheme highly depends on instant information of these metrics for each candidate PoA, sent through the serving PoA. However, in the highly dynamic vehicular environment, this information may not be accurate, which is likely to cause the undesired HO result. Similarly with [36], the authors in [37] also considered four types of user preference in selecting a target network, such as longer connection duration with network, higher network bandwidth, lower communication cost of network, and service orientation.

The authors in [38] proposed two algorithms for the horizontal HO between IEEE 802.11 networks, namely longest distance algorithm (LDF) and least HO frequency algorithm (LHFA). In LDF, a user selects a target AP which can provide the longest service time by utilizing the location information and service radius of neighboring APs. Meanwhile, the LHFA intends to minimize the number of HO events. For this purpose, it firstly constructs a handover-graph according to the current trajectory of the user. Then, using Dijkstra's algorithm, it finds a set of APs to be connected along the trajectory which minimizes the number of HO events. However, it did not discuss the trajectory prediction clearly and assumed that the user always moves on a straight road.

When considering various topology of road such as U-turn, bend and intersection, the assumption of straight road is very tight and unpractical, and thus the HO algorithm can be ineffective in real environment.

The scheme in [39] schedules the scanning order of candidate APs on the predicted moving path of the user, based on the expected connection duration. The AP which can provide the longest connection time is scanned first. Since the moving path is predicted with the last three records on general positioning system (GPS) information of the user, if the location information is not accurate, the user may connect to the undesired AP.

Ref. [40] utilizes the road topology to predict the trajectory of a user. The road is divided into several road sections, which are the parts of road bounded by two intersections or a dead end. A set of candidate APs within a road section is prior decided, based on the received signal strength indicator (RSSI) calculated according to a predefined channel model. However, this scheme did not discuss how to select the target APs from candidate APs. Furthermore, since the selection of candidate APs is merely based on the calculated RSS, it still may cause frequent HO events.

The authors in [41] proposed a crowdsourcing-based HO scheme, in which the moving pattern of a user on a particular road is derived from the past moving data of the user reported by itself, and this moving pattern is again used to predict the future moving direction of the user. Then, the AP of which the longest connection time is expected on the predicted future moving path of the user is selected as a HO target. However, this technique requires a complex information configuration of the road, which introduces a significant overhead.

In [11], the authors proposed an AP clustering scheme for a fast HO, where some adjacent APs are grouped into a cluster with the same basic service set identifier (BSSID) and then the handover between APs belonging to the same cluster requires only the channel switching. This clustering-based HO scheme was implemented on mobile devices and APs, and its performance was evaluated through real experiments.

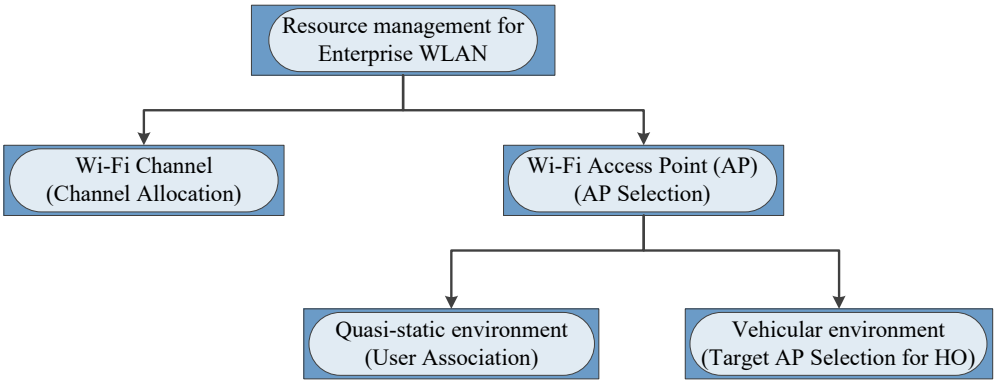


Figure 1.1: Diagram of the proposed resource management schemes for enterprise WLANs.

1.3 Research Scope and Proposed Schemes

The scope of this dissertation are as follows. First, we address the channel sharing and interference problem between APs by devising a channelization algorithm which exploits channel bonding. The channel allocation is executed in-long-term time manner. This is because the channel allocation result usually will not change if the network does not change (i.e., no new AP added or deleted from the system). On the other hand, the user association should be periodically executed in a shorter time interval. This is because the number of users, as well as their traffic load and channel condition, changes over time. For the user association problem, we proposed two schemes which are dedicated for quasi-static and vehicular environments. The reason is that the characteristic of quasi-static environment and vehicular environment is completely different so that a different approach should be taken during UA process. The diagram of the proposed resource management schemes are depicted in Fig. 1.1. The detail of the proposal will be elaborated in the following Subsections.

1.3.1 Centralized Channelization Scheme for Wireless LANs Exploiting Channel Bonding

First, we propose a centralized channelization scheme for WLANs, which exploits the channel bonding to improve the network performance. We present an interference graph generation technique used as the foundation of the channel allocation process. It includes detection of the direct interference (i.e., the APs located within the carrier sensing range) and the hidden interferences (i.e., the APs located out of carrier sensing range of each other, but their service area overlaps). Moreover, we also present the derivation technique of cost value for each edge in the graph. We propose channel allocation scheme by considering the channel bonding to maximize the total network throughput. The allocation process takes account both the controlled and uncontrolled APs. We also propose the primary channel selection scheme given the channel allocation for each AP by considering the static and dynamic channel bonding strategy to maximize the channel utilization. Then, we assess the proposed channel allocation scheme through the experiments utilizing the commercial hardware.

1.3.2 User Association for Load Balancing and Energy Saving in Enterprise WLANs

Second, we develop a UA scheme for enterprise WLAN (quasi-static environment) with two goals: one is the performance enhancement by load balancing and efficient multicasting; the other is the reduction of energy consumption by AP sleeping. We devise a UA scheme which is simultaneously targeting high throughput performance, load balancing among working APs, and the energy efficiency enhancement by AP sleeping, under complicated environment where time-division and space-division multiple access techniques are considered together with channel sharing and MU-MIMO and the multicast traffic as well as unicast traffic is supported. We claim that by considering multicast transmission, MU-MIMO transmission, and channel sharing together, we can increase the chance to switch more APs off without deteriorating network

performance. As a result, considering all those factors together may increase energy-saving. To the best of our knowledge, the UA scheme taking account of these all factors has not been reported yet in literature. In addition, we verify the feasibility and practicality of the proposed UA scheme by implementing it, using the commodity hardware currently available. Furthermore, we demonstrate the outstanding performance of the proposed scheme by comparing with other existing UA schemes through extensive simulation with various settings.

1.3.3 A Graph-Based Handover Scheduling for Heterogenous Vehicular Networks

Finally, we address the HO problems of the WiFi in order to provide a low-cost and high-speed Internet access in the vehicular environment. It is obvious that the frequent HO occurrence and long HO latency may cause severe connection interruption to users. Accordingly, in providing satisfactory services to the users accessing Internet through WiFi inside a moving vehicle, the fast handover with short latency between APs can be essential [8]. In this paper, we design a new HO scheme for effectively realizing fast handover between WiFi APs, by exploiting the characteristic that the movement of a user is predictable with high accuracy on the basis of road topology.

The deployment of APs on a road can be usually modeled as a graph, when the location and service coverage of each AP are known. Since the WiFi network may not cover the whole road area, it is obvious that not only the APs but also the cellular base stations (BSs) should be included in this graph modeling. For convenience of description, when there is no need to explicitly specify AP and BS, we use a point of attachment (PoA) instead of AP/BS. Note that the PoAs, of which service areas are partially overlapped, are referred to as being adjacent to each other. In the graph modeling for PoA deployment on road, each PoA becomes a vertex and the adjacency of two PoAs is represented as an edge between the corresponding vertexes. Since the

moving path of a user may be predicted beforehand on the basis of road topology, the PoAs that the user probably passes through while moving on a road also can be easily identified from the PoA adjacency graph for the road. Thus, it is possible to determine a sequence list of target PoAs for HO, in advance before the vehicle of user enters the road. In this paper, we regard a sequence list of target PoAs for HO as a HO schedule. Such a prior HO scheduling can lead to a fast HO because a relatively long time is usually needed to search a next target PoA for HO.

On the other hand, since a real topology of road is actually various and somewhat complex and, furthermore, the moving path of a user changes over time, a single graph modeling of the entire road is ineffective from a HO scheduling perspective. Thus, partitioning a road into several small segments being appropriate to efficient HO scheduling is very important. We, in this paper, suggest a road partitioning scheme and systematically develop a PoA adjacency graph for each road segment. In such a graph, the edge from a PoA a to another PoA b means that the handover from PoA a to PoA b is possible and the weight value of the edge represents the cost of the corresponding HO. Note that the HO cost should be well mapped onto the objectives of HO scheduling, so that the objectives can be achieved by merely selecting a minimal cost path for the given PoA adjacency graph. In this paper, we also transform our HO scheduling objectives to the HO cost, which is assigned to each edge of the graph.

One of our HO scheduling objectives is to minimize the number of HO events for avoiding frequent HO occurrence. We also aim at minimizing the total HO latency that a user will experience while passing through a road segment. To this end, we assign the lower weight to the edge of the HO type with shorter execution process. Note that the HO process between APs with the same IP subnet address (so-called the L2 HO) has fewer execution steps than the HO between APs with respective different subnet addresses (so-called the L3 HO), because the L3 HO requires an additional DHCP process for getting a new IP address. As a result, the L3 HO events need longer

execution time and we try to reduce the L3 HO events in HO scheduling. In addition, the AP with fewer users is preferred as a target AP for HO, because the users can get the higher opportunity of accessing the channel from mitigated competition. Also, the APs rather than BS are preferred so that the users can enjoy mobile Internet access with cheaper fee. Thus, we assign the lower weight to the edge for any AP than BS and to the edge for the AP with fewer users.

Based on a directed graph representing both the adjacency relationship of PoAs and the HO cost between them, the proposed HO scheduling problem is formulated as a mixed integer linear programming (MILP) for finding a sequence list of PoAs with a minimum cost and is solved by a central scheduling server.

Before entering a particular road segment, a user requests its HO schedule for the segment to the scheduling server and receives a sequence list of target PoAs together with their working channels. While moving on the road segment, as soon as the connection with its current serving PoA becomes unstable, the user can try a connection to the next PoA, without a long-time HO step for searching a target PoA. This realizes a very fast and efficient handover in high-speed vehicular environment. Moreover, since the cental sever is responsible for HO scheduling of all users, the users are free from the burden calculating their HO schedules, with no need of individually storing the graphs for all road segments.

The contributions can be summarized as follows. The HO scheduling scheme for the heterogeneous vehicular network is proposed. The proposed scheme considers the IEEE 802.11b/g/n/ac-based WiFi as an alternative of the cellular network, instead of the IEEE 802.11p. Thus, it requires a relatively low deployment cost and it is directly applicable. Then, we systematically develop a graph model for PoA deployment around a road considering on a complex structure of real roads, which include straight, curve, intersection, and u-turn. Based on the generated graph, a HO schedule of a user is determined in advance before the vehicle of the user enters the road. Moreover, the

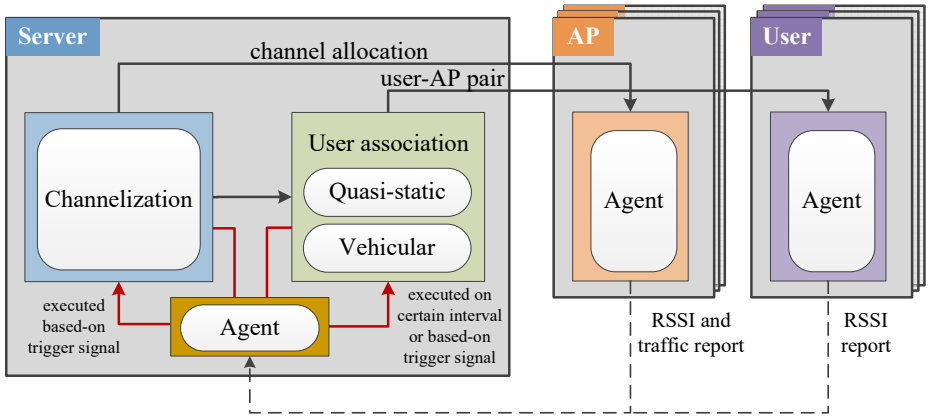


Figure 1.2: Resource allocation procedure for enterprise WLANs.

HO scheduling problem is formulated so as to maximize the connection time on WiFi and minimize the total HO latency while taking account of local load balancing. This makes the fast handover possible and, in turn, enables the users in a moving vehicle to enjoy Internet access with cheap fee. We also provide the detail HO operation utilizing the computed HO schedule in the vehicular environment which include a delivery of HO schedule and a HO event triggering. Finally, we assessed the effectiveness of the proposed HO scheme under a practical scenario of heterogeneous WiFi/cellular network on a real road topology.

Finally, Fig. 1.2 depicts the complete procedure of the proposed resource allocation for enterprise WLANs. The channelization scheme should be executed based on certain trigger. For example during the network initialization phase or when there is network changes (i.e., new APs being added or deleted). On the other hand, the user association for quasi-static environment is executed per-certain interval of time. Moreover, it can be also executed when certain AP experiences overloaded condition so that the load will be balanced. Meanwhile, the HO scheduling scheme in the vehicular environment should be executed based on the request from the users. When the user joins the network or approach end of the road segment, the user sends HO request message.

Then, the server calculates the HO schedule to be sent the corresponding user. Note that the communication between the server and user are always conducted through the serving AP of the corresponding user.

1.4 Organization

The reminder of the thesis is organized as follows. Chapter 2 address describes the propose centralized channelization scheme for wireless LANs exploiting channel bonding. Section 2.1 elaborates the system model and Section 2.2 discuss the channel sharing and bonding condition. The proposed channelization scheme is given in Section 2.3. Then, Section 2.4 presents the detail implementation of the proposed channelization scheme.

In Chapter 3, we present the proposed user association for load balancing and energy saving in enterprise wireless LANs. We start by giving the system model in Section 3.1. Section 3.2 elaborates the detail of the user association probem and the proposed user association algorithm is presented in Seciont 3.3. We give the detail implementation of the proposed user association scheme in Section 3.4.

The graph-based handover scheduling for heterogenous vehicular networks is presented in Chapter 4. The system model and graph-based modeling technique are presented in Section 4.1 and 4.2, respectively. Then, Section 4.3 describes the handover scheduling problem for target AP selection. The proposed handover scheduling algorithm is described in Section 4.4.

In Chapter 5, we discuss the performance evaluation for the proposed schemes. The performance evaluation metrics is presented in Section 5.1. In section, 5.2 we shows the results from experiment utilizing the test beds. The performance results from the simulation are presented in Sectoin 5.3. Finally, Chapter 6 concludes this dissertation.

Chapter 2

Centralized Channelization Scheme for Wireless LANs Exploiting Channel Bonding

2.1 System Model

We consider the WLANs which consist of a set of controlled APs denoted by $\mathcal{M}$, and a set of uncontrolled APs denoted by $\mathcal{E}$, randomly deployed within a particular service area. The controlled APs are connected with the wired link to and managed by a centralized controller. On the other hand, the uncontrolled APs are the independent APs besides the controlled APs, which exist within the deployment area of the controlled APs.

Furthermore, $\mathcal{C}$ denotes the set of all available channels, where $\mathcal{C}$ is a subset of the set of allowable channels defined by the 802.11 standards including all the basic channels and possible channel bonding options[1]. We number the channels in $\mathcal{C}$ with the index between 1 and $|\mathcal{C}|$ in consecutive order, starting with all the basic channels,

[1]The 802.11 standard defines a fixed possible combinations of channel with width 20 MHz, 40 MHz, 80 MHz, and 160 MHz. Thus, the set of channels for allocation should be drawn from these combinations.

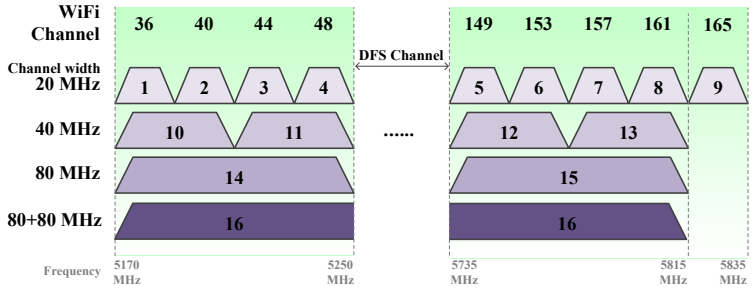


Figure 2.1: Channel indexing given the example set of available channels ($\mathcal{C}$) in 5 GHz.

the channel with width 40 MHz, 80 MHz, and 160 MHz. Fig. 2.1 shows the channel indexing (index $1 \sim 16$), given the example set of available channels ($\mathcal{C}$) in 5 GHz. A channel-c, where $c \in \mathcal{C}$, consists of one or more basic channels. The starting and ending basic channel index of channel-c are i_c^s and i_c^e, respectively. $i_c^s = i_c^e$, if channel-c consists of a single basic channel. Two channels in $\mathcal{C}$ may overlap with each other if those channels have at least one common basic channel. The overlapping indicator for channel-c and channel-d is denoted by λ_{cd}, where $\lambda_{cd} = 1$ if the channel-c and channel-d are overlapped; otherwise $\lambda_{cd} = 0$.

The channel assigned to an AP-a is denoted by c_a. The allocation indicator of channel-c to AP-a, $a \in \mathcal{M}, c \in \mathcal{C}$ is denoted by x_{ac}. $x_{ac} = 1$, if AP-a is assigned channel-c; otherwise $x_{ac} = 0$. On the other hand, the allocation indicator of channel-c to AP-b, $b \in \mathcal{E}, c \in \mathcal{C}$ is denoted by y_{bc}. $y_{bc} = 1$, if AP-b is assigned channel-c; otherwise $y_{bc} = 0$. Moreover, the primary channel index of the uncontrolled AP-b is denoted by p_b. For the uncontrolled AP-b, the c_b, p_b, $i_{c_b}^s$, $i_{c_b}^e$ and y_{bc} are known to the centralized controller in advance. This information can be observed from the beacon frame transmitted periodically by the corresponding AP. Therefore, the c_b, p_b, $i_{c_b}^s$, $i_{c_b}^e$, and y_{bc}, where $b \in \mathcal{E}$, are all given parameters. Table 2.1 presents the summary of some notations used within the paper.

Table 2.1: Notations

Symbol	Description
$\mathcal{M}, \mathcal{E}$	Set of controlled and uncontrolled APs, respectively
$\mathcal{C}$	Set of all available channels
i_c^s, i_c^e	Starting and ending index of the basic channel which construct the channel-c, respectively
λ_{cd}	Overlapping indicator for channel-c and channel-d
c_a	Channel assigned to the AP-a
$\mathrm{W}(c)$	Number of basic channels in channel-c
p_b	Primary channel index of the AP-b, $b \in \mathcal{E}$
x_{ac}, y_{bc}	Assignment indicator of the channel-c to the AP-$a, a \in \mathcal{M}$ and AP-$b, b \in \mathcal{E}$, respectively
β_{aj}	Primary channel assignment indicator of the basic channel with index-j to the AP-a
ρ_{ab}	Direct interference indicator of AP-a and AP-b
υ_{ab}	Hidden interference indicator of AP-a and AP-b
δ_{ab}	Cost value of directed edge from AP-a to AP-b
r_c	Estimated transmission rate using channel-c
SF_a^c	Sharing factor of the AP-a assigned channel-c
$\mathcal{I}_{a,\mathcal{M}}$	Set of interference sources of AP-a which belong to the controlled APs
$\mathcal{I}_{a,\mathcal{E}}$	Set of interference sources of AP-a which belong the uncontrolled APs

2.2 Channel Sharing and Bonding

In this section, we discuss the channel sharing and bonding which affect the channelization process for WLANs.

2.2.1 Interference between APs

To assign the channel to the APs accurately, we have to observe the relation between APs regarding the interference which an AP can cause to the other APs. This relationship can be learned through an interference graph. In this work, the interference graph is a directed graph, in which a vertex represents the AP and a directed edge from AP-a to AP-b means that AP-a is the interference source of AP-b. We classify the interference relation between APs into two categories: direct and hidden interference. The AP-a is a direct interference source of the AP-b if the AP-a is located within the carrier sensing range of the AP-b. On the other hand, the AP-a is a hidden interference source of the AP-b if the AP-a is not located within the carrier sensing range of the AP-b, but some of the service area of the AP-a and AP-b overlaps each other.

The indicators for direct and hidden interference relation between AP-a and AP-b are ρ_{ab} and υ_{ab}, respectively. $\rho_{ab} = 1$, if the AP-a is the direct interference source of AP-b; otherwise $\rho_{ab} = 0$. Moreover, $\upsilon_{ab} = 1$, if the AP-a is the hidden interference source of AP-b; otherwise $\upsilon_{ab} = 0$. The directed edge from the AP-a to the AP-b is assigned a cost value denoted by δ_{ab}. $\delta_{ab} = 1$, if the AP-a is the direct interference source of the AP-b. On the other hand, if the AP-a is the hidden interference source of the AP-b, $0 < \delta_{ab} < 1$. The derivation of the exact cost value for hidden interference case will be discussed in the next section.

2.2.2 Channel Sharing

The implication of assigning the same or overlapped channel to the APs, which are direct or hidden interference source of each other, is the channel sharing condition. The

AP should share the channel with the other APs which are its direct or hidden interference source. The probability of an AP to access the channel depends on the number of interference sources (APs) with the same or the overlapped channel. The sum of total cost contribution of these interference sources including the cost contribution of the respective AP will also be referred to as the sharing factor (SF).

Now let us calculate the SF of an AP when it is assigned a particular channel. The direct and hidden interference source make a different contribution to the SF. For the case of the direct interference, an AP can not transmit (perform backoff) if one of its direct interference sources is transmitting. On the other hand, for the case of hidden interference source, only the users located in the overlapped service area of two APs are affected. This contribution is reflected through the cost value of the directed edge. Therefore, the SF of the AP-a assigned with channel-c is calculated as in Eq. (2.1).

$$SF_a^c = 1 + \sum_{b \in \mathcal{M}, a \neq b} (\rho_{ba} + \upsilon_{ba}) \sum_{d \in \mathbf{C}} x_{bd} \lambda_{cd} \delta_{ba} + \sum_{b \in \mathcal{E}} (\rho_{ba} + \upsilon_{ba}) \sum_{d \in \mathbf{C}} y_{ba} \lambda_{cd} \delta_{ba}. \tag{2.1}$$

The value 1 in Eq. (2.1) accounts for the AP-a itself. The second term accounts for the total contribution of the direct and hidden interference sources of the other controlled APs. The third terms account for the total contribution of the direct and hidden interference sources of the uncontrolled APs. Thus, the access probability of AP-a on the channel-c can be estimated as $\frac{1}{SF_a^c}$. To improve the performance, an AP should be assigned a channel with the least value of SF.

2.2.3 Channel Bonding

There are two channel bonding strategies defined by IEEE 802.11ac standard, i.e., static channel bonding (SCB) and dynamic channel bonding (DCB). In the SCB, the AP and station always use the whole allocated basic channels for every data transmis-

sion [16]. On the other hand, the DCB allows the AP and station to use the available contiguous free basic channels which includes the primary channel for transmission [42]. In DCB, the channels for transmission are decided dynamically right before each data transmission.

In the channel bonding, the location of the primary channel affects the performance of channel bonding. Thus, besides the allocation of a channel with appropriate width, the selection of the primary channel location also becomes an important issue. Depending on the channel bonding strategy, the primary channel location influences the performance differently.

Static Channel Bonding (SCB)

The authors in [2] reported that two issues arise in SCB, i.e., total and partial invading. They are caused by the channel sensing/reserving asymmetry due to the inappropriate location of the primary channel and the different clear channel assessment (CCA) threshold between primary and secondary channel. The reserving range is defined as a maximum distance from the transmitter, in which the receiver can still sense the transmission of the transmitter over its channel. If the receiver listens to the transmission over its primary channel, the distance is referred to as a primary reserving range. Otherwise, the distance is referred to as a secondary reserving range. Thus, if the sensing range is defined from the perspective of the receiver, the reserving range is defined from the perspective of the transmitter.

Fig. 2.2 depicts the example of the total invading case. AP-*a* uses 80 MHz channel (4 basic channels) and sets its primary channel location on the first basic channel. On the other hand, AP-*b* uses 40 MHz channel (2 basic channels) and sets its primary channel location on the third channel with respect to the channel of AP-*a*. According to the standard [1], the primary channel has lower CCA threshold than that of the secondary channel. Thus, the primary and secondary reserving range are differ-

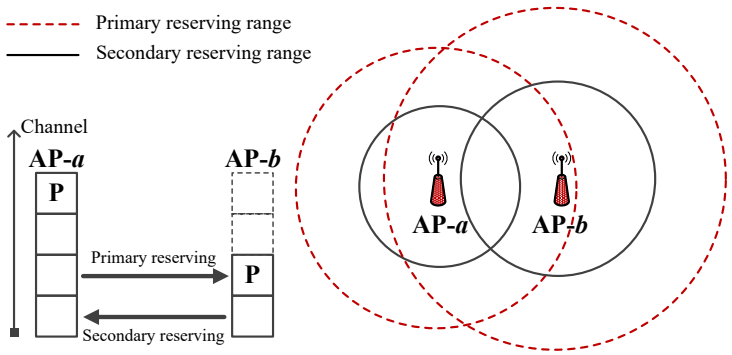


Figure 2.2: Example of the total invading case.

ent. Moreover, the reserving range of an AP using wider bandwidth is usually smaller than that of the AP using narrower bandwidth. The reason is that considering a fixed total transmission power, the transmission power per-basic channel of the AP using wider bandwidth is lower. In Fig. 2.2, the AP-*a* should sense the transmission of AP-*b* using its secondary channel, but the AP-*b* can sense the transmission of AP-*a* using its primary channel. Since the primary reserving range of AP-*a* covers the AP-*b*, the transmission by AP-*a* can always be detected by AP-*b*. However, the secondary reserving range of AP-*b* does not cover the AP-*a*. Thus, the transmission by AP-*b* cannot be detected by AP-*a*. As a result, the AP-*a* is always success in reserving the channel, while the AP-*b* may fail.

Meanwhile, Fig. 2.3 shows the partial invading case. The channel width allocation is the same with that of the Fig. 2.2, but the location is different. The primary channel of AP-*b* is set on the first channel with respect to the channel of AP-*a*. Since the primary reserving range of AP-*b* covers AP-*a* and the AP-*a* can sense the transmission of AP-*b* over its primary channel, the AP-*a* can always detect the transmission of AP-*b*. On the other hand, the primary reserving range of AP-*a* does not cover AP-*b*. Thus, the AP-*b* cannot detect the transmission of AP-*a*. This condition causes the channel

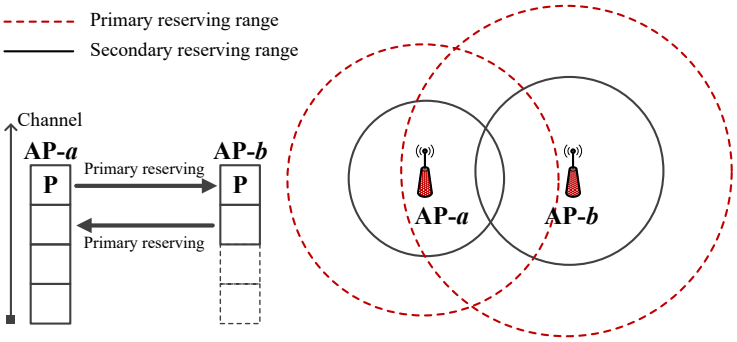


Figure 2.3: Example of the partial invading case.

access probability of AP-*a* is always lower than that of the AP-*b*.

To solve the total invading issue, the authors in [2] suggested setting the primary channel on the same basic channel. Thus, both APs can detect each other transmission. For the partial invading case, it was suggested that the primary channel of both APs are set on different basic channels with the maximum gap of their primary channel index, thus both APs get the same probability to access the channel.

Dynamic Channel Bonding (DCB)

To understand the effect of primary channel location in DCB, let us observe the following example. The AP-*a* and AP-*b* are located within each other carrier sensing range. AP-*a* and AP-*b* are assigned 80 MHz and 40 MHz channel, respectively, as shown in Fig. 2.4.

Fig. 2.4a and 2.4b show a different combination location of the primary channel of AP-*a* and AP-*b*. In the case of Fig. 2.4a, if AP-*b* is transmitting on its 40 MHz channel, AP-*a* cannot transmit at all, even though the rest 40 Mhz channel is free. The reason is that the channel for transmission should always include the primary channel. For the case of Fig. 2.4b, even though AP-*b* is transmitting on its 40 MHz, AP-*a* can still transmit on the rest 40 MHz. Thus, the combination of primary channel location in

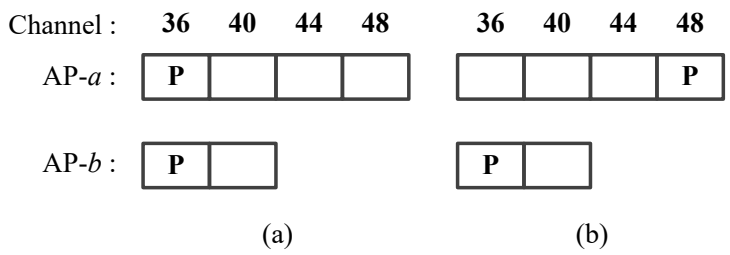


Figure 2.4: Example of two combinations of primary channel location.

the Fig. 2.4b increases the channel utilization. This condition suggests that, for DCB strategy, the primary channel of both APs are also set on different basic channels which maximizes the gap of their channel index to increase the channel utilization.

2.3 Channelization Scheme

The proposed channelization mechanism consists of three main steps. The first step is to build the interference graph as the foundation of the channel allocation process. The second step is to assign particular channel with appropriate width to each AP with the objective of maximizing the total network throughput. The last step is to decide the primary channel location, given the channel allocation of each AP, with the objective of maximizing the channel utilization.

2.3.1 Building Interference Graph

The interference graph is generated by detecting the direct and hidden interference relation between APs. Each AP can observe its direct interference relationship with other APs by listening to the periodic beacon frame of other APs. For this purpose, the central controller may request the controlled APs to initially transmit beacon frame on a particular channel. Then, the neighbor APs which are sensed with the received signal strength indicator (RSSI) higher than a predefined threshold value becomes the direct interference sources. On the other hand, the hidden interference relation can

be detected by the help of the user device. The user device, by default, periodically scans all the available channels. Thus, it has the RSSI information about the neighboring APs. Then, the user reports this information to the central controller through its current serving AP. The user may explicitly send a RSSI report frame periodically to the central controller, or the report frame may be attached to the probe request frame. A single RSSI report frame contains the RSSI information of several neighbor APs scanned by the user from its current location at a particular time. Moreover, a single RSSI information in the report frame contains $\{BSSID,\ RSSI,\ timestamp\}$. Then, the central controller may observe the reporting results to detect the hidden interference relation between the APs. The AP-a and AP-b is hidden interference source of each other if they are both located out of their sensing range, but there is a user experienced the RSSI above the threshold value from both AP-a and AP-b at the same time.

The direct interference edge is assigned a cost equal to one. On the other hand, the cost for hidden interference edge can be calculated by utilizing the RSSI information reported by the user. For this purpose, the central controller divides the time into several intervals with a period of τ. The user may send more than one RSSI report frame within a single time interval. For each user on each time interval, the central controller selects only one report frame randomly. Then, for time interval-t and a pair of AP-a and AP-b, the central controller calculates the total number of users which experiences the RSSI above threshold only from the AP-a denoted by $N^t_{a,-b}$, the total number of user which experiences RSSI above threshold only from the AP-b denoted by $N^t_{b,-a}$, and the total number of users which experiences the RSSI above the threshold value from both the AP-a and AP-b denoted by $N^t_{a,b}$. Then, if the total number of time interval is T, the cost is calculated as in (2.2) and (2.3).

$$\delta_{ab} = \frac{\sum_{t=1}^{T} N^t_{a,b}}{\sum_{t=1}^{T} N^t_{b,-a} + \sum_{t=1}^{T} N^t_{a,b}} \tag{2.2}$$

$$\delta_{ba} = \frac{\sum_{t=1}^{T} N_{a,b}^{t}}{\sum_{t=1}^{T} N_{a,-b}^{t} + \sum_{t=1}^{T} N_{a,b}^{t}} \tag{2.3}$$

Expressions (2.2) and (2.3) account for the portion of the number of users located within the service area of both AP-a and AP-b at the same time over the number of all users of AP-a and AP-b, respectively. The graph and the corresponding cost value should be periodically updated (e.g., once every several hours). For the cost, the moving average technique can be used to update its value. Then, the channel allocation scheme should be re-run after the graph is updated.

2.3.2 Channel Allocation

Problem Formulation

The objective of the proposed channel allocation scheme is to maximize the total network throughput. Therefore, the channel allocation problem is formulated as in (3.8)–(4.18).

$$\max \sum_{a \in \mathcal{M}} \frac{\sum_{c \in \mathcal{C}} x_{ac} r_c}{\sum_{c \in \mathcal{C}} x_{ac} SF_a^c} \tag{2.4}$$

$$s.t. \sum_{c \in \mathcal{C}} x_{ac} = 1, \qquad \forall a \in \mathcal{M} \tag{2.5}$$

$$x_{ac} \in \{0,1\}, \qquad \forall a \in \mathcal{M}, c \in \mathcal{C} \tag{2.6}$$

where r_c is the transmission rate when using the channel-c, $c \in \mathcal{C}$, which is derived from [16][2]. The objective (3.8) maximizes the total network throughput. The constraint (4.17) states that a single AP should be assigned exactly one of channel-c, $c \in \mathcal{C}$. The constraint (4.18) is already clear.

The problem in (3.8)–(4.18) is a nonlinear fractional problem which can not be solved directly. Thus, we transform the problem into its equivalent linear form using

[2]In [16], the estimated transmission rate using the channel with width: 20 MHz = 65 Mbps, 40 MHz = 121.5 Mbps, 80 MHz = 175.5 Mbps, and 160 MHz = 232 Mbps.

a parametric linearization technique as in [43]. For this purpose, for each $a \in \mathcal{M}$, we introduce variables q_a. Furthermore, let $\rho_{ba}^{cd} := x_{ac}x_{bd}$ and $\Psi_{ba}^{cd} := \lambda_{cd}\delta_{ba}\big(\rho_{ba} + \upsilon_{ba}\big)$, thus we have the transformed problem as in (2.7)–(3.16).

$$\max \sum_{a\in\mathcal{M}} \Big(\sum_{c\in\mathcal{C}} x_{ac}r_c - q_a \sum_{c\in\mathcal{C}} \sum_{b\in\mathcal{M},a\neq b} \sum_{d\in\mathcal{C}} \rho_{ba}^{cd}\Psi_{ba}^{cd} - q_a \sum_{c\in\mathcal{C}} \sum_{b\in\mathcal{E}} \sum_{d\in\mathcal{C}} x_{ac}y_{bd}\Psi_{ba}^{cd} \Big) \tag{2.7}$$

$$s.t. \sum_{c\in\mathcal{C}} x_{ac} = 1, \qquad \forall a \in \mathcal{M} \tag{2.8}$$

$$\rho_{ba}^{cd} \geq x_{ac} + x_{bd} - 1, \qquad \forall a,b \in \mathcal{M}, a \neq b, c,d \in \mathcal{C} \tag{2.9}$$

$$\rho_{ba}^{cd} \leq x_{ac}, \qquad \forall a,b \in \mathcal{M}, a \neq b, c,d \in \mathcal{C} \tag{2.10}$$

$$\rho_{ba}^{cd} \leq x_{bd}, \qquad \forall a,b \in \mathcal{M}, a \neq b, c,d \in \mathcal{C} \tag{2.11}$$

$$\rho_{ba}^{cd} \in \{0,1\}, \qquad \forall a,b \in \mathcal{M}, a \neq b, c,d \in \mathcal{C} \tag{2.12}$$

$$x_{ac} \in \{0,1\}, \qquad \forall a \in \mathcal{M}, c \in \mathcal{C} \tag{2.13}$$

The constraints (4.20)–(3.15) are introduced to ensure $\rho_{ba}^{cd} = x_{ac}x_{bd}$. Given the value of $q_a, \forall a \in \mathcal{M}$, the problem in (2.7)–(3.16) is mixed-integer linear programming (MILP) which can be optimally solved by using CPLEX [44] or LPSolve [45] solver.

Allocation Algorithm

According the discussion above, we need to define $q_a, \forall a \in \mathcal{M}$ to solve the problem in (2.7)–(3.16). We initially set the value of $q_a, \forall a \in \mathcal{M}$ to zero. Then, we use the Dinkelbach's Algorithm to update their value until the convergence (i.e., the gap between the old and new values of $q_a, \forall a \in \mathcal{M}$ are not larger than a predefined threshold ε_1) [46]. Its convergence and optimality are proven in [46] and [47]. The Algorithm 1 presents the detail of the algorithm.

Algorithm 1 Channel Allocation Algorithm

$q_a \leftarrow 0$, for $a \in \mathcal{M}$
$oldQ_a \leftarrow 1$, for $a \in \mathcal{M}$
maxITR $\leftarrow 10$, itr $\leftarrow 1$
$\varepsilon_1 \leftarrow 10^{-5}$, stop $\leftarrow false$
while stop = $false$ and itr $\leq$ maxITR **do**
    Solve (2.7)–(3.16) using MILP solver, given all q_a.
    stop $\leftarrow true$
    for all $a \in \mathcal{M}$ **do**
        if $|q_a - oldQ_a| > \varepsilon$ **then**
            $oldQ_a \leftarrow q_a$, $q_a \leftarrow \frac{\sum_{c \in \mathcal{C}} x_{ac} r_c}{\sum_{c \in \mathcal{C}} x_{ac} SF_a^c}$
            stop $\leftarrow false$
        end if
    end for
    itr $\leftarrow$ itr + 1
end while
return The channel allocation (x_{ac})

2.3.3 Primary Channel Selection

Problem Formulation

Given the channel allocation result (i.e., c_a is known, $\forall a \in \mathcal{M} \bigcup \mathcal{E}$), the primary channel location should be decided for each controlled AP. The strategy for selecting primary channel location is: if AP-a and AP-b have the total invading relation, their primary channel should be set on the same basic channel. Otherwise, for partial invading and dynamic channel bonding case, their primary channel should be set on the different basic channels which maximize the gap of their index. First, we populate the interference sources of AP-a given the channel allocation result. The set of interference sources which belong to the controlled and uncontrolled APs are denoted by $\mathcal{I}_{a,\mathcal{M}}$ and $\mathcal{I}_{a,\mathcal{E}}$, respectively.

$$\mathcal{I}_{a,\mathcal{M}} = \bigcup_{b \in \mathcal{M}, a \neq b} \left\{ b \;\middle|\; \lambda_{c_a c_b} = 1 \wedge \left(\rho_{ba} + \upsilon_{ba}\right) = 1 \right\} \tag{2.14}$$

$$\mathcal{I}_{a,\mathcal{E}} = \bigcup_{b\in\mathcal{E}} \left\{ b \;\middle|\; \lambda_{c_a c_b} = 1 \wedge \left(\rho_{ba} + \upsilon_{ba}\right) = 1 \right\} \tag{2.15}$$

Then, $\mathcal{I}_a = \mathcal{I}_{a,\mathcal{M}} \bigcup \mathcal{I}_{a,\mathcal{E}}$. Let β_{aj} be the selection indicator of the basic channel with index-j as a primary channel of the AP-a. $\beta_{aj} = 1$, if the AP-a is assigned basic channel with index-j as its primary channel; otherwise $\beta_{aj} = 0$. Moreover, the total number of basic channels allocated to the AP-a is denoted by $\mathrm{W}(c_a)$, and the maximum possible gap of the index of the primary channels is Δ_{m}. The primary channel selection problem is formulated as in (2.16)–(2.18).

$$\max \sum_{a\in\mathcal{M}} \sum_{b\in\mathcal{I}_a} \left(w_{ba}\delta_{ba}\phi_{ba} + \frac{w_{ba}\delta_{ba}\phi_{ba}}{\max_{d\in\mathcal{I}_a} w_{da}\delta_{da}\phi_{da}} \right) \tag{2.16}$$

$$s.t. \sum_{j=i^s_{c_a}}^{i^e_{c_a}} \beta_{aj} = 1, \qquad \forall a \in \mathcal{M} \tag{2.17}$$

$$\beta_{aj} \in \{0,1\}, \qquad \forall a \in \mathcal{M},\ i^s_{c_a} \le j \le i^e_{c_a} \tag{2.18}$$

where ϕ_{ab} and w_{ab} in (2.16) are defined as follows:

$$\phi_{ab} = \begin{cases} \left|\sum_{j=i^s_{c_a}}^{i^e_{c_a}} j\beta_{aj} - \sum_{k=i^s_{c_b}}^{i^e_{c_b}} k\beta_{bk}\right|, & \forall b \in \mathcal{I}_{a,\mathcal{M}} \\ \left|\sum_{j=i^s_{c_a}}^{i^e_{c_a}} j\beta_{aj} - p_b\right|, & \forall b \in \mathcal{I}_{a,\mathcal{E}} \end{cases} \tag{2.19}$$

$$w_{ab} = \begin{cases} -\Delta_{\mathrm{m}}, & \text{if AP-}a\text{ \& AP-}b \\ & \text{are total invading} \\ |1 + \mathrm{W}(c_a) - \mathrm{W}(c_b)|, & \text{otherwise} \end{cases} \tag{2.20}$$

The objective (2.16) minimizes the gap between primary channel index of AP-a and AP-b (ϕ_{ab}) if AP-a and AP-b are total invading. The reason is that, for the total invading case, we assign negative weight which equals to $-\Delta_{\mathrm{m}}$. This value ensures that the primary channel of the APs, which have the total invading relation, to be set on the same basic channel. Otherwise, the objective (2.16) maximizes the gap between the

primary channel index of the APs. The higher priority (higher weight value) is given to the pair of APs whose the total number of basic channel difference is larger than that of the others. The constraint (2.17) states that exactly a single basic channel should be set as a primary channel for an AP. The constraint in (2.18) is self-explanatory.

The problem in (2.16)–(2.18) should be converted into a solvable linear problem. For this purpose, we introduce variables $z_{ab}, \forall a \in \mathcal{M}, \forall b \in \mathcal{I}_a$. Furhtermore, let $\mu_a := \max_{d \in \mathcal{I}_a} w_{da}\delta_{da}\phi_{da}$, $\psi_{ba} := |\sum_{j=i^s_{c_a}}^{i^e_{c_a}} j\beta_{aj} - \sum_{k=i^s_{c_b}}^{i^e_{c_b}} k\beta_{bk}|$, and $\sigma_{ba} := |\sum_{j=i^s_{c_a}}^{i^e_{c_a}} j\beta_{aj} - p_b|$. Then, the transformed problem becomes:

$$\max \sum_{a\in\mathcal{M}} \Big(\sum_{b\in\mathcal{I}_{a,\mathcal{M}}} 2w_{ba}\delta_{ba}\psi_{ba} + \sum_{b\in\mathcal{I}_{a,\mathcal{E}}} 2w_{ba}\delta_{ba}\sigma_{ba} - \sum_{b\in\mathcal{I}_a} z_{ba}\mu_a \Big) \tag{2.21}$$

$$s.t. \sum_{j=i^s_{c_a}}^{i^e_{c_a}} \beta_{aj} = 1, \qquad \forall a \in \mathcal{M} \tag{2.22}$$

$$\mu_a \geq w_{ba}\delta_{ba}\psi_{ba}, \qquad \forall a \in \mathcal{M},\ \forall b \in \mathcal{I}_{a,\mathcal{M}} \tag{2.23}$$

$$\mu_a \geq w_{ba}\delta_{ba}\sigma_{ba}, \qquad \forall a \in \mathcal{M},\ \forall b \in \mathcal{I}_{a,\mathcal{E}} \tag{2.24}$$

$$\psi_{ba} - \sum_{j=i^s_{c_a}}^{i^e_{c_a}} j\beta_{aj} + \sum_{k=i^s_{c_b}}^{i^e_{c_b}} k\beta_{bk} \leq 2\Delta_{\mathrm{m}} d^1_{ba}, \qquad \forall a \in \mathcal{M},\ \forall b \in \mathcal{I}_{a,\mathcal{M}} \tag{2.25}$$

$$\psi_{ba} + \sum_{j=i^s_{c_a}}^{i^e_{c_a}} j\beta_{aj} - \sum_{k=i^s_{c_b}}^{i^e_{c_b}} k\beta_{bk} \leq 2\Delta_{\mathrm{m}} d^2_{ba}, \qquad \forall a \in \mathcal{M},\ \forall b \in \mathcal{I}_{a,\mathcal{M}} \tag{2.26}$$

$$d^1_{ba} + d^2_{ba} = 1, \qquad \forall a \in \mathcal{M},\ \forall b \in \mathcal{I}_{a,\mathcal{M}} \tag{2.27}$$

$$d^1_{ba}, d^2_{ba} \in \{0,1\}, \qquad \forall a \in \mathcal{M},\ \forall b \in \mathcal{I}_{a,\mathcal{M}} \tag{2.28}$$

$$\sigma_{ba} - \sum_{j=i^s_{c_a}}^{i^e_{c_a}} j\beta_{aj} + p_b \leq 2\Delta_{\mathrm{m}} e^1_{ba},$$

$$\forall a \in \mathcal{M},\ \forall b \in \mathcal{I}_{a,\mathcal{E}} \tag{2.29}$$

$$\sigma_{ba} + \sum_{j=i^s_{c_a}}^{i^e_{c_a}} j\beta_{aj} - p_b \leq 2\Delta_{\mathrm{m}} e^2_{ba},$$

$$\forall a \in \mathcal{M},\ \forall b \in \mathcal{I}_{a,\mathcal{E}} \tag{2.30}$$

$$e^1_{ba} + e^2_{ba} = 1, \qquad \forall a \in \mathcal{M},\ \forall b \in \mathcal{I}_{a,\mathcal{E}} \tag{2.31}$$

$$e^1_{ba}, e^2_{ba} \in \{0,1\}, \qquad \forall a \in \mathcal{M},\ \forall b \in \mathcal{I}_{a,\mathcal{E}} \tag{2.32}$$

$$\beta_{aj} \in \{0,1\}. \qquad \forall a \in \mathcal{M},\ i^s_{c_a} \leq j \leq i^e_{c_a} \tag{2.33}$$

The constraints (2.23) and (2.24) ensure that maximizing $\left(-z_{ba}\mu_a\right)$ is equivalent to maximizing $\left(-z_{ba}\max_{b\in\mathcal{I}_a} w_{ba}\delta_{ba}\phi_{ba}\right)$ [48]. The constraints in (2.25)–(2.28) guarantee that the value of $\psi_{ba} = |\sum_{j=i^s_{c_a}}^{i^e_{c_a}} j\beta_{aj} - \sum_{k=i^s_{c_b}}^{i^e_{c_b}} k\beta_{bk}|$ [48]. Finally, the constraints in (2.29)–(2.32) guarantee that the value of $\sigma_{ba} = |\sum_{j=i^s_{c_a}}^{i^e_{c_a}} j\beta_{aj} - p_b|$ [48].

Selection Algorithm

Given $z_{ab}, \forall a \in \mathcal{M}, \forall b \in \mathcal{I}_a$, the problem in (2.21)–(2.33) is a mixed-integer linear programming (MILP). Thus, to solve the problem, we initially set $z_{ab} = 0, \forall a \in \mathcal{M}, \forall b \in \mathcal{I}_a$. Then, utilizing the Dinkelbach's Algorithm [46, 47], we update the value of z_{ab} until it converges (i.e., the gap between the old and new values of $z_{ba}, \forall a \in \mathcal{M}, \forall b \in \mathcal{I}_a$ are not larger than a predefined threshold ε_2). The complete algorithm is presented in Algorithm 2.

2.4 Implementation

For the implementation purpose, we use a desktop computer as a central server. This computer uses Intel i7 CPU and it has 8 GB of RAM. The channel allocation algorithm is implemented using Java programming language. CPLEX [27] is used for the MILP solver of the channel allocation and primary channel selection problem. For the wireless AP, we use the TP-LinkAP(TL-WDR4300) and TP-Link C7 AC1750 [36] with

Algorithm 2 Primary Channel Selection Algorithm

$z_{ba} \leftarrow 0$, for $a \in \mathcal{M}, b \in \mathcal{I}_a$
$oldZ_{ba} \leftarrow 1$, for $a \in \mathcal{M}, b \in \mathcal{I}_a$
maxITR $\leftarrow 10$, itr $\leftarrow 1$
$\varepsilon_2 \leftarrow 10^{-3}$, stop $\leftarrow false$
while stop = $false$ and itr $\leq$ maxITR **do**
    Solve (2.21)–(2.31) using MILP solver, given all z_{ba}.
    stop $\leftarrow true$
    for all $a \in \mathcal{M}$ **do**
        for all $b \in \mathcal{I}_a$ **do**
            if $|z_{ba} - oldZ_{ba}| > \varepsilon_2$ **then**
                $oldZ_{ba} \leftarrow z_{ba}$, $z_{ba} \leftarrow \frac{w_{ba}\delta_{ba}\phi_{ba}}{\max_{d \in \mathcal{I}_a} w_{da}\delta_{da}\phi_{da}}$
                stop $\leftarrow false$
            end if
        end for
    end for
    itr $\leftarrow$ itr + 1
end while
return The primary channel position (β_{aj})

the OpenWRT as its operating system [37]. We utilize the *iw* tool available for OpenWRT distribution to perform channel scanning on each AP. We also develop an agent for the AP to send the channel scanning result to the server. The scanning operation is controlled by the server. During initiation phase, the central controller ask all APs to set its working channel on particular frequency through the agent installed in the AP. However, this frequency will be only assigned for collecting RSSI information not as a final allocation. Then, the controller requests the APs to perform channel scanning one-by-one with certain order. Thus, the RSSI of an AP can be measured by other APs. Upon receiving the RSSI information from all APs, the server store it on its local database for interference graph generation. In addition, for the station we use Galaxy Tab SM-P900 and Galaxy Note SM-P600. We also develop an agent for each station to report channel scanning result to the server.

Chapter 3

User Association for Load Balancing and Energy Saving in Enterprise Wireless LANs

3.1 System Model

3.1.1 IEEE 802.11 ESS-based Enterprise WLAN

We consider an IEEE 802.11 based WLAN, which consists of a set of wireless APs denoted by $\mathcal{A}$ and a set of users (stations) denoted by $\mathcal{S}$. The APs in $\mathcal{A}$ form an Extended Service Set (ESS) which is controlled by a centralized controller. Each AP serves a certain service area which may overlap with the service areas of other APs. The APs are connected to the Internet and to the centralized controller using wired links.

We assume that a proper channelization algorithm has assigned an appropriate channel to each AP so that the interference from the APs of other networks (i.e., the external APs being out of the ESS) can be minimized.[1] Within the ESS under consid-

[1] This assumption is acceptable and practical in the enterprise WLANs where the APs are managed by a centralized controller. Previous works on UA problem in [19], [22], and [30] also consider a similar assumption.

eration, each AP uses a single 20 MHz channel and multiple APs may share the same channel. The channel sharing between AP l and AP n being located within the carrier sensing range of each other is represented by a binary variable $\vartheta_{l,n}$: if the AP l and the AP n share the same channel, $\vartheta_{l,n} = 1$; otherwise, $\vartheta_{l,n} = 0$. Note that $\vartheta_{n,n} = 1$. Each user measures and reports the channel condition of all surrounding APs, periodically by overhearing the beacons of the APs.

There are two types of users, the unicast users and the multicast users. It is assumed that the type of a user can be changed but, at a particular time, the user merely becomes a unicast user or a multicast user. The multicast users who have to receive the same traffic form a "multicast group." We assume that a multicast user can belong to only one multicast group. A multicast group can be served by one or more APs. We also assume that the APs transmit the multicast traffic with a fixed rate.[2] The APs collect the statistics regarding the traffic arrival rates of users and report them to the centralized controller. As a result, the statistics regarding the traffic arrival rate of each user is assumed to be known to the centralized controller which performs the proposed UA scheme. The unicast user i may have downlink and uplink traffic, of which the arrival rates are denoted by λ_i^d and λ_i^u, respectively. On the other hand, the multicast users have only downlink traffic. The downlink traffic arrival rate of the users in multicast group k is denoted by β_k.

Each user is associated with only a single AP at a time.[3] The association of user i with AP n is represented by a binary variable $x_{i,n}$: $x_{i,n} = 1$ if the user i is associated with the AP n; $x_{i,n} = 0$, otherwise. On the other hand, $y_{k,n}$ represents the association of multicast group k with AP n: $y_{k,n} = 1$ if at least one user of the multicast group k

[2] In IEEE 802.11 based WLANs, although there are a lot of transmission techniques for multicast traffic, the commercial APs being widely available in the market select the lowest supported transmission rate for multicast transmission. This simple technique can guarantee the robustness of multicast transmission with less overhead.

[3] In the enterprise WLANs, the typical users have tendency to stay at the same location for a relatively long time period before moving to another location. Thus, the association of a user with an AP is usually held for a while.

is associated with the AP n; $y_{k,n}$ = 0, otherwise. Although the user i belongs to the multicast group k and $y_{k,n} = 1$, it is noted that $x_{i,n}$ may be 0 since the multicast group k may be served by several APs. The APs with no associated user are put into sleep.

It is assumed that all APs are equipped with N transmit antennas and a single receive antenna, whereas the users are equipped with a single transmit and receive antenna. With multiple transmit antennas, the transmission scheme which exploits the transmit diversity such as space time block coding (STBC) [49] can be better, in improving the reliability of multicast transmission, than MU-MIMO. Accordingly, we apply the MU-MIMO transmission only to serve the downlink traffic of unicast users. The unicast users of which downlink traffic is concurrently transmitted by an AP are regarded as belonging to the same "MU-MIMO group." Since all transmissions of downlink unicast traffic are done using MU-MIMO, it is noted that there can exist the MU-MIMO groups with only a single unicast user, which corresponds to non MU-MIMO transmission.

Let $\mathbb{M}$ be a set of all feasible MU-MIMO groups. An MU-MIMO group should have one or more unicast users. Since each AP is equipped with N transmit antennas, the number of unicast users in an MU-MIMO group cannot exceed N. Thus, $\mathbb{M}$ is composed of the subsets of $\mathcal{U}$ which include at least one and at maximum N unicast users. For instance, if $N = 3$ and $\mathcal{U} = \{1,2,3,4\}$, then $\mathbb{M} = \Big\{\{1\},\{2\},\{3\},\{4\},\{1,2\},\{1,3\},\{1,4\},\{2,3\},\{2,4\},\{3,4\},\{1,2,3\},\{1,2,4\},\{1,3,4\},\{2,3,4\}\Big\}$. Let $\mathcal{M}_j$ represent the MU-MIMO group j. When indexing the MU-MIMO groups according to their respective positions in the set $\mathbb{M}$, with the above example, $\mathcal{M}_1 = \{1\}, \mathcal{M}_2 = \{2\}, \cdots, \mathcal{M}_{14} = \{2,3,4\}$. Let $\omega_{i,j}$ be a binary variable representing that the unicast user i belongs to the MU-MIMO group j. That is, if $i \in \mathcal{M}_j$, $\omega_{i,j} = 1$; otherwise, $\omega_{i,j} = 0$.

Let $z_{j,n}$ be a binary variable indicating whether the MU-MIMO group j is associated with the AP n. That is, when the MU-MIMO group j is associated with the AP n, $z_{j,n} = 1$; otherwise, $z_{j,n} = 0$. If $z_{j,n} = 1$ and $\omega_{i,j} = 1$, it is obvious that $x_{i,n} = 1$. Also,

since any user is served by only one AP, the selected MU-MIMO groups are disjoint with each other and their union set should include all unicast users.

3.1.2 Downlink Achievable Rate for MU-MIMO Groups

The lower the channel correlation between the users of the same MU-MIMO group is, the higher sum-rate is achieved by the corresponding group [32], [50]. Thus, when designing a UA algorithm, it is very necessary to assess the expected achievable rates of feasible MU-MIMO groups at each AP [32]. Now, we estimate the downlink achievable sum-rate of the MU-MIMO group j at AP n.

When an AP transmits the traffic concurrently to multiple users of an MU-MIMO group, a practical signal processing technique such as zero-forcing beamforming (ZFBF) [50] may be used to realize this concurrent transmission. The ZFBF uses the channel state information (CSI) to compute the beamforming vector for the transmission of a given MU-MIMO group. The beamforming vector depends on the channel correlation between the users of an MU-MIMO group and, as a result, it affects per-user achievable rate. Thus, not only the real traffic transmission of an AP to each its MU-MIMO group but also the sum-rate estimation of any feasible MU-MIMO group at each AP for UA can be based on the CSIs of the users. However, for any feasible MU-MIMO group in UA process, estimating its sum-rate by deriving the channel correlation among the users with the reported CSIs can incur excessive computational overhead and reporting overhead. Furthermore, since CSIs are time-varying, it may be inappropriate to estimate the sum-rate of MU-MIMO group by using merely instantaneous CSIs in the UA process conducted on a long-term basis.

In the proposed UA algorithm, for low complexity, the sum-rate of any feasible MU-MIMO group is estimated based on the beacon signal strength received at each user device, like in [51]. To this end, the user reports periodically this received beacon signal strength to the controller. We will describe this reporting process in Subsection

IV-A, in more detail. Let $\bar{\zeta}_{i,n}$ be the average of the beacon signal strengths that the unicast user i has received from AP n, during the interval of UA process. We assume that each AP transmits its beacon signal through one omni-directional antenna with much shorter period than the UA interval. According to [51], when the user i belongs to the MU-MIMO group j, the average channel quality for estimating the transmission rate of AP n for the unicast user i of MU-MIMO group j, denoted by $\alpha_{i,j,n}$, is assessed as follows.

$$\alpha_{i,j,n} = \frac{N - |\mathcal{M}_j| + 1}{|\mathcal{M}_j|} \times \frac{\bar{\zeta}_{i,n}}{N}, \tag{3.1}$$

where $|\mathcal{M}_j|$ is the cardinality of the set $\mathcal{M}_j$, i.e., the number of unicast users belonging to the MU-MIMO group j.

The average downlink achievable rate of the unicast user i belonging to the MU-MIMO group j at AP n, denoted by $c^{\mathrm{d}}_{i,j,n}$, can be determined by using $\alpha_{i,j,n}$. For example, there may be a mapping table of the received signal strength (RSS) into a transmission rate for the corresponding AP (see Table II). Then, $c^{\mathrm{d}}_{i,j,n}$ can be determined by using $\alpha_{i,j,n}$ as the RSS of the user i belonging to the MU-MIMO j at AP n.

Next, let us estimate the average channel time that the AP n takes to serve the downlink traffic of MU-MIMO group j, denoted by $\nu^{\mathrm{d}}_n(\mathcal{M}_j)$. With the downlink traffic arrival rate λ^{d}_i of the unicast user i,

$$\nu^{\mathrm{d}}_n(\mathcal{M}_j) = \max_{i \in \mathcal{M}_j} \left(\lambda^{\mathrm{d}}_i / c^{\mathrm{d}}_{i,j,n} \right). \tag{3.2}$$

3.1.3 Candidate MU-MIMO Groups

It is noted that $|\mathbb{M}| = \sum_{k=1}^{N} \binom{|\mathcal{U}|}{k}$. Since a typical enterprise WLAN is expected to serve many users, the set $\mathbb{M}$ of all feasible MU-MIMO groups for the entire ESS may have a very big size. It is apparent that the computational complexity of UA depends on the number of all feasible MU-MIMO groups which should be considered in the UA algorithm. Accordingly, we need to reduce the number of candidate MU-MIMO groups

for mitigating the complexity of UA process.

Algorithm 3 Generating Candidate MU-MIMO Groups.

$\mathbb{G}_n \leftarrow \emptyset,\ \mathcal{J}_n \leftarrow \emptyset,\ \forall n \in \mathcal{A}$
$\mathcal{Y} \leftarrow \emptyset$
for $j = 1$ to $|\mathbb{M}|$ **do**
  for all $n \in \mathcal{A}$ **do**
    if $\Big(v_n^{\mathrm{d}}(\mathcal{M}_j) \le v_n^{\mathrm{d}}(\mathcal{M}_j \backslash \{i\}) + v_n^{\mathrm{d}}(\{i\})$
      && $c_{i,j,n}^{\mathrm{d}} \ge \phi_{\min} \Big),\ \forall i \in \mathcal{M}_j$ **then**
      $\mathbb{G}_n \leftarrow \mathbb{G}_n \cup \{\mathcal{M}_j\}$
      $\mathcal{J}_n \leftarrow \mathcal{J}_n \cup \{j\}$
      $\mathcal{Y} \leftarrow \mathcal{Y} \cup \mathcal{M}_j$
    end if
  end for
  $\mathcal{Z} \leftarrow \mathcal{U} \backslash \mathcal{Y}$
end for
if $\mathcal{Z} \neq \emptyset$ **then**
  for all $i \in \mathcal{Z}$ **do**
    $\mathcal{X}_i \leftarrow \{\ j \mid i \in \mathcal{M}_j, 1 \le j \le |\mathbb{M}|\}$
    $(\hat{j}, \hat{n}) \leftarrow \underset{\forall j \in \mathcal{X}_i, \forall n \in \mathcal{A}}{\operatorname{argmin}} \Big(v_n^{\mathrm{d}}(\mathcal{M}_j) - v_n^{\mathrm{d}}(\mathcal{M}_j \backslash \{i\}) \Big)$
    $\mathbb{G}_{\hat{n}} \leftarrow \mathbb{G}_{\hat{n}} \cup \{\mathcal{M}_{\hat{j}}\}$
    $\mathcal{J}_{\hat{n}} \leftarrow \mathcal{J}_{\hat{n}} \cup \{\hat{j}\}$
  end for
end if
return $\mathbb{G}_n,\ \mathcal{J}_n,\ \forall n \in \mathcal{A}$

Note that a unicast user out of the service range of a specific AP is unlikely to belong to any MU-MIMO group associated with the AP. Thus, it is more effective to manage the set of feasible MU-MIMO groups for each AP, not for the entire ESS. We check the following two simple conditions, in constituting the set of all possible MU-MIMO groups of any AP n, $\mathbb{G}_n$. One is that all members of MU-MIMO group j should be able to achieve at least the minimum rate $\phi_{\min}$ at AP n, i.e., $c_{i,j,n}^{\mathrm{d}} \ge \phi_{\min}$ for all $i \in \mathcal{M}_j$. The other is that, for each user i in $\mathcal{M}_j$, the AP should be able to

get the benefit (i.e., shorter channel time) when serving the user i as a member of the MU-MIMO group j, rather than serving it separately from the group j. That is, for all $i \in \mathcal{M}_j$, $v_n^{\mathrm{d}}(\mathcal{M}_j) \leq v_n^{\mathrm{d}}(\mathcal{M}_j \backslash \{i\}) + v_n^{\mathrm{d}}(\{i\})$. When $\mathcal{M}_j = \{i\}$, it is obvious that $v_n^{\mathrm{d}}(\mathcal{M}_j \backslash \{i\}) = v_n^{\mathrm{d}}(\emptyset) = 0$. Constituting the feasible MU-MIMO groups for each AP by checking these two conditions is summarized in lines 3 – 11 of Algorithm 1, where $\mathcal{J}_n$ denotes the index sets of all possible MU-MIMO groups for AP n.

On the other hand, when a unicast user has downlink transmission rate lower than $\phi_{\min}$ from all APs, it does not belong to any feasible MU-MIMO group after the grouping of lines 3 – 11. The set $\mathcal{Z}$ (line 12) contains such users. For each user i in $\mathcal{Z}$, among any MU-MIMO groups which has the user i as its member, the AP n and the group j having the minimum difference between $v_n^{\mathrm{d}}(\mathcal{M}_j)$ and $v_n^{\mathrm{d}}(\mathcal{M}_j \backslash \{i\})$ are selected. Then, the MU-MIMO group j is newly inserted to the set of feasible MU-MIMO groups of AP n (lines 16–19).

3.2 User Association Problem

In this section, we formulate our UA problem into a multi objective optimization problem.

3.2.1 Factors of UA Objective

The goal of our UA process is to balance the load among working APs while supporting as high transmission rate as possible to each user group (MU-MIMO or multicast) with as few working APs as possible. Accordingly, the UA objective is composed of three factors: the number of working APs, the average achievable rate per user group supported by AP, and the load of AP.

Number of Working APs

Let a_n be a binary indicator representing that the AP n is a working AP (WAP), i.e., the existence of user(s) associated with the AP n. Since the number of users associated with the AP n is $\sum_{i\in\mathcal{S}} x_{i,n}$,

$$\mathrm{a}_n = \min\Big(\sum\nolimits_{i\in\mathcal{S}} x_{i,n},\ 1\Big). \tag{3.3}$$

Then, the number of WAPs within ESS is $\sum_{n\in\mathcal{A}} \mathrm{a}_n$.

Per-Group Downlink Achievable Transmission Rate

Note that the association of user groups (MU-MIMO and multicast) with AP is the key point of UA process. It is actually desirable for an user group to be associated with the AP that can provide the highest transmission rate to itself. Thus, we take account of the achievable transmission rate that AP can provide to each user group, when being associated with itself. Since the MU-MIMO and the multicast transmission are applied only to downlink and furthermore the downlink traffic is more dominant than the uplink traffic in a typical WLAN, we consider only the downlink transmission rate of AP.

Let $\Upsilon_{j,n}$ $(j \in \mathcal{J}_n)$ denote the aggregate downlink rate of AP n for the MU-MIMO group j, i.e., $\Upsilon_{j,n} = \sum_{i\in\mathcal{M}_j} c^{\mathrm{d}}_{i,j,n}$, and let $r^{\mathrm{mcst}}_{k,n}$ be the transmission rate of AP n for the multicast group k. Then, the average (i.e., the arithmetic mean) of the downlink transmission rates of AP n for its associated MU-MIMO and multicast groups is

$$\bar{\mathfrak{R}}^{\mathrm{d}}_n = \Big(\sum_{j\in\mathcal{J}_n} z_{j,n}\Upsilon_{j,n} + \sum_{k\in\mathcal{K}} y_{k,n} r^{\mathrm{mcst}}_{k,n}\Big) \Big/ \Big(\sum_{j\in\mathcal{J}_n} z_{j,n} + \sum_{k\in\mathcal{K}} y_{k,n}\Big). \tag{3.4}$$

On the other hand, we assume that the AP n may share the channel with any neighboring APs within its carrier sensing range. Since the APs sharing a channel can use the channel only during some time fraction, the channel sharing has the effect of lower-

ing the transmission rates of APs. Thus, in estimating the downlink achievable rate per user group at AP n under channel sharing, we should consider the fraction of channel time utilized by AP n.

When several APs share the same channel, the fraction of channel time for an AP may be influenced by various components such as the number of APs sharing the channel, the number of associated users within the APs, and the load of each AP, etc. Among these, the number of APs sharing the channel is likely to be a main factor having a great effect on the fraction of channel time that an AP gets. Furthermore, the number of APs sharing the channel is known already before the UA work because the channel allocation for all APs is given, whereas the number of associated users within each AP is the result of UA work and the load of each AP is a real-time parameter varied during operation. Although there may be various ways to estimate the fraction of channel time for an AP,[4] for the sake of simplicity and trackable solution, we consider only the number of APs using the same channel. The more APs share the channel, the less fraction of channel time an AP gets. Thus, we estimate the fraction of channel time at AP n to be $1/\left(\sum_{l\in\mathcal{A}} \mathrm{a}_l \vartheta_{l,n}\right)$, where $\sum_{l\in\mathcal{A}} \mathrm{a}_l \vartheta_{l,n}$ is the total number of APs within carrier sensing range of AP n which use the same channel including AP n itself. Note that this estimation is merely used for designing the UA scheme, not assessing the fraction of channel time that an AP really gets during operation.

When Γ_n is per-group downlink achievable rate of AP n under channel sharing, we estimate Γ_n as

$$\Gamma_n = \frac{\bar{\mathfrak{R}}_n^{\mathrm{d}}}{\sum_{l\in\mathcal{A}} \mathrm{a}_l \ \vartheta_{l,n}}. \tag{3.5}$$

[4]For example, the load of AP which is estimated in a history-based manner may be exploited to estimate the channel time faction of AP.

Load of AP

We define the load of an AP as the channel utilization similarly to [19], which is the portion of channel busy time for serving all the associated users. Since the APs have different transmission capability according to their 802.11 specifications and the traffic arrival rate is also different for each user, the load of AP is likely to be expressed appropriately by the channel utilization, not the traffic volume. Let $c_{i,n}^{\mathrm{u}}$ be the uplink transmission rate of unicast user i associated with AP n. Remind that the traffic arrival rate of multicast group k and the uplink traffic arrival rate of unicast user i are denoted by β_k and λ_i^{u}, respectively. When μ_n is the portion of channel busy time by the AP n,

$$\mu_n = \sum_{j \in \mathcal{J}_n} z_{j,n} \left(v_n^{\mathrm{d}}(\mathcal{M}_j) + \lambda_i^{\mathrm{u}} / c_{i,n}^{\mathrm{u}} \right) + \sum_{k \in \mathcal{K}} y_{k,n} \, \beta_k / r_{k,n}^{\mathrm{mcst}}. \tag{3.6}$$

We should consider the channel sharing in assessing the load of AP, since the APs can utilize not the whole channel time but some portion of channel time. Thus, we estimate the load of AP n, ρ_n, by the total utilization for the corresponding channel, which is the sum of channel utilizations of all APs sharing the same channel with the AP n.

$$\rho_n = \sum_{l \in \mathcal{A}} \vartheta_{l,n} \, \mu_l. \tag{3.7}$$

Table 3.1 summarizes the notations for convenience of readers.

Table 3.1: Notations

Symbol	Description
$\mathcal{A}$	The set of APs in the ESS
N	The number of transmit antennas of an AP
$\mathcal{S}$	The set of all users
$\mathcal{U}$	The set of unicast users
$\mathcal{K}$	The set of multicast groups
$\mathcal{G}_k$	The set of multicast users in the multicast group k

$\mathbb{G}_n$	The set of all possible MU-MIMO groups of AP n
$\mathcal{J}_n$	The index sets of all possible MU-MIMO groups of AP n
$\mathcal{M}_j$	The set of all users in the MU-MIMO group j
λ_i^{u}	Uplink traffic arrival rate for the unicast user i
λ_i^{d}	Downlink traffic arrival rate for the unicast user i
β_k	Traffic arrival rate for the multicast group k
$\omega_{i,j}$	A binary variable indicating whether the unicast user i belongs to the MU-MIMO group j
a_n	A binary variable indicating whether there are one or more users associated with AP n
$x_{i,n}$	A binary variable representing the association of user i with AP n
$y_{k,n}$	A binary variable representing the association of multicast group k with AP n
$z_{j,n}$	A binary variable representing the association of MU-MIMO group j with AP n
$\vartheta_{l,n}$	A binary variable indicating whether the AP l and AP n within the carrier sensing range of each other share the same channel; $\vartheta_{n,n} = 1$ for $n \in \mathcal{A}$
$c_{i,n}^{\mathrm{u}}$	Uplink physical transmission rate of the unicast user i associated with AP n
$c_{i,j,n}^{\mathrm{d}}$	Downlink physical transmission rate of AP n for the unicast user i belonging to the MU-MIMO group j
$r_{k,n}^{\mathrm{mcst}}$	Traffic transmission rate of AP n to the multicast group k
$c_{i,n}^{\mathrm{mcst}}$	Downlink physical transmission rate of AP n for the multicast user i
$v_n^{\mathrm{d}}(\mathcal{M}_j)$	The downlink service time of MU-MIMO group j by AP n
Γ_n	Per-group downlink achievable rate of AP n
ρ_n	The load of AP n
$R_n^{\max}$	The maximum physical transmission rate of AP n

3.2.2 Problem Formulation

We formulate the UA problem as multi-objective optimization. The multi-objective optimization is multi criteria decision making which involves more than one objective functions. First, let us discuss the objective functions. Recall that the objectives of the proposed UA scheme are to maximize the sum of achievable transmission rate, to balance the load among the working APs, and to minimize the energy consumption by minimizing the number of working APs, we have three objective functions. The first function is related to transmission rate, i.e., $f_1 = \sum_{n\in\mathcal{A}} \frac{\Gamma_n}{R_n^{\max}}$. The second function addresses load balancing, i.e., $f_2 = \sum_{n\in\mathcal{A}} \frac{\rho_n}{\max_{l\in\mathcal{A}} \rho_l}$. Finally, the last function represents AP on/off, i.e., $f_3 = \sum_{n\in\mathcal{A}} \mathrm{a}_n$. To solve the multi-objective optimization problem, scalarizing the objective functions is a priori method [52]. The linear scalarization form of the above multi-objective functions is given in (3.8). The corresponding constraints for the formulated UA problem are given in (3.9)–(3.16)

$$\max \quad w_1 \sum_{n\in\mathcal{A}} \frac{\Gamma_n}{R_n^{\max}} + w_2 \sum_{n\in\mathcal{A}} \frac{\rho_n}{\max_{l\in\mathcal{A}} \rho_l} - w_3 \sum_{n\in\mathcal{A}} \mathrm{a}_n \tag{3.8}$$

$$s.t. \quad x_{i,n} = \sum_{j\in\mathcal{J}_n} \omega_{i,j} z_{j,n} \qquad \forall i \in \mathcal{U},\ \forall n \in \mathcal{A} \tag{3.9}$$

$$\sum_{n\in\mathcal{A}} x_{i,n} = 1, \qquad \forall i \in \mathcal{S} \tag{3.10}$$

$$\sum_{n\in\mathcal{A}} x_{i,n} c_{i,n}^{\mathrm{mcst}} \geq r_{k,n}^{\mathrm{mcst}}, \qquad \forall i \in \mathcal{G}_k,\ \forall k \in \mathcal{K} \tag{3.11}$$

$$y_{k,n} = \min\left(\sum_{i\in\mathcal{G}_k} x_{i,n}, 1\right), \qquad \forall k \in \mathcal{K},\ \forall n \in \mathcal{A} \tag{3.12}$$

$$\mathrm{a}_n = \min\left(\sum_{i\in\mathcal{S}} x_{i,n},\ 1\right), \qquad \forall n \in \mathcal{A} \tag{3.13}$$

$$x_{i,n} \in \{0,1\}, \qquad \forall i \in \mathcal{S},\ \forall n \in \mathcal{A} \tag{3.14}$$

$$z_{j,n} \in \{0,1\}, \qquad \forall j \in \mathcal{J}_n,\ \forall n \in \mathcal{A} \tag{3.15}$$

$$w_1 + w_2 + w_3 = 1, \tag{3.16}$$

where $R_n^{\max}$ is the maximum transmission rate of AP n, which is merely adopted as a normalization factor for constituting a multi-objective optimization function.

The objective function in (3.8) consists of three terms. In the first term, we intend to enhance the performance of user groups from the viewpoint of not a single AP but the entire ESS system, by maximizing the sum of per-group downlink achievable rates of WAPs. This sum rate maximization can be achieved in two ways. One is to select the AP which can offer the best downlink achievable rate to each user group. The other is to minimize the number of APs to serve multicast users which belong to the same group in order to reduce the performance anomaly effect.[5] The second term in (3.8) balances the loads of APs while considering the channel sharing. When all WAPs have the same load, i.e., the system load is completely balanced among WAPs, the second term is maximized. Finally, the last term in (3.8) is for improving the energy efficiency by operating the ESS system with as few APs as possible. The weights (w_1, w_2, w_3) are assigned to three terms of the objective function, respectively. The value of weight specifies the priority of its corresponding term.

The constraint (4.17) implies that a unicast user is associated with the AP which its MU-MIMO group is associated with. The constraint (4.18) expresses that a user can be associated with only one AP at a particular time. The constraint (4.19) requires that the physical downlink transmission rate of a multicast user should be higher than the multicast transmission rate used by the AP. Note that most of commercial APs use a fixed transmission rate to serve the multicast traffic. A robust multicast transmission can be realized with the constraint (4.19). The constraint (4.20) implies that the AP should serve the traffic of the group which the multicast user associated with itself belongs to. The constraints (3.13) – (3.16) are likely to be self-evident.

[5] In viewpoint of the entire ESS, more channel resource for serving a multicast group can be required when each user of the group is associated with its best AP, than when the multicast group is associated with as few APs as possible although some users of the group are not served by their respective best APs. Such anomalistic phenomenon is called “performance anomaly.”

3.3 User Association Scheme

3.3.1 Equivalent Linear Problem

The multi-objective optimization problem in (3.8) is a mixed integer non-linear fractional problem which is known to be NP-hard. To get the optimal solution, we convert the original problem into an equivalent solvable linear problem, through the following transformation process.

The objective function (3.8) of the original problem has a fractional form. It is well known that the Dinkelbach's algorithm in [46] allows a fractional problem to be solved as a sequence of parameterized non-fractional problems. Accordingly, as the first step for transforming the original problem into a solvable equivalent problem, we convert a fractional objective function into its corresponding non-fractional function, by using parametric technique. The fractional form appears on the first and second terms of the objective function. Thus, we introduce the parameters f and h for these terms, respectively. Following the conversion in [47], we have the non-fractional form of objective function as (3.17).

$$\begin{aligned} \max \quad & \sum_{n\in\mathcal{A}} \left(w_1 \Lambda_n - f_n \sum_{l\in\mathcal{A}} \vartheta_{l,n} \xi_l \right) + \sum_{n\in\mathcal{A}} \left(w_2 \rho_n - h_n \max_{l\in\mathcal{A}} \rho_l \right) \\ & - w_3 \sum_{n\in\mathcal{A}} \mathrm{a}_n, \end{aligned} \tag{3.17}$$

where $\Lambda_n = \left(\sum_{j\in\mathcal{J}_n} z_{j,n} \frac{\Upsilon_{j,n}}{R_n^{\max}} + \sum_{k\in\mathcal{K}} y_{k,n} \frac{r_{k,n}^{\mathrm{mcst}}}{R_n^{\max}} \right)$ and $\xi_l = \mathrm{a}_l \times \left(\sum_{j\in\mathcal{J}_l} z_{j,l} + \sum_{k\in\mathcal{K}} y_{k,l} \right)$.

The second step is to linearize a non-linear form, either on the transformed non-fractional objective function or on the constraints. Let $\varphi := \max_{l\in\mathcal{A}} \rho_l$. Then, the complete equivalent solvable problem is shown in (3.18)–(3.35).

$$\begin{aligned} \max \quad & \sum_{n\in\mathcal{A}} \left(w_1 \Lambda_n - f_n \sum_{l\in\mathcal{A}} \vartheta_{l,n} \xi_l \right) \\ & + \sum_{n\in\mathcal{A}} \left(w_2 \rho_n - h_n \varphi \right) - w_3 \sum_{n\in\mathcal{A}} \mathrm{a}_n, \end{aligned} \tag{3.18}$$

$$s.t.\ x_{i,n} = \sum_{j \in \mathcal{J}_n} \omega_{i,j}\, z_{j,n} \qquad \forall i \in \mathcal{U},\ \forall n \in \mathcal{A} \tag{3.19}$$

$$\sum_{n \in \mathcal{A}} x_{i,n} = 1, \qquad \forall i \in \mathcal{S} \tag{3.20}$$

$$\sum_{n \in \mathcal{A}} x_{i,n} c^{\text{mcst}}_{i,n} \geq r^{\text{mcst}}_{k,n}, \qquad \forall k \in \mathcal{K},\ \forall i \in \mathcal{G}_k \tag{3.21}$$

$$\varphi \geq \rho_l, \qquad \forall l \in \mathcal{A} \tag{3.22}$$

$$y_{k,n} \leq \sum_{i \in \mathcal{G}_k} x_{i,n}, \qquad \forall k \in \mathcal{K},\ \forall n \in \mathcal{A} \tag{3.23}$$

$$|\mathcal{G}_k|\, y_{k,n} \geq \sum_{i \in \mathcal{G}_k} x_{i,n}, \qquad \forall k \in \mathcal{K},\ \forall n \in \mathcal{A} \tag{3.24}$$

$$\mathrm{a}_n \leq \sum_{i \in \mathcal{S}} x_{i,n}, \qquad \forall n \in \mathcal{A} \tag{3.25}$$

$$|\mathcal{S}|\, \mathrm{a}_n \geq \sum_{i \in \mathcal{S}} x_{i,n}, \qquad \forall n \in \mathcal{A} \tag{3.26}$$

$$\xi_l \leq |\mathcal{S}|\, \mathrm{a}_l, \qquad \forall l \in \mathcal{A} \tag{3.27}$$

$$\xi_l \leq \sum_{j \in \mathcal{J}_l} z_{j,l} + \sum_{k \in \mathcal{K}} y_{k,l}, \qquad \forall l \in \mathcal{A} \tag{3.28}$$

$$\xi_l \geq \Big(\sum_{j \in \mathcal{J}_l} z_{j,l} + \sum_{k \in \mathcal{K}} y_{k,l} \Big) - (1 - \mathrm{a}_l)|\mathcal{S}|, \qquad \forall l \in \mathcal{A} \tag{3.29}$$

$$\xi_l \geq 0, \qquad \forall l \in \mathcal{A} \tag{3.30}$$

$$x_{i,n} \in \{0,1\}, \qquad \forall i \in \mathcal{S},\ \forall n \in \mathcal{A} \tag{3.31}$$

$$y_{k,n} \in \{0,1\}, \qquad \forall k \in \mathcal{K},\ \forall n \in \mathcal{A} \tag{3.32}$$

$$z_{j,n} \in \{0,1\}, \qquad \forall j \in \mathcal{J}_n,\ \forall n \in \mathcal{A} \tag{3.33}$$

$$\mathrm{a}_n \in \{0,1\}, \qquad \forall n \in \mathcal{A} \tag{3.34}$$

$$w_1 + w_2 + w_3 = 1. \tag{3.35}$$

We add the constraint (3.22) to represent that $\varphi = \max_{l \in \mathcal{A}} \rho_l$. In addition, to ensure $y_{k,n} = \min\left(\sum_{i \in \mathcal{G}_k} x_{i,n}, 1\right)$, we add the constraints (3.23), (3.24), and (3.32). Similarly, the constraints (3.25), (3.26), and (3.34) enforce $\mathrm{a}_n = \min\left(\sum_{i \in \mathcal{S}} x_{i,n}, 1\right)$. Finally, we

add the constraints (3.27) – (3.30) for guaranteeing $\xi_l = \mathrm{a}_l \left(\sum_{j \in \mathcal{J}_l} z_{j,l} + \sum_{k \in \mathcal{K}} y_{k,l} \right)$.

At this point, we have a non-fractional and linear form of optimization problem. When the values of f_n and h_n for all n are given, our transformed problem turns into a mixed-integer linear programming (MILP) problem. It is well known that such a problem can be optimally solved using branch and bound technique. There are plenty of solvers widely available to solve such MILP problems, for example, CPLEX [44] and lp_solve [45].

3.3.2 Solution Algorithm

To solve the problem in (3.18) – (3.35) by using an MILP solver, we should specify the values of parameters f_n and h_n for all $n \in \mathcal{A}$. Furthermore, the optimal solution of original problem can be obtained when solving the problem (3.18) with the optimal values of these parameters. As stated before, the Dinkelbach's algorithm in [46] allows a fractional problem to be solved as a sequence of parameterized non-fractional problems, where the values of parameters are updated iteratively and finally converge to their respective optimal values. The UA algorithm which solves the problem (3.18) – (3.35) based on the Dinkelbach method is presented in Algorithm 4.

In the algorithm, the values of parameters f_n and h_n for all $n \in \mathcal{A}$ are initialized to 0. The MILP solver calculates the solution of the problem (3.18) – (3.35), with given values of f_n and h_n. The algorithm repeats the process that updates the values of all f_n and h_n and re-solves the problem by using the MILP solver with the updated values, until the convergence condition of parameter values is satisfied. That is, the iteration is stopped when the difference between the new and old values of f_n and the difference between the new and old values of h_n for all $n \in \mathcal{A}$ are smaller than a predefined threshold ε.

Algorithm 4 User Association Algorithm.

$f_n \leftarrow 0,\ h_n \leftarrow 0 \quad$ for all $n \in \mathcal{A}$
$f_n^{\text{old}} \leftarrow 1,\ h_n^{\text{old}} \leftarrow 1 \quad$ for all $n \in \mathcal{A}$
$\varepsilon \leftarrow 10^{-5},$ stop $\leftarrow false$
while stop = $false$ **do**
    Solve (3.18) using MILP solver, given f_n and $h_n, \forall n \in \mathcal{A}$.
    stop $\leftarrow true$
    for all $n \in \mathcal{A}$ **do**
        if $|f_n - f_n^{\text{old}}| > \varepsilon$ **then**
            $f_n^{\text{old}} \leftarrow f_n, \quad f_n \leftarrow \frac{w_1 \Lambda_n}{\sum_{l \in \mathcal{A}} \vartheta_{l,n} \xi_l}$
            stop $\leftarrow false$
        end if
        if $|h_n - h_n^{\text{old}}| > \varepsilon$ **then**
            $h_n^{\text{old}} \leftarrow h_n, \quad h_n \leftarrow \frac{w_2 \rho_n}{\mu}$
            stop $\leftarrow false$
        end if
    end for
end while
return the association results $(x_{i,n}, y_{k,n}, z_{j,n})$

3.3.3 Computational Complexity (Execution Time)

In order to assess the computational complexity of the proposed UA algorithm under enterprise WLAN environment where the APs may be densely deployed, we test the convergence speed of the algorithm according to the number of APs (i.e. 50, 100, and 150). The number of unicast users is set to four times the number of APs and the number of multicast users is fixed to 100. The RSS value and the traffic arrival rate of each user are randomly set. Fig. 3.1 shows that the Dinkelbach's algorithm converges after a maximum of five iterations. The UA objective function in y-axis refers to objective function in (3.8). When taking account of such quick convergence of the Dinkelbach's algorithm, we can expect that the computational complexity of the proposed UA algorithm will mainly depend on the execution time of MILP problem using CPLEX. According to [53], in solving very large-scale linear problems (including MILP) which

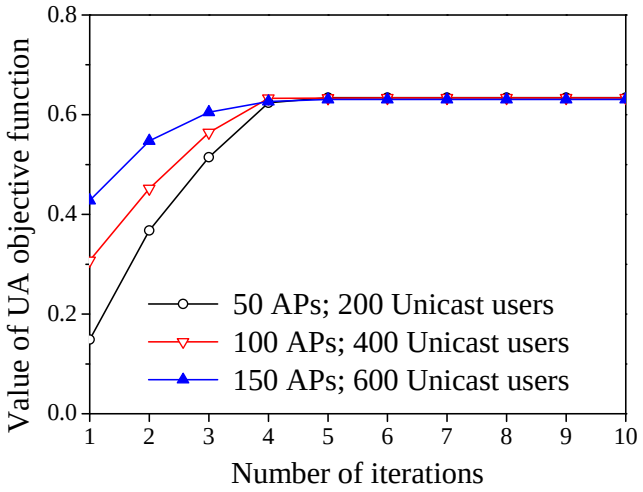


Figure 3.1: Convergence of the proposed UA Algorithm.

contain about two millions of constraints and variables, the average execution time of CPLEX is merely about two seconds. Since the formulated MILP problem in (3.18) – (3.35) has much fewer variables and constraints, the execution time of our algorithm is expected to be much shorter.[6] Considering that the UA algorithm is activated once every several minutes, the computational complexity is reasonably acceptable and this confirms the feasibility and practicality of the proposed UA scheme.

3.4 Implementation

Samsung Galaxy Note 2014 Edition is used as a user device. We have developed an agent on each user device to perform channel scanning and to execute handover command utilizing *Android API*. Meanwhile, TP-Link AP (TL-WDR4300) [54] which uses OpenWRT as its operating system [55] is utilized as a wireless AP. We also have de-

[6] According to our numerical evaluation using MATLAB, which is run on a desktop computer with Intel i7 processor and 16 GByte of RAM, the total execution time of the proposed UA algorithm to solve the problem with 150 APs is less than five seconds. This implies that the average execution time of CPLEX for solving the corresponding MILP problem is under one second. Note that the execution time can get shorter by using a more advanced implementation technique on a server with higher computing power.

veloped an agent on the AP to gather the statistics of users utilizing IPtraf [56] and to gather multicast group information utilizing TCPDump [57]. CPLEX [44] is used for the MILP solver. Note that, since our implementation is purely done in application software level, no lower level implementation and hardware changes are required. Thus, the proposed scheme can be easily implemented using the current commodity hardware.

A centralized controller performs the UA work according to Algorithm 2, with the period of T_{UA}. It is noted that the traffic arrival rates for all unicast users and multicast groups $\left(\lambda_i^{\mathrm{u}}, \lambda_i^{\mathrm{d}}, \beta_k, \ \forall i,k\right)$ and the physical transmission rates of WAPs for all users $\left(c_{i,j,n}^{\mathrm{d}}, c_{i,n}^{\mathrm{u}}, c_{i,n}^{\mathrm{mcst}}, \ \forall i,j,n\right)$ are used as the input parameters of Algorithm 2. The controller gets these input parameter values from WAPs and users, as follows. During the interval of length T_{arr}, each WAP monitors the uplink and downlink traffic rates of its each associated unicast user and the downlink traffic rate of its each associated multicast group, and reports them to the controller at the end of interval. On the other hand, a user measures the RSS of beacon signal from each surrounding WAP during the interval of T_{CQ}, and reports the arithmetic mean of the measured RSS values of each WAP to the controller via its serving WAP, with the period of T_{CQ}. The controller estimates the physical transmission rates of WAPs for users, based on the RSS values of WAPs reported by the users. Note that, by default, the WiFi-enabled user device performs background scanning to measure the RSS values of beacon signals from the surrounding APs. Whenever receiving the measured traffic arrival rates or channel quality information from WAPs or users, the centralized controller updates the value of the corresponding parameter by calculating its exponentially weighted moving averages.[7]

[7]For example, if the WAP n received the uplink traffic of B bytes from the unicast user i during the mth interval of length T_{arr} and reported B/T_{arr} as the instantaneous uplink data rate of user i, the centralized controller updates $c_{i,n}^{\mathrm{u}}(m)$ as $\delta B/T_{\mathrm{arr}} + (1-\delta)c_{i,n}^{\mathrm{u}}(m-1)$, where δ is called a moving average weight and $0 < \delta \leq 1$.

Even though the UA algorithm is executed with the period of T_{UA}, a user can join the network, leave the network, or perform handover to other AP at any time. A newly activated user and a handover user are associated with the WAP having the highest beacon RSS. As a result, the load of a certain WAP may exceed the predetermined overload threshold, Φ. In such case, even before the UA execution timer is expired, the centralized controller wakes-up the nearest sleeping AP of the overloaded WAP and requests that the users measure and report the beacon RSS value of the new WAP. Then, the centralized controller re-executes the UA algorithm. This work can be repeated several times until the loads of all WAPs are below Φ or there is no more AP which can be switched-on.

On the other hand, by switching some APs off, the coverage hole may be formed within the service area of ESS. The coverage hole can be removed by increasing the transmit power of the neighboring WAPs. Another option is to let a sleeping AP intermittently wake-up to check if there is a user being not served by any WAP within its service area. If it finds a new user, the AP is switched on for serving the user and reports its switched-on state to the centralized controller.

Chapter 4

A Graph-Based Handover Scheduling for Heterogenous Vehicular Networks

4.1 System Model

In this paper, we consider heterogeneous networks where WiFi systems coexist with cellular networks such as Long Term Evolution (LTE) [1], as depicted in Fig. 4.1. Since we are interested in efficiently supporting the Internet access of users in vehicles on roads, we concentrate on communication in the areas around roads. The WiFi APs or simply APs are randomly and densely deployed, which causes the service areas of many APs to overlap with each other. On the other hand, the cellular BSs or simply BSs are deployed under a certain cell planning strategy, so that the service area of the cellular network covers the whole road area. Unlike the cellular networks, the service area of the WLANs may not cover some part of road areas. The APs and BSs are connected to the Internet using wired links. In this paper, the AP and BS are referred to as a point of attachment (PoA). In addition, we consider that the whole area is divided

[1] The cellular networks may also include GSM and UMTS network but, for simplicity we only mention the LTE.

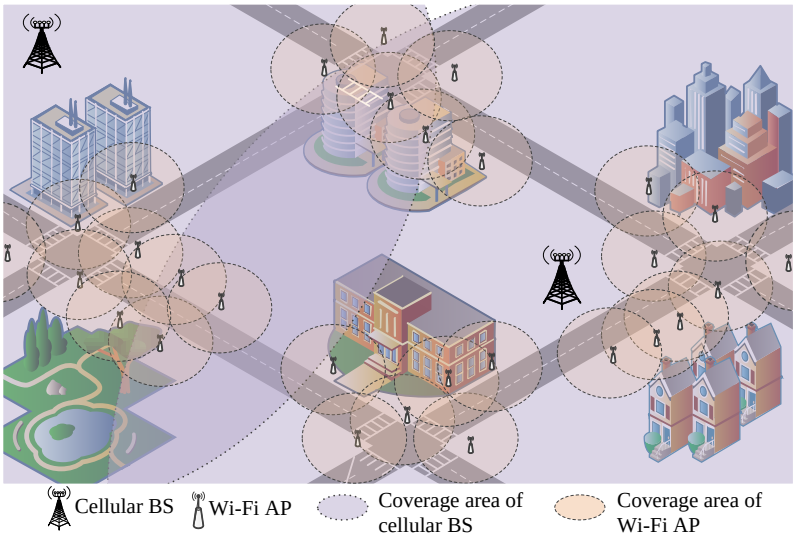


Figure 4.1: Heterogenous networks in vehicular environments.

into disjoint sub-areas. Therefore, from this point, we can focus on a particular sub-area.

Each sub-area is managed by a server, possibly in a similar manner to cloudlet system [62]. The server stores information regarding all PoAs within its sub-area as well as information of the roads belonging to the corresponding sub-area. The set of all PoAs within a sub-area is denoted by $\mathcal{P}$, which consists of the set of APs, $\mathcal{A}$, and the set of BSs, $\mathcal{B}$. Thus, $\mathcal{P} = \mathcal{A} \cup \mathcal{B}$. $\mathbf{p}_a$ and R_a respectively denote the location and service radius of PoA-a ($\in \mathcal{P}$). For a given coordinate point $\mathbf{p}_a$, its latitude and longitude components are denoted by $\mathbf{p}_{a.x}$ and $\mathbf{p}_{a.y}$, respectively.

On the other hand, the roads within a sub-area are divided into several road portions. A road portion is started from an intersection or u-turn area and ended in the next nearest intersection or u-turn area. Thus, a road portion contains no intersection and u-turn areas. The road portion-p is represented by a list of curve points, $\mathcal{L}_p$. When the road portion-p has n curve points and its ith curve points is denoted by $\mathbf{l}_{p_i}$, $\mathcal{L}_p = \left(\mathbf{l}_{p_1}, \mathbf{l}_{p_2}, ..., \mathbf{l}_{p_n}\right)$. Moreover, a straight line between two consecutive curve points

is referred to as a moving path. The north azimuth value formed by ith the moving path $(\mathbf{l}_{p_i}, \mathbf{l}_{p_{i+1}})$ of the road portion-p is denoted by θ_p^i.

A user device or simply a user, is equipped with both WiFi interface and cellular network interface, moves in vehicular speed along a road. The user can track its current position and moving direction using GPS and can connect to both WiFi and cellular network while moving on the road. However, to maximize the connection time on WiFi, the user prefers to transmit data through WiFi whenever possible. To maintain a network connection, the user needs to perform handover to either AP or BS while moving. Thus, horizontal and vertical HO may take place.[2] In the case of horizontal HO between WiFi APs, the HO can be further categorized into two types of L2 and L3 HO, depending on the subnet configuration of the APs, like in [11]. The handover between two APs with the same subnet address is referred to as the L2 HO, whereas the L3 HO means the handover between the APs with respective different subnet addresses. The HO latency from PoA-a to PoA-b is denoted by τ_{ab}. And, $\eta_a(t)$ denotes the number of users connected to PoA-a at time t.

4.2 Graph-Based Modeling

Our HO scheduling design is based on a directed graph representing the PoA deployment on a road, where each PoA is represented as a graph node and the adjacency of two PoAs becomes an edge. Let us examine this graph modeling.

4.2.1 Division of Road Portion into Road Segments

We first define a road segment as the part of a road portion satisfying the condition that, if the HO from AP-a to AP-b is possible while moving on the road segment, the HO from AP-b to AP-a reversely on the same road segment is not possible. This condition

[2] In the case of vertical HO, the user does not actually perform the HO operation. The user can only switch the currently active network interface for data transmission either from LTE to WiFi interface or vice versa.

is referred to as a *single direction HO*. With a single direction HO condition, a road segment can be modelled as a graph which has no loop between any two nodes (APs), i.e., the graph cannot have simultaneously the edge from vertex-a to vertex-b and the edge from vertex-b to vertex-a. This no-loop characteristic can lead to more simple but efficient HO scheduling on the road segment.

Let us explain with an example that the *single direction HO* condition may not be held within a road portion. Although a road portion does not contain any intersection or u-turn area (refer to the definition of a road portion in Section 4.1), it may contain a sharp turnout section which excessively changes the moving direction of a user. For instance, a user, which firstly moves toward South direction at the beginning of a road portion, may finally move toward North direction at the end of road portion due to the sharp turnout section. As the implication, the user, which performs handover from AP-a to AP-b at the beginning of road portion, may possibly perform handover from AP-b to AP-a at the end of road portion.

Now, we partition a road portion into one or more road segments, each of which satisfies the *single direction HO* condition. Let $\mathcal{M}_s$ be the list of curve points which construct the road segment-s. $\mathcal{M}_s = \left(\mathbf{m}_{s_1}, \mathbf{m}_{s_2}, ..., \mathbf{m}_{s_k}\right)$, where $\mathbf{m}_{s_1}$ and $\mathbf{m}_{s_k}$ are the first and last curve points, respectively. In addition, the north azimuth value formed by the moving path $(\mathbf{m}_{s_i}, \mathbf{m}_{s_{i+1}})$ of the road segment-s is denoted by θ_s^i. We assume that, since a road segment does not contain any sharp turn-out section, the difference between the north azimuth value of any two moving paths within the road segment should not be greater than a certain threshold value, Γ. This threshold is used to detect the sharp turn-out section within a road portion and to divide the road portion into disjoint road segments. The detail is presented in Algorithm 5, where EmptyList() is a function which generates an empty list and InSertList(L, e) is a function which inserts e as the last element of a list L.

We start by setting the first curve point of road portion as the first curve point of

Algorithm 5 Construction of Road Segments

1: A list of curve points $\mathcal{L}_p = \left(\mathbf{l}_{p_1}, \mathbf{l}_{p_2}, ..., \mathbf{l}_{p_n}\right)$
2: A list of north azimuth values $\left(\theta_p^1, \theta_p^2, ..., \theta_p^{n-1}\right)$
3: Threshold of north azimuth value difference, Γ
4: $\mathcal{M}_s \leftarrow$ EmptyList()
5: InsertList$\left(\mathcal{M}_s, \mathbf{l}_{p_1}\right)$, $\ell \leftarrow 1$, $C \leftarrow \{\}$
6: **for** $i = 2$ to $n-1$ **do**
7: **if** $|\theta_p^\ell - \theta_p^i| \leq \Gamma$ **then**
8: InsertList$(\mathcal{M}_s, \mathbf{l}_{p_i})$
9: **else**
10: $C \leftarrow C \cup \left\{ \text{InsertList}(\mathcal{M}_s, \mathbf{l}_{p_i}) \right\}$
11: $\mathcal{M}_s \leftarrow$ EmptyList()
12: InsertList$\left(\mathcal{M}_s, \mathbf{l}_{p_i}\right)$, $\ell \leftarrow i$
13: **end if**
14: **end for**
15: $C \leftarrow C \cup \left\{ \text{InsertList}(\mathcal{M}_s, \mathbf{l}_{p_n}) \right\}$
16: **return** C

the first road segment on the corresponding road portion (line 4, 5). Then, we continuously record all the other curve points which belong to the current road segment, while observing the north azimuth value difference to detect the next road segment. This is done until all curve points of the corresponding road portion are processed (lines 6 – 14). The algorithm returns a set of all road segments for given road portion.

Note that since the HO scheduling problem on each road segment is independent of each other, we initially treat the HO scheduling problem on a single road segment. Then, we will merge all pieces together in Section 4.4.

4.2.2 Relation between PoAs on a Road Segment

Let us consider a road segment-s with a list of curve points, $\mathcal{M}_s = \left(\mathbf{m}_{s_1}, \mathbf{m}_{s_2}, ..., \mathbf{m}_{s_k}\right)$. To formulate the HO scheduling problem, we need to discuss the relation between PoAs on road segment-s. We define the *initial PoAs* of a road segment as the PoAs whose service area overlaps with the first curve point. Also, the *final PoAs* of a road

segment is defined as the PoAs whose service area overlaps with the last curve point. When $\mathcal{S}_s$ and $\mathcal{E}_s$ respectively denote the set of *initial PoAs* and the set of *final PoAs* for the road segment-s,

$$\mathcal{S}_s = \bigcup_{\forall a \in \mathcal{P}} \Big\{ a \;\Big|\; \|\mathbf{p}_a - \mathbf{m}_{s_1}\| \leq R_a - \Delta \Big\}, \tag{4.1}$$

$$\mathcal{E}_s = \bigcup_{\forall a \in \mathcal{P}} \Big\{ a \;\Big|\; \|\mathbf{p}_a - \mathbf{m}_{s_n}\| \leq R_a - \Delta \Big\}, \tag{4.2}$$

where $\|\mathbf{x} - \mathbf{y}\|$ is the Euclidean distance between point $\mathbf{x}$ and $\mathbf{y}$, and Δ is a distant margin.

The perpendicular distance from the location of PoA-a, $\mathbf{p}_a$, to a straight line formed by the moving path $(\mathbf{m}_{s_i}, \mathbf{m}_{s_{i+1}})$ is

$$d_a = \frac{\Phi_a}{\|\mathbf{m}_{s_{i+1}} - \mathbf{m}_{s_i}\|}, \tag{4.3}$$

where $\Phi_a = |(\mathbf{m}_{s_{i+1}.y} - \mathbf{m}_{s_i.y})\mathbf{p}_{a.x} - (\mathbf{m}_{s_{i+1}.x} - \mathbf{m}_{s_i.x})\mathbf{p}_{a.y} + \mathbf{m}_{a_{i+1}.x}\mathbf{m}_{s_i.y} - \mathbf{m}_{s_{i+1}.y}\mathbf{m}_{s_i.x}|$ and $|\cdot|$ is the absolute value. If $d_a < R_a$, it implies that the service area of PoA-a intersects the straight line formed by the moving path $(\mathbf{m}_{s_i}, \mathbf{m}_{s_{i+1}})$. $\mathbf{q}_a^1$ and $\mathbf{q}_a^2$ denote the first and second intersection points, respectively, and are calculated as in [64]. Fig. 4.2 depicts all possible cases for the intersection between the service area of PoA-a and a straight line formed by the moving path $(\mathbf{m}_{s_i}, \mathbf{m}_{s_{i+1}})$.

Based on the intersection points between the service area of the PoA and the straight line formed by the moving path, we define the moving path being covered by PoA.

Definition 1. A PoA-a covers a moving path $(\mathbf{m}_{s_i}, \mathbf{m}_{s_{i+1}})$, which is written as COVER$\big(a$, $(\mathbf{m}_{s_i}, \mathbf{m}_{s_{i+1}})\big)$, if some part of the path $(\mathbf{m}_{s_i}, \mathbf{m}_{s_{i+1}})$ is located inside the service area of PoA-a. It is satisfied if one of the following conditions is fulfilled.

$$\|\mathbf{m}_{s_i} - \mathbf{p}_a\| \leq R_a, \tag{4.4}$$

$$\|\mathbf{m}_{s_{i+1}} - \mathbf{p}_a\| \leq R_a, \tag{4.5}$$

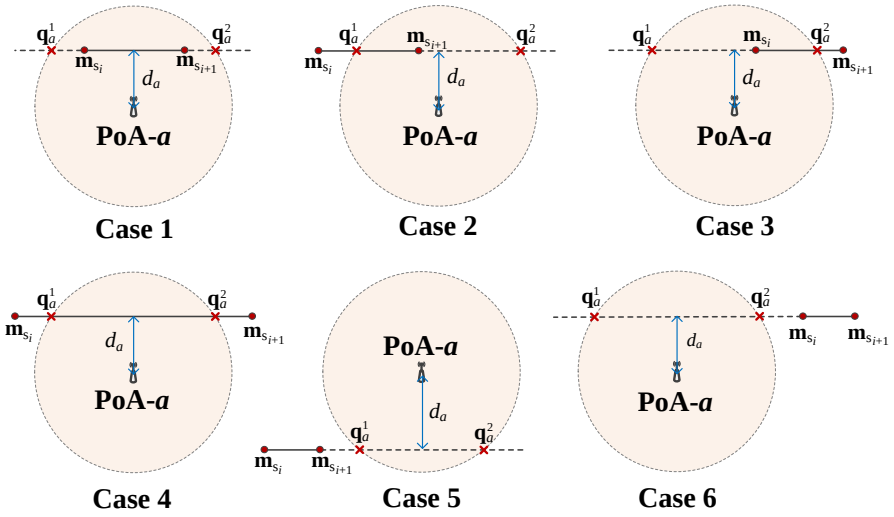


Figure 4.2: All possible cases of intersection between service area of PoA-a and the straight line of the moving path ($\mathbf{m}_{s_i}$, $\mathbf{m}_{s_{i+1}}$).

$$\Big(d_a < R_a\Big) \wedge \Big(\|\mathbf{m}_{s_i} - \mathbf{q}_a^1\| + \|\mathbf{q}_a^1 - \mathbf{q}_a^2\| + \|\mathbf{q}_a^2 - \mathbf{m}_{s_{i+1}}\| = \|\mathbf{m}_{s_i} - \mathbf{m}_{s_{i+1}}\|\Big). \tag{4.6}$$

The equation (4.4) and (4.5) indicate whether $\mathbf{m}_{s_i}$ or $\mathbf{m}_{s_{i+1}}$ is located in the service area of PoA-a (Case 1, 2, and 3 in Fig. 4.2). The equation (4.6) holds when both $\mathbf{m}_{s_i}$ and $\mathbf{m}_{s_{i+1}}$ are located outside the service area of PoA-a, but the straight line formed by those points intersects the service area of PoA-a (Case 4 in Fig. 4.2). A PoA covering any moving path of the road segment-s (Cases 1 – 4 in Fig. 4.2) becomes one of the candidate PoAs for HO on the road segment-s.

When the PoA-a covers one or more moving paths of the road segment-s, there are one or two intersection points between the PoA and the covered moving paths, marked by the red-crosses in an example of Fig. 4.3. The first and second intersection points are denoted by $\mathbf{u}_{s,a}$ and $\mathbf{v}_{s,a}$, respectively. If the first curve point of road segment-s is within the service area of PoA-a, i.e., if $\|\mathbf{p}_a - \mathbf{m}_{s_1}\| < R_a$, $\mathbf{u}_{s,a} = \mathbf{m}_{s_1}$. Also, if the PoA-

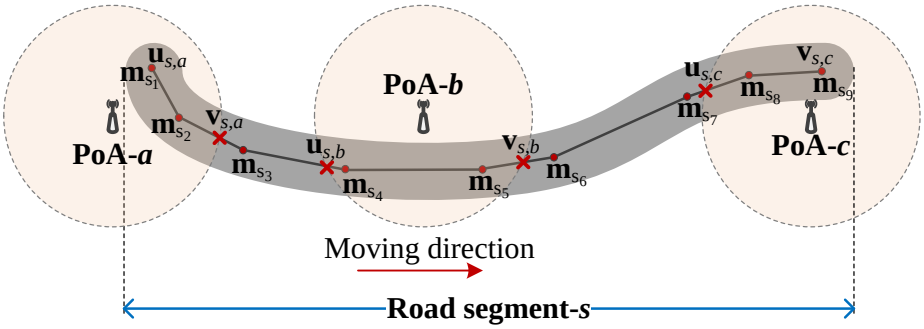


Figure 4.3: The intersection points between the PoAs and the road segment-s where $\mathcal{M}_s = \left(\mathbf{m}_{s_1}, \mathbf{m}_{s_2}, ..., \mathbf{m}_{s_9}\right)$.

a contains the last curve point of the road segment-s ($\|\mathbf{p}_a - \mathbf{m}_{s_k}\| < R_a$), $\mathbf{v}_{s,a} = \mathbf{m}_{s_k}$. Based on the above discussion, we describe the relationship between PoAs on the road segment-s.

Definition 2. A PoA-b is the next neighbor of PoA-a on the road segment-s, which is written as $\text{NEXT}_s\left(a,b\right)$, if there exists a moving path $(\mathbf{m}_{s_i}, \mathbf{m}_{s_{i+1}})$, such that PoA-$a$ and PoA-b satisfy the conditions in (4.7) and (4.8).

$$\|\mathbf{q}_a^1 - \mathbf{q}_a^2\| + \|\mathbf{q}_b^1 - \mathbf{q}_b^2\| > \max_{\phi,\psi \in \{\mathbf{q}_a^1, \mathbf{q}_a^2, \mathbf{q}_b^1, \mathbf{q}_b^2\}} \|\phi - \psi\|, \tag{4.7}$$

$$\Big(h_s(\mathbf{u}_{s,a}, \mathbf{m}_{s_k}) > h_s(\mathbf{u}_{s,b}, \mathbf{m}_{s_k})\Big) \vee \Big(h_s(\mathbf{m}_{s_1}, \mathbf{v}_{s,a}) < h_s(\mathbf{m}_{s_1}, \mathbf{v}_{s,b})\Big), \tag{4.8}$$

where $h_s(\mathbf{x}, \mathbf{y})$ is the total length of all moving paths from the point $\mathbf{x}$ to $\mathbf{y}$.

The condition (4.7) states that the service areas of two PoAs should continuously cover the moving path $(\mathbf{m}_{s_i}, \mathbf{m}_{s_{i+1}})$. The condition (4.8) is that the PoA-a should be followed by the PoA-b with regard to the moving direction on the road segment-s. The HO from PoA-a to PoA-b in the road segment-s may take place if the PoA-b is the next neighbor of the PoA-a.

4.2.3 Directed Graph Representation

For given PoA deployment on the road segment-s, we construct a directed graph $\mathbf{G}_s = (\mathbf{V}_s, \mathbf{E}_s)$, where $\mathbf{V}_s$ is a set of PoAs whose service area covers at least one moving path of the road segment-s and $\mathbf{E}_s$ is a set of all feasible directed edges between PoAs in $\mathbf{V}_s$. Remind that $\mathcal{A}$ is a set of all APs and $\mathcal{B}$ is a set of all BSs. Let $\mathbf{V}_{\mathcal{A},s}$ and $\mathbf{V}_{\mathcal{B},s}$ respectively denote a set of APs and a set of BSs, covering at least one moving path of the road segment-s.

$$\mathbf{V}_{\mathcal{A},s} = \bigcup_{\forall a \in \mathcal{A}} \Big\{ a \;\Big|\; \exists \mathbf{m}_{s_i}, \mathbf{m}_{s_{i+1}} \in \mathcal{M}_s \text{ such that } \mathrm{COVER}\big(a, (\mathbf{m}_{s_i}, \mathbf{m}_{s_{i+1}})\big) \Big\}, \tag{4.9}$$

$$\mathbf{V}_{\mathcal{B},s} = \bigcup_{\forall b \in \mathcal{B}} \Big\{ b \;\Big|\; \exists \mathbf{m}_{s_i}, \mathbf{m}_{s_{i+1}} \in \mathcal{M}_s \text{ such that } \mathrm{COVER}\big(b, (\mathbf{m}_{s_i}, \mathbf{m}_{s_{i+1}})\big) \Big\}. \tag{4.10}$$

Then, $\mathbf{V}_s = \mathbf{V}_{\mathcal{A},s} \bigcup \mathbf{V}_{\mathcal{B},s}$. For the sake of convenience, we will also refer a vertex in $\mathbf{V}_{\mathcal{A},s}$ as AP and a vertex in $\mathbf{V}_{\mathcal{B},s}$ as BS.

Now let us derive all feasible edges in $\mathbf{E}_s$. When the PoA-b is the next neighbor of the PoA-a, there exists a directed edge, e_{ab}, from PoA-a to PoA-b. Let $\mathbf{E}_{\mathcal{A},s}$ be a set of all directed edges between two PoAs (i.e., APs) in $\mathbf{V}_{\mathcal{A},s}$, and let $\mathbf{E}_{\mathcal{B},s}$ be a set of all directed edges between two BSs in $\mathbf{V}_{\mathcal{B},s}$.

$$\mathbf{E}_{\mathcal{A},s} = \bigcup_{\forall a,b \in \mathbf{V}_{\mathcal{A},s}} \Big\{ e_{ab} \;\Big|\; \mathrm{NEXT}_s(a,b) \Big\}, \tag{4.11}$$

$$\mathbf{E}_{\mathcal{B},s} = \bigcup_{\forall a,b \in \mathbf{V}_{\mathcal{B},s}} \Big\{ e_{ab} \;\Big|\; \mathrm{NEXT}_s(a,b) \Big\}, \tag{4.12}$$

If there exists a directed edge from the PoA-a to the PoA-b, it implies that the HO may take place from the PoA-a to the PoA-b. The edges in $\mathbf{E}_{\mathcal{A},s}$ or $\mathbf{E}_{\mathcal{B},s}$ only specify

the horizontal HO either between APs or between BSs. In addition to the horizontal HO, the vertical HO between AP and BS may also take place. Thus, we also should consider the edges which specify the vertical HO. To maximize the connection time on WiFi, the user is allowed to perform the HO from AP to BS only when the current serving AP does not have any other AP as its next neighbor. Thus, the AP can have an outgoing edge to the BS only when the corresponding AP has no outgoing edge to any other AP. Note that an AP belonging to the final PoAs does not have any outgoing edge to the other BSs on the same road segment. Let $\mathbf{V}_{\mathcal{A},s}^{-\mathcal{E}_s}$ denote a set of all APs in $\mathbf{V}_{\mathcal{A},s}$ excluding the APs in $\mathcal{E}_s$. When $\mathbf{E}_s^{\text{vo}}$ is a set of all feasible outgoing edges from AP in $\mathbf{V}_{\mathcal{A},s}^{-\mathcal{E}_s}$ to BS in $\mathbf{V}_{\mathcal{B},s}$,

$$\mathbf{E}_s^{\text{vo}} = \bigcup_{\substack{\forall a \in \mathbf{V}_{\mathcal{A},s}^{-\mathcal{E}_s}, \forall b \in \mathbf{V}_{\mathcal{B},s}, \\ \forall c \in \mathbf{V}_{\mathcal{A},s}}} \left\{ e_{ab} \;\middle|\; e_{ac} \notin \mathbf{E}_{\mathcal{A},s},\ \text{NEXT}_s(a,b) \right\}. \tag{4.13}$$

Note that a user should always perform handover from BS to AP whenever the AP becomes available, This implies that an AP can have an incoming edge from the BS only when having no incoming edge from any other AP. Let $\mathbf{V}_{\mathcal{A},s}^{-\mathcal{S}_s}$ be a set of all APs in $\mathbf{V}_{\mathcal{A},s}$ excluding the APs in $\mathcal{S}_s$. Then, a set of all feasible outgoing edges from BS in $\mathbf{V}_{\mathcal{B},s}$ to AP in $\mathbf{V}_{\mathcal{A},s}^{-\mathcal{S}_s}$ is

$$\mathbf{E}_s^{\text{vi}} = \bigcup_{\substack{\forall b \in \mathbf{V}_{\mathcal{B},s}, \forall a \in \mathbf{V}_{\mathcal{A},s}^{-\mathcal{S}_s}, \\ \forall c \in \mathbf{V}_{\mathcal{A},s}}} \left\{ e_{ba} \;\middle|\; e_{ca} \notin \mathbf{E}_{\mathcal{A},s},\ \text{NEXT}_s(b,a) \right\}. \tag{4.14}$$

Finally, a set of all feasible edges, $\mathbf{E}_s$, is as

$$\mathbf{E}_s = \mathbf{E}_{\mathcal{A},s} \bigcup \mathbf{E}_{\mathcal{B},s} \bigcup \mathbf{E}_s^{\text{vo}} \bigcup \mathbf{E}_s^{\text{vi}}. \tag{4.15}$$

Fig. 4.4 depicts an example for the PoA deployment on road segment-s and Fig. 4.5 shows the corresponding directed graph. Each edge in the graph has a weight representing a cost of the corresponding HO, which will be discussed in the next section.

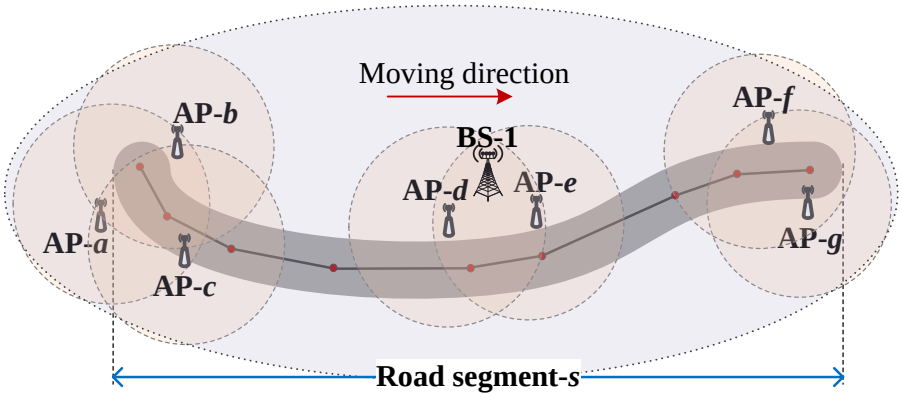


Figure 4.4: Example deployment of PoAs on a road segment.

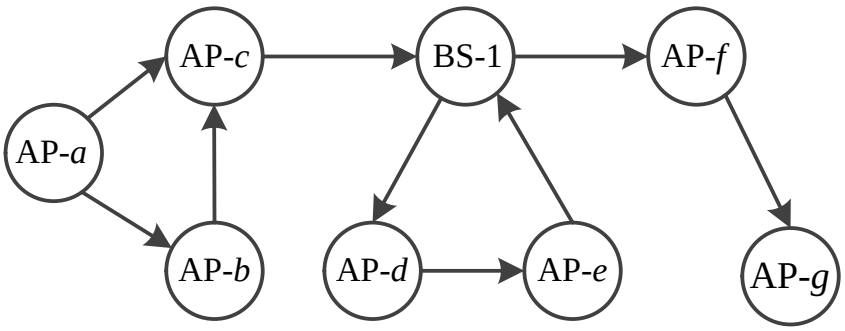


Figure 4.5: A directed graph for the road segment of Fig. 4.4.

4.3 Handover Scheduling Problem

In this section, we formulate the HO scheduling problem and propose the HO scheduling algorithm.

4.3.1 Problem Formulation

The primary goal of our scheme is to find the best HO schedule for each road segment, which minimizes the total HO latency while having network connection through WiFi (not cellular) whenever possible and preferring the handover to the AP with fewer

users. The HO schedule is a sequence list of target PoAs in the order of the HO actions to be conducted by a user when it moves along a particular road segment. The HO schedule is derived by solving the following HO scheduling problem.

We formulate the HO scheduling problem of road segment-s, using the directed graph $\mathbf{G}_s = (\mathbf{V}_s, \mathbf{E}_s)$. Let z_{ab} be a binary variable indicating whether the edge e_{ab} $(\in \mathbf{E}_s)$ is selected. Selecting e_{ab} means that the HO from the PoA-a to the PoA-b is scheduled. If e_{ab} is selected, $z_{ab} = 1$; otherwise $z_{ab} = 0$. Let $w(e_{ab}, t)$ be a weight assigned to edge e_{ab} at time t. The HO scheduling problem for the road segment-s at a time t, started from the PoA-υ ($\upsilon \in \mathcal{S}_s$) and ended at any PoA in $\mathcal{E}_s$, is formulated as (4.16) – (4.20).

$$\min \quad \sum_{e_{ab} \in \mathbf{E}_s} z_{ab}\, w(e_{ab}, t) \tag{4.16}$$

$$s.t. \quad \sum_{a \in \mathbf{V}_s, e_{\upsilon a} \in \mathbf{E}_s} z_{\upsilon a} \geq 1, \tag{4.17}$$

$$\sum_{b \in \mathbf{V}_s, e_{ab} \in \mathbf{E}_s} z_{ab} - \sum_{c \in \mathbf{V}_s, e_{ca} \in \mathbf{E}_s} z_{ca} = 0, \quad \forall a \in \Omega_s \tag{4.18}$$

$$\sum_{a \in \mathbf{V}_s} \sum_{b \in \mathcal{E}_s, e_{ab} \in \mathbf{E}_s} e_{ab} \geq 1, \quad b \neq \upsilon \tag{4.19}$$

$$z_{ab} \in \{0, 1\}, \quad \forall a, b \in \mathbf{V}_s \tag{4.20}$$

where Ω_s is the set of all PoAs in $\mathbf{V}_s$ excluding the starting PoA and the final PoAs.

The objective (4.16) seeks for the HO schedule that minimizes the sum of weights. The constraint (4.17) ensures that at least one of the outgoing edges of PoA-υ should be included since it is a starting PoA. The constraint (4.18) ensures that if the PoA-a in Ω_s is included in the HO schedule, both its incoming and outgoing edges should be selected together since it is a connecting PoA. The constraint (4.19) ensures that the HO schedule is ended at one of the PoA in $\mathcal{E}_s$, thus at least one incoming edge to one of these PoAs should be selected. The constraint (4.20) is self-explanatory.

4.3.2 Weight of Edge

To solve the above HO scheduling problem, we first should design a weight of each edge, i.e., $w(e_{ab},t)$ for $e_{ab} \in \mathbf{E}_s$. Recall that the objectives of the HO scheduling problem are: (1) maximizing the connection time on WiFi; (2) minimizing the total HO latency taken to perform all HO operations during moving on roads; and (3) preferring the handover to the AP with fewer users. The weight value of each edge should be assigned so that the objectives are fulfilled.

To encourage that a user selects WiFi (not cellular) for communication whenever possible, the highest preference should be given to a directed edge from BS to AP ($\in \mathbf{E}_s^{\text{vi}}$). Thus, the edge belonging to $\mathbf{E}_s^{\text{vi}}$ should have the smaller weight value than the edges with the other types, because of a minimization problem. Next, to achieve the objective of minimizing the total HO latency, the second highest preference is given to a directed edge between APs ($\in \mathbf{E}_{\mathcal{A},s}$), since the horizontal HO between APs has much shorter latency than the vertical HO between AP and BS. Lastly, the lowest preference is given to an edge from AP to BS ($\in \mathbf{E}_s^{\text{vo}}$) and an edge between BSs ($\in \mathbf{E}_{\mathcal{B},s}$), since the connection to BS is unfavorable. Based on the above discussion, the weight of e_{ab} at a time t, denoted by $w(e_{ab},t)$, is.

$$w(e_{ab},t) = \begin{cases} -\text{card}(\mathbf{V}_s), & \text{if } e_{ab} \in \mathbf{E}_s^{\text{vi}} \\ \alpha \frac{\tau_{ab}}{T} + \beta \frac{\eta_b(t)}{N_b}, & \text{if } e_{ab} \in \mathbf{E}_{\mathcal{A},s} \\ 1, & \text{if } e_{ab} \in \mathbf{E}_s^{\text{vi}} \bigcup \mathbf{E}_{\mathcal{B},s} \end{cases} \tag{4.21}$$

In (4.21), $\text{card}(\mathbf{V}_s)$ is the cardinality of the set $\mathbf{V}_s$, τ_{ab} is the HO latency from AP-a to AP-b, and $\eta_b(t)$ is the number of users connected to AP-b at time t. In addition, T is the maximum possible HO latency and N_b is the maximum allowable number of users at PoA-b.

Now, we examine more specifically the weight value according to each edge type,

starting from the edges in $\mathbf{E}_{\mathcal{A},s}$. The edge from AP-$a$ to AP-b has the weight of $\alpha\frac{\tau_{ab}}{T} + \beta\frac{\eta_b(t)}{N_b}$. The first term accounts for the HO latency and the second term reflects the number of users associated with the target AP, where α and β ($\alpha+\beta=1$) are the relative importance factors of the first and second terms. Note that any edge in $\mathbf{E}_{\mathcal{A},s}$ has a positive value smaller than one, since each term is normalized by its maximum value. The weight is designed with the purpose of reducing the total HO latency (the first term) while preferring the handover to the AP with fewer associated users (the second term). Since the handover to the AP with fewer associated users can help to relieve the channel contention among users within AP, each user may have an opportunity to get the higher throughput. Furthermore, since the number of associated users at each AP is changed over time, this term can give a randomness effect to the HO schedule. Note that the graph representing the PoA deployment on a road segment is a fixed factor. In addition, because the HO latency between PoAs mainly depends on its HO type which is a pre-determined parameter, this term reflecting the number of associated users at each AP is only the time-varying factor in (4.21). Therefore, without this term, the HO schedule on a particular road segment may always contain the same set of PoAs. This situation is very undesirable since the users are concentrated to some particular APs on a road segment and such load unbalancing may incur the performance degradation.

On the other hand, for maintaining the network connection through WiFi whenever possible, any directed edge from BS to AP should have a weight smaller than zero, since an edge between two APs has a positive value smaller than one. We simply set the weight of any edge in $\mathbf{E}_s^{\text{vi}}$ to $-\text{card}(\mathbf{V}_s)$. As a result, whenever an AP becomes available, the user is forced to connect to the AP.

An edge from BS to other BS is selected only when the BS does not have an AP as its next neighbor. The HO from AP to BS is also scheduled only when the AP has no neighboring AP. Thus, when considering a positive weight value smaller than one for an edge between APs, the weight of an edge belonging to these types can be set to

any positive value equal to or larger than one. We simply assign the weight of one to the edges from AP to BS and from BS to BS.

4.3.3 HO Scheduling Algorithm

The problem in (4.16) – (4.20) is a mixed integer linear programming (MILP) problem, which can be efficiently and quickly solved using any MILP solver such as CPLEX [44] or LPsolve [45]. Then, the HO schedule is derived from the solution of the problem, which is a set of selected edges. Algorithm 6 returns the HO schedule of a user for the road segment-s, which is a sequence list of target PoAs in order of the HO actions to be conducted by the user while moving along the road segment. Like in Algorithm 5, the function EmptyList() generates an empty list and the function InSertList(L, e) inserts e as the last element of a list L.

Algorithm 6 HO scheduling on road segment-s, started at PoA-υ

1: Solve the problem (4.16) – (4.20) using MILP solver
2: Let $\mathbb{E}$ be the set of edges returned by MILP solver
3: Let $\mathbb{V}$ be the set of vertices whose edge is in $\mathbb{E}$
4: $L_{s,\upsilon} \leftarrow$ EmptyList()
5: InSertList$(L_{s,\upsilon},\ \upsilon)$, $a \leftarrow \upsilon$
6: **while** $\mathbb{E} \neq \emptyset$ **do**
7: **if** $a \in \mathcal{B}$ **then**
8: $b \leftarrow \text{argmin}_{\hat{b} \in \mathbb{V}, e_{a\hat{b}} \in \mathbb{E}} h_s(\mathbf{m}_{s_1}, \mathbf{v}_{s,\hat{b}})$
9: **else**
10: find b in $\mathbb{V}$ such that $e_{ab} \in \mathbb{E}$
11: **end if**
12: InSertList$(L_{s,\upsilon},\ b)$, $\mathbb{E} \leftarrow \mathbb{E} \backslash \{e_{ab}\}$, $a \leftarrow b$
13: **end while**
14: **return** $L_{s,\upsilon}$

Firstly, the starting PoA-υ is inserted to an empty list $L_{s,v}$ as the first element (lines 4 – 5). After that, the HO schedule is derived by repeatedly identifying the next PoA to be connected through the already selected edge until all edges are processed (lines 6 – 13). If the current PoA-a is a BS (line 7), there may be several possible outgoing edges

(see Fig. 4.5). Thus, to avoid skipping any edge in $\mathbb{E}$, we should select the outgoing edge to a PoA which is the closest to the starting curve point of the corresponding road segment (line 8). However, if the current PoA-a is an AP, there is only one outgoing edge from AP-a in $\mathbb{E}$ (line 10). Then, the PoA-b is inserted to the list of target PoAs. An example of the HO schedule for the graph with starting PoA-a in Fig. 4.5 derived by the Algorithm 6 is $\big($AP-a, AP-c, BS-1, AP-d, AP-e, BS-1, AP-$f\big)$.

4.4 Handover Scheduling Operation

The proposed HO scheduling is composed of an offline phase and an online phase. In the offline phase, the server generates a directed graph for each road segment, by utilizing the pre-inputted information about network configurations and road topologies. It is possible since the graph generation requires only the static information about PoAs and road segments. Meanwhile, since the weight values of edges are periodically updated with the number of associated users reported by each PoA, the HO schedules cannot be computed in advance and should be generated on demand. In the online phase, when receiving the request for HO schedule from a user, the server computes the HO schedule for the user by using Algorithm 6. Note that the HO schedule delivered from the server to the user is merely an ordered list of HO target PoAs. Accordingly, the user should decide the HO triggering points (i.e., when) by itself while moving along a road segment. The block diagram in Fig. 4.6 summarizes the steps in the proposed HO scheme. Now, we describe the HO schedule delivery and the HO triggering in the online phase.

4.4.1 HO Schedule Delivery

When a user connects to the last PoA of the current HO schedule or when it initially joins the network, the user request a new HO schedule to the server through its serving

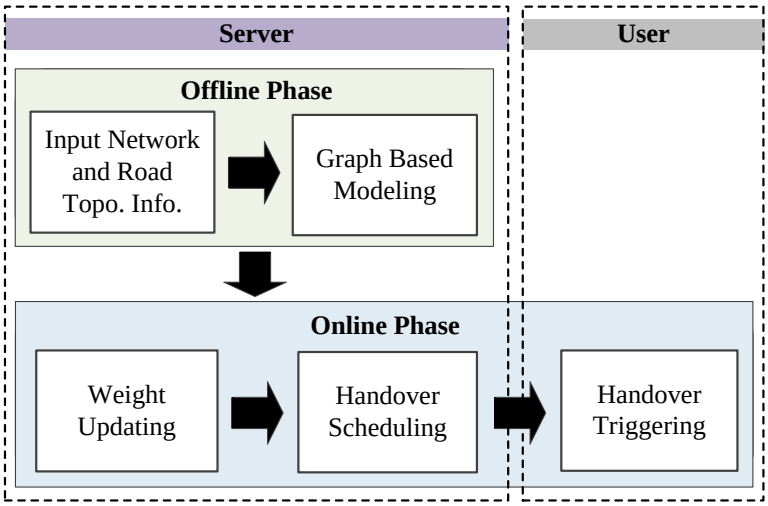


Figure 4.6: Block diagram of the proposed HO scheduling scheme.

PoA.[3] For this, the user should send a HO request (HO_REQ) packet, which contains the information about its current location and its serving PoA, to the server. Upon receiving this request, the server checks the information in the HO_REQ packet and identifies the possible future road segments to which the user may move, as follows. If a user is currently connected to the final PoA, it implies that the user is close to the end of the road segment. Thus, by observing the current location of the user, the server may figure out the next possible road segments to which the user may move. On the other hand, if a user is not currently connected to the final PoA, it implies that the user has newly joined the network in the middle of a road segment and its next road segment is the current one.

When recognizing the next possible road segments, the server should compute the HO schedules for all possible road segments since it cannot know the road segment of which the user will finally move to. Assume that the PoA-υ is the current serving PoA of the user. For any possible road segment-s, the server gets the HO schedule $L_{s,\upsilon}$, by executing the Algorithm 6 with the starting PoA υ. Then, the server generates

[3] A user joining the network may select its serving PoA based on RSSI.

the information of (*PoA ID, Channel ID, HO Type*) for each PoA in $L_{s,\upsilon}$, where the *PoA ID* is the unique identifier of the corresponding PoA, the *Channel ID* identifies the working channel of the PoA, the *HO type* indicates whether the type of the HO event from the corresponding PoA to the next PoA is L2, L3, or vertical HO. Let $\mathcal{I}_{s,\upsilon}(i)$ be the information for the ith PoA of $L_{s,\upsilon}$ and let $n_{s,\upsilon}$ be the number of PoAs in $L_{s,\upsilon}$. Then, the HO schedule information for road segment-s is configured as $\mathbb{Z}_{s,\upsilon} = \left(\theta_s^1, \mathcal{I}_{s,\upsilon}(1), \mathcal{I}_{s,\upsilon}(2), \cdots, \mathcal{I}_{s,\upsilon}(n_{s,\upsilon})\right)$, where θ_s^1 is the north azimuth value of the first moving path of the road segment-s. After that, the server replies with a HO reply (HO_REP) packet, which contains the HO schedule information $\mathbb{Z}_{S1,\upsilon}, \mathbb{Z}_{S2,\upsilon}, \cdots, \mathbb{Z}_{S\kappa,\upsilon}$ for all possible road segments $S1, S2, \cdots, S\kappa$.

Upon receiving the HO_REP packet, if there are more than one HO schedule (e.g., at the intersection), the user firstly should select an appropriate HO schedule based on its final moving direction, before updating the HO schedule. It may be better to make the selection as late as possible until the next road segment is manifested, as long as the user does not get out of the service area of the current serving PoA. Based on this reasoning, when the serving PoA of the user is an AP, the user performs the HO schedule selection when it approaches the service area boundary of AP (i.e., when the RSSI value of the serving AP falls below the HO RSSI threshold). Meanwhile, the user being currently served by the BS performs the HO schedule selection when it first detects a new AP, or approaches the cell edge.

When θ_c is the north azimuth value of the current moving direction of the user, the user selects the HO schedule whose north azimuth value is equal to θ_s^*.

$$\theta_s^* = \underset{\theta_i \in \{\theta_{S1}^1, \theta_{S2}^1, \ldots, \theta_{S\kappa}^1\}}{\operatorname{argmin}} |\theta_i - \theta_c|, \tag{4.22}$$

4.4.2 HO Triggering and Execution

Since a user gets the HO schedule in advance before entering a road segment, the user knows that it should perform the handover to which PoAs while moving on the road segment. Additionally, the user has to decide when the handover should be tried (HO triggering). We basically consider an RSSI-based HO triggering.

Firstly, when the current PoA is a BS and the next PoA in the current HO schedule is an AP, the user continually checks the accessibility to the next AP while communicating with the serving BS, by intermittently scanning the working channel of the AP. Note that such channel scanning does not affect the data transmission since the user uses different network interfaces for the LTE and WiFi. When detecting that the RSSI value of the next AP is higher than a predefined minimum RSSI threshold, the user should immediately trigger the handover to the AP, regardless of the RSSI status of its serving BS. It allows the user to connect to the WiFi network whenever possible. When the current PoA is a BS and there is no neighboring AP, the horizontal HO mechanism of LTE can be used. One the other hand, if the serving PoA is an AP, the handover operation to the next PoA (either AP or BS) is started when the RSSI of serving AP falls below a predefined threshold.

Once the HO is triggered, the user and the target PoA execute the HO protocol steps (e.g., horizontal HO between WiFi APs, horizontal HO between BSs, or vertical HO between AP and BS). The readers can refer to [11] for the execution steps of horizontal HO between WiFi APs and [63] for horizontal HO between LTE BSs. For the vertical HO in our scheme, because the user uses different network interfaces to access WiFi and LTE, it is merely needed to switch the active network interface for data transmission either from LTE to WiFi or vice versa.

Note that if the handover to a particular target PoA fails after a certain number of trials, the user marks the PoA as being unavailable. Then, the user scans the working

channels of neighboring PoAs and selects the PoA with the highest RSSI as a HO target PoA. Afterwards, the user executes the handover to the new target PoA by following the proposed HO procedure. In addition, the user may report the unavailable PoA to the server. The server can make a further decision regarding the unavailable PoA, for example by removing it from the graph. Then, it will be excluded in any future HO schedules until it is available.

4.4.3 Communication Overhead

In the proposed HO scheme, the burden for computing the HO schedule is moved from user to server, whereas a little communication overhead for delivering the schedule information from server to user is added (the overhead of HO request is negligible). Let us estimate this delivery overhead.

Whenever a user reaches an intersection, the user receives the HO schedules for all possible road segments. For each PoA in a HO schedule, we use the PoA ID of 6 bytes, the channel ID of 1 byte, and the HO type of 1 byte (totally 8 bytes). The HO schedule which contains N_p PoAs uses $(8N_p + 1)$ bytes, including the north azimuth value of 1 byte. If the number of possible road segments at the intersection is N_m, the whole HO schedule information delivered to the user is $(8N_p + 1)N_m$ bytes. Note that the value of N_p depends on the length of road segment. Since the length of a road segment can be designed not to be too long, N_p is likely to be not large. The intersection is also commonly constructed by a few of road segments, implying that the value of N_m is also small. Note that the HO schedule information of a user has the largest size at intersection. As a result, the information amount of any HO schedule is reasonably so small (a few hundred bytes) as to be carried by a single MAC frame. Moreover, since the server sends such the HO schedule information to a user only when it connects to the last PoA of the current HO schedule or initially joins the network, the information delivery overhead of the proposed scheme is expected to be not high.

Chapter 5

Performance Evaluation

5.1 Centralized Channelization Scheme for Wireless LANs Exploiting Channel Bonding

5.1.1 Experiment Settings

Fig. 5.1 shows the experiment environment and the deployment of the APs and user devices. In the experiment, we consider three scenarios, i.e., the scenario I: controlled APs deployment only and the scenario II: controlled and uncontrolled APs deployment. Note that to realize scenario II, we only need to additionally deploy the uncontrolled APs to the scenario I. For the scenario III, the APs and users deployment are exactly the same with scenario II. In Fig. 5.1, the controlled APs are numbered from AP-1 to AP-4 and the uncontrolled APs are numbered from AP-5 to AP-8. For both scenarios, there are only four available channels, i.e., channel 36, 40, 44, and 48. Thus, we have four basic channels, two channels with a width of 40 MHz, and a single channel with a width of 80 MHz. All uncontrolled APs are assigned 20 Mhz channel as follows: channel 44 for AP-5, channel 40 for AP-6, channel 44 for AP-7, and channel 36 for AP-8. In the scenario III, we set all uncontrolled APs use channel bonding as

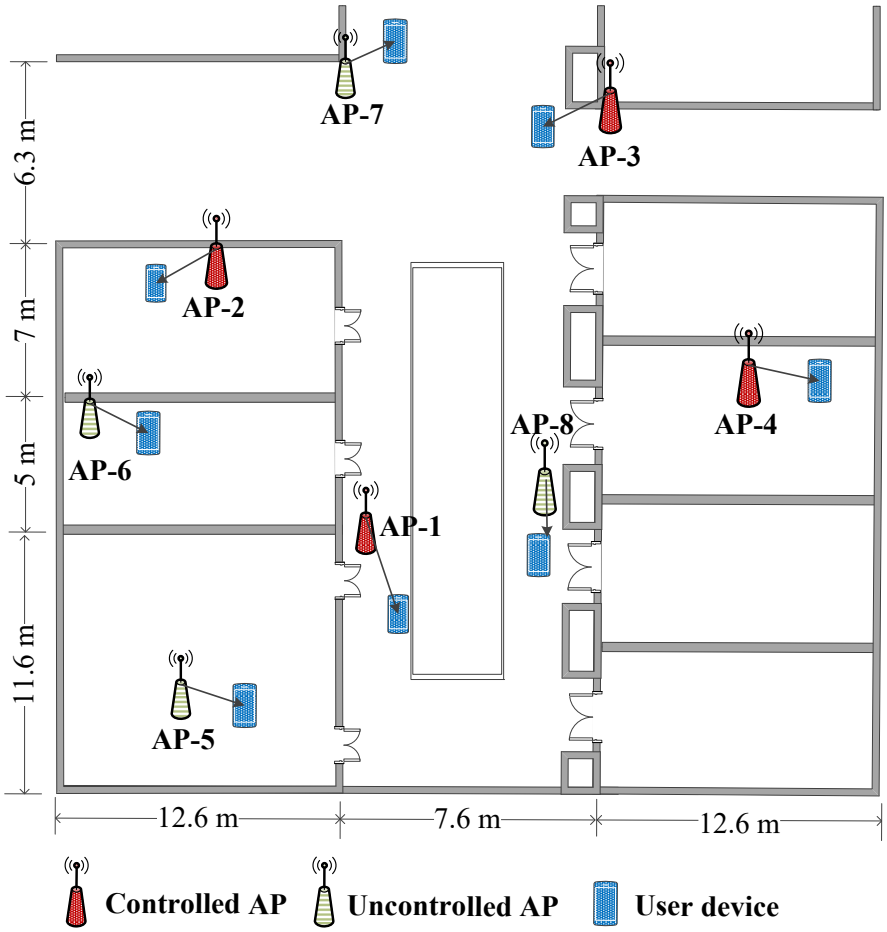


Figure 5.1: Experiment environment for evaluation channelization schemes.

follows: channel 36 and 40 for AP-5, channel 44 and 48 for AP-6, channel 36 and 40 for AP-7, and channel 44 and 48 for AP-8. Thus, in this scenario, all channels are being used by at least two APs.

A single 802.11 client is connected to each AP to receive downlink traffic. Moreover, a notebook is connected using the wired link to each AP which generates a UDP traffic for the client using *Iperf tool*. The traffic generation rate of the notebook connected to each controlled AP is 200 Mbps. This amount of traffic is enough to emulate

the saturation condition for the controlled AP. On the other hand, the traffic generation rate of the notebook connected to each uncontrolled AP is 50 Mbps.

5.1.2 Comparison Schemes

For the comparison schemes, we consider the centralized version of [13] (namely LIC) and the channel allocation scheme in [16]. In LIC, the controller uses the interference graph to assign 20 MHz channel to the AP. The controller assigns the channel with the least number of interference sources to each AP in consecutive order. On the other hand, Scheme in [16] considers only controlled APs and direct interference relation between them. The APs which can be assigned the same channel are group together. Thus, the APs, which are in a single group, do not have direct interference relation with each other. The channel assignment is done in the group manner in the consecutive order. Firstly, each group is assigned a single 20 MHz channel. After the groups are assigned a channel, and there are still available free basic channels, the channel for the first group is doubled, then continued with the second group, and so forth. This process is continued until there is no free basic channel.

5.1.3 Preliminary Experiment for Building Interference Graph

In the real environment, the central controller can build the interference graph automatically. Moreover, the controller should also update the graph periodically (e.g., once every several hours) before the execution of the channelization algorithm. For our experiment, we conduct the following measurement to derive the interference graph. We first set the working channel of all controlled APs to the channel-36, and make all controlled APs perform channel scanning to derive the direct interference relation. Moreover, we develop a simple Android application for the client devices to scan the channel and collect the RSSI status of the APs to populate the hidden interference relation between APs and calculate its corresponding cost value. Then, five users each with

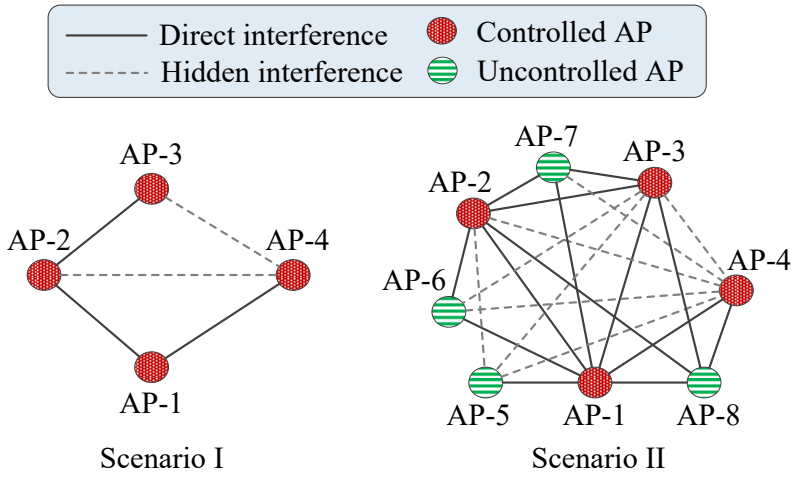


Figure 5.2: The interference graphs.

a single client device measure the RSSI of the APs within the experiment environment for a total duration of one hour. The movement model of a user is as follows, in the corridor the user continuously moves with the walking speed. Once the user arrives in the class room or seminar room, the user stays for quite a while before the user finally moves again. The duration of the user stays inside the room is much longer than that of in the corridor.

The device automatically scans and stores the RSSI information locally with the period of 5 seconds. At the end of the experiment, we gather all the RSSI report information stored locally in each client device. We observe the hidden interference relation and calculate the corresponding cost value. The RSSI threshold value for building interference graph is set to -82 dBm. Fig. 5.2 depicts the interference graph for the scenario I, II, and III (for simplicity the directed edge is replaced with the normal edge). The corresponding cost values are presented in Table 5.1. Note that the cost value for the same edge on the scenario I, II, and III is the same. The APs in the table header are the interference sources.

Table 5.1: The cost of edge on the interference graph

AP	1	2	3	4	5	6	7	8
1	-	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2	1.00	-	1.00	0.13	0.36	1.00	1.00	1.00
3	1.00	1.00	-	0.21	0.11	0.24	1.00	1.00
4	1.00	0.15	0.23	-	0.09	0.13	0.25	1.00

5.1.4 Experiment Results

The channel allocation and the interference graph results of the proposed scheme for the scenario I and II are depicted in Fig. 5.3. For scenario I, the proposed scheme successfully exploits the fact that the AP-2 and AP-4 can be assigned the same channel without deteriorating their performance. As a result, there are only three sets of channel required for the allocation. Since we have four available basic channels, thus by assigning AP-2 and AP-4 with the same 40 MHz channel maximizes the total network throughput (optimal allocation). Note that the asterisk mark in the table means that the corresponding channel is selected as the primary channel. Moreover, for scenario II, since there are more interference sources, the proposed scheme only assigns 40 MHz channel to the AP-4. This assignment is possible since the AP-4 has a relatively small total number of interference sources.

Fig. 5.4 depicts the channel allocation and interference graph results of the LIC in the scenario I and II. Since in the scenario I the LIC assign only 20 MHz, it ends in assigning a distinct channel to the AP. In the case of scenario II, the LIC barely minimizes the total number of interference sources of each AP. It is confirmed by the number of edges on the resulted interference graph. However, due to the limited number of channel, it is unavoidable that there still exists a direct interference relation in the graph.

The channel allocation and interference graph results of the Scheme in [16], are

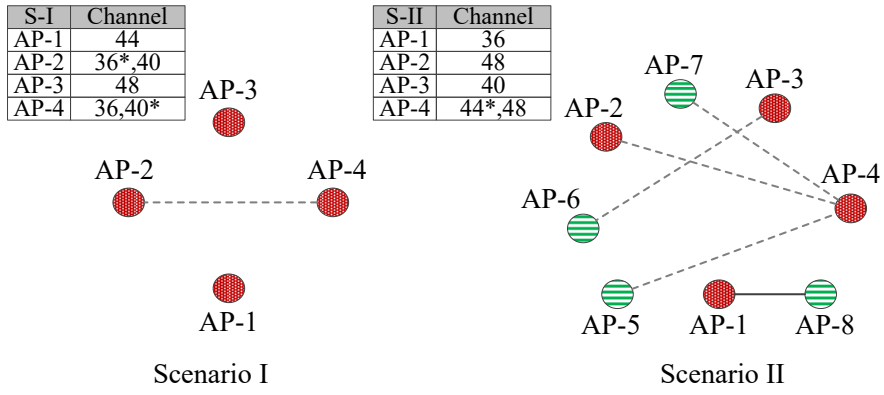


Figure 5.3: The results of channel allocation and interference graph by the proposed scheme.

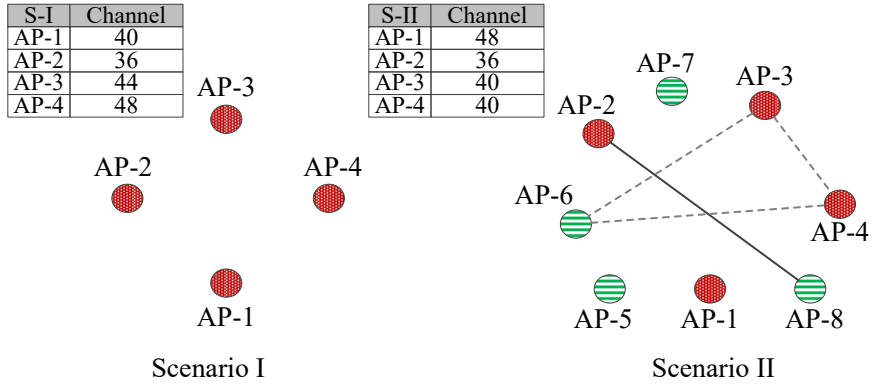


Figure 5.4: The results of channel allocation and interference graph by the LIC.

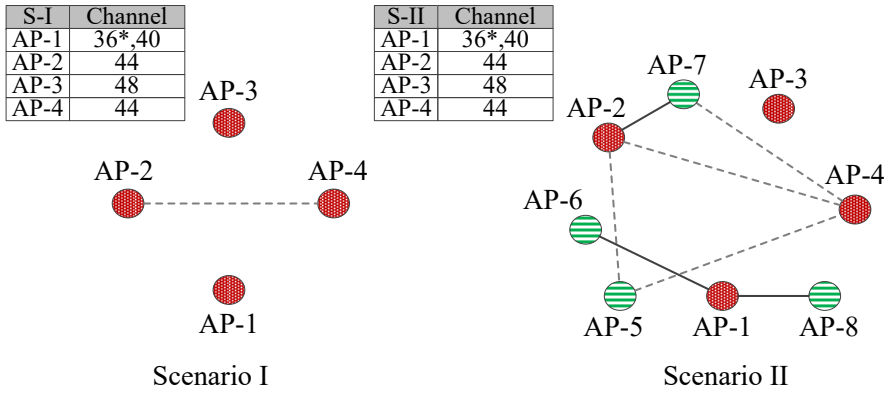


Figure 5.5: The results of channel allocation and interference graph by the Scheme in [16].

presented in Fig. 5.5. The Scheme in [16] assigns the channel in the consecutive order. Thus, even though it can exploit channel bonding, it does not result in the optimal allocation. As a result, for the scenario I, only a single AP is assigned 40 MHz channel. For the scenario II, since the Scheme in [16] does not consider the existence of the uncontrolled APs, thus its allocation result does not change as compared to that of in the scenario I. This causes AP-1 and AP-2 to share the channel with some of uncontrolled APs, which degrades the performance severely.

The allocation result for the scenario III are as follows. By running the proposed scheme, the AP-1 is assigned 20 MHz channel i.e., channel-44. The AP-2 and AP-4 are assigned the same 40 MHz channel, i.e., channel 36 and 40. The AP-3 is assigned channel-48. Next, executing the LIC scheme, the channel allocation are as follows: AP-1 is assigned channel-36, AP-2 is assigned channel-44, AP-3 is assigned channel-40, and AP-4 is assigned channel-48. Meanwhile, scheme in [16] has the same allocation result as that it scenario II. This is because the scheme in [16] does not consider the existence of uncontrolled APs.

Next, we want to evaluate the performance of the system as a result of channel

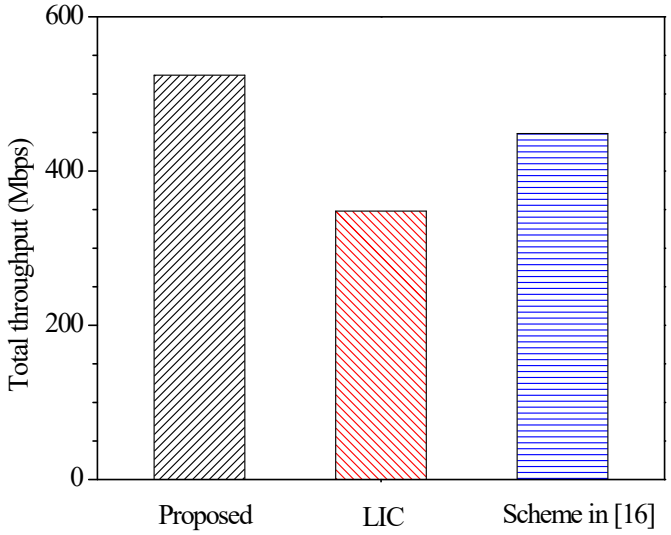


Figure 5.6: Total network throughput (scenario I).

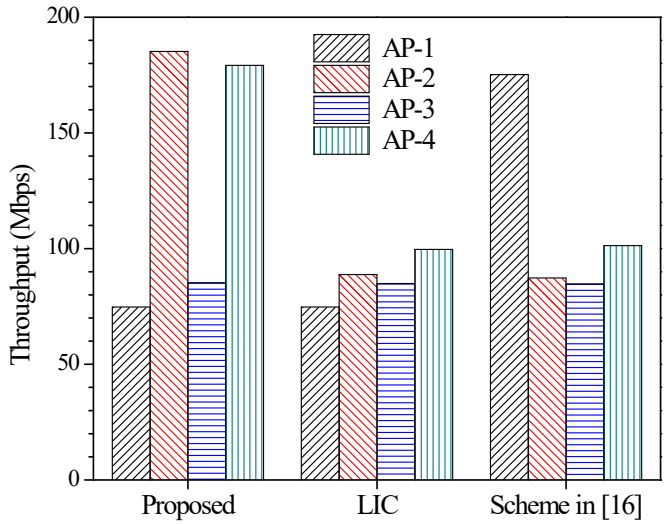


Figure 5.7: Throughput per-AP (scenario I).

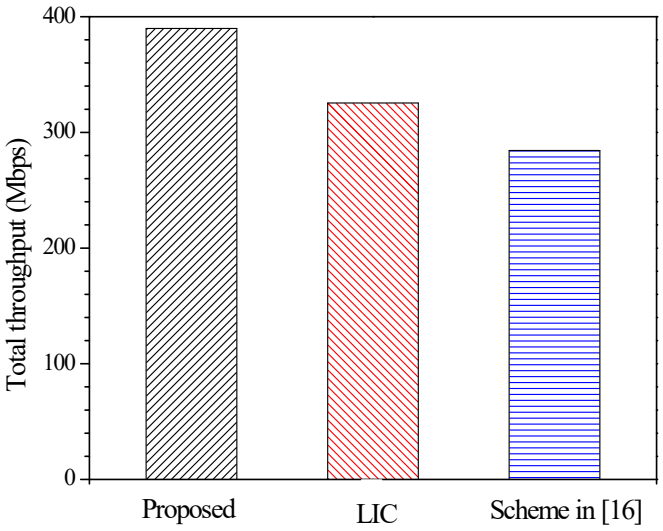


Figure 5.8: Total network throughput (scenario II).

allocation by observing the throughput. For this purpose UDP traffic is tramitted by AP to the clients. Fig. 5.6 and 5.7 show the comparison of throughput in the scenario I. As the proposed scheme results in optimal channel allocation results (i.e., it assigns 40 MHz to AP-2 and AP-4), it achieves the highest total system throughput. It is also shown by the throughput per-AP (Fig. 5.7), in which for the proposed scheme, the AP-2 and AP-4 get the highest value. The Scheme in [16] only assigns 40 MHz channel to a single AP. Thus, its total system throughput is lower than that of the proposed scheme. The LIC assigns only 20 MHz channel, which results in the lowest total system throughput. Note that in Fig. 5.7, the throughput of the APs with the same channel width is slightly different. This difference is mainly caused by the existence of the external interference which can not be controlled during the experiment.

The results of throughput for scenario II are depicted in Figs. 5.8 and 5.9. The proposed scheme results in the highest total system throughput. The reason is that

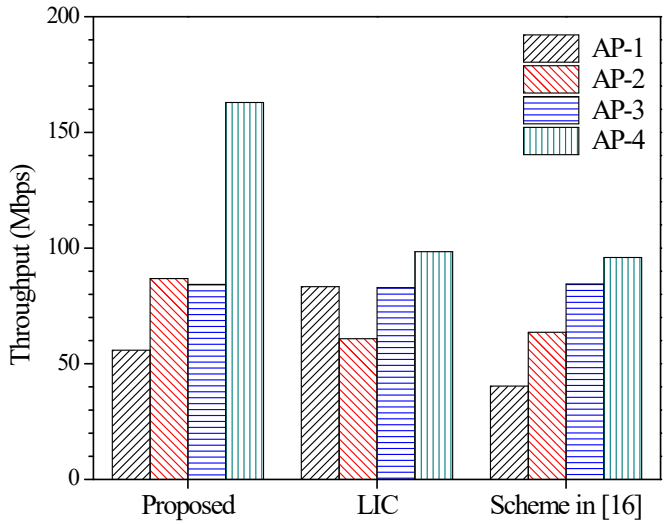


Figure 5.9: Throughput per-AP (scenario II).

it produces the optimal allocation results. The Scheme in [16] results in the worst throughput as it does not consider the existence of the uncontrolled APs. In this case, the uncontrolled APs causes some APs to share the channel, which significantly degrades the network performance. It shows that if the channel bonding is not assigned carefully, the performance of the WLANs using channel bonding is worse than that of the WLANs using basic channel only.

Figs. 5.10 and 5.11 show the throughput for the scenario III. Similarly, the proposed scheme achieve significantly higher total throughput than the compared schemes. This demonstrate that the proposed scheme successfully exploits channel bonding even in the dense WLAN environment. The proposed scheme starts to assign channel bonding to the controlled APs when it found that the uncontrolled APs also use channel bonding. This is because using single channel will not give any advantage since all channels have been used by the uncontrolled APs. Thus, since the interference is non-

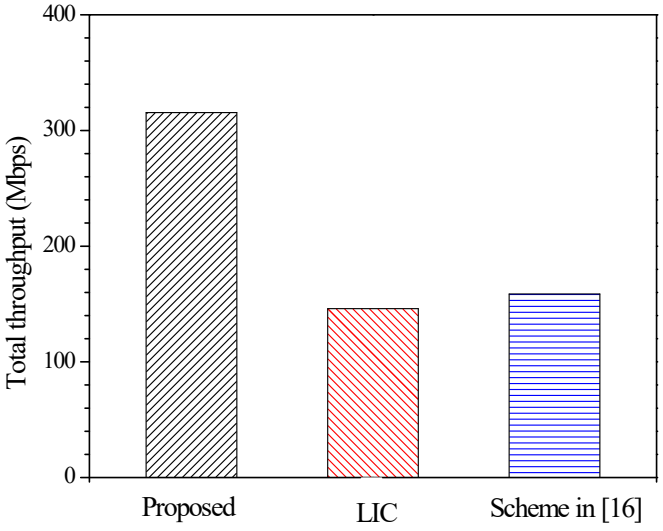


Figure 5.10: Total network throughput (scenario III).

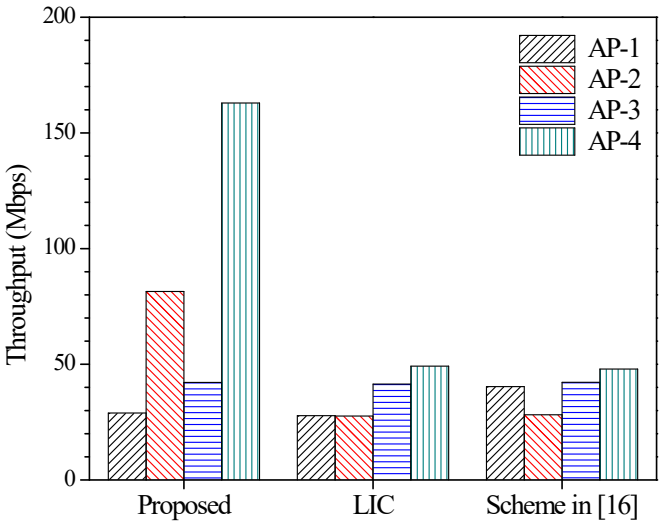


Figure 5.11: Throughput per-AP (scenario III).

negligible, the proposed scheme prefers channel bonding to get the advantage of higher transmission rate. This phenomena can be also observed from the result of LIC scheme and scheme in [16]. In the scenario III, unlike scenario II, scheme in [16] achieves higher total throughput than LIC scheme. The reason is exactly the same with the above discussion. However, since the scheme in [16] does not consider interference from the uncontrolled APs, its total throughput is not significantly higher than that of LIC scheme.

The experiments show that the proposed channel allocation scheme successfully exploits the channel bonding strategy either in less dense or dense WLAN deployment. According to the results, the channel bonding is applied if the AP has a relatively small total number of direct and hidden interference sources. Meanwhile, even though we cannot evaluate the performance of the proposed scheme under DCB due to the limited capability of the experiment equipments, we believe that the performance of the proposed scheme under SCB provides the baseline results. In which, the proposed scheme will provide a higher performance gain under DCB than that of under SCB. The reason is that the primary channel selection under DCB is more critical than under SCB, thus a proper primary channel assignment can be expected to offer a higher performance improvement in DCB.

5.2 User Association for Load Balancing and Energy Saving in Enterprise Wireless LANs

5.2.1 Performance Metrics

As the metrics for evaluating the performance of UA scheme, we adopt a load balancing factor (LBF), the offered load of a WAP, the ratio of WAP, the aggregate throughput of the entire ESS, and the throughput per WAP.

Offered Load of WAP

Let $\mathfrak{L}_n$ be the offered load of WAP n, defined as the sum of uplink and downlink traffic arrival rates to the WAP n as a result of UA process. Then,

$$\mathfrak{L}_n = \sum_{j \in \mathcal{J}_n} z_{j,n} \sum_{i \in \mathcal{M}_j} \left(\lambda_i^{\mathrm{u}} + \lambda_i^{\mathrm{d}}\right) + \sum_{k \in \mathcal{K}} y_{k,n} \beta_k. \tag{5.1}$$

Load Balancing Factor

We define the load balancing factor η as the offered load per WAP, normalized by the biggest value among the offered loads of all WAPs. When Ω denotes the set of WAPs, i.e., $\Omega = \{n \mid \mathrm{a}_n = 1,\ n \in \mathcal{A}\}$,

$$\eta = \frac{1}{|\Omega|} \sum_{n \in \Omega} \frac{\mathfrak{L}_n}{\max_{l \in \Omega} \mathfrak{L}_l}. \tag{5.2}$$

If η is close to 1, the system traffic load is well balanced among WAPs. Note that the load balancing factor tells us how much the total traffic arrival rate of an AP differ with those of other APs.

Ratio of WAP

The ratio of WAP is calculated as the average number of WAPs during each UA interval, divided by the total number of APs. The smaller ratio of WAP means that more APs are switched-off, which implies more energy saving.

Throughput

The throughput is defined as the traffic amount (bits) received successfully per second. Note that the throughput of multicast group is calculated as the average of the throughputs of all multicast users within the group. Unless stated otherwise, the unit of throughput is *Mbps*.

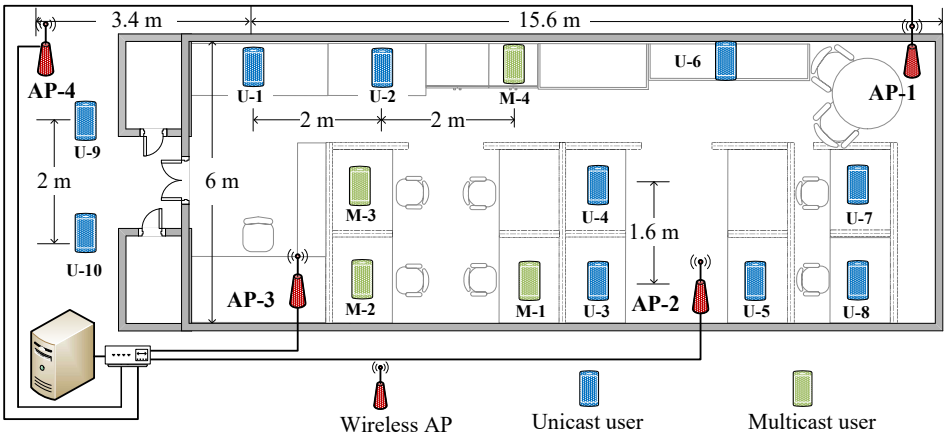


Figure 5.12: Deployment of APs and users for evaluating the proposed UA scheme.

5.2.2 Experiment Settings

Our WLAN testbed consists of three entities, i.e., 802.11n enabled user devices, 802.11n based wireless APs, and a server as centralized controller. Fig. 5.12 shows the deployment of those equipments in our testbed environment. We conducted the experiment in non MU-MIMO system, since most of the current APs and user devices do not support a MU-MIMO function.

The experiment parameters are set as follows: $w_1 = 0.3$, $w_2 = 0.3$, $w_3 = 0.4$, $r_{k,n}^{\text{mcst}} =$ 6.5 Mbps for any multicast group k and any AP n, $R_n^{\max} = 65$ Mbps for any AP n, $\phi_{\min} = 6.5$ Mbps, $\Phi = 0.7$, $T_{\text{UA}} = 60$ s, $T_{\text{arr}} = 2$ s, $T_{\text{CQ}} = 5$ s, and $\delta = 0.5$. The transmit power of AP is set to 7 dBm. Table 5.2 shows mapping of RSS into the data rate where a long guard interval (GI) is assumed to be used in IEEE 802.11n. The MCS-I is the modulation and coding scheme index used for the corresponding RSS.

5.2.3 Experiment Results

We first investigate the performance of the proposed scheme when the distinct channel is assigned to each AP and the UDP traffic is sent to the unicast users. At the beginning

Table 5.2: Conversion of RSS to Data Rate [58].

MCS Index	Min. RSS (dBm)	Data Rate (Mbps)	
		20 MHz Channel Width	40 MHz Channel Width
0	-82	6.50	13.50
1	-79	13.00	27.00
2	-77	19.50	40.50
3	-74	26.00	54.00
4	-70	39.00	81.00
5	-66	52.00	108.00
6	-65	58.50	121.50
7	-64	65.00	135.00

($t = 0$), four multicast users (M-1, M-2, M-3, M-4) belonging to the same group and three unicast users (U-1, U-2, U-3) are deployed. The downlink UDP traffic for the multicast group is generated at the rate of 3 Mbps[1], and each unicast user has downlink UDP traffic with arrival rate of 10 Mbps. Moreover, each unicast user has downlink UDP traffic with arrival rate of 10 Mbps. At $t = 51$ s, seven unicast users (U-4, U-5, U-6, U-7, U-8, U-9, U-10) are additionally activated. At $t = 111$ s, all unicast users are deactivated. We set $T_{UA} = 60$ s so that the UA algorithm is run at $t = 60$ s and $t = 120$ s.

By examining the result in the time range $[0, 50]$ of Fig. 5.13, we can observe that the proposed scheme switches two APs on for serving the given users. During this time range, two unicast users are associated with the AP 2, and the remaining one unicast user and all multicast users are associated with AP 3. Furthermore, the value of LBF seems to be quite high, which implies that the load is well balanced. Note that seven unicast users are newly activated at $t = 51$ s but the UA work with the period of

[1]The mulsticast transmission is usually used for audio or video streaming. According to [59], the bitrate of high-quality MPEG-4 video is 3 Mbps.

60 s is performed at $t = 60$ s. During the time duration $[51, 60]$, since each activated unicast user is associated with its best WAP based on RSS, the offered loads of AP 2 and AP 3 are increased (see Fig. 5.13) and the throughput per WAP is also increased (see Fig. 5.14). At $t = 60$, after executing the UA algorithm with only the existing WAPs, the controller finds that the loads of AP 2 and AP 3 exceed the threshold Φ. Accordingly, the controller switches AP 4 on, and gathers the channel status from all users, and then re-executes the UA. At this point, U-5, U-6, U-7, and U-8 are associated with AP 2; all multicast users, U-2, U-3, and U-4 with AP 3; U-1, U-9, and U-10 with AP 4. This UA result still causes the load of AP 2 and AP 3 to exceed the threshold Φ. Thus, the controller switches AP 1 on and re-executes the UA algorithm after gathering the channel state information from all users. As a result, U-6 and U-7 are associated with AP 1; U-4, U-5, U-8 with AP 2; all multicast users, U-2, and U-3 with AP 3; U-1, U-9, and U-10 with AP 4. Fig. 5.14 shows that the total throughput increases as the users are newly added but the throughput per WAP is firstly increased, and then it is decreased because AP 1 and AP 4 are newly switched on. On the other hand, at $t = 111$ s, all the unicast users are deactivated and the controller executes the UA algorithm at $t = 120$ s. As a result, only a single AP (i.e., AP 3) has been turned on to serve the multicast users and this leads to less energy consumption. The throughput decreases since there are only the multicast users.

Next, we examine the performance of system when the server sends TCP traffic to the unicast users. The other experiment settings are exactly the same as those of Fig. 5.13 and 5.14. Fig. 5.4 shows the offered load, LBF, total throughput, and throughput per WAP. The results are almost similar with those of the previous experiment in Fig. 5.13 and 5.14. This is because the traffic type has little effect on the UA result, thus the association between users and APs in this experiment is the same as those of the previous experiment. However, the slight difference is observed when the offered load is relatively high. During the time range $[60, 110]$, the graphs of Fig. 5.15

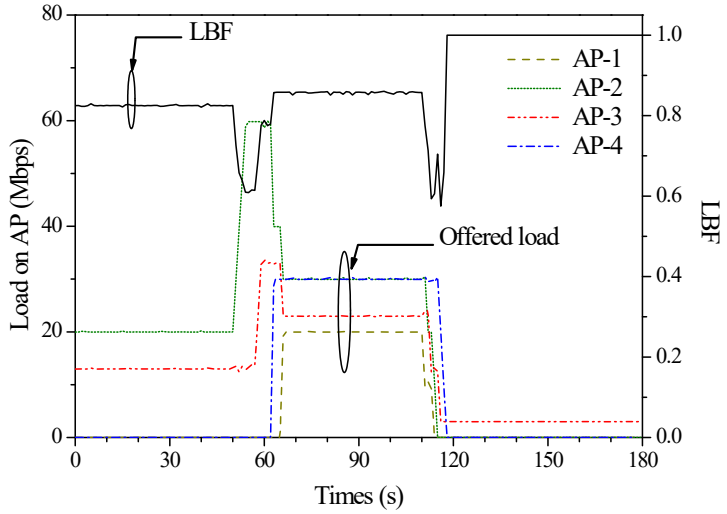


Figure 5.13: Offered load (without channel sharing and UDP traffic).

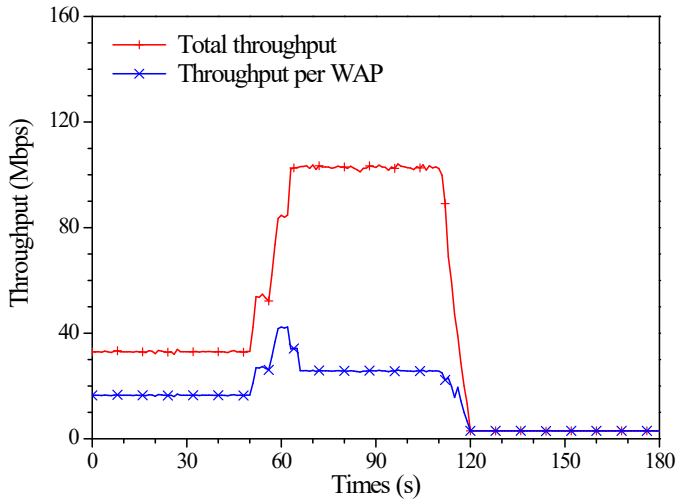


Figure 5.14: Throughput (without channel sharing and UDP traffic).

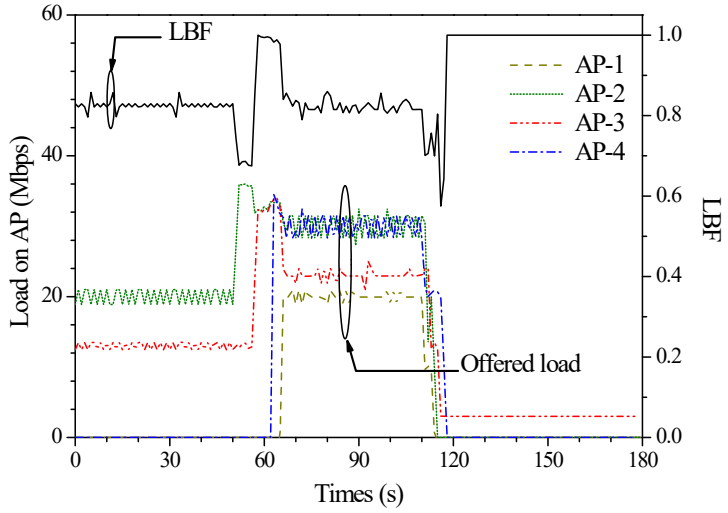


Figure 5.15: Offered load (without channel sharing and TCP traffic).

and 5.16 fluctuate more than the corresponding graphs of Fig. 5.13 and 5.14. This is mainly caused by the congestion control and reliable transmission mechanism of TCP. The TCP ACKs for reliable transmission create uplink traffic from the users, which leads to the contention among AP and users. This contention causes the TCP to adjust its window size dynamically, which affects the offered load on each AP. However, this phenomenon does not reduce the average throughput, because there is still enough resource to serve the offered load.

Now, let us demonstrate the channel sharing effect through experiment. To this end, we assign the same channel for AP 1 and AP 2, whereas AP 3 and AP 4 respectively use channels being different from that of AP 1. All unicast users receive only the UDP traffic from the server, like in Fig. 5.13 and 5.14. The other parameter settings such as the traffic arrival rate of a user and the scenario of user activation/deactivation are the same as those of Fig. 5.13 and 5.14. The results are depicted in Fig. 5.17 and 5.18. During the time range $[0,50]$ in which three unicast users and four users of a single

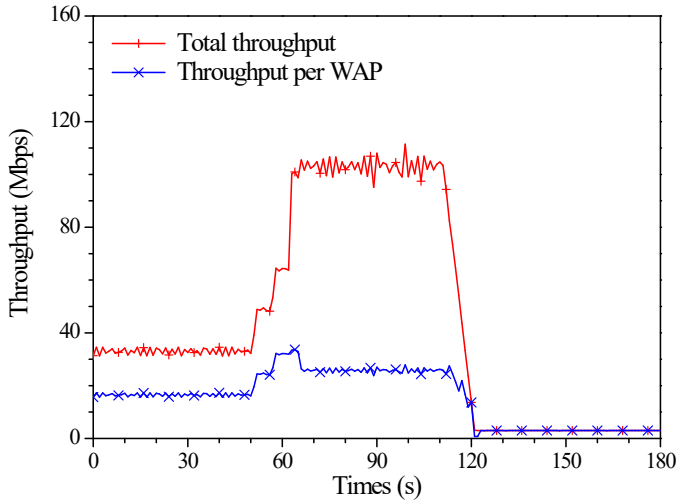


Figure 5.16: Throughput (without channel sharing and the TCP traffic).

multicast group have been deployed, the graph shows a similar trend with Fig. 5.13 and 5.14. Starting at $t = 51$ s, seven other unicast users are additionally activated and at $t = 60$, the UA algorithm is re-executed. Unlike Fig. 5.13 and 5.14, the AP 1 still remains inactive (switch off) state although the loads of AP 2 and AP 3 increase above the threshold. All users are served using AP 2, AP 3, and AP 4. This is because AP 1 and AP 2 share the same channel and thus they should share the air time when being switched on at the same time. Based on the current parameter values, the proposed scheme decides that there is no significant gain achieved by switching-on both AP 1 and AP 2 together. Furthermore, by switching AP 1 off, energy consumption of the entire ESS system can be reduced. On the other hand, at $t = 111$ s, all unicast users are deactivated and only the multicast users are left. At $t = 120$, the UA algorithm is executed. Similarly to the first experiment result in Fig. 5.13 and 5.14, the proposed scheme switches a single AP on (i.e., AP 3) and puts the other APs off, and this leads to less energy consumption for the entire ESS.

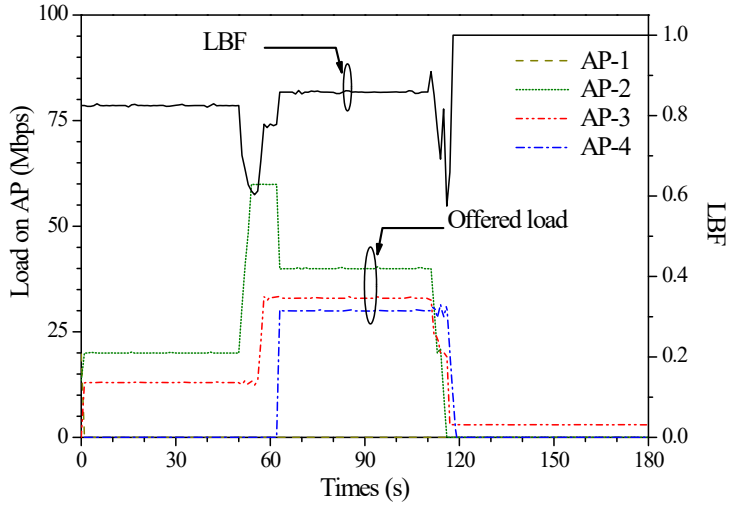


Figure 5.17: Offered load (with channel sharing and UDP traffic).

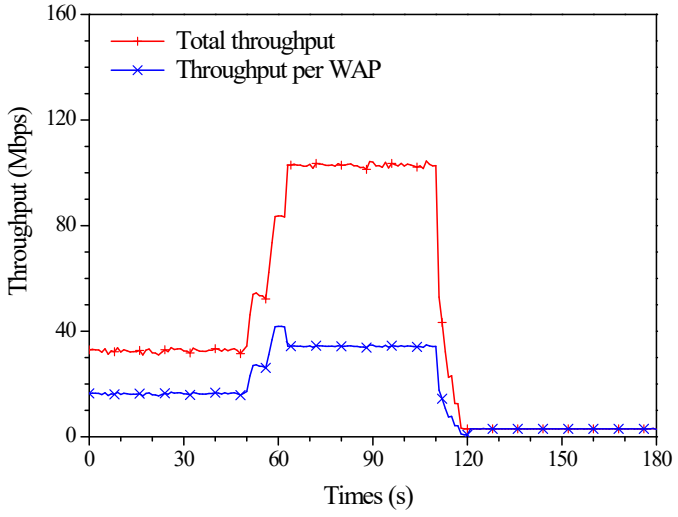


Figure 5.18: Throughput (with channel sharing and UDP traffic).

Fig. 5.19 and 5.20 illustrates the experiment results obtained with the same experiment settings as those of Fig. 5.17 and 5.18, except that all unicast users receive the TCP (not UDP) traffic. As mentioned already, since the traffic type does not affect the UA result, the graphs in Fig. 5.19 and 5.20 have a similar trend with the corresponding graphs in Fig. 5.17 and 5.18. We also have observed this phenomenon between Fig. 5.13–5.14 and Fig. 5.15–5.16. The reason is basically the same. However, we can see much more severe fluctuation in the graphs of Fig. 5.19 and 5.20 compared with those of Fig. 5.15 and 5.16, particularly during the time range $[60, 110]$. This is because, although the system is highly loaded in channel sharing environment, some APs can remain inactive (AP 1 in the experiment of Fig. 5.19 and 5.20) and then each working AP should serve more users. Accordingly, the contention among devices (i.e., AP and users) for sending uplink TCP ACKs as well as downlink traffic gets higher, which makes more traffic to be congested in the AP. This incurs more frequent congestion control events and results in more severe fluctuation. In addition, such high contention leads to the degradation of throughput performance. As observed in Figs. 5(b) and 6(b), under heavy load condition, each WAP gets the lower throughput when sending the TCP traffic (33.7 Mbps), than when sending the UDP traffic (34.3 Mbps).

In summary, the experiment results of Figs. 5.13–5.20 demonstrate that the proposed scheme well adapts to the environmental change, while effectively handling not only the UDP traffic but also the TCP traffic. The APs are switched on/off adaptively according to the system load without deteriorating the performance of WLANs. In addition, the proposed scheme well distributes the traffic load among WAPs.

5.2.4 Simulation Settings

For evaluating the performance of proposed UA scheme under various scenarios and parameter settings while comparing with some existing schemes, we conduct extensive simulations. For the simulation, we have developed a simulator by using MATLAB.

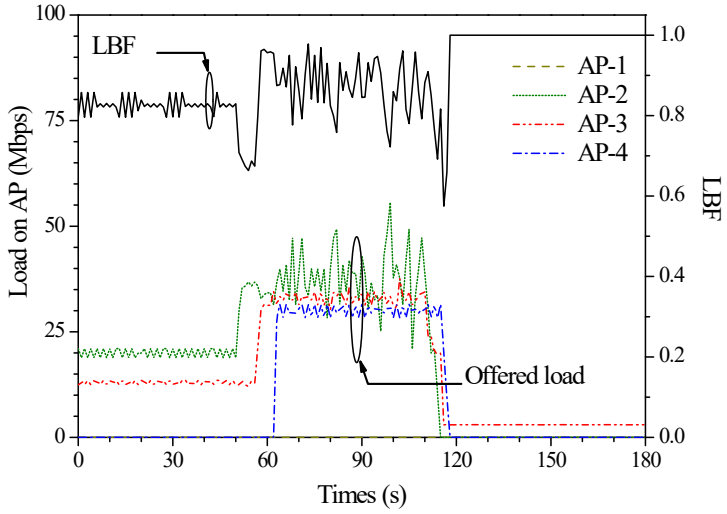


Figure 5.19: Offered load (with channel sharing and TCP traffic).

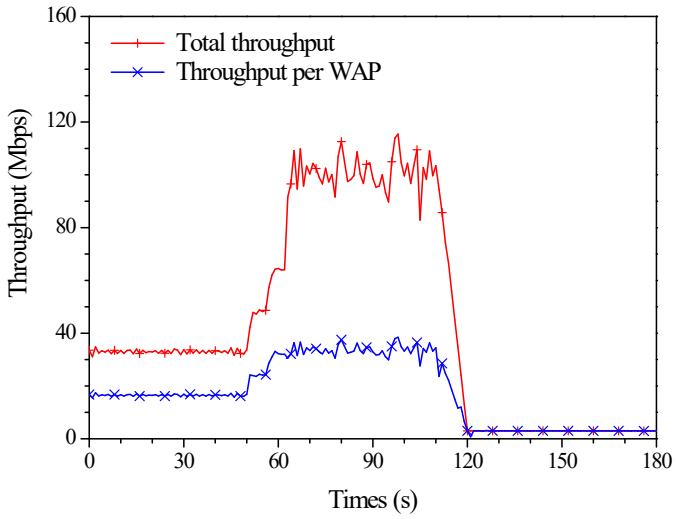


Figure 5.20: Throughput (with channel sharing and TCP traffic).

As a default simulation setting, we deploy four APs, one at each corner of L m $\times$ L m square area, which are numbered in clockwise order. Each AP is assigned a channel with the bandwidth of 20 MHz. The channel model in [60] is used.[2] The transmit power of an AP is 18 dBm and the noise power is set to -90 dBm. The data frame size is set to 1500 bytes and each AP transmits the downlink traffic of users in round-robin fashion. The data transmission rate is decided by referring to Table 5.2, with given RSS value. The physical transmission rate of multicast traffic is set to 6.5 Mbps. A multicast group generates only downlink UDP traffic at the rate of 3 Mbps. A unicast user has uplink UDP traffic as well as downlink UDP traffic. The uplink traffic arrival rate of a unicast user is 1 Mbps and its downlink arrival rate is 5.6 Mbps unless mentioned otherwise. The other parameter values are the same as those in the experiment.

5.2.5 Comparison Schemes

For performance comparison, we select the RSSI-based scheme, the scheme in [22], and the scheme in [28]. In the RSSI-based UA scheme, each user is associated with the AP from which the received beacon signal strength is the strongest. As mentioned already in Introduction, the scheme in [22] intends to maximize the minimum of the MAC efficiencies of users. In simulation, the throughput weight factor is set to 0, like in [22]. The scheme in [28] tries to switch some APs off according to the current traffic load condition. That is, an AP with light load tries to relocate its users to other AP in light or normal load state so that it can go to sleep, and the AP having heavy load relocates its users to the neighbor AP in light or normal load state for load balancing. In simulation, the heavy load and light load thresholds are set to 0.7 and 0.4, respectively, like in [28].

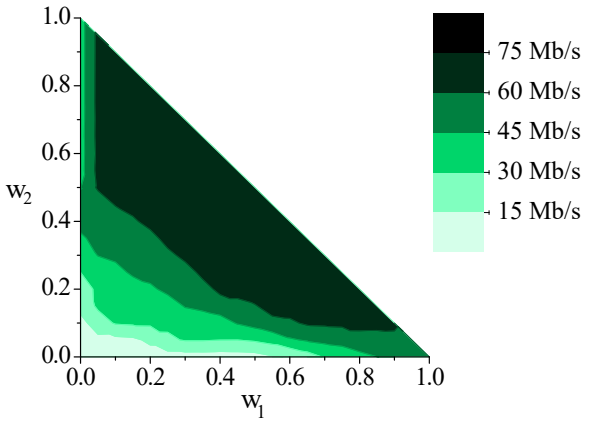
[2] Let d (meters) denote the distance between AP and user. If $d \leq 5$, the path loss is 53.03 dB and the lognormal shadowing has the standard deviation of 3 dB; otherwise, the path loss is $29.57 + 35\log_{10}(d)$ dB and the standard deviation of lognormal shadowing is set to 4 dB.

5.2.6 Simulation Results

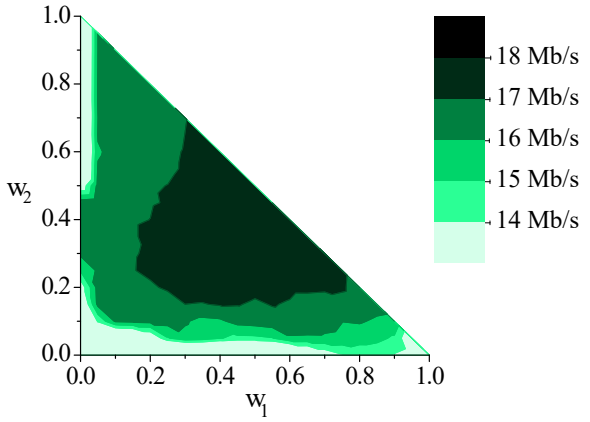
Fig. 5.21 shows the influence that the weights w_1, w_2, and w_3 ($w_1 + w_2 + w_3 = 1$) of our UA objective function have on the performance of the proposed scheme. Remind that the weight w_1 is related to throughput, w_2 is the load balancing weight, and w_3 is the weight for using the fewer WAPs. The size parameter L of service area is set to 80 m, and twelve unicast users and eight users of a single multicast group are randomly located within the service area. All APs are assigned different channels without channel sharing, and the transmission range of each AP covers the entire service area, modeling a hot spot area of real environment.

As shown in Fig. 5.21(a), the total system throughput increases with larger w_1 and/or w_2, which means the smaller w_3. The smaller w_3 implies more WAPs and it naturally results in the higher system throughput. The throughput per WAP is high when both w_1 and w_2 are relatively high, as shown in Fig. 5.21(b). When only the w_1 is high, the proposed scheme behaves like the RSSI-based scheme, where a user individually tries to associate with the AP of highest RSS. This causes some APs to be excessively underloaded, which decreases the throughput per WAP. If only the w_2 is high, the multicast users are distributed to all APs, which aggravates the performance anomaly effect and, as a result, decreases the throughput per WAP. When both w_1 and w_2 are low, WAPs are too few to accommodate traffic load. On the other hand, Fig. 5.21(c) shows that the load balancing factor has a high value when the number of WAPs is small with a high w_3. It is because few WAPs make a sufficiently large amount of traffic load be assigned to each WAP so that the difference between the traffic loads of APs is small. Based on all these results, we suggest the range value of $0.3 \sim 0.4$ for w_1 and w_2, for a reasonable good performance. In the following all simulations of non MU-MIMO system, we set $w_1 = 0.3$, $w_2 = 0.3$, and $w_3 = 0.4$.

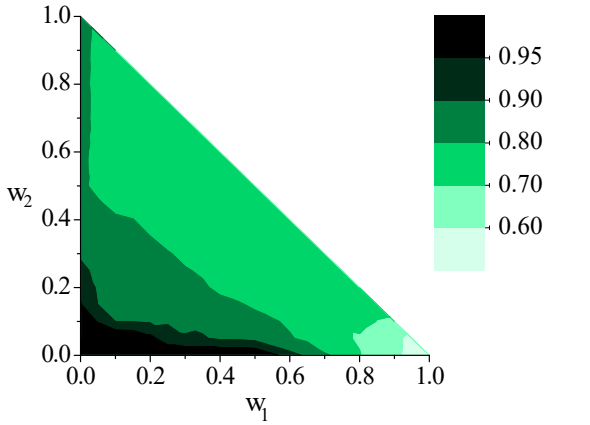
Next, we examine the effect that the number of unicast users has on the perfor-



(a) Total throughput (Mbps)



(b) Throughput per WAP (Mbps)

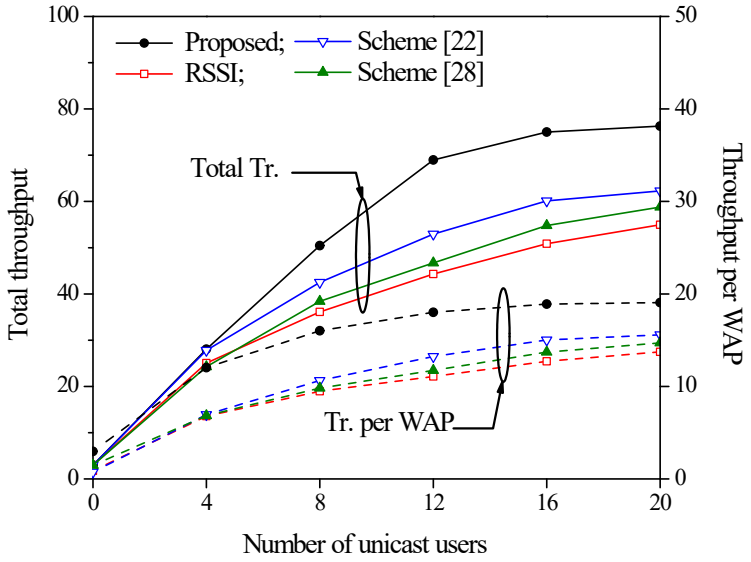


(c) Load balancing factor

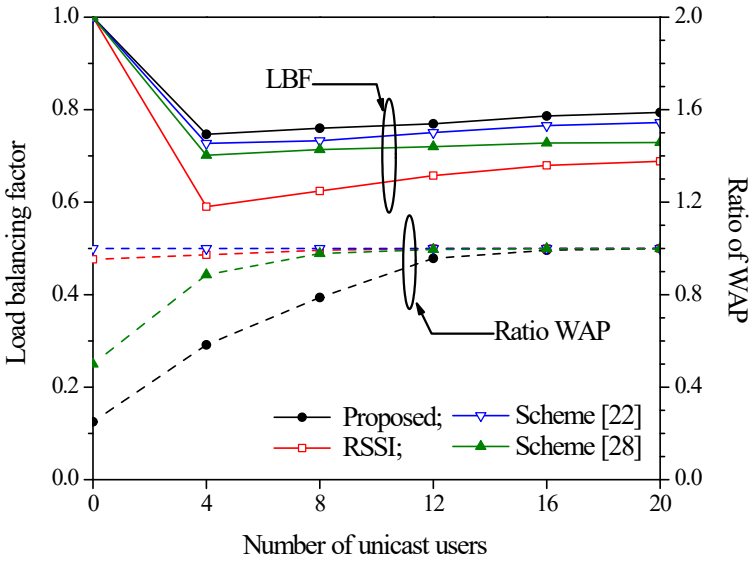
Figure 5.21: Effect of weight combinations on performance (non MU-MIMO system without channel sharing).

mance, in Fig. 5.22. The other parameter settings are the same as those of Fig. 5.21. As shown in Fig. 5.22(a), the proposed scheme achieves much higher total throughput than the other schemes. This is because the proposed scheme tries to assign the multicast users to as few APs as possible, whereas the comparison schemes distribute multicast users to almost all APs without consideration for the multicast group and it incurs the performance anomaly phenomenon which reduces the achievable throughput. In addition, since the proposed scheme controls the number of WAPs adaptively in proportion to the system load, it outperforms the other schemes in the throughput per WAP (Fig. 5.22(a)). Furthermore, as shown in Fig. 5.22(b), the loads of WAPs are also well balanced under the proposed scheme. Especially under low load, the proposed scheme can save energy greatly without deteriorating the performance because of being able to serve all users with few WAPs (Fig. 5.22(b)). On the other hand, the scheme [22] provides better performance compared with the scheme [28], because the scheme [28] does not appropriately assess the load status of AP by not considering the instantaneous channel condition of user when estimating the saturation throughput.

In Fig. 5.23, we examine the effect of channel sharing among the APs in proximity where the AP 1 and AP 3 share the same channel, whereas the AP 2 and the AP 4 use distinct channels with other APs. The other simulation settings are the same as those of Fig. 5.22. We can observe that the results for channel sharing in Fig. 5.23 have a very similar tendency to the results for no channel sharing in Fig. 5.22. However, the throughput difference between the proposed scheme and the other schemes is larger rather than that without channel sharing of Fig. 5.22. This is because the proposed scheme prefers to switch-off one among two APs using the same channel for performance improvement, as compared with the other schemes that do not care channel sharing. By comparing Fig. 5.22(b) and Fig. 5.23(b), we can observe this feature of the proposed scheme, i.e., the proposed scheme takes the policy of utilizing the fewer WAPs under channel sharing situation.

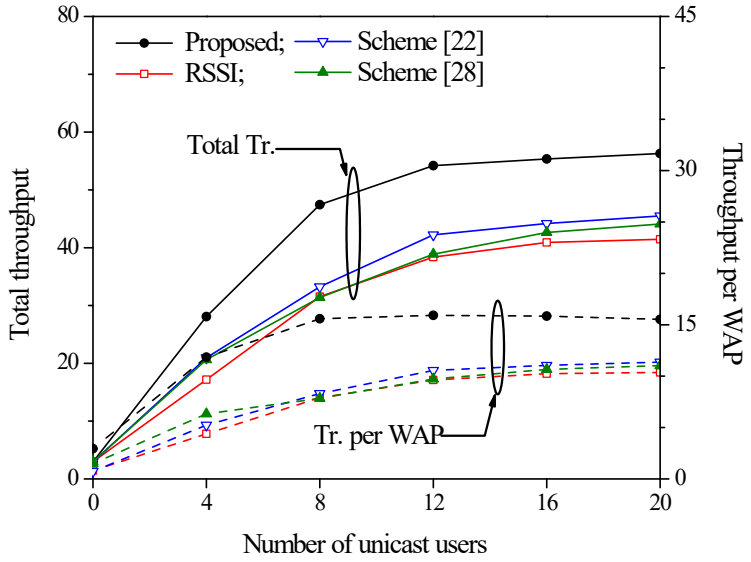


(a) Throughput

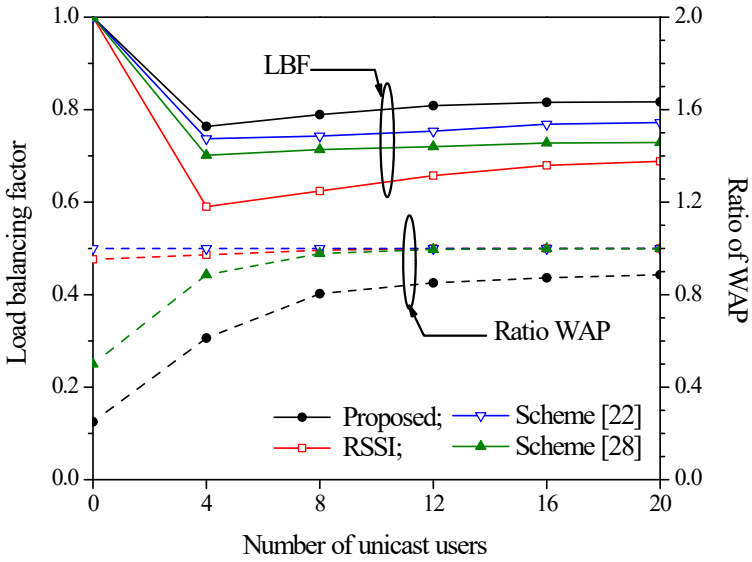


(b) LBF and ratio of WAP

Figure 5.22: Performance according to the number of unicast users in non MU-MIMO system without channel sharing.



(a) Throughput

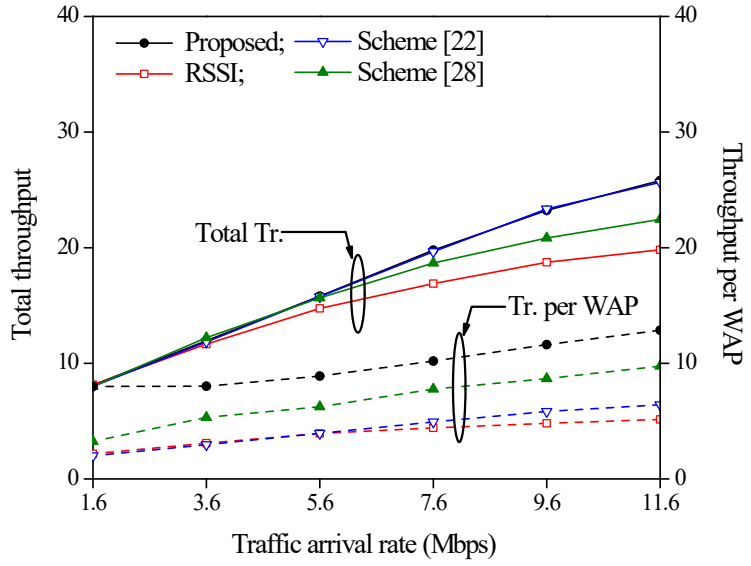


(b) LBF and ratio of WAP

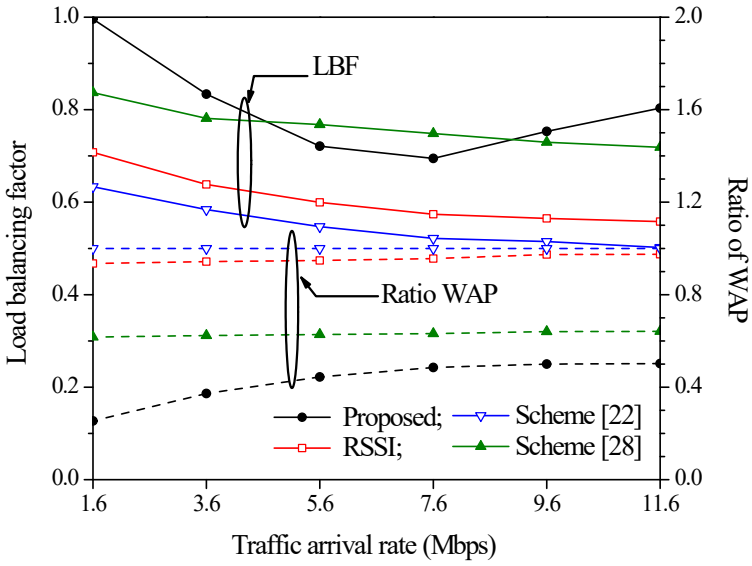
Figure 5.23: Performance according to the number of unicast users in non MU-MIMO system with channel sharing.

Fig. 5.24 depicts the performances of UA schemes according to the downlink traffic arrival rate of a unicast user. Two unicast users and a single multicast group of eight users are deployed in the service area, and the other settings are the same as those in Fig. 5.22. As shown in Fig. 5.24(a), the proposed scheme outperforms the other comparison schemes in throughout, especially in the throughput per WAP. This is because the proposed scheme determines the number of WAPs adaptively to given load. Accordingly, the proposed scheme has much fewer WAPs, i.e., the lower ratio of WAPs, as compared with the other schemes (see Fig. 5.24(b)). On the other hand, as also depicted in Fig. 5.24(b), the load balancing factors of all UA schemes decrease as the traffic arrival rate of a unicast user increases. It is because, with the fixed number of unicast users, increasing their traffic arrival rates causes the larger load difference between WAPs. The proposed scheme and the scheme [28] have higher LBF than the RSSI-based scheme and the scheme [22], because both of them switch some APs off according to system load and this reduces the load difference between WAPs. When the traffic arrival rate of a unicast user is relatively small, since the proposed scheme tries to assign the multicast users to one AP and two unicast users to another AP, it has the lower LBF than the scheme [28] which distributes the multicast users to several APs, but when the traffic arrival rate is relatively large, it ends in assigning unicast users into different APs while leading to the improved load balancing among WAPs.

Fig. 5.25 depicts the performances of UA schemes according to the number of multicast groups, where the number of users per multicast group is fixed to 8 and there are two unicast users. The other simulation settings are the same as those in Fig. 5.22. The effect of adding a new multicast group in the proposed scheme is similar to the effect of adding a new unicast user, shown in Fig. 5.22. The total throughput of the proposed scheme increases with more multicast groups, but the total throughput in other schemes slightly decreases as the number of multicast groups increases (see Fig. 5.25(a)). The reason is as follows. Since the comparison schemes treat the

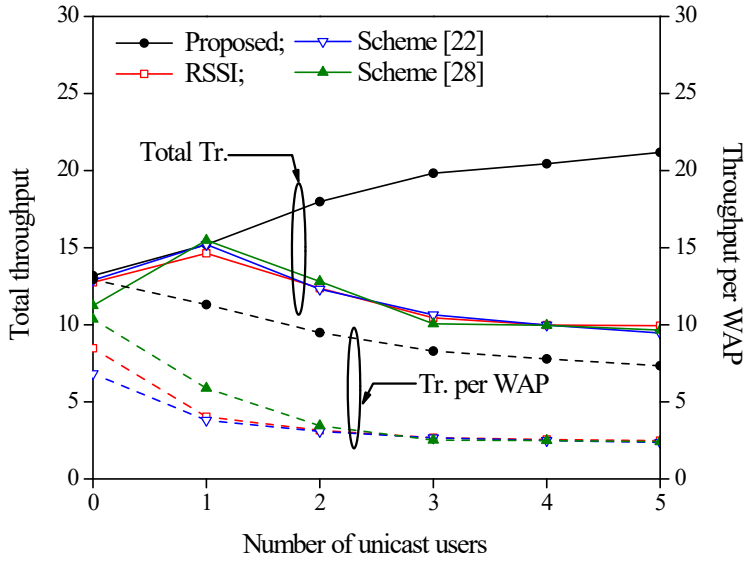


(a) Throughput

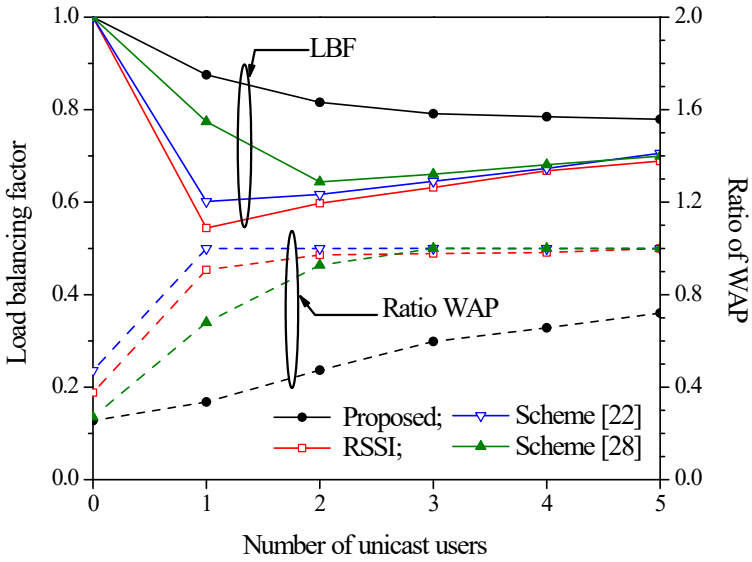


(b) LBF and ratio of WAP

Figure 5.24: Performance according to the downlink traffic arrival rate of a unicast user in non MU-MIMO system without channel sharing.



(a) Throughput



(b) LBF and ratio of WAP

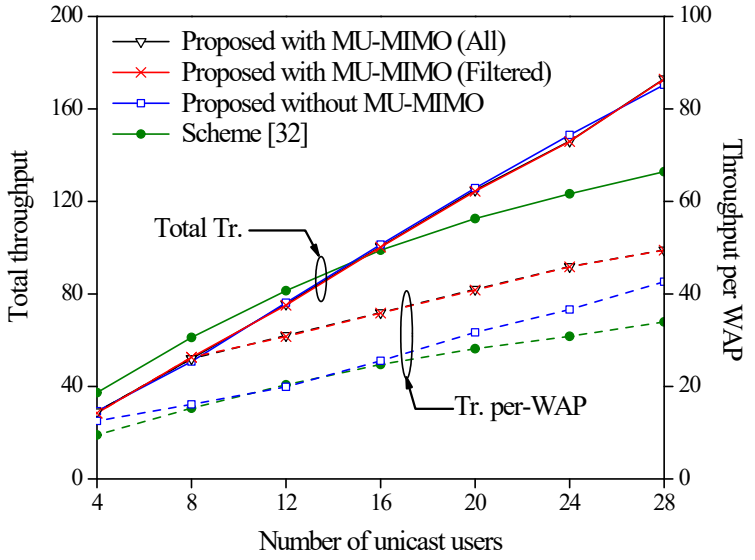
Figure 5.25: Performance according to the number of multicast groups in non MU-MIMO system without channel sharing.

multicast user individually (not as a group), the multicast traffic from each group is injected to almost all APs. Accordingly, the performance anomaly phenomenon caused by the multicast traffic gets worse and, in turn, this leads to the decrease of both total throughput and throughput per WAP and the activation of more APs. Thus, as shown in Fig. 5.25(b), the ESS under the proposed scheme consumes less energy with fewer WAPs in comparison to the other schemes, in serving all users.

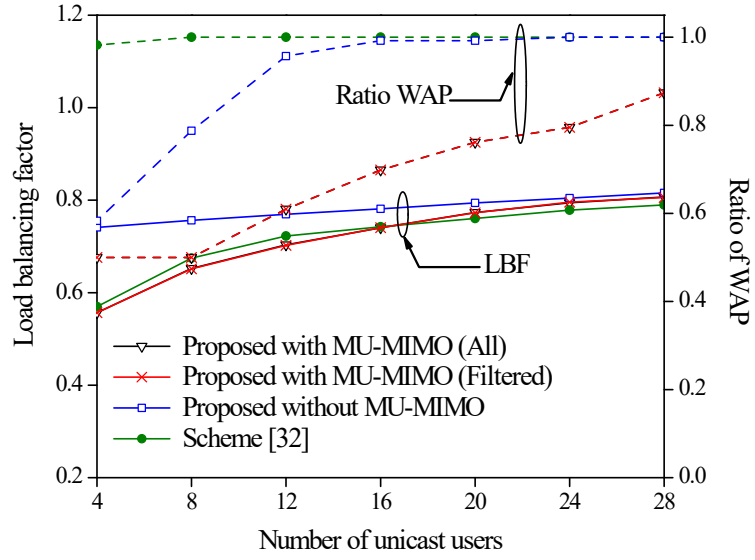
5.2.7 Simulation for MU-MIMO System

We assess the performance of the proposed UA scheme in the system where MU-MIMO technique is applied to downlink traffic transmission of unicast users. In simulation, all APs are assumed to be equipped with four transmit antennas, and the weights are set as $w_1 = 0.4$, $w_2 = 0.4$, $w_3 = 0.2$. The other settings except these weights are the same as those of Fig. 5.22. For performance comparison, we select the scheme in [32], which performs the UA while taking account of channel correlation between users for exploiting MU-MIMO. In addition, to assess the performance degradation effect due to filtered MU-MIMO grouping in Algorithm 1, we compare the proposed UA scheme which considers all feasible MU-MIMO groups, i.e., $\mathbb{M}$, without Algorithm 1. The performance of the proposed scheme performing UA without considering MU-MIMO is also presented. For this case, the weights are the same as those of Fig. 5.22. We modified our MATLAB simulation code so as to support MU-MIMO transmission by utilizing the code provided by [61].

The comparison results are depicted in Fig. 5.26. Note that, irrespective of the used UA schemes, whenever transmitting downlink unicast traffic, each AP newly forms MU-MIMO groups and calculates their beamforming vectors, depending on the channel correlation between its all associated unicast users. As shown in Fig. 5.26, the performance difference between the proposed UA scheme based on filtered MU-MIMO groups (Algorithm 3) and the proposed scheme considering all groups is not



(a) Throughput



(b) LBF and ratio of WAP

Figure 5.26: Performance according to the number of unicast users in MU-MIMO system without channel sharing.

noticeable. This confirms the effectiveness of our "Generating MU-MIMO Candidate Groups" algorithm (Algorithm 3). On the other hand, under the proposed UA scheme considering MU-MIMO, most unicast users are concentrated to some APs for exploiting simultaneous transmission. Accordingly, the UA scheme considering MU-MIMO can put more APs into sleep (Fig. 5.26(b)) and this results in much higher throughput per WAP while saving more energy without deteriorating the total throughput performance (Fig. 5.26(a)). Meanwhile, in the scheme performing UA without consideration of MU-MIMO, since the association is treated individually for each unicast user, the load can be more well balanced among WAPs (Fig. 5.26(b)). However, as the number of unicast users increases, the unicast users are distributed to more APs although the association of unicast users is done in group manner. As a result, the performance difference between the proposed UA schemes in LBF is decreased with more users (Fig. 5.26(b)).

On the other hand, as shown in Fig. 5.26(b), the scheme in [32] has similar LBF performance to the proposed UA schemes considering MU-MIMO, because both of them deal with the association of unicast users in group manner. However, the scheme in [32] does not treat the association of multicast users in group manner. Thus, since the multicast users of the same group are usually distributed to several APs, more APs are switched on and this results in the lower throughput per WAP and more energy consumption for the entire ESS.

In summary, the comparison results from extensive simulation in Figs. 5.22 – 5.26 demonstrates that the proposed UA scheme achieves remarkably outstanding performance while well adapting to various enterprise WLAN environments, i.e., coexistence of multicast and unicast traffic, dramatic variation over time in the number of users and traffic load, and supporting for MU-MIMO and channel sharing.

5.3 A Graph-Based Handover Scheduling for Heterogenous Vehicular Networks

5.3.1 Performance Metrics

The following metrics are used for performance evaluation.

- **Number of HO events** is the total number of L2 and L3 HO events performed by a user moving along the roads during its travel time (i.e., a simulation run time).
- **Ratio of effective connection time** is defined as the ratio of the effective data transmission time excluding the total time taken to finish all of the HO events, for the whole travel time.
- **Fairness index** is calculated using Jain's equation for the total throughput of AP.

$$\mathcal{F} = \frac{\left(\sum_{i=1}^{M} X_i\right)^2}{M \cdot \sum_{i=1}^{M} (X_i)^2}, \tag{5.3}$$

  where M is the total number of APs and X_i is the total throughput of AP-i during the whole travel time. The higher fairness index implies the smaller differences between APs in their throughputs. In other words, the APs are utilized more fairly.
- **Packet drop ratio** is the ratio of the packets dropped by a user during its travel time due to HO events.
- **Average throughput** refers to the average throughput per user, including uplink and downlink one.

5.3.2 Simulation Settings

We consider a particular region in the Manhattan city (red-marked area) as shown in Fig. 5.27 whose size is approximately 1000 m x 1000 m. The detail information for the

Figure 5.27: Simulation environment: Manhattan area.

roads on this area is got from OpenStreetMap [65]. The default number of vehicles is 300. We utilize the SUMO tool [66] to generate the traffic model of the vehicles inside the red-marked area. The moving speed of a vehicle is uniformly distributed in the range of $0 \sim 50$ km/h. Furthermore, we set 8 blocks within the area to be clean blocks, in which there is no installed AP. Then, 600 APs are randomly deployed according to a uniform distribution within the area, excluding the clean blocks. It creates several road segments which are not fully covered by the WiFi. On the other hand, we deploy five LTE BSs with the service radius of 500 m (one on each corner and one in the middle of the red-marked area). Thus, the LTE network covers all the road segments.

While moving along a road, the vehicles are connected to PoAs for wireless Internet access. During simulation, downlink UDP traffic with the default arrival rate of 4 Mbps is sent to a user. At the same time, the user also sends uplink UDP traffic whose arrival rate is 0.1 Mbps. The APs within the same block are assigned into the same subnet with the probability of 0.8. Thus, the L2 HO takes place only between APs in the same block which belong to the same subnet. The L2 HO execution steps include a single channel scanning, authentication and association process, which take 80 ms to

finish. The L3 HO includes L2 HO with an additional DHCP process and is assumed to takes 2 s to finish.[3] Accordingly, $T = 2$ s in (4.21). In addition, the maximum allowable number of users at each AP is set to 10, i.e., $N_b = 10$, $\forall b \in \mathcal{A}$ in (4.21). All packets in the buffer are assumed to be dropped when a user executes the HO.

In the simulation, we consider the IEEE 802.11n for WiFi, of which the parameters are presented in Table 5.3. The APs in proximity are assigned a different operating channel. Thus, co-channel interference between the neighbor APs is assumed to be negligible. For the channel model, we consider pathloss, lognormal shadowing, and multipath fading. The path loss is $33.3 + 36.7 \log_{10} d$ (dB), where d (meters) is the distance between AP and user [67]. The standard deviation of shadowing is 8 dB and the multipath fading follows Jake's spectrum model [68]. The average service radius of WiFi, R_a, is approximately 80 m [69]. For the performance comparison with the proposed scheme, we consider the handover schemes in [36] and [39]. Similarly to the proposed scheme, in the comparison schemes, a user connects to the BS only when there is no available AP. Note that since all schemes access the cellular network in the same manner, the performance differences among three schemes are merely originated from WiFi. Therefore, for the purpose of performance comparison, we measure only the performance of WiFi. The total simulation time is 1100 s.

5.3.3 Simulation Results

Table 5.4 shows the average number of HO events according to the HO type, and the ratio of effective connection time. The default parameter settings are used. We fist discuss the effect of α on the performance of the proposed scheme. Note that α is a

[3] Actually, these L2 and L3 HO latency values of 80 ms and 2 s respectively were derived through experiment in the real vehicular network environment. The experiment settings were the same with those in [11]. To measure the L2 HO latency, we set the APs to be in the same subnet, thus there was no DHCP process required during the HO. On the other hand, for the L3 HO, we set the APs to be in different subnet and enabled the DHCP server on each AP. The HO was triggered when the RSSI of the current serving AP was equal or less than -80 dBm.

Table 5.3: Simulation parameters for WiFi.

Parameters	**Value**
Channel bandwidth	20 MHz
Tx power	20 dBm
Minimum RSSI threshold	-82 dBm
HO RSSI threshold	-80 dBm
Data frame length	1500 bytes
ACK frame length	14 bytes
MAC header	38 bytes
Slot time	9 μs
SIFS time	16 μs
DIFS time	34 μs

weight of HO latency (see (4.21)). Thus, with the larger α, the proposed scheme tries to reduce the HO latency by preferring the L2 HO with shorter latency compared to the L3 HO, and this allows a user to get the longer effective connection time. Accordingly, in the proposed scheme, as the α gets larger, the average number of the L2 HO events and the ratio of effective connection time increase.

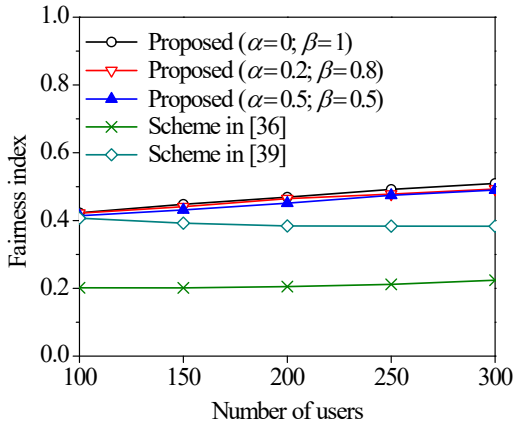
On the other hand, we can observe in Table 5.4 that the proposed scheme generates the fewer HO events than the schemes in [36] and [39]. It implies that the proposed scheme more accurately selects the target APs. The scheme in [36] selects a target AP, based on the instant information regarding the current network status sent by the serving AP. This information may not be accurate because of time delay, when it is received by the users. Furthermore, the users in a certain location at a particular time may receive the same information, which causes them to select the same inappropriate AP for HO. The scheme in [39] merely tries to maximize the sojourn time of a user within the AP without considering the HO latency between APs. Thus, it results in the fewer HO events compared to the scheme in [36]. However, since its target PoA selection highly depends on the record of previous moving direction which may not

Table 5.4: Comparison of the average number of HO events and the ratio of effective connection time.

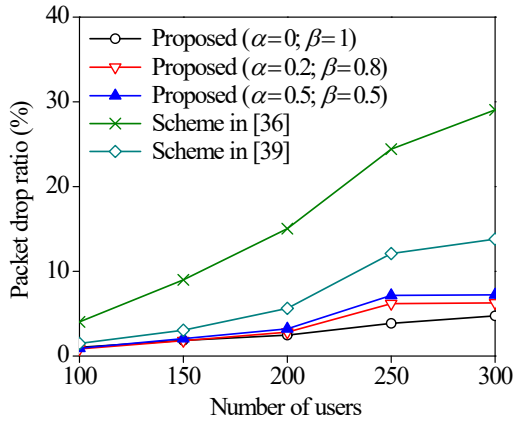
Scheme	**Average Number of HO Events (L2, L3)**	**Ratio of Effective Connection Time**
Proposed ($\alpha = 0$)	(16.44, 70.15)	86.80 %
Proposed ($\alpha = 0.2$)	(35.30, 51.98)	89.97 %
Proposed ($\alpha = 0.5$)	(35.49, 51.78)	90.00 %
Scheme in [36]	(48.87, 89.66)	83.01 %
Scheme in [39]	(23.18, 72.53)	86.32 %

accurately represent the future moving direction, it may still make an inaccurate AP selection.

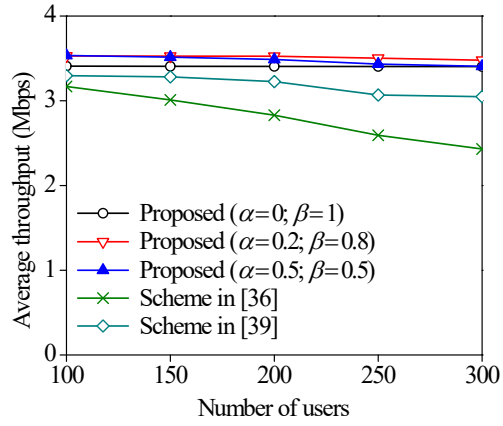
Figs. 5.28 and 5.29 depict the fairness index, the packet drop ratio, and the average throughput, according to the number of users and traffic arrival rate, respectively. First, we discuss the influence of β on the performance of the proposed scheme ($\alpha + \beta = 1$). Note that β is a weight parameter for preferring the AP with fewer associated users, in selecting a HO target AP. Accordingly, as shown in Figs. 5.28(a) and 5.29(a), it is obvious that the highest fairness index is achieved when $\beta = 1$, because distributing the users among APs gets the highest priority when $\beta = 1$. Since the contention among users gets severer with the smaller β, the number of packets in the buffer increases as β decreases. As a result, with the smaller β, the more packets are dropped during the handover operation and the packet drop ratio increases (see Figs. 5.28(b) and 5.29(b)). Meanwhile, in Figs. 5.28(c) and 5.29(c), we observe that the throughput of the proposed scheme slightly decreases when α is zero or too high. The reason is that when $\alpha = 0$, the effective connection time ratio decreases (see Table 5.4) and this causes the decrease of user throughput. A too high α also incurs the high contention and leads to the decrease of throughput. From the above simulation results, it is observed that the proposed scheme achieves better performance when α is not zero while being much



(a) Average fairness index



(b) Average dropped packet of the user



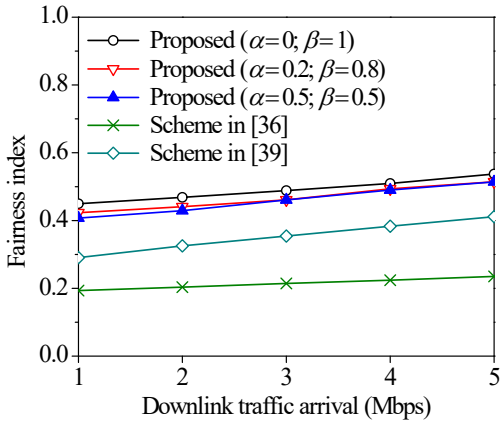
(c) Average throughput of the user

Figure 5.28: Performance of the system for various number of users.

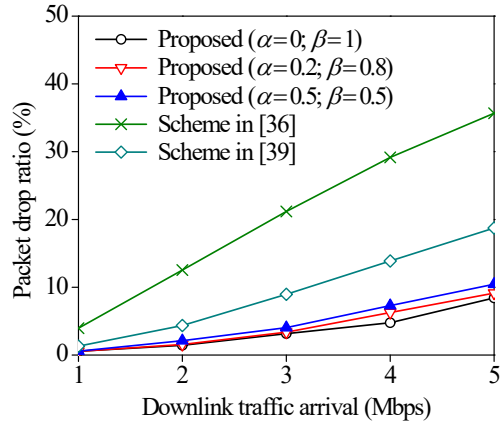
smaller than β. Next, let us compare the performances of the proposed scheme and the existing schemes. We firstly discuss in Fig. 5.28(a) the throughput fairness performance according to the number of users. In general, the fairness index increases with more users due to averaging effect, because the users are well distributed among the available road segments and more APs have a chance to serve the users. However, regardless of this condition, the fairness index of the scheme in [39] slightly decreases with more users. The reason is that the scheme in [39] merely selects a target AP to maximize the sojourn time of the user within the AP. Thus, since the users moving on the same road segment have a tendency to select the same target APs, as the number of users increases, the users are concentrated to only some particular APs in each road segment, while degrading the fairness performance. On the other hand, as already explained in Table 5.4, the scheme in [36] may select a target AP based on old information. We observed in simulation that many APs are left unused. As a result, the scheme in [36] has the worst performance among the schemes in throughput fairness.

Fig. 5.28(b) shows the packet drop ratio. Since the proposed scheme aims at reducing the HO latency while preferring the APs with fewer associated users, the users under the proposed scheme experience the short HO period and less contention among users within AP. As a result, the proposed scheme has the lower packet drop ratio than the schemes in [36] and [39]. Also, as we can observe in Fig. 5.28(c), the features of the proposed scheme, such as the less contention, longer effective connection time, and lower packet drop ratio, provide the higher throughput to a user, compared to the schemes in [36] and [39].

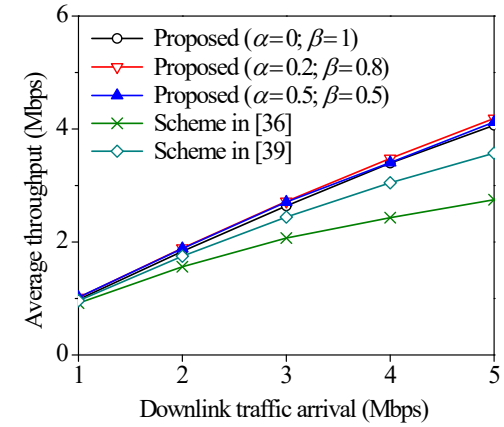
Next, we examine in Fig. 5.29 the performances of three schemes according to downlink traffic load. With the higher traffic arrival rate, the fairness in all schemes is improved because the throughput difference between APs decreases (see Fig. 5.29(a)). In addition, as the downlink arrival rate of a user increases, it is natural that the packet drop ratios of all schemes get higher because each user has more packets in the buffer



(a) Average fairness index

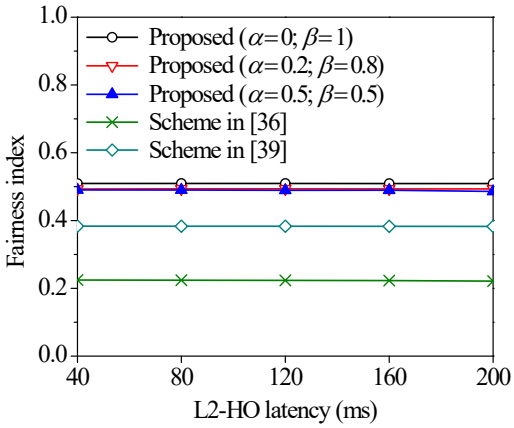


(b) Average dropped packet of the user

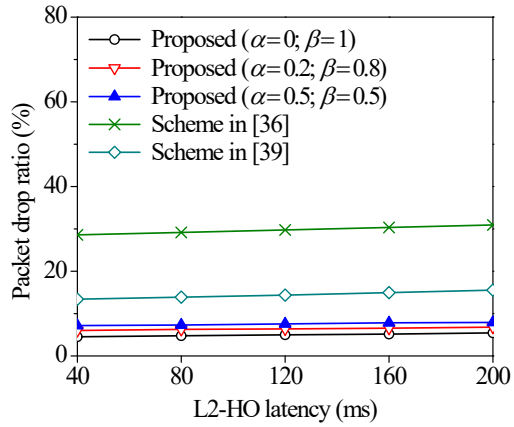


(c) Average throughput of the user

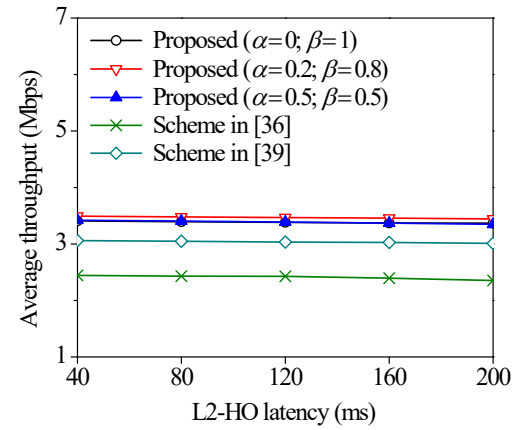
Figure 5.29: Performance of the system for various downlink traffic arrival rate.



(a) Average fairness index



(b) Average dropped packet of the user



(c) Average throughput of the user

Figure 5.30: Performance of the system for various L2-HO latency.

(see Fig. 5.29(b)), and the total throughput of user increases because of its increased downlink throughput (see Fig. 5.29(c)). On the other hand, we can observe in Fig. 5.29 that the proposed scheme achieves much better performance in all of fairness, packet drop ratio, and the user throughput, than the schemes in [36] and [39]. As already described in Fig. 5.28, it is caused by the desirable features of the proposed scheme, which are the less contention, longer effective connection time, and shorter HO latency.

Finally, we present the performance of the system for various L2-HO latency in Fig. 5.30. It obvious that the packet drop ratio increases and throughput decrease when L2-HO latency increase. This is because a longer L2-HO latency will reduce the effective connection time which increases the connection loss. It causes more packets to be dropped during HO and the throughput will decrease. Meanwhile, the fairness index does not change significantly when the L2-HO latency is varied between 40 ms $\sim$ 200 ms. Overall, the simulation results also show that the proposed scheme outperforms the comparison schemes. The reason is exactly the same with that of in Figs. 5.28 and 5.29. This confirms the advantage of the proposed HO scheme.

Chapter 6

Conculsion

In this dissertation, we have proposed resource management schemes for improving the performance of enterprise WLANs. The proposed resource management schemes include a centralized channelization and AP selection schemes also known as UA schemes. The centralized channelization scheme exploits the channel bonding to improve the throughput. The channelization scheme consists of three main steps, i.e., interference graph generation, channel allocation, and primary channel selection. The interference graph is built by considering direct and hidden interference relation. Thus, it gives a more accurate prediction regarding the effect that an AP can cause to the other APs. Furthermore, the channel allocation is dedicated to maximizing the total network throughput by allocating a wider channel to the AP which has a relatively small total number of interference sources. Finally, the primary channel is selected to maximize the channel utilization.

On the other hand, the UA scheme for the enterprise environment (quasi-static mobility) targets simultaneously (1) high throughput performance, (2) load balancing among working APs, and (3) the energy efficiency enhancement by AP sleeping, un-

der complicated environment where time-division and space-division multiple access techniques are considered together with channel sharing and MU-MIMO and the multicast traffic as well as unicast traffic is supported. We have designed the UA scheme by using an optimization technique. Regardless the complexity of the problem, our develop technique transforms the problem into solvable problem so that the solution can be efficiently derived. Even though the optimization approach itself is a standard one, since we took account of all technical issues mentioned above, which can be raised in real enterprise WLANs, the proposed scheme can be a very *practical* solution.

Our implementation and experiment of the proposed channelization and UA scheme demonstrate the feasibility and practicality of the scheme. The experiment results reveal that the proposed channelization scheme successfully exploits the channel bonding even in dense WLAN environment. Thus, the total network throughput can be increased. On the other hand, the experiment results also demonstrate that the proposed UA scheme well adapts to the dynamic change of the offered load on WLAN. Moreover, the extensive simulations show that the proposed UA scheme is superior to the existing schemes in almost all performance measures including the system throughput, the load balancing factor, and the energy efficiency (i.e., the ratio of WAP). In summary, since the proposed scheme well adapts to various practical WLAN environments and significantly improve the performance, the scheme can be used in realizing the enterprise WLANs with high performance-to-cost ratio.

Finally, for vehicular environments, we have also proposed a graph-based HO scheduling scheme. The proposed scheme exploits the characteristic that the movement of a user is predictable with high accuracy on the basis of road topology. It enables the computation of the HO schedule by a central server in advance considering only the information about the network configurations and road topologies. Using the computed HO schedule, the user only needs to decide when to initiate the connection to the next PoA while moving on the road segment. This technique realizes a very

fast and efficient HO in the vehicular environment. Moreover, since the HO scheduling is conducted by the central server, the proposed scheme may take advantage of easily utilizing abundant computing and storage resources in the cloud, and the users are free from the burden of calculating their HO schedules. Regardless of its centralized feature, the proposed scheme introduces very low communication overhead. The simulation results also show that the proposed scheme outperforms the comparison schemes on all performance metrics. Therefore, we believe that the proposed scheme can help strengthening the position of WiFi as an alternative technology of cellular network for providing a low-cost Internet access in the vehicular environment.

Bibliography

[1] IEEE, "Enhancement for very high throughput operation in bands below 6 GHz," *IEEE standard amendement 802.11ac,* 2011.

[2] S. Jang and S. Bahk, "A channel allocation algorithm for reducing the channel sensing/reserving asymmetry in 802.11ac networks," *IEEE Trans. Mob. Comput.,* vol. 14, no. 3, pp. 458–472, Mar. 2015.

[3] O. Ercetin, "Association games in IEEE 802.11 wireless local area networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 5136–5143, Dec. 2008.

[4] W. Na, N. N. Dao, and S. Cho, "Mitigating WiFi interference to improve throughput for in-vehicle infotainment networks," *IEEE Wireless Commun.*, vol. 23, no. 1, Feb. 2016, pp. 22-–28.

[5] N. Lu, N. Cheng, N. Zhang, X. Shen, J. W. Mark, F. Bai, "Wi-Fi hotspot at signalized intersection: Cost-Effectiveness for vehicular Internet access," *IEEE Trans. Veh. Technol.*, vol. 65, no. 5, May 2016, pp. 3506–3518.

[6] "Wigle Net", [Online]. https://wigle.net/. [Accessed: 2017-07-11].

[7] K. Grochia, "LTE or WiFi? Client-side Internet link selection for smartphones," in *Proc. Int. Conf. Compt. Net.*, Brunow, Poland, May 2015.

[8] P. Lv, X. Wang, X. Xue, and M. Xu, "Swimming: Seamless and efficient WiFi-based Internet access from moving vehicles," *IEEE Trans. Mobile Comput.*, vol. 14, no. 5, May 2015, pp. 1085–1097.

[9] R. Mahajan, J. Zahorjan, and B. Zill, "Understanding WiFi-based connectivity from moving vehicles," in *Proc. 7th ACM SIGCOMM IMC '07*, California, USA, Oct. 2007.

[10] W. Xu, H. A. Omar, W. Zhuang, X. Shen, "Delay analysis of in-vehicle Internet access via on-road WiFi access points," *IEEE Access*, vol. 5, Feb. 2017, pp. 2736–2746.

[11] M. H. Dwijaksara, M. Hwang, W. S. Jeon, and D. G. Jeong, "Design and implementation of a fast handoff scheme supporting vehicular mobility over IEEE 802.11 WLAN," in *Proc. ACM SAC 2017*, Marrakesh, Morocco, Apr. 2017.

[12] S. Chieochan, E. Hossain, and J. Diamond, "Channel assignment schemes for infrastructure-based 802.11 WLANs: A survey," *IEEE Commun. Surv. Tut,* vol. 15, no. 1, pp. 124–136, Feb. 2010.

[13] Linux wireless,"ACS: Automatic channel selection," [Online]. Available: https://wireless.wiki.kernel.org/en/ users/documentation/acs, Accessed: 2017-08-10.

[14] P. Kulkarni, Z. Zhong, and F. Cao, "Moving away from the crowd: Channel selection in uncoordinated unplanned dense wireless LANs," in *Proc. 32th ACM Sym. Applied Comput. '17,* Marakesh, Morocco, Apr. 2017.

[15] T. H. Lim, W. S. Jeon, and D. G. Jeong, "Centralized channel allocation scheme in densely deployed 802.11 wireless LANs," in *Proc. 18th Int. Conf. Adv. Commun. Technol. '16,* Pyeongchang, Korea, Mar. 2016.

[16] B. Bellalta, A. Checco, A. Zocca, and J. Barcelo, "On the interactions between multiple overlapping wlans using channel bonding," *IEEE Trans. Veh. Tehcnol.,* vol. 65, no. 2, pp. 796–812, Feb. 2015.

[17] S. Han, X. Zhang, and K. G. Shin, "Fair and efficient coexistence of heterogeneous channel widths in next-generation wireless LANs," *IEEE Trans. Mob. Comput.,* vol. 15, no. 11, pp. 2749–2760, Nov. 2016.

[18] R. Murty, J. Padhye, R. Chandra, A. Wolman, and B. Zill, "Designing high performance enterprise Wi-Fi networks," in *Proc. NSDI*, Berkeley, CA, Apr. 2008, pp. 73–78.

[19] Y. Bejerano, S. J. Han, and L. Li, "Fairness and load balancing in wireless LANs using association control," *IEEE/ACM Trans. Netw.*, vol. 15, no. 3, pp. 560–573, June 2007.

[20] W. Li, S. Wang, Y. Cui, X. Cheng, R. Xin, M. A. Al-Rodhaan, and A. Al-Dhelaan, "AP association for proportional fairness in multirate WLANs," *IEEE/ACM Trans. Netw.*, vol. 22, no. 1, pp. 191–202, Feb. 2014.

[21] 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, July 2017.

[22] D. Gong and Y. Yang, "On-line AP association algorithm for 802.11n WLANs with heterogeneous clients," *IEEE Trans. Comput.*, vol. 63, no. 11, pp. 2772–2786, Nov. 2014.

[23] Y. H. Yang, Y. Chen, C. Jiang, and K. J. R. Liu, "Wireless networks association game with data-driven statistical modeling," *IEEE Trans. Wireless Commun.*, vol. 15, no. 1, pp. 512–524, Jan. 2016.

[24] M. Touati, R. E. Aouzi, M. Coupechoux, E. Altman, and J. M. Kelif, "A controlled mathcing game for WLANs," *IEEE J. Sel. Areas Commun.,* vol. 35, no. 3, pp. 707–720, Mar. 2017.

[25] M. Amer, A. Busson, and I. G. Lassous, "Association optimization based on access fairness for Wi-Fi networks," *Computer Networks,* vol. 137, pp. 173–188, June 2018.

[26] J. M. Vella and S. Zammit, "A survey of multicasting over wireless access networks," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 3, pp. 718–753, Second Quarter, 2013.

[27] A. Chen, D. Lee, and P. Sinha, "Efficient multicasting over large-scale WLANs through controlled association," *Computer Networks*, vol. 53, no. 1, pp. 45–59, Oct. 2008.

[28] C. Rossi, C. Casetti, C. F. Chiasserini, and C. Borgiattino, "Cooperative energy-efficient management of federated WiFi networks," *IEEE Trans. Mobile Comput.*, vol. 14, no. 11, pp. 2201–2215, Nov. 2015.

[29] C. Xu, Z. Han, G. Zhao, and S. Yu, "A sleeping and offloading scheme for energy-efficient WLANs," *IEEE Commun. Lett.,* vol. 21, no. 4, pp. 877–880, Apr. 2017.

[30] O. B. Karimi, J. Liu, and J. Rexford, "Optimal collaborative access point association in wireless networks," in *Proc. IEEE Infocom*, Toronto, Canada, May 2014, pp. 1141–1149.

[31] D. Gesbert, M. Kountouris, R. W. Heath Jr., C.-B. Chae, and T. Sälzer, "Shifting the MIMO Paradigm," *IEEE Signal Proc. Magazine*, vol. 24, no. 5, pp. 36–46, Sept. 2007.

[32] Y. C. Hsu, K. C. J. Lin, and W. T. Chen, "Client-AP association for multiuser MIMO networks," in *Proc. IEEE ICC*, London, UK, June 2015, pp. 2154–2159.

[33] B. V. Quang, R. V. Prasad, and I. Niemegeers, "A survey on handoffs – Lessons for 60 GHz based wireless systems," *IEEE Commun. Survey and Tut.*, vol. 14, no. 1, 2012, pp. 64–86.

[34] A. Ahmed, L. M. Boulahia, and D. Gaiti, "Enabling vertical handover decisions in heterogeneous wireless networks: A state-of-the-art and a classification," *IEEE Commun. Survey and Tut.*, vol. 16, no. 2, 2014, pp. 776–811.

[35] K. Xu, K. C. Wang, R. Amin, J. Martin, R. Izard, "A fast cloud-based network selection scheme using coalition formation games in vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 64, no. 1, Nov. 2015, pp. 5327–5339.

[36] J. M. M. Barja, H. Ahmadi, S. M. Tornell, C. T. Calafate, J. C. Cano, P. Manzoni, and L. A. DaSilva, "Breaking the vehicular wireless communications barriers: Vertical handover techniques for heterogeneous networks," *IEEE Trans. Veh. Technol.*, vol. 64, no. 12, Dec. 2015, pp. 5878–5890.

[37] S. Wang, C. Fan, C. H. Hsu, Q. Sun, and F. Yang, "A vertical handoff method via self-selection decision tree for Internet of vehicles," *IEEE Sys. Journal*, vol. 10, no. 3, Sep. 2016, pp. 1183–1192.

[38] S. C. Huang and Y. S. Tzeng, "Novel handoff frequency algorithms for vehicular area networks," *Information Technol. Convergence*, LNEE, Springer, vol. 253, July 2013, pp. 3–12.

[39] M. Almulla, Y. Wang, A. Boukerche, and Z. Zhang, "Design of a fast location-based handoff scheme for IEEE 802.11 vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 63, no. 8, Oct. 2014, pp. 3853–3866.

[40] M. Mouton, G. Castignani, R. Frank, and T. Engel, "Enabling vehicular mobility in city-wide IEEE 802.11 networks through predictive handovers," *Veh. Commun.*, Elsevier, vol. 2, no. 2, Apr. 2015, pp. 59–69.

[41] M. H. Dwijaksara, H. S. Oh, M. Hwang, W. S. Jeon, and D. G. Jeong, "Crowdsourcing-based handover scheme for wireless LAN in vehicular environments," in *Proc. IEEE APWCS 2017*, Incheon, Korea, Aug. 2017.

[42] A. Faridi, B. Bellalta, and A. Checco, "Analysis of dynamic channel bonding in dense networks of WLANs," *IEEE Trans. Mob. Comput.,* vol. 16, no. 8, pp. 2118–2131, Aug. 2017.

[43] M. H. Dwijkasara, W. S. Jeon, and D. G. Jeong, "A joint user association and load balancing scheme for wireless LANs supporting multicast transmission," in *Proc. ACM SAC*, Pisa, Italy, Apr. 2016.

[44] IBM, IBM ILOG CPLEX, http://www-01.ibm.com/software/commerce/optimization/cplex-optimizer, Accessed: 2018-03-29.

[45] "lp_solve 5.5," http://lpsolve.sourceforge.net/5.5/, Accessed: 2018-03-29.

[46] W. Dinkelbach, "On nonlinear fractional programming," *Management Science*, vol. 13, no. 7 pp. 492–498, Mar. 1967.

[47] Y. Almogy and O. Levin, "A class of fractional programming problems," *Operations Research* vol. 19, no. 1, pp. 57–67, Feb. 1971.

[48] FICO Xpress Optimization Suite, "MIP formulations and linearizations," June 2009, [Online]. Available: http://www.fico.com/, Accessed: 2017-08-10.

[49] V. Tarokh, M. Jafarkhani, and A. R. Calderbank, "Space–Time Block Codes from Orthogonal Designs," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp.1456–1466, July 1999.

[50] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE J. Sel. Areas Commun.* vol. 24, no. 3, pp. 528–541, Mar. 2006.

[51] N. Anand, J. Lee, S. J. Lee, and E. W. Knightly, "Mode and user selection for multi-user MIMO WLANs without CSI," in *Proc. IEEE Infocom*, Hongkong, Apr. 2015, pp. 451–459.

[52] C. L. Hwang and A. S. M. Masud, "Multiple objective decision making, methods and applications: A state-of-the-art survey," *Springer-Verlag*, ISBN978-0-387-09111-2, Berlin, Germany, 1979.

[53] J. L. Gearhart, K. L. Adair, R. J. Detry, J. D. Durfee,K. A. Jones, and N. Martin, "Comparison of open-source linear programming solvers," Sandia Report, SAND2013-8847, Oct. 2013, available from http://prod.sandia.gov/techlib/access-control.cgi/2013/138847.pdf.

[54] TP-Link, "TL-WDR4300," http://www.tp-link.com/en/download/TL-WDR4300_V1.html, Accessed: 2018-03-29.

[55] "OpenWrt," http://openwrt.org, Accessed: 2018-03-29.

[56] G. Paul, "IP Network Monitoring Software," http://iptraf.seul.org/, Accessed: 2018-03-29

[57] "TCPDump and Libcap," http://www.tcpdump.org/, Accessed: 2018-03-29.

[58] D. Ferro and B. Rink, "Understanding technology options for deploying Wi-Fi," Ubee Interactive, white paper, available from http://www.ubeeinteractive.com/newsroom/press-releases/understanding-technology-options-deploying-wifi-presentation, Accessed: 2018-03-29.

[59] C. W. Kong and J. Y. B. Lee, "Slice-and-patch An algorithm to support VBR video streaming in a multicast-based video-on-demand system," in *Proc. IEEE ICPADS*, Taiwan, China, Dec. 2002.

[60] V. Erceg, L. Schumacher, P. Kyritsi, A. Molisch, D. S. Baum, et al., "TGn channel models," IEEE 802.11-03/940r4, May 2004.

[61] N. Ravindran and N. Jindal, "Matlab implementation for optimizing throughput using zero-forcing (ZF), PU2RC or Single-user beamforming (SUBF) by varying the number of feedback bits per user, for a given total feedback load," http://people.ece.umn.edu/∼nihar/mud_csi_code.html, Accessed: 2018-03-29.

[62] Y. Lue, M. J. Lee, and Y. Zheng, "Adaptive multi-resource allocation for cloudlet-based mobile cloud computing," *IEEE Trans. Mobile Comput.*, vol. 15, no. no, Oct. 2016, pp. 2398–2410.

[63] V. S. Rao and R. Gajula, "Interoperability in LTE," available from https://www.slideshare.net/allabout4g/interoperability-in-lte. Accessed: 2017-11-24.

[64] Mathworld, "Circle line intersection," available from http://mathworld.wolfram.com/Circle-LineIntersection.html. Accessed: 2017-11-24.

[65] S. Coast, Open street map (OSM), available from https://www.openstreetmap.org. Accessed: 2017-11-24.

[66] German Aerospace Center, "Simulation of urban mobility," available from http://www.sumo.dlr.de/userdoc/. Accessed: 2017-11-24.

[67] R. Porat, S. K. Yong, and K. Doppler, "TGah Channel Model," IEEE 802.11-11/0968r4, Mar. 2015.

[68] W. C. Jakes, “Microwave mobile communications,” Piscataway, NJ, USA:IEEE Press, 1994.

[69] H. Nishiyama, M. Ito, and N. Kato, “Relay-by-smartphone: Realizing multi-hop device-to-device communications,” *IEEE Commun. Mag.*, vol. 52, no. 4, Apr. 2014, pp. 56–65.

Acknowledgements

First and foremost, I would like to thank God Almighty, “Ida Sang Hyang Widhi Wasa”, for his abundant blessing which gives me the strength, knowledge, and ability to undertake this research study and to complete it satisfactorily. Without his blessings, this achievement would not have been possible.

I would like to also express my sincere gratitude to my advisor Professor Wha Sook Jeon for her the continuous support during my Ph.D. study as well as for her patience, encouragement, and immense knowledge. Her thorough guidance helped me in all the time of research and writing of this dissertation. Furthermore, I also want to express my deep gratitude to Professor Dong Geun Jeong for his invaluable guidance. He had co-advised me throughout my Ph.D. study. His comments and directions had made a tremendous contribution to the completion of the dissertation.

Besides my advisors, I would like to thank the rest of my dissertation committee: Professor Chong-kwon Kim, Professor Taekyoung Kwon, and Professor Kaewon Choi, not only for their insightful comments and encouragement, but also for the sharp questions which pushed me to widen my research from various perspectives.

My sincere thank also goes to all of the Mobile Computing and Communications Laboratory (MCCL) alumnus and members, who had been in touch with me during my Ph.D. study. Dr. Dong Heon Lee and Dr. Juhee Kim had been such good seniors who

always gave advise and motivated me at the beginning of my study. My ex-roommate, Mr. Hyun Seob Oh, who always thinks positively, had tirelessly motivated me during my study. Mr. Seung Bom Seo was always ready to help me anytime. Mr. Min Sik Hwang was one of the best co-workers I have ever met. Mr. Geun Yeop Ha always kindly supported me while working on the project. Also the rest of alumnus and members: Mr. Young Il Han, Mr. Jong Ki Cho, Mr. Chan Ki Cho, Mr. Won Seon Heo, Mr. Kwang Im, Mr. Tae Hun Lim, Mr. Jung Gwon Shin, Mr. Hyo Won Lee, Mr. Seung Hak Lee, Ms. Anastasia, Mr. Gang To Lee, and Mr. Hand Ul Lee were such good and supportive lab-mates.

My acknowledgement would be incomplete without thanking the biggest source of my strength, my family. The love and prayer of my parents, I Wayan Pageh and Ni Luh Suastiasih had given me the strength to endure all the difficulties. My elder brother Gde Harta Wijaya, his wife Dewa Ayu Eka, and of course my cute and lovely nieces Putu Ayuna Wijaya and Made Ayura Wijaya had also helped me to reach this stage in my life. Finally, this last word of acknowledgment, I have saved for my dear wife Desi Kusumaningtyas and our beloved daughter Putu Hyerin Dwijaksara, who have been with me all these years and have made them the best years of my life.