Mobility Management for Heterogeneous Wireless Networks

Graduate School of Systems and Information Engineering
University of Tsukuba

July 2016

Qiang, Li

Mobility Management for Heterogeneous Wireless Networks

Qiang, Li

Submitted in partial fulfillment of
the requirements for the degree of
Doctor of ENGINEERING

Graduate School of Systems and Information Engineering
University of Tsukuba

July 2016

Abstract

Over the past decades, wireless techniques have developed rapidly. Nowadays, there always be a lot of different wireless techniques, *e.g.*, WiFi, WiMAX, 2G, 3G, 4G and LTE, in our living spaces. These different kinds of wireless techniques co-exist and construct the heterogeneous wireless networks. Moving in the heterogeneous wireless networks, users will face with various mobility management problems. An efficient and effective mobility management scheme for heterogeneous wireless networks becomes more and more important that has attracted extensive attention. The objectives of mobility management consist of two aspects: (1) *location management* that is to track where the mobile devices are; (2) *handoff management* that is to ensure the sessions of mobile devices are continuous when the devices are moving around. From these two aspects, the dissertation studies the mobility management for heterogeneous wireless networks.

First, we investigate the location management. Specifically, we focus on the pinball routing problem in location management for nested mobile networks. After analyzing the impact of pinball routing problem on both inter-domain and intra-domain communications, we conclude the root cause of pinball routing problem is that the location information of mobile devices is equipped by a few agents. In order to break this limitation and alleviate the pinball routing problem for inter-domain communications, we propose a self-adaptive route optimization scheme. The proposed scheme can adaptively adopt the most appropriate sub-schemes for different situations. Then, we extend the self-adaptive route optimization scheme for intra-domain communications. By making use of the standard location update process, the extended scheme can effectively limit the intra-domain communications in a relatively small region. Through extensive simulation results, we validate that our proposed self-adaptive route optimization scheme significantly reduce the overheads of both inter-domain and intra-domain communications for nested mobile networks.

Second, we study the network selection problem in handoff management for hybrid 5G environments. Motivated by the limitation in real world that users are usually unwilling to share their private information for privacy preservation, we propose a multi-objective distributed network selection scheme for hybrid 5G environments. In our proposed scheme, each user will select a new network based on limited local information. Due to the lack of global information, users have no choice but to make some inferences or estimations during their network selection processes. In order to assist users in making reasonable inferences, we study the relations between any two users and define the correlation degree metric. Then, we exploit two performance attributes to evaluate networks. One is the channel capacity which indicates the profit a user can get from a certain network. The other one is the blocking probability which reflects the risk a user will undertake for a network. After this, we formulate the network selection problem as a multi-objective optimization problem which maximizes the channel capacity and minimizes the blocking probability simultaneously. Then we transform the formulated multi-objective optimization problem into an equivalent maximization problem. Through a distributed method, we solve the transformed maximization problem in polynomial time and linear space. The solution of the transformed maximization problem is also approved to be a Pareto Optimal result of the original multi-objective optimization problem. Extensive experiment results validate that our proposed scheme promotes the total throughput and user served ratio effectively.

Third, we investigate both the network selection problem and the handoff timing problem in handoff management for Software-Defined Networking (SDN) based heterogeneous wireless networks. Due to the lack of global information, most of exiting handoff schemes failed to be global optimal. The emergence of SDN technique makes it possible to break this limitation. In SDN architecture, a SDN controller separates the control planes from the data planes of network devices, and provides the centralized control for the whole system. We formulate the network selection problem as a 0-1 inter programming problem which maximizes the sum of channel capacities that mobile devices can obtain from their new networks. The SDN controller solves the formulated problem and gets the network selection results. After the network selection process is finished, we let each mobile device wait for a time period and then make a decision. Only if the newly selected network is consistently more appropriate than the current network during this time period, will the mobile device transfer its network connection to the new network. The proposed SDN-based vertical handoff scheme ensures that a mobile device will transfer to the most appropriate network at the most appropriate time. Comprehensive simulation results reveal that the proposed scheme reduces the number of vertical handoffs, maximizes the total throughput and user served ratio significantly.

Overall system performances (*e.g.*, through, end-to-end delay, user served ratio) have been improved by applying our proposed schemes. Our work also has potential contributions to the mobility management for future heterogeneous wireless networks.

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my advisor, Prof. Jie Li. Without his excellent guidance and continuous support, it would have been impossible for me to carry out research smoothly. He is a strict researcher who has very rigorous academic attitude, he is also a venerable elder who gives me full patience and encouragement. His high expectations for students have inspired me to challenge myself much more than I ever had. It is my great honor to be a student of such wise man in this very important stage of my life.

Besides my advisor, I would like to thank Prof. Hiroyuki Kitagawa, Prof. Kazuhiko Kato, Prof. Kazuki Katagishi, and Prof. Shigetomo Kimura for being my dissertation committee members and providing valuable advices and comments on evaluating this dissertation in its final form. I also would like to thank the Department of Computer Science, the Graduate School of Systems and Information Engineering at University of Tsukuba, for their continuous supports on my study in Japan.

I cannot thank my advisor during my postgraduate study, Prof. Baohua Zhao enough for his supervision, guidance, and valuable help, without which it would have been difficult for me to have an opportunity to pursue the PhD degree in Japan. I would also like to thank the China Scholarship Council (CSC) Foundation, for offering the scholarship to support my study in University of Tsukuba.

I would like to acknowledge the memorable helps from the members of the Operating System and Distributed Processing (OSDP) lab, Prof. Hisao Kameda, Dr. Huang Lu, Dr. Xiaoyan Wang, Dr. Biao Han, Dr. Xiucai Ye, Shuai Fu, Zichen Jin, Yujie Hu, Zhengxu Li, Cheng Sun, Ke Tang, Naoto Ishiduka, Serigne Mbacke Ndiaye, Myagmar Batsergelen, Wenchao Li, Yicong Zhang, Gaoxuan Cao, Jiaxing Li, Di Wu, Baojun Zhou, Yifan Li, Yifan Ren, Qin Xiang, Wei Zhang, Ruoran Shi, Wilson Edwin, Shutan Wang, Yangfei Lin. It has been my great honor to work with them. I would like to thank Prof. Mohsen Guizani (Qatar University), Prof. Yusheng Ji (National Institute of Informatics), Prof. Changcheng Huang (Carleton University), Prof. Eitan Altman

(INRIA Sophia-Antipolis), Prof. Corinne Touati (INRIA-LIG, France), and Prof. Wenzhong Guo (Fuzhou University, China) for their kind suggestions and comments on my work.

Finally, and most importantly, I would like to acknowledge the immeasurable support given to me by my parents and my boyfriend. Words cannot describe my gratitude towards them.

My apologies if I have inadvertently omitted anyone to whom acknowledge is due.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This chapter begins with introducing the background and the motivation of our research work. A summary of the main contributions is presented in Section 1.2. Section 1.3 provides the organization of the dissertation.

1.1 Research Background and Motivation

Over the past few years, there has been tremendous growth in wireless techniques. Different kinds of wireless techniques co-exist and construct the heterogeneous wireless networks. As the number of mobile terminals increases, mobility management in heterogeneous wireless networks becomes more and more important [1]. The objective of mobility management in heterogeneous wireless networks is to track where the mobile users are, and ensure the sessions of mobile users are continuous when users are moving around [2]. The study on mobility management mainly consists of *location management* and *handoff management* as shown in Fig.1.1. In this section, we will explain the meaning of each component in detail.

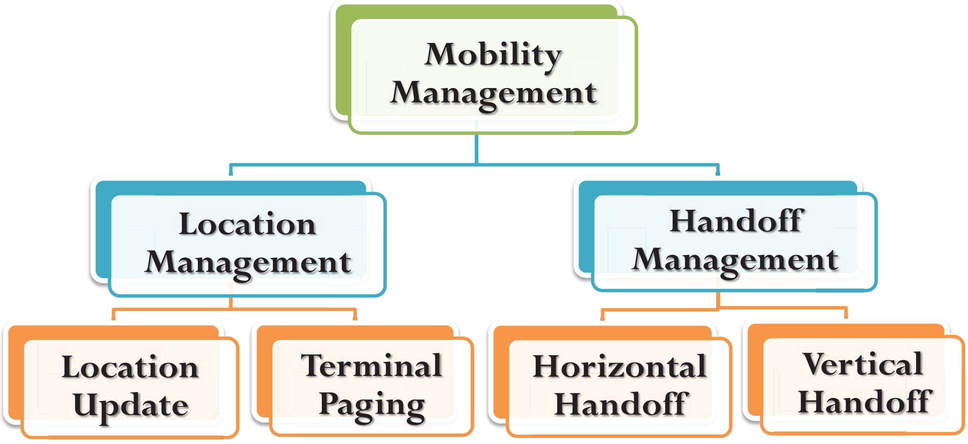


FIGURE 1.1: Classification of mobility management.

1.1.1 Location Management

The location management enables the network track the locations of mobile users [3]. When a connection needs to be established for a particular user, the network has to determine the location of the mobile user exactly. The operation of informing the network about the current location of mobile user is called *location update*. The operation of determining the locations of mobile users is called *terminal paging*.

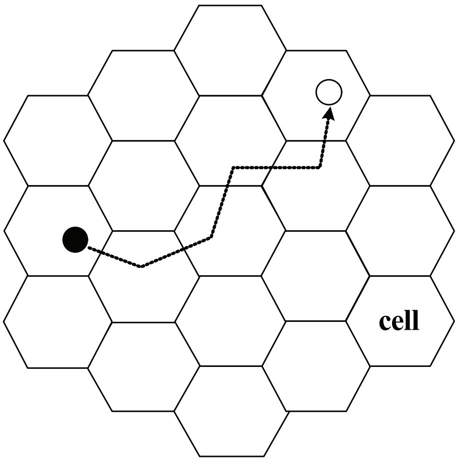


FIGURE 1.2: An example of location management.

There is a trade-off between the location update and terminal paging. As the example shown in Fig.1.2, the coverage area of networks is divided into numerous cells. If a mobile user announces its location whenever it crosses a cell boundary, the system can maintain the location of the mobile user precisely. In this situation, the terminal paging can be obviated. However, if the call arrival rate is low, such frequent location update process will cost a lot of signal overhead. On the other hand, if the mobile user does not perform the location update timely, it will take longer time to determine the location of a mobile user in the large scale scenarios.

1.1.2 Handoff Management

The handoff management aims to maintain the active connections of mobile users when they are moving around [4]. In heterogeneous wireless networks, various wireless techniques will co-exist and overlap. Moving in heterogeneous wireless networks, mobile users often need to switch their inter-network connections from a network to another network as shown in Fig.1.3. If the original network and the target network support the same kind of wireless technique, the switching between them is called the horizontal handoff. Otherwise, if the original network and the target network support different kinds of wireless techniques, the switching between them is called the vertical handoff.

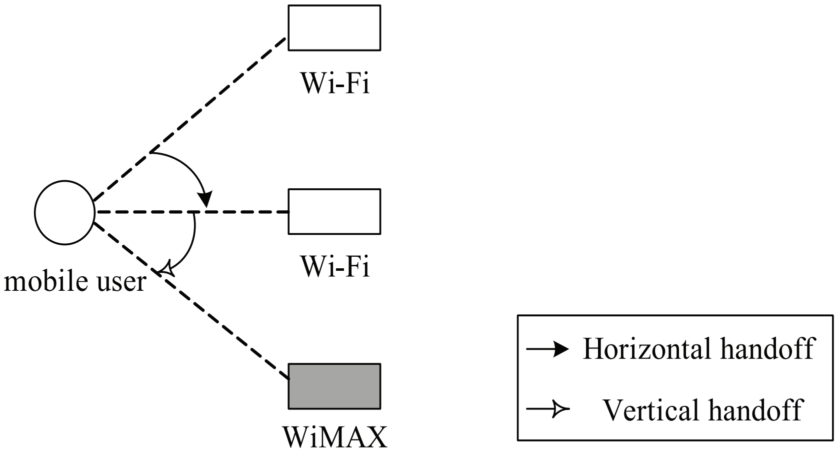


Figure 1.3: An example of handoff management.

In the horizontal handoff, since the networks have the same media features, network evaluation mainly depends on the Received Signal Strength (RSS) [5]. While in the vertical handoff, networks have different capabilities, code methods, bit error rates and other features. Research on vertical handoff will be more difficult than horizontal handoff. If a problem in vertical handoff can be solved, solution to the same problem in horizontal handoff can be derived easily.

1.2 Contributions

Even a lot of work has been conducted in the mobility management for heterogeneous wireless networks, there still exists some open problems. Specifically, we focus on the pinball routing problem, network selection problem and handoff timing problem in this dissertation. We carry out research following the roadmap as shown in Fig. 1.4. The main contributions of this dissertation are summarized as follows.

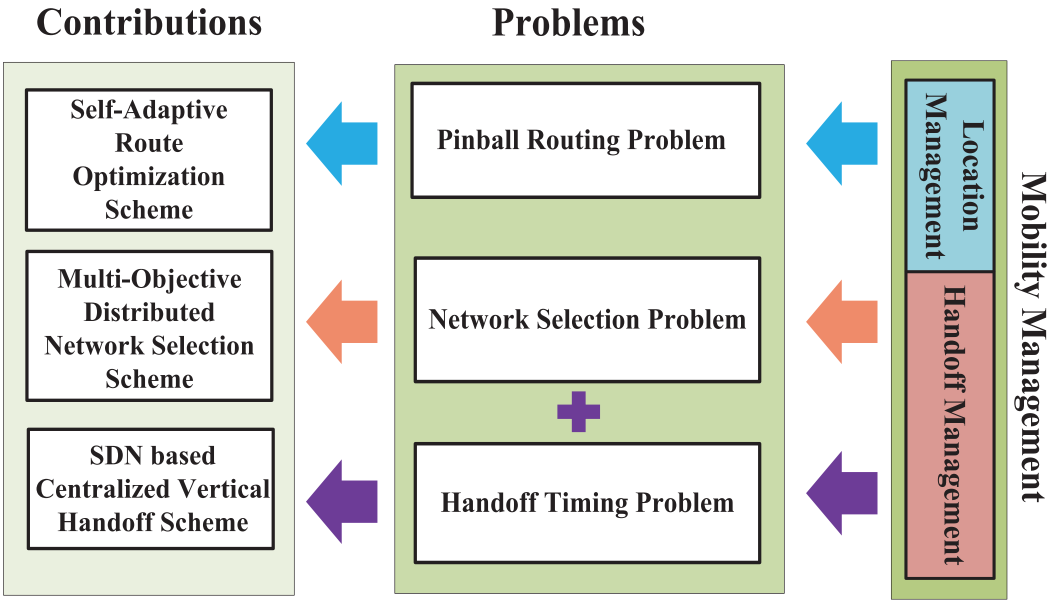


FIGURE 1.4: A roadmap of the research work performed in this dissertation.

1.2.1 A Self-Adaptive Route Optimization Scheme for Nested Mobile Networks

In IP-based wireless networks, each mobile user has two kinds of addresses: Home-Address (HoA) which is an identifier, never changes once be obtained; Care-of-Address (CoA) which is locator, updates as user moves [6]. Each mobile user has a home agent. The home agent of a mobile user maintains the relationship between its HoA and CoA. Packets which are sent to a mobile user, should be delivered to the home agent of this mobile user at first (Fig.1.5 (a)). This triangular routing method will incur the Pinball Routing Problem [7] in a nested mobile network. As the example shown in Fig.1.5 (b), mobile user A is attaching to mobile user B. Packets which are sent to mobile user A have to pass through the mobile user B. In order to reach mobile user B, packets should be delivered to the home agent of mobile user B at first. After this, the home agent of mobile user B will deliver the received packets to the home agent of mobile user A. Each time when the packets pass through a home agent, these packets will be encapsulated once, then decapsulated by the corresponding mobile user. As a result, packets will experience several times encapsulation and decapsulation in a nested structure. This process increases the end-to-end delay and transmission overhead dramatically. Situation will get worse as the nesting level increases.

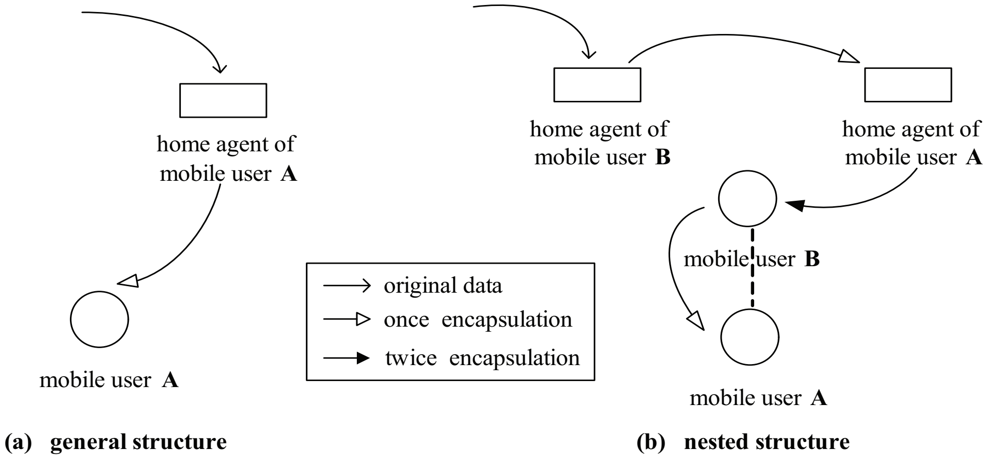


FIGURE 1.5: An example of pinball routing problem.

In order to solve the pinball routing problem of the location management, we propose a self-adaptive route optimization scheme in Chapter 3. We carefully study the pinball routing problem and conclude the root cause of pinball routing problem is that only the corresponding Home Agent (HA) equips with the location information of a Mobile Router (MR). In order to optimize the routing process of inter-domain communication, we propose a self-adaptive route optimization scheme which consists of two sub-schemes: time-saving sub-scheme and mobility-transparency sub-scheme. The time-saving sub-scheme is suitable for low mobility and large communication traffic scenario, and the mobility-transparency sub-scheme is suitable for high mobility scenario. Given a scenario, our proposed scheme can adaptively adopt the most appropriate sub-scheme. Furthermore, we extend the self-adaptive scheme to optimize the routing process of intra-domain communication. The extended scheme ensures that the intra-domain communication is limited in a relatively small region.

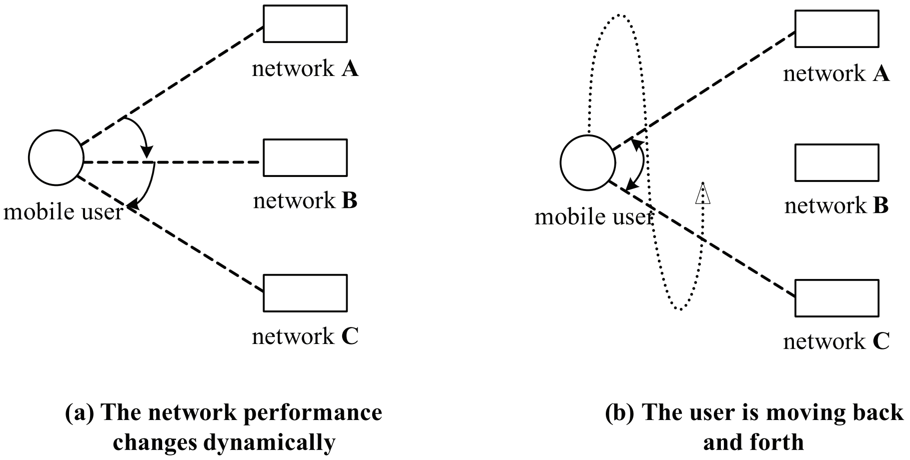
1.2.2 A Multi-Objective Distributed Network Selection Scheme for Hybrid 5G Environments

In hybrid 5G environments, 5G technique will integrate with other existing techniques to provide ubiquitous high-rate and seamless communication service [8]. Moving in hybrid 5G environments, a user always needs to switch its inter-network connection from the current network to another network. For a single user, it may have several available networks. A user has to select a network from its available networks, to which the inter-network connection should be switched. In most of the existing research on network selection problem, mobile users are assumed to be selfish. That is to say, users always choose the best performance networks as their selected networks [9]. As we known, the resources of networks are limited. If a large number of users choose the same network simultaneously, resources of the selected network will be exhausted, and the users will be blocked. Therefore, the objective of network selection is to select a network as well as possible and avoid being blocked.

In order to solve the network selection problem of the handoff management, we propose a multi-objective distributed network selection scheme for hybrid 5G environments in Chapter 4. We consider the distributed scenario in which there is no centralized control entity. Each user has to make its network selection during the vertical handoff process by itself. We also consider a general limitation in the real world that is users do not share their private information with each others. Under this limitation, we firstly study the relations between any two users, and define the correlation degree which could efficiently distinguish the categories of relations, and sufficiently reflect the association strength. Base on the correlation degree, private information of itself and two pieces of public information, a user can estimate it blocking probability for each available network device. Then we formulate the network selection problem as a multi-objective optimization problem which maximizes the channel capacity and minimizes the blocking probability simultaneously. After that, we transform the formulated multi-objective optimization problem into a maximization problem by taking the throughput metric into consideration. At last, we solve the transformed maximization problem in polynomial time and liner space. Moreover, we prove that the solution of the transformed maximization problem is a Pareto Optimal result of the original multi-objective optimization problem. Through extensive experiments we validate that our proposed scheme promotes the total throughput and user served ratio effectively.

1.2.3 A Software-Defined Networking based Centralized Vertical Handoff Scheme for Heterogeneous Wireless Network Environments

After the network selection process is finished, a user needs to determine the time when its inter-network connection should be switched to the selected network [10]. If the inter-network connection switching is implemented too early, there will be a lot of unnecessary handoff. For example as shown in Fig.1.6 (a), a mobile user is connecting to the network A, and its selected network is network B. Suppose that, the mobile user switches its inter-network connection from A to B immediately. Unfortunately, the performance of network B gets worse within a very short period of time. As a result, this mobile user needs to handoff once again. Besides of this, we consider another situation as shown in Fig.1.6 (b). A mobile user is moving back and forth between network A and network C. If the handoff is performed in an improper time, this user may require frequent connection switching between A and C. From these simple examples we find that improper handoff timing will incur a lot of unnecessary handoffs. There are several existing vertical handoff schemes try to solve both the network selection problem and the handoff timing problem. However due to lack of the global view, most of existing vertical handoff schemes failed to be global optimal. The emergence of Software-Defined Networking (SDN) technique provides a chance to break this limitation.



(a) The network performance changes dynamically

(b) The user is moving back and forth

Figure 1.6: Examples of handoff timing problem.

Based on the SDN technique, we propose a centralized vertical handoff scheme which solves both the network selection and the handoff timing problems of the handoff management for heterogeneous wireless network environments in Chapter 5. We firstly divide users into two classes: non-handoff users and handoff users. Non-handoff users will stay in the connections with their current networks. While handoff users will send their handoff requests to an SDN controller. Then we formulate the network selection of handoff users as a 0-1 integer programming problem with the objective of maximizing the sum of channel capacities. The SDN controller solves the formulated 0-1 integer programming problem to calculate the optimal network selection results. After the network selection process is finished, we propose a handoff timing algorithm to determine the time when the network selection results should be implemented. Based on limited information and simple calculation, the proposed handoff timing algorithm can predict the movement directions of handoff users. Only if handoff users are certain to move away from their current network devices, will the network selection results be implemented.

1.3 Dissertation Organization

This dissertation is organized as follows. In Chapter 1, we introduce the motivation and the background of our research, outline the main contributions in this dissertation. In Chapter 2, we present an overview of the mobility management for heterogeneous wireless networks. An adaptive route optimization scheme for nested mobile networks is presented in Chapter 3. In Chapter 4, we propose a multi-objective distributed network selection scheme for hybrid 5G environments. Chapter 5 presents a software-defined networking based centralized vertical handoff scheme for heterogeneous wireless networks. Finally we conclude this dissertation and points out the future work in Chapter 6.

Chapter 2

Overview of Mobility Management for Heterogeneous Wireless Networks

This chapter presents an overview of the mobility management for heterogeneous wireless networks based on different Open System Interconnection (OSI) layers. Specifically, we focus on the existing work about pinball routing problem, network selection problem and handoff timing problem based on cross-layer (*i.e.*, link layer and network layer).

2.1 Current Research Status of Mobility Management

A variety of mobility management schemes have been proposed to deal with mobility related problems at different OSI layers. Depending on the layers of communication protocol they primarily use, these mobility management schemes can be divided into four categories: link layer mobility management schemes, network layer mobility management schemes, transport layer mobility management schemes and cross-layer mobility management schemes [1].

Link layer mobility management is responsible for establishing a radio link between a mobile node and its new access point when the mobile user is changing the point of attachment [11]. Link layer mobility management schemes provide mobility related features in the underlying radio systems which tightly couple with specific wireless technologies. Link layer mobility management schemes always work following four phases: recognizing the loss of the wireless connection, searching for new available networks, authentication with a new network and association with the new network. In order to reduce the management overhead caused by the link layer, researchers always try to accelerate the searching phase. Mobile users are required to scan the medium while they are still connecting with the current networks. In some cases, channel mask will be used in order to further reduce the number of channels that must be scanned.

Network layer mobility management schemes make use of the features at the IP layer, and are agnostic of the underlying information [12]. Since network layer mobility is transparent to higher layer, applications on mobile users can maintain working without any modifications. The study on network layer mobility management mainly focuses on the routing problem. The triangular routing method that is corresponding node sends all packets via home agent to a mobile node will incur higher latency and network load. There are several kinds of ideas have been studied to optimize the triangular routing, such as let the corresponding node maintain the current location of a mobile user, let home agent actively inform the corresponding node about the current location of a mobile user and so on.

Transport layer mobility management schemes intent to maintain the reliability of TCP link and the correctness of semantic while users are moving around. Study on transport layer mobility management can be divided into four groups: transport layer handoff, transport layer connection migration, gateway-based mobility and complete transport layer mobility management [13]. The transport layer handoff aims at improving the system performance during handoff on transport layer. The transport layer connection migration schemes are the schemes that can migrate multiple connections. The gateway-based mobility schemes provide mobility by putting a infrastructure between corresponding node and mobile users and splitting the connection. The first three groups of schemes are incomplete mobility management schemes while the last group of schemes provide complete mobility management with both handoff management and location management.

Cross-layer mobility management schemes mainly concern the handoff management aspect. Most of this kind of schemes are using the features gathered from link layer to make an efficient handoff decision in network layer.

2.2 Existing Work about Pinball Routing Problem

Some related work has been conducted in addressing the pinball routing problem for nested mobile networks. S. Pack *et al.* [14] proposed an adaptive network mobility support scheme which jointly optimizes the binding update traffic and the tunneling overhead. When the Session-to-Mobility Ratio (SMR) is lower than a pre-defined threshold, intermediary Mobile Routers (MRs) and terminal MRs will perform the Regional Care-of-Address (RCoA) and the on-Link Care-of-Address (LCoA) binding update procedures respectively. Consequently, the number of binding updates will be reduced. When the SMR is higher than the pre-defined threshold, intermediary MRs and terminal MRs will perform the LCoA and the RCoA binding update procedures respectively. In this case, the number of tunnelings will be reduced. However, this work needs to change the core architecture defined by Network mobility Basic Support Protocol (NBSP). A Mobility Anchor Point (MAP) will be introduced in order to support this work.

H. Kim *et al.* [15] proposed the Simple Route Optimization (S-RO) Scheme with NEtwork MObility (NEMO) Transparency. Similar to Mobile IPv6 [16], S-RO scheme requires that when an MR receives an encapsulated data, the MR will send a Binding Update (BU) message to the original sender of the received data. After receiving a BU message, the original sender will extract and cache the Care-of-Address (CoA) of destination MR from this BU message. Through this process, a soft route between the original sender and the destination MR can be established. After that, the information exchange between the original sender and the destination MR can be forwarded through the established soft route directly. Although this work effectively reduces the data transmission overhead, it will increase the location update overhead.

An efficient route optimization scheme for nested-NEMO is proposed by K. Humayun *et al.* [17], in which each MR is related to two CoAes: root-CoA and local-CoA. The root-CoA is equal to the CoA of Top Level Mobile Router (TLMR) in the nested-NEMO.

While the local-CoA refers to the regular CoA of an MR in NBSP. Accordingly, there are two types of entries in the routing table in each MR: fixed and visiting. The fixed entry in the routing table stores the information of homed networks, and the visiting entry stores the information of temporarily attached MRs. Authors declared that their proposed scheme can remove the tunnels completely from the nested NEMO. Unfortunately, the pinball routing problem in intra-domain communication has not been noticed.

An interesting scheme is proposed by H. Cho *et al.* [16] called the Route Optimization scheme using Tree Information Option (ROTIO). In basic ROTIO, each HA maintains the Home-Address (HoA) of TLMR. Data sent from CN to destination MR will pass through only two HAs (HA of TLMR, and HA of destination MR). Therefore, the inter-domain communication can be optimized in the basic ROTIO scheme. For intra-domain communication, ROTIO requires TLMR to maintain the topology of the whole domain. Data of intra-domain communication will be sent to TLMR at first. Then, TLMR forwards the received data to its destination. The routing process of intra-domain communication in ROTIO still has improvement room.

Another ingenious scheme is proposed by G. S. Kuo *et al.* [18] named the Hierarchical Mobility Support for Route Optimization (HMSRO). Similar to ROTIO, HMSRO also requires each HA maintains the information of TLMR. However, the maintained information is the CoA of TLMR rather than the HoA of TLMR. Therefore, data of inter-domain communication will pass through one and only one HA (HA of destination MR). After the data arrives at TLMR, TLMR forwards the received data according to its routing table. The routing table maps the prefix of destination address to the next hop address. Data will be relayed to the destination hop by hop. The performance of this work is unsatisfactory in high mobility scenarios.

Our proposed scheme does not need to change the core architecture defined by NBSP. Furthermore, we consider both the mobility transparency overhead and the location update overhead. The overhead of inter-domain and intra-domain communications can be effectively reduced by using our proposed scheme that will be presented in Chapter 3.

2.3 Existing Work about Network Selection Problem

A fuzzy logic based network selection scheme is proposed by J. Hou *et al.* [19], in which three input fuzzy variables are considered. At first, these three fuzzy variables are fuzzified and converted into some input fuzzy sets by a singleton fuzzifier. Then the input fuzzy sets are mapped into output fuzzy sets by an algebraic product operation. Finally, the output fuzzy sets are defuzzified into a crisp decision point which indicates a network selection result. The proposed fuzzy logic based network selection scheme is difficult to be implemented since it lacks of explicit definitions for the “high” and “low” probabilities

A. Roy et al. [20] formulated the network selection problem as a multi-objective optimization problem which maximizes the Reference Signal Received Power (RSRP) and the available Resource Blocks (RBs) simultaneously. By using the weighted linear sum approach, the formulated multi-objective optimization problem is simply merged into a single aggregate objective function. The solution of this single aggregate objective function is the network selection result. Authors also analyzed the performance of their proposed scheme by using a three-dimensional Markov chain. Authors only cared about how to maximize the profits of network selection results from two aspects, but failed to consider the risks of such network selection behaviors.

In order to improve the Quality of Experience (QoE), a Multiplicative Utility based Automatic Network Selection (MU-ANS) scheme is proposed by Nguyen-Vuong Q.t. *et al.* [21]. In MU-ANS scheme, the network selection problem is formulated as a Multiple Attribute Decision Making (MADM) problem. After calculating the multi-criteria utility function value for each available network device, a user selects the highest scoring network device as the new network device. This kind of approach cannot avoid the collision situation. As a result, the performance of the proposed scheme heavily depends on the situations of network devices.

Chao *et al.* [22] proposed a two-step network selection scheme. The first step is pre-decision progress, in which a filtering function is used to evaluate the performance of network devices. If no network device can pass the pre-decision, a user will stay in the connection with its current network device. If there is only one network device passing the pre-decision, a user will handoff to the sole network device. If there are several network devices passing the pre-decision and a user has insufficient power, the user will randomly handoff to a network device. If there are several network devices passing the pre-decision and a user has sufficient power, the user will execute the second step. In the second step, the network selection scheme is also formulated as an MADM problem and the highest scoring network device will be selected. The complex procedure of two-step network selection scheme dissatisfies the fast decision requirement of network selection.

We consider both the benefit and the risk of the network selection behaviors in our work which will be presented in Chapter 4. The proposed scheme can satisfy the new challenges of network selection in hybrid 5G environments. We also propose a global optimal network selection scheme by using the Software Defined Networking (SDN) technique in Chapter 5.

2.4 Existing Work about Handoff Timing Problem

Mobile users will wait for an appropriate time to perform the handoff. The waiting time before handoff is called the stability period. H.J. Wang *et al.* [23] proposed a Policy-Enable Handoff Scheme (PEHS) and pointed out that, the length of stability period should be proportional to the handoff latency. At first, authors calculated out the request time for making up the loss due to handoff latency (T_{makeup}). Then they proved that, the stability period equals to T_{makeup} add the length of handoff latency.

Based on PEHS, Chen *et al.* [24] introduced two adaptive decision methods to adjust the length of stability period. The first adaptive decision method adjusts the stability period, relying on the utility of selected network to the utility of current network ratio. If the ratio increases, a user will perform handoff at once. Otherwise, if the ratio decreases, this user will observe selected network for a long time. The second adaptive decision method is based on the ratio of two measured utility ratios. When the utility ratio is fast decreasing, the second adaptive decision method performs better.

Lee *et al.* [25] carried out research from the aspect of optimization, and proposed an Enhanced Group Handoff Scheme (EGHS) which combined solves the network selection and the handoff timing problems. In EGHS, each user evaluates its available access points on the remaining bandwidth. The network selection is formulated as a convex optimization problem. The objective of formulated problem is to minimize the handoff blocking probability. By using the Karush-Kuhn-Tucker condition, the formulated convex optimization problem is solved and a new access point can be determined. After the new access point is determined, a user transfers its inter-network connection to the new access point after an adjusted delay.

An interesting scheme proposed by Ciubotaru *et al.* [26] is called Smooth Adaptive Soft Handoff Algorithm (SASHA). Similar to EGHS, SASHA combines the network selection problem with the handoff timing problem. In SASHA, a user obtains a weighted sum of various performance parameters, and calculates the Quality of Service (QoS) values of its available access points. The user allocates its traffic according to the QoS values. As the user leaves an access point and gets closer to another, the QoS value of the leaving access point gets lower, and the QoS value of the approaching access point gets higher. As a result, traffic on the leaving access point is transferred to the approaching access point gradually. In order to apply SASHA, a user needs to keep all of its ports working which will cost a lot of energy.

2.5 Summary

In this chapter, we provided an overview of mobility management for heterogeneous wireless networks, and specifically presented the existing work about pinball routing problem, network selection problem and the handoff timing problem. The experiences of existing work will guide our research work in the following chapters.

Chapter 3

A Self-Adaptive Route Optimization Scheme for Nested Mobile Networks

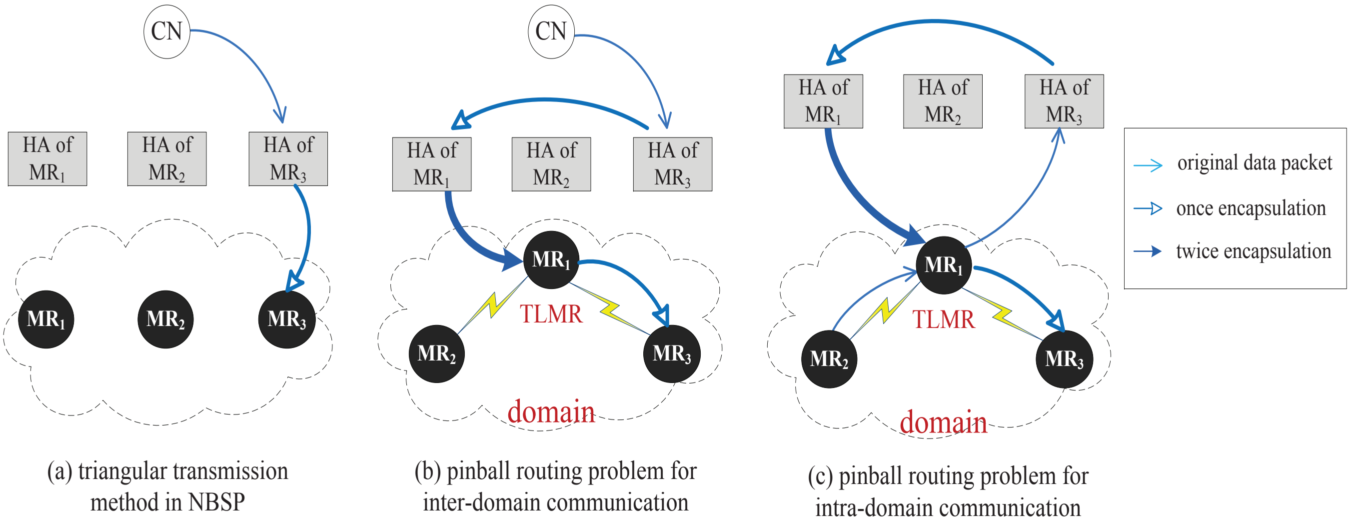
This chapter investigates the route optimization problem for nested mobile NEtwork MObility (NEMO) environment. We firstly propose a self-adaptive scheme which optimizes the routing process of inter-domain communication. The self-adaptive scheme consists of two sub-schemes: time-saving sub-scheme and mobility-transparency sub-scheme. The time-saving sub-scheme can minimize the data transmission overhead of inter-domain communication, which is suitable for low mobility and large communication traffic scenario. The mobility-transparency sub-scheme can reduce the location update overhead to the greatest extent, which is suitable for high mobility scenario. A threshold is used to adaptively adopt the most appropriate sub-scheme for a given scenario. Furthermore, we extend the self-adaptive scheme for intra-domain communication. The extended scheme ensures that the intra-domain communication is limited in a relatively small region. Theoretical analysis and simulation results demonstrate that the proposed scheme can reduce the overheads of both inter-domain and intra-domain communications for nested mobile networks significantly.

This chapter is organized as follows. Section 3.1 introduces the core architecture of nested mobile networks and the pinball routing problem. System description and problem formulation are presented in Section 3.2. Section 3.3 presents a self-adaptive route optimization scheme for inter-domain communication. In section 3.4, the proposed self-adaptive route optimization scheme is extended for intra-domain communication. Section 3.5 is the performance evaluation. Finally, Section 3.6 summarizes this chapter.

3.1 Introduction

As ubiquitous computing proliferate, more and more mobile devices emerge in wireless networks. Some mobile devices could connect together to construct a small wireless network, which moves as an unit. The mobility of this unit is called NEtwork MObility (NEMO) [27]. The Network mobility Basic Support Protocol (NBSP) [7] is the primary protocol of NEMO, which defines the core architecture of NEMO environment. In this chapter, we investigate the route optimization problem for nested mobile NEMO environment based on NBSP.

In the core architecture defined by NBSP, a mobile device which equips with the storing and forwarding functions is called Mobile Router (MR). Some MRs may connect together to construct a tree-configured nested structure. The root of the tree is called Top Level Mobile Router (TLMR), and the region of TLMR is called domain. Only the mobility of domain is considered in this chapter. The structure of domain is relatively stable, MRs in domain will move as an unit. Each MR has a Home Agent (HA) which maintains its current location information. Meanwhile, each HA is assumed to deal with only one MR in most of research on pinball routing problem. When a Corresponding Node (CN) tires to communicate with an MR (*i.e.*, inter-domain communication), the data destined for the MR will be sent to its corresponding HA at first. After receiving the data, the corresponding HA encapsulates the received data with the current location information of MR, and then sends the encapsulated data out (Fig. 3.1 (a)). This kind of triangular transmission method will incur the pinball routing problem [28] in nested structure. Take the scenario shown in Fig. 3.1 (b) as an example, in which CN attempts to communicate with the mobile device MR_3. Since MR_3 is appending to the mobile device MR_1, in order to determine the current location of MR_3, CN has to determine



(a) triangular transmission method in NBSP

(b) pinball routing problem for inter-domain communication

(c) pinball routing problem for intra-domain communication

Figure 3.1: Illustration of the pinball routing problem

the current location of MR_1 at first. Hence, the data destined for MR_3 will pass through two HAs. Each time when the data passes through a HA, the data will be encapsulated once. Then, the encapsulated data will be decapsulated by corresponding MR later. This process will increase the data packet size and end-to-end delay dramatically. Situation gets worse as nesting levels increase. The pinball routing problem is more serious for intra-domain communication. Take the scenario shown in Fig. 3.1 (c) as an example, mobile device MR_2 tries to communicate with mobile device MR_3 that is inside the same domain. Since NBSP requires that data destined for an MR has to be sent to the corresponding HA at first, data of intra-domain communication will be sent outside the domain then sent back.

In order to alleviate the impact of pinball routing problem for both inter-domain and intra-domain communications, we propose a self-adaptive route optimization scheme on the basis of our previous work [29] in this chapter. The proposed self-adaptive route optimization scheme consists of two sub-schemes: time-saving sub-scheme and mobility-transparency sub-scheme. The time-saving sub-scheme can reduce the data transmission overhead of inter-domain communication to the greatest extent, but incurs more location update overhead. On the contrary, the mobility-transparency sub-scheme can minimize the location update overhead of inter-domain communication, at the cost of twice encapsulation. In a word, the time-saving sub-scheme performs well for the low mobility and large communication traffic scenario, while the mobility-transparency sub-scheme performs well for the high mobility scenario. A threshold is used to adaptively adopt the most appropriate sub-scheme for a given scenario. Furthermore, the proposed self-adaptive route optimization scheme is extended for intra-domain communication. In the extended scheme, each MR is required to maintain the structure of its subtree. After receiving data, an MR will check whether the destination of received data belongs to its subtree or not. If the destination is inside its subtree, the MR will relay the received data to the destination directly. Otherwise, the received data will be forwarded upwards. Through this process, the intra-domain communication can be limited in a pretty small area.

The main contributions of this chapter are summarized as follows:

- We analyze the impact of pinball routing problem on both inter-domain and intra-domain communications for nested mobile NEMO environment. We conclude the root cause of pinball routing problem is that only the corresponding HA equips with the location information of an MR.
- We propose a self-adaptive route optimization scheme to alleviate the pinball routing problem for inter-domain communication. The proposed scheme can adaptively adopt the most appropriate sub-scheme for different scenarios.
- We extend the proposed self-adaptive route optimization scheme to alleviate the pinball routing problem for intra-domain communication. By making use of the standard location update process, the extended scheme efficiently works with just minor storage overhead.
- We design and carry out the comparison experiments to evaluate the performance of our proposed scheme.

3.2 System Description and Problem Formulation

We investigate the pinball routing problem in the architecture defined by NBSP [7]. Consider a general nested mobile network which consists of n Mobile Routers (MRs).

Let M be the set of MRs, where $M = \{m_1, m_2, \cdots, m_n\}$. Given an arbitrary element m_i ($m_i \in M$), it can be characterized by an ordered triple $\langle h_i, c_i, l_i \rangle$ which corresponds to its three attributes: Home-Address (HoA), Care-of-Address (CoA) and nesting levels [30]. The HoA is the identifier of an MR, never change once obtained. The CoA indicates the current location of an MR, which is generated by inheriting the prefix of parent MR. Since new prefix will be assigned to an MR when it moves into a new Access Point (AP) region, the CoA of an MR will update accordingly. The nesting levels refer to the tree level of an MR in the tree-configured structure. If m_i is the Top Level Mobile Router (TLMR) of the tree-configured structure, its nesting level $l_i = 1$. Each MR has a Home Agent (HA) to maintains its location information, and each HA deals with only one MR. Therefore, n MRs correspond to n HAs. Let A be the set of HAs, where $A = \{a_1, a_2, \cdots, a_n\}$, and a_i is the HA of m_i. Let $S(a, b)$ denote the communication between a and b. Then the inter-domain communication can be represented by $S(CN, m_i)$, and the intra-domain communication is represented by $S(m_i, m_j)$. For convenience, the notations and abbreviations used in this chapter are summarized in Table 3.1.

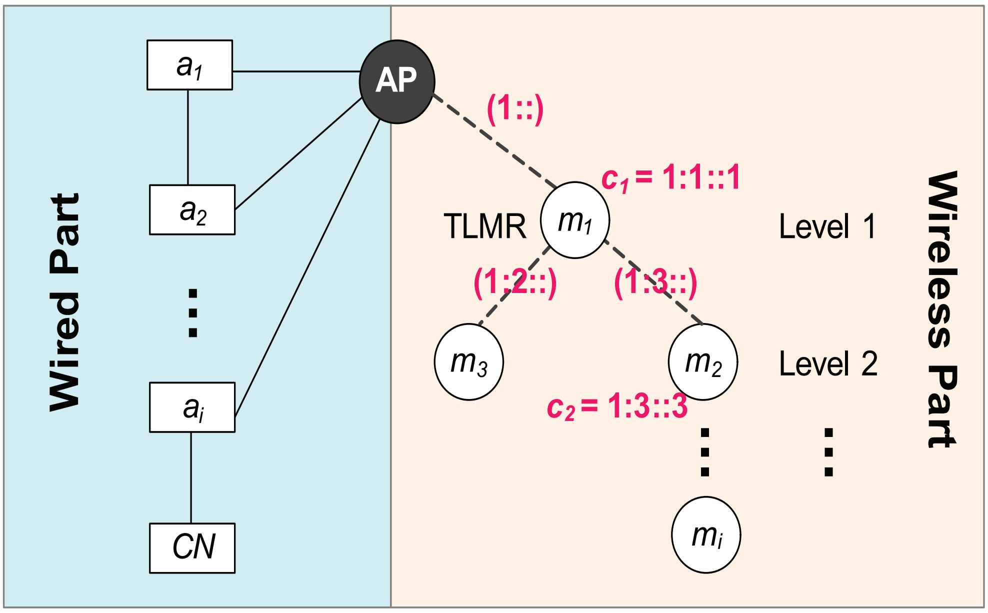


FIGURE 3.2: An example of inter-domain communication

The overhead of a communication consists of two parts: data transmission overhead

Table 3.1: Notation and Abbreviation Summary in Chapter 3

M	Set of mobile routers $\{m_i\}$, $i = 1, 2, \cdots, n$
h_i	Home-address of mobile router m_i
c_i	Care-of-address of mobile router m_i
l_i	Nesting levels of mobile router m_i
A	Set of home agents $\{a_i\}$, $i = 1, 2, \cdots, n$
$S(CN, m_i)$	Inter-domain communication between corresponding node CN and mobile router m_i
$C(CN, m_i)$	Overhead of inter-domain communication $S(CN, m_i)$
$CT(CN, m_i)$	Data transmission overhead of inter-domain communication $S(CN, m_i)$
$CU(CN, m_i)$	Location update overhead of inter-domain communication $S(CN, m_i)$
$S(m_i, m_j)$	Intra-domain communication between mobile routers m_i and m_j
$C(m_i, m_j)$	Overhead of intra-domain communication $S(m_i, m_j)$
$CT(m_i, m_j)$	Data transmission overhead of intra-domain communication $S(m_i, m_j)$
$CU(m_i, m_j)$	Location update overhead of intra-domain communication $S(m_i, m_j)$
S_D	Size of data in Bytes
S_H	Size of an IPv6 header in Bytes
S_B	Size of a binding update message in Bytes
B_w	Bandwidth of wired channel in Bps
B_{wl}	Bandwidth of wireless channel in Bps
$\alpha(t_i)$	Number of coming calls destined for a domain in time slot t_i
$\beta(t_i)$	Number of passed AP regions of a mobile network within time slot t_i
$CMR(t_i)$	Call-to-mobility ratio of a mobile network within time slot t_i
μ	Threshold of call-to-mobility ratio
AP	access point
BU	binding update
CMR	call-to-mobility ratio
CN	corresponding node
CoA	care-of-address
HA	home agent
MR	mobile router
NBSP	network mobility basic support protocol
NEMO	network mobility
SMR	session-to-mobility ratio
TLMR	top level mobile router
HoA	home-address

and location update overhead. In this chapter, overhead specifically refers to the time cost [31]. The data transmission overhead is the end-to-end delay of data transfer. The location update overhead is the required time for updating the location information maintained in HAs, when MRs move into a new AP region [32]. Denote the overhead of an inter-domain communication $S(CN, m_i)$ as $C(CN, m_i)$. Based on previous discussion we note that, $C(CN, m_i)$ is equal to the sum of the data transmission overhead $CT(CN, m_i)$ and the location update overhead $CU(CN, m_i)$ as follows,

$$C(CN, m_i) = CT(CN, m_i) + CU(CN, m_i). \tag{3.1}$$

As Fig. 3.2 shows, an AP divides the route of inter-domain communication $S(CN, m_i)$ into wired part and wireless part. From CN to AP is the wired part, there are a lot of HAs in this part. Meanwhile, the wireless part refers to the region between AP and m_i, there are a lot of MRs in the wireless part. Let $CT(CN, AP)$ and $CT(AP, m_i)$ be the data transmission overhead in wired and wireless parts respectively. Then, the data transmission overhead of the inter-domain communication $CT(CN, m_i)$ can be calculated as follows,

$$CT(CN, m_i) = CT(CN, AP) + CT(AP, m_i). \tag{3.2}$$

Since the destination m_i locates at the l_i th ($l_i \in \mathbb{Z}^+$) level of the nested structure, data sent to m_i will pass through $l_i - 1$ MRs before reaching m_i. Correspondingly, data will pass through l_i HAs in the wired part. Each time when data passes through a HA, the HA will add an IPv6 header [33] to the data. Let S_D denote the size of data in Bytes, S_H denote the size of an IPv6 header in Bytes, B_w denote the bandwidth of the wired channel in Bps. As a result, the data transmission overhead in the wired part $CT(CN, AP)$ is calculated as follows,

$$CT(CN, AP) = \sum_{j=0}^{l_i} \left(\frac{S_D + j \cdot S_H}{B_w} \right). \tag{3.3}$$

In the wireless part, data will be forwarded from AP to m_i. Each time when data passes through an MR, the MR will remove an IPv6 header from the data. Let B_{wl} denote the bandwidth of wireless channel in Bps. As a result, the data transmission overhead in the wireless part $CT(AP, m_i)$ is calculated as follows,

$$CT(AP, m_i) = \sum_{j=1}^{l_i} \left(\frac{S_D + j \cdot S_H}{B_{wl}} \right). \tag{3.4}$$

As the domain moves into a new AP region (i.e., the TLMR connects to a new AP), the MRs inside this domain will obtain new CoAes. As soon as an MR obtains a new CoA, it sends a Binding Update (BU) message [34] to its HA to notify the location update. According to the received BU message, the HA will update the location information maintained in its binding cache. Note that the BU message has to be forwarded by several MRs in the wireless part. While in the wired part, the BU message can be sent from AP to the corresponding HA directly. Let S_B denote the size of a BU message in Bytes. Assume that the domain will pass through β new AP regions within the duration of communication $C(CN, m_i)$. As a result, the location update overhead $CU(CN, m_i)$ can be calculated as follows,

$$CU(CN, m_i) = \beta \cdot \left(\frac{S_B \cdot l_i}{B_{wl}} + \frac{S_B}{B_w} \right). \tag{3.5}$$

The root cause of pinball routing problem is that only the corresponding HA equips with the location information of an MR. The pinball routing problem will incur multiple encapsulation and decapsulation [35], which will increase the data packet size and end-to-end delay dramatically. For an inter-domain communication, the times of encapsulation is equal to the nesting levels of destination MR in the nested structure. In order to alleviate the pinball routing problem for inter-domain communication, we propose a self-adaptive route optimization scheme.

3.3 Proposed Scheme for Inter-domain Communication

We propose a self-adaptive route optimization scheme for inter-domain communication. In the proposed scheme, data experiences at most twice encapsulation regardless of the nesting levels of destination MR. In this section, we will provide an overview of the self-adaptive scheme at first. Then, we will explain the two sub-schemes respectively. Finally, we will discuss the calculation method of the optimal threshold.

3.3.1 Self-Adaptive Route Optimization Scheme

The self-adaptive route optimization scheme consists of two sub-schemes: *time-saving sub-scheme* and *mobility-transparency sub-scheme*. If the time-saving sub-scheme is applied, data will experience only once encapsulation in the wired part and once decapsulation in the wireless part. The time-saving sub-scheme can reduce the data transmission overhead to a great extent. However, since the movement of domain is not transparent to HAs, the location update overhead of time-saving sub-scheme is uncontrollable. This drawback can be overcame by using the mobility-transparency sub-scheme, at the cost of twice encapsulation. Given a scenario, the proposed scheme will adaptively adopt the most appropriate sub-scheme through the following three steps.

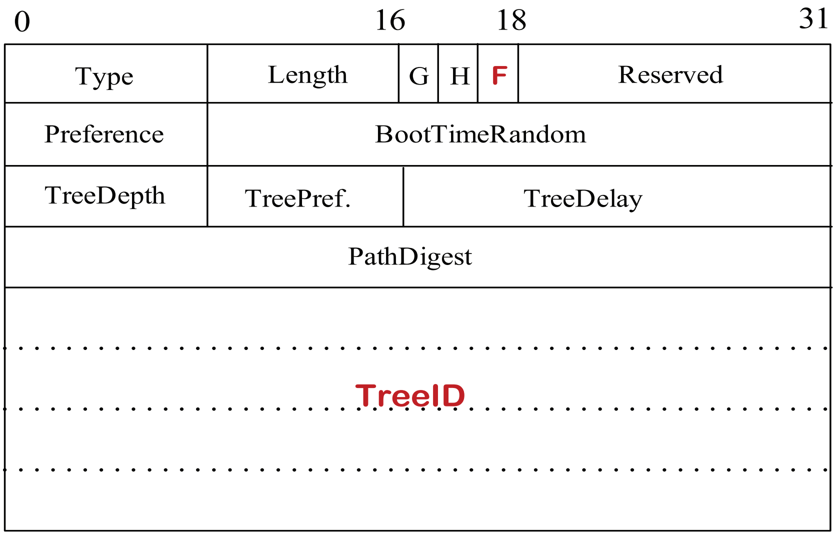


FIGURE 3.3: Structure of the modified tree information option

Step 1: Initialization. We modify the structure of the Tree Information Option (TIO) [36] in order to cooperate with the proposed scheme. The modified structure of the TIO is shown in Fig. 3.3, where the **TreeID** field is redefined, and the **F** field is new added. The value of **TreeID** can be the CoA of TLMR or the HoA of TLMR. Meanwhile, the value of **F** indicates the address categories of the **TreeID** field. If the value of **TreeID** is the CoA of TLMR, **F** equals to 1. Otherwise, **F** is 0.

Step 2: Calculate the Call-to-Mobility Ratio. We extend the definition of Call-to-Mobility Ratio (CMR) [37] from a mobile device [38] to a domain as Definition 3.3.1. By introducing the time-slotted idea, a continuous period of time is divided into a discrete time samples. In each time slot, TLMR counts the number of coming calls and the number of passed AP regions, then calculates the CMR of the domain.

Definition 3.3.1 (*call-to-mobility ratio of a domain*). In the time slot t_i ($i = 1, 2, \cdots$), let $\alpha(t_i)$ denote the number of coming calls destined for a domain, $\beta(t_i)$ denote the number of AP regions that the domain passes through within t_i. The call-to-mobility ratio of this domain in t_i is $CMR(t_i) = \alpha(t_i)/\beta(t_i)$.

Step 3: Choose a Sub-Scheme. After calculating the $CMR(t_i)$, TLMR compares the $CMR(t_i)$ with a threshold μ. According to the comparison result, TLMR sets the values of the **TreeID** field and **F** field in TIO. Then TLMR appends this TIO to a Route Advertisement (RA) message [39]. As the RA message is propagated downwards, each MR can get a copy of the **TreeID** from the appended TIO. During the location update process, each MR will replace the CoA information in the BU message with the **TreeID**, then send this modified BU message to its HA. As a result, each HA will maintain the correspondence between HoA of MR and **TreeID**. Different values of **TreeID** direct to the different sub-schemes.

(i) If $CMR(t_i) > \mu$, **TreeID** is set to be the CoA of TLMR, and the *time-saving sub-scheme* will be adopted. In this case, the domain is relatively stable and a large amount of calls come to it. Data transmission overhead is the major overhead of inter-domain communication. Reduce the communication overhead can improve the performance of inter-domain communication greatly.

(ii) If $CMR(t_i) \leq \mu$, **TreeID** is set to be the HoA of TLMR, and the *mobility-transparency sub-scheme* will be adopted. In this case, the domain moves fast. Each time when the domain moves into a new AP region, it leads to once location update. Since the standard location update process [40] will cost a lot time, the mobility transparency becomes more urgent.

3.3.2 Time-Saving Sub-Scheme

In order to explain how the time-saving sub-scheme works, we take the scenario shown in Fig. 3.4 as an example, where a corresponding node CN attempts to communicate with the MR m_5. As the HA of m_5, a_5 will intercept the data at first. In time-saving sub-scheme, information maintained by each home agent directs to the TLMR m_1 directly. In our example that is, the information maintained by a_5 is the CoA of TLMR c_1. According to this indication, a_5 encapsulates the received data and forwards the encapsulated data to m_1.

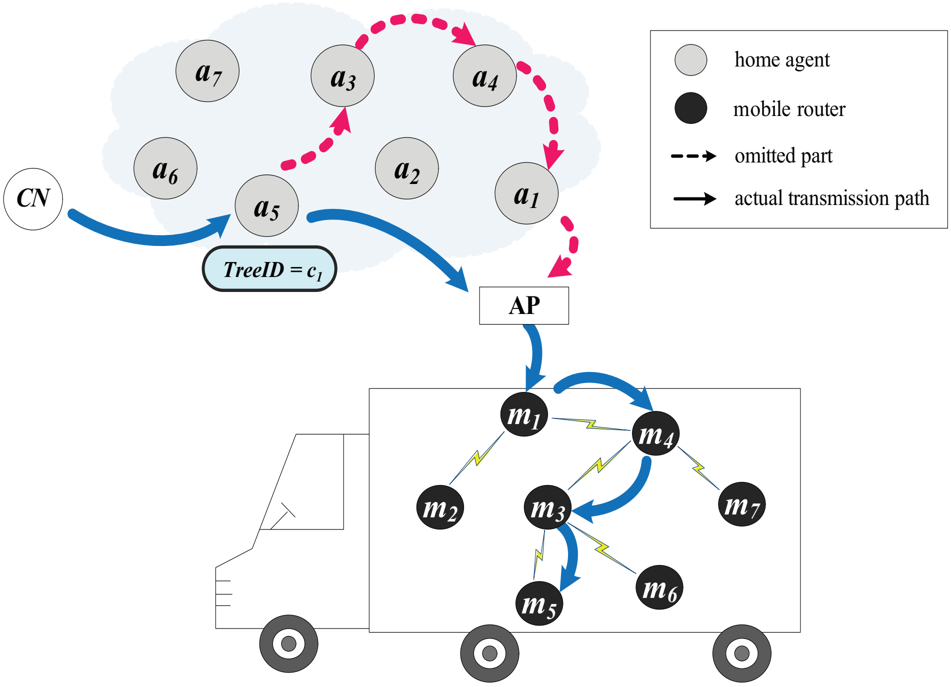


FIGURE 3.4: Illustration of the time-saving sub-scheme

After arriving at the domain, data will be forwarded to its destination m_5 level by level. At last, m_5 decapsulates the received data and get the original information. In this example, the nesting levels of destination MR m_5 are 4. If this inter-domain communication is performed in NBSP, data will experience 4 times encapsulation. However in our proposed time-saving sub-scheme, data needs only once encapsulation. This advantage will be more obvious as the nesting levels of destination MR increase. As this example demonstrates, the time-saving sub-scheme can reduce the data transmission overhead to the greatest extent. When the communication traffic is heavy, the time-saving sub-scheme improves the performance of inter-domain communication greatly. However, it does not means the time-saving sub-scheme is the optimal choice. Since each HA maintains the current location information of TLMR in the time-saving sub-scheme, this information needs to be updated whenever the domain moves into a new AP region. If the domain moves fast, the location update overhead of time-saving sub-scheme will be large.

3.3.3 Mobility-Transparency Sub-Scheme

In order to optimize the route of inter-domain communication for the high mobility scenarios, we propose the mobility-transparency sub-scheme. Consider the same scenario as the time-saving sub-scheme has, and the data transmission process of the mobility-transparency sub-scheme is shown in Fig. 3.5. In the mobility-transparency sub-scheme, each HA maintains the identifier of the TLMR. In our example that is, the information maintained by a_5 is the HoA of TLMR h_1. This information indicates that the destination MR m_5 belongs to the domain of m_1. In order to find m_5, we have to find the TLMR m_1 at first. Hence, a_5 will intercept the data destined for m_5, and encapsulate the received data with h_1, then send the encapsulated data out. As the HA of m_1, a_1 will intercept the encapsulated data and search its binding cache [41] to determine the location of m_1. Data will be encapsulated by a_1 once again, then sent to m_1. After arriving at the TLMR m_1, m_1 decapsulates the received data and forwards the data to the destination MR m_5.

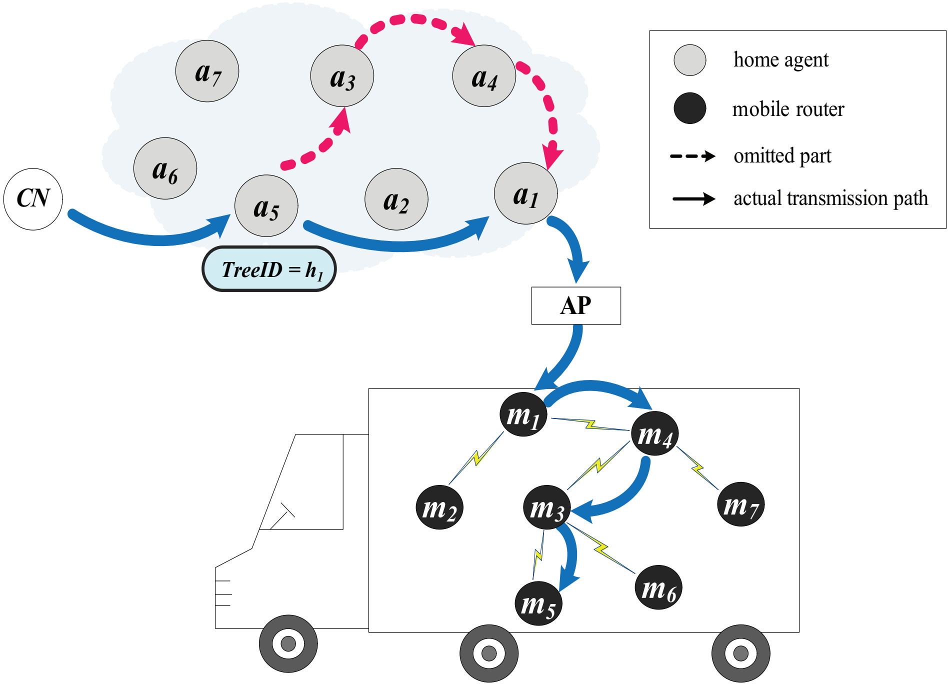


Figure 3.5: Illustration of the mobility-transparency sub-scheme

Compared with the time-saving sub-scheme, data of inter-domain communication needs once more encapsulation and decapsulation in the mobility-transparency sub-scheme. However the times of encapsulation and decapsulation can remain unchanged at 2, no matter how much are the nesting levels of destination MR. Moreover, since the information maintained by each HA (except for the HA of TLMR) dose not change when the domain moves, the location update overhead in mobility-transparency is pretty small. We make a comparison among the NBSP, the time-saving sub-scheme and the mobility-transparency sub-scheme in Table 3.2. Comparison results indicate that the proposed two sub-schemes can effectively reduce the data transmission and location update overheads of inter-domain communication. Moreover, the time-saving sub-scheme performs well for the low mobility and large communication traffic scenario, while the mobility-transparency sub-scheme performs well for the high mobility scenario. Given a scenario, the proposed self-adaptive route optimization scheme will make use of a threshold μ to determine which sub-scheme should be adopted.

TABLE 3.2: Comparison among Two Proposed Sub-Schemes and NBSP

	Data Transmission Overhead	**Location Update Overhead**
NBSP	the times of encapsulation and decapsulation are proportional to the nesting levels of destination MR	all of HAs need information update
Time-Saving Sub-Scheme	once encapsulation and once decapsulation	all of HAs need information update
Mobility-Transparency Sub-Scheme	twice encapsulation and twice decapsulation	only the HA of TLMR needs information update

3.3.4 Optimal Threshold μ

We proposed the *time-saving sub-scheme* and the *mobility-transparency sub-scheme* for different scenarios. These two sub-schemes have their own advantages and drawbacks. Given a scenario, the most appropriate sub-scheme should be adaptively adopted. The goals of the proposed self-adaptive scheme are to minimize both the data transmission overhead and the location update overhead. However, these two goals are conflicting. If we want to reduce the data transmission overhead of inter-domain communication, more detailed information (i.e., location information) has to be provided for HAs, which inevitably incurs larger location update overhead, and vice versa. In this subsection, we will discuss how to achieve the trade-off between these two goals and calculate the optimal CMR threshold.

For simplicity but without loss of generality, we assume that the time interval between any two coming calls follows an exponential distribution [42] with rate λ_c. Therefore, the mean of the time interval between any two calls is $\int_0^{+\infty} x\lambda_c e^{-\lambda_c x} d_x = 1/\lambda_c$. Let *ncall* denote the average number of coming calls per unit time, which is calculated as follows,

$$ncall = \frac{1}{1/\lambda_c} = \lambda_c. \tag{3.6}$$

For a whole domain, let CT_{TS} and CT_{MT} denote the data transmission overhead of inter-domain communication in the time-saving sub-scheme and the mobility-transparency sub-scheme respectively. Note that data of inter-domain communication in the mobility-transparency sub-scheme experiences once more encapsulation and decapsulation than it in the time-saving sub-scheme. For $ncall$ coming calls, the difference of data transmission overhead between the mobility-transparency sub-scheme and the time-saving sub-scheme is calculated as follows,

$$CT_{MT} - CT_{TS} = ncall \cdot \left(\frac{S_D + 2S_H}{B_w} + \frac{S_H}{B_{wl}} \right). \tag{3.7}$$

Furthermore we assume the number of AP regions which are crossed by the domain per unit time follows a Possion distribution [43] with rate λ_m. Let $nmob$ denote the average number of crossed AP regions per unit time, which is calculated by following equation,

$$nmob = \sum_{x=0}^{\infty} \left(x \frac{\lambda_m^x}{x!} e^{-\lambda_m} \right) = \lambda_m. \tag{3.8}$$

Each time when the domain crosses an AP region, it will incur once location update. For the whole domain, let CU_{TS} and CU_{MT} denote the location update overhead of inter-domain communication in the time-saving sub-scheme and the mobility-transparency sub-scheme respectively. If the mobility-transparency sub-scheme is adopted, only the HA of TLMR needs information update. TLMR sends a BU message to its HA to notify the new location information. For $nmob$ times location update, the location update overhead in the mobility-transparency sub-scheme CU_{MT} is calculated as follows,

$$CU_{MT} = nmob \cdot \left(\frac{S_B}{B_w} + \frac{S_B}{B_{wl}} \right). \tag{3.9}$$

If the time-saving sub-scheme is adopted, all of HAs (i.e., $a_i \in A, 1 \leq i \leq n$) need location update. For $nmob$ times location update, the location update overhead in the time-saving sub-scheme CU_{TS} is calculated as follows,

$$CU_{TS} = nmob \cdot \sum_{a_i \in A} \left(\frac{S_B}{B_w} + \frac{S_B \cdot l_i}{B_{wl}} \right). \tag{3.10}$$

According to Eqn. (3.9) and Eqn. (3.10), the difference of location update overhead for a whole domain between two sub-schemes is calculated as follows,

$$CU_{TS} - CU_{MT} = nmob \cdot \sum_{a_i \in A, i \neq 1} \left(\frac{S_B}{B_w} + \frac{S_B \cdot l_i}{B_{wl}} \right). \tag{3.11}$$

In the system description section 3.2, we defined a metric CMR to evaluate a scenario. When the CMR of a domain denoted by $CMR(t_i)$ is small, the mobility-transparency sub-scheme is more suitable.

As the value of $CMR(t_i)$ increases, the advantage of mobility-transparency sub-scheme decreases gradually. When the $CMR(t_i)$ increases to a certain value μ, the overhead of mobility-transparency sub-scheme is equal to the overhead of time-saving sub-scheme. That means compared with time-saving sub-scheme, the advantage in location update of mobility-transparency sub-scheme is offset by its disadvantage in communication. In this situation, we can get the following relation,

$$CT_{MT} - CT_{TS} = CU_{TS} - CU_{MT}. \tag{3.12}$$

After the $CMR(t_i)$ increases to bigger than the certain value μ, the huge data transmission overhead in mobility-transparency sub-scheme will dramatically degrade the performance of inter-domain communication. Therefore, the time-saving sub-scheme should be adopted.

According to above analysis we find that μ is the threshold of CMR. When $CMR(t_i)$ is small than μ, the mobility-transparency sub-scheme is more appropriate. As the $CMR(t_i)$ exceeds the threshold μ, the time-saving sub-scheme becomes more appropriate. Based on Eqn. (3.7), Eqn. (3.11) and Eqn. (3.12), the value of threshold μ can be calculated as follows,

$$\mu = \frac{ncall}{nmob} = \frac{B_w \cdot B_{wl} \sum\limits_{a_i \in A, i \neq 1} (S_B/B_w + S_B \cdot l_i/B_{wl})}{(S_D + 2S_H) \cdot B_{wl} + S_H \cdot B_w}. \tag{3.13}$$

3.4 Extension for Intra-domain Communication

The route of inter-domain communication has been optimized in the self-adaptive route optimization scheme. As we have discussed in Introduction, the pinball routing problem is more serious for intra-domain communication [44]. It is an unwise behavior that even the source and destination of an intra-domain communication are inside the same domain, data has to be sent out of the domain at first. Consider a scenario, there is a data exchange between two devices of a Personal Area Network (PAN) [45]. The privacy data should be limited inside the PAN, rather than sent to a device (HA) in the Internet. The root cause of the pinball routing problem is that only HAs equip with the location information of MRs. Hence, the breakthrough point of the pinball routing problem for intra-domain communication is to let more devices obtain the location information of MRs. In this section, we will extend the self-adaptive route optimization scheme for intra-domain communication.

In a domain, if an MR can receive a RA message from other MR, the MR which sent out this RA message is called the parent MR [46]. The extended scheme requires that each MR maintains the topology of its subtree. We make use of just once standard location update process to help MRs to construct the topologies of their subtrees. In the standard location update process, each MR will send a BU message to the corresponding HA to notify its new CoA. The BU message will be forwarded out of the domain by several MRs. Whenever an MR received a BU message, the MR will create an entry to record the received BU message in its binding cache. Then, the MR will check the size of received BU message. If the size of BU message shows that there are two CoAes, the MR will replace the first CoA with its own CoA then send the revised BU message upwards. If there is only one CoA, the MR will add its own CoA to this BU message then send the revised BU message upwards. After the BU message reaches TLMR, TLMR will remove the extra CoA from this BU message and send it to the corresponding HA. Through this process, each MR can construct the topology of its subtree. In the data

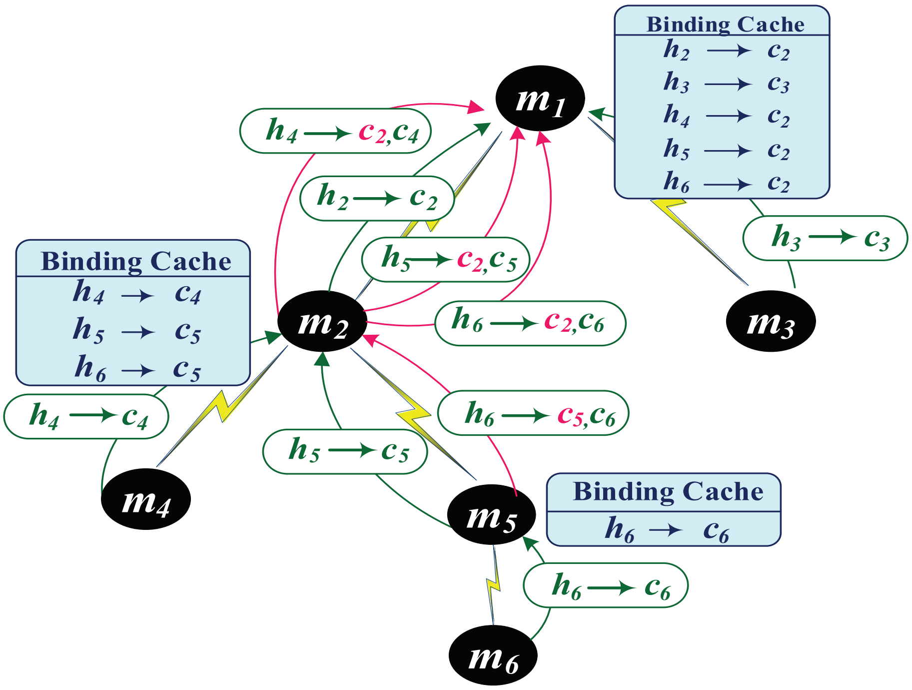


FIGURE 3.6: Route optimization for intra-domain communication

transmission process, when an MR receives data, the MR searches its binding cache for destination. If there exists a corresponding entry, the MR will forward the received data according to this corresponding entry. Otherwise, the MR sends the received data to its parent MR.

In order to explain the extended scheme more clearly, we considered a scenario shown in Fig. 3.6. Take the MR m_6 for instance, m_6 sends out a BU message in the location update process. As the parent MR of m_6, m_5 will receive this BU message. According to the received BU message, m_5 creates an entry (i.e., $h_6 \rightarrow c_6$) in its binding cache. This entry indicates that MR m_6 can be reached through c_6. Then, m_5 checks the size of this BU message and finds out that there is only one CoA in the BU message. Thus, m_5 adds its own CoA c_5 to the BU message, and sends the revised BU message upwards. The revised BU message will be received by m_2. Similar to m_5, m_2 creates an entry (i.e., $h_6 \rightarrow c_5$) in its binding cache which means the MR m_6 can be reached through c_5. After this, m_2 checks the size of this BU message and finds out that there are two CoAes in the BU message. As a result, m_2 replaces the first CoA (i.e., c_5) with its own CoA (i.e.,

c_2) and then sends the BU message upwards. After the BU message reaching TLMR, TLMR will remove the extra CoA filed (i.e., c_2) from this BU message and send the message to its corresponding HA. Through this kind of process, each MR can construct the topology of its subtree.

Consider an intra-domain communication between m_4 and m_6, where m_4 tries to send data to m_6. At first, m_4 will set the destination field of data to be the HoA of m_6 (i.e., h_6). As the parent MR of m_4, m_2 will receive the data. After receiving the data, m_2 checks its binding cache and finds out that the destination m_6 can be reached through c_5. Thus, m_2 forwards the received data to m_5. After the data reaching m_5, m_5 also checks its binding cache and finds out that the destination m_6 can be reached through c_6. Then, the data will be forwarded to its destination according to this indication. In order to explain the situation when there is no corresponding entry in the binding cache of an MR, we consider another intra-domain communication between m_4 and m_3. After receiving the data sent from m_4, m_2 finds out that there is no entry directing to the destination m_3 in its binding cache. As a result. m_2 will forward the received data to its parent MR m_1. m_1 searches its binding cache and finds that the destination m_3 can be reached through c_3. According to this indication, m_1 will forward the received data to the destination.

From these two examples we observe that the intra-domain communication can be limited inside a small scope in our extended scheme. This achievement not only alleviates the pinball routing problem for intra-domain communication, but also preserves the privacy of data [47]. Furthermore, the overhead of the extended scheme is very low. Once the nested structure is built, as long as the structure does not change, the information maintained by MRs does not need to be updated no matter how the domain moves. Moreover, the extended scheme just makes use of once standard location update process in NBSP, it dose not need to change the work flow of the standard location update process. A more prominent advantage is that the extended scheme will not introduce any extra control message.

TABLE 3.3: Experimental Parameters for Optimal Route Optimization Scheme

Parameter	**Value**
Simulator	MATLAB
Number of mobile nodes	10
Number of coming packets	2000
Size of a data packet S_D	512 Bytes
Bandwidth of wired channel B_w	100 Mbps
Bandwidth of wireless channel B_{wl}	10 Mbps
Size of a data packet S_D	500 Bytes
Size of a Binding Update (BU) message S_B	68 Bytes
Size of an IPv6 header S_H	40 Bytes
Threshold of Call-to-Mobility Ratio (CMR) μ	11.5
Confidence interval	95%

3.5 Performance Evaluation

In this section, we provide the performance evaluation of the proposed Self-Adaptive Route Optimization Scheme (AROS). We compare the proposed AROS with the Network mobility Basic Support Protocol (NBSP) [7] and another classic scheme the Mobile IPv6 [16] in the overheads of inter-domain and intra-domain communications. In this chapter, overhead refers to the time consumption that consists of both the data transmission overhead and location update overhead. The considered scenario is shown in Fig. 3.7. There are ten MRs (i.e., m_1, m_2, $\cdots$, m_{10}) and ten corresponding HAs (i.e., a_1, a_2, $\cdots$, a_{10}) in the considered scenario. According to the theoretical analysis made in Section 3.3, the CMR threshold μ of the considered scenario approaches to 11.5. When the CMR is smaller than or equal to the threshold μ, the mobility-transparency sub-scheme will be adopted. When the CMR is bigger than the threshold μ, the time-saving sub-scheme will be adopted. Some important experimental parameters are given in Table 3.3.

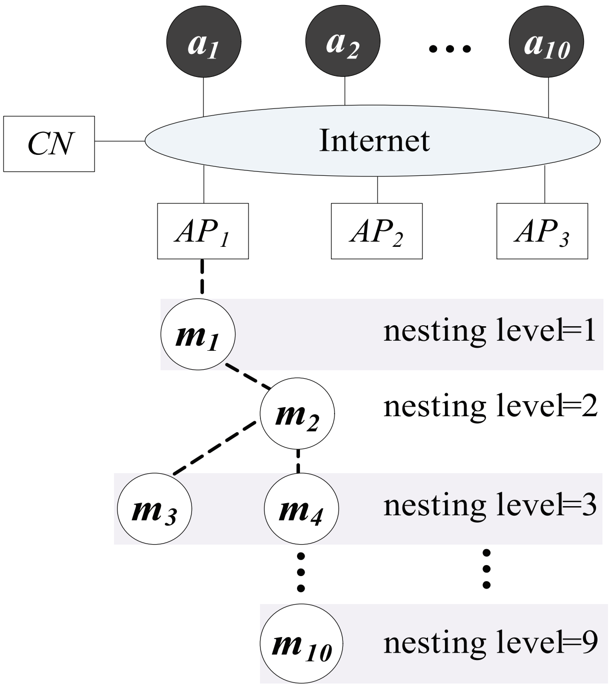
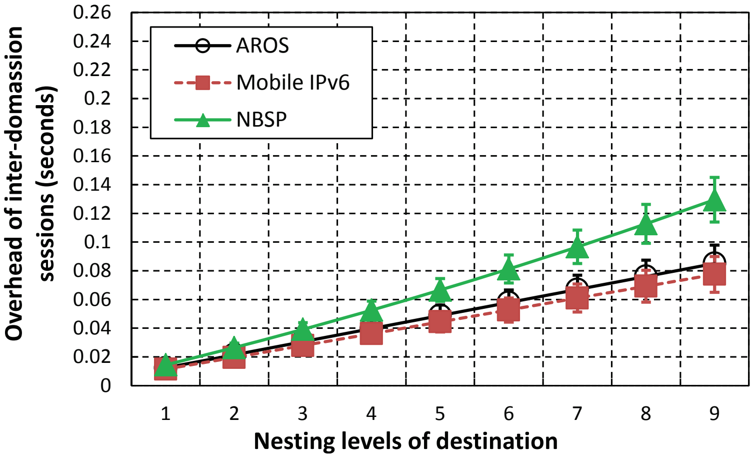


Figure 3.7: Network topology for experiment

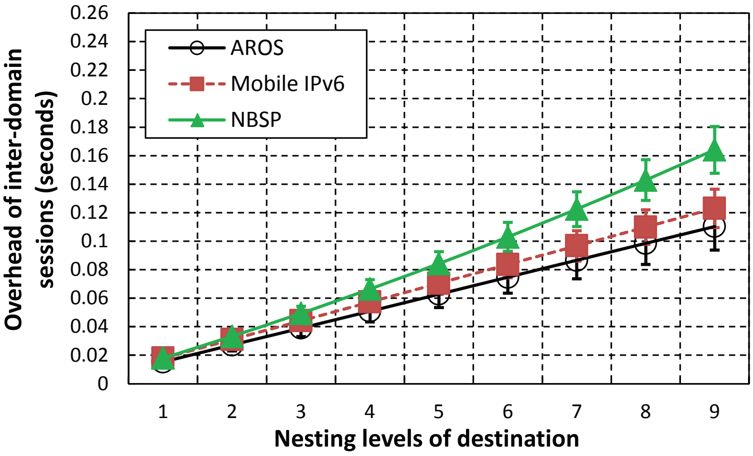
3.5.1 Overhead of inter-domain communication

In the considered scenario, we study the overheads of inter-domain communications between CN and MRs under different CMRs. From the experiment results shown in Fig. 3.8 we observe that our proposed scheme AROS always has the least overhead. As the nesting levels of destination increases, the overheads of inter-domain communication in three schemes will all increase. However, the proposed AROS has the slowest increment speed.

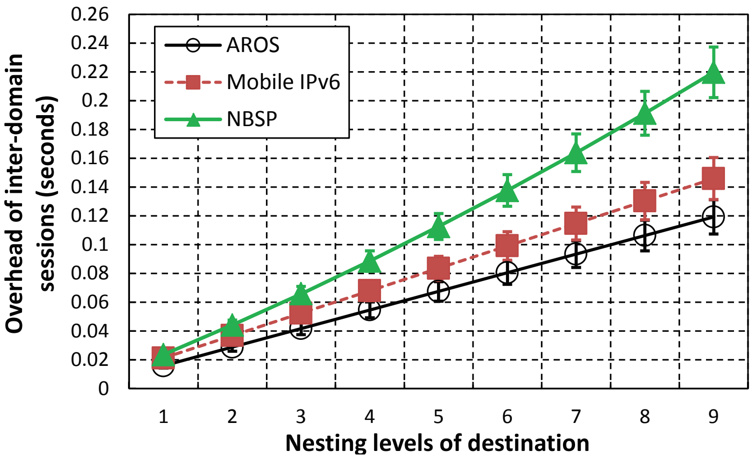
Furthermore, we maintain the number of coming packets per unit time to be 2000, and change the number of passing domains per unit time. From crosswise comparison we observe that when the CMR is smaller than or equal to 11.5, the overhead of inter-domain communication in Mobile IPv6 is very close to that in AROS. While the difference of inter-domain data transmission overhead between Mobile IPv6 and AROS will raise when the CMR increases from 11.5 to 13.5. The reason behind this phenomenon is the sub-scheme switching. When the CMR is smaller than or equal to the threshold



(a) call-to-mobility ratio = 9.5



(b) call-to-mobility ratio = 11.5



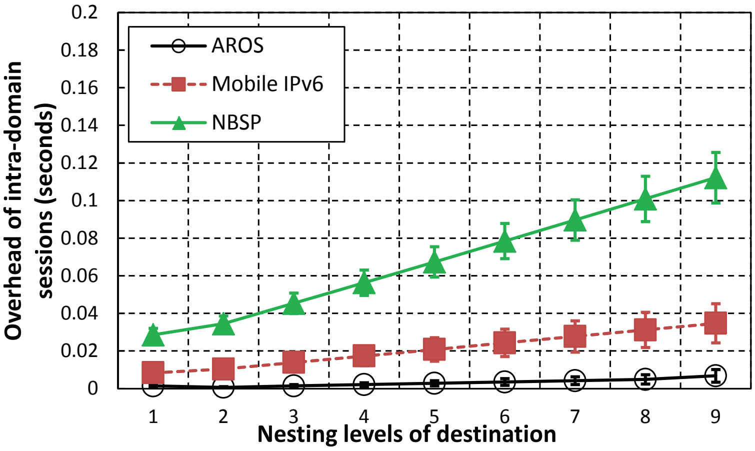
(c) call-to-mobility ratio = 13.5

FIGURE 3.8: Nesting levels vs. overhead of inter-domain communications.

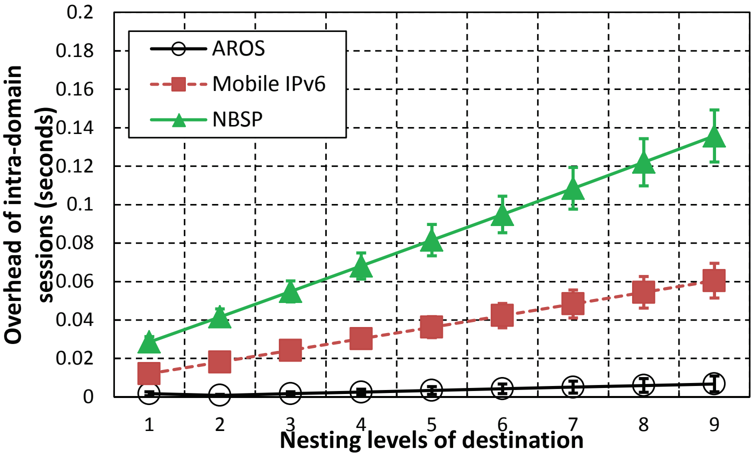
μ ($\mu = 11.5$), the mobility-transparency sub-scheme will be adopted in AROS. The data transmission processes in mobility-transparency sub-scheme is similar to Mobile IPv6, and their location update processes has a slight difference. There are two kinds of BU messages (*i.e.*, normal BU message and local BU message) will be sent out in the location update process of Mobile IPv6, while mobility-transparency sub-scheme works based on the standard location update process. When the CMR is bigger than the threshold μ, the time-saving sub-scheme will be adopted in AROS. Since data of inter-domain communication will experience twice encapsulation in Mobile IPv6 and once encapsulation in the time-saving sub-scheme, the data transmission overhead of inter-domain communication in the time-saving sub-scheme is less than in the Mobile IPv6. As the number of inter-domain communications increases, the advantage of AROS becomes more obvious.

3.5.2 Overhead of intra-domain communication

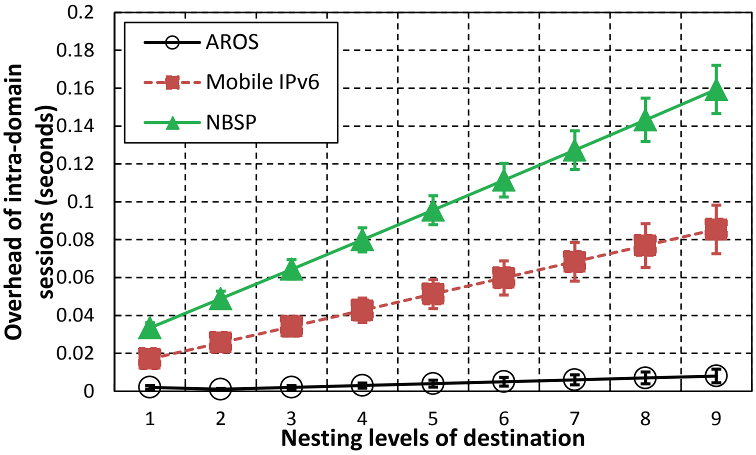
In the considered scenario, we study the overheads of intra-domain communications between m_3 and MRs under different CMRs. The experiment results are shown in Fig. 3.9. The main trend is that the overhead of intra-domain communication increases as the nesting levels of destination increase. Furthermore, the overheads of intra-domain communication in Mobile IPv6 and AROS are much less than in NBSP. In NBSP, data of intra-domain communication will be sent out of the domain at first, then sent back. While in AROS and Mobile IPv6, data of intra-domain communication can be limited inside the domain. In Mobile IPv6, data of intra-domain communication will be sent to the TLMR (i.e., m_1) at first, then sent to the destination according to the routing table in TLMR. In AROS, data of intra-domain communication only needs to be sent to the common parent MR of sender and destination at first. For example, when the nesting levels of destination are 9 (i.e., intra-domain communication between m_3 and m_{10}), data will be sent from m_3 to the common parent MR m_2 at first, then forwarded to the destination m_{10} level by level. The advantage of AROS will be more obvious if the sender and the destination of intra-domain communication are closer to their common parent MR.



(a) call-to-mobility ratio = 9.5



(b) call-to-mobility ratio = 11.5



(c) call-to-mobility ratio = 13.5

Figure 3.9: Nesting levels vs. overhead of intra-domain communications.

3.6 Summary

In this chapter, we investigated the pinball routing problem for nested mobile network mobility environment. In order to alleviate the heavy data transmission overhead and location update overhead caused by the pinball routing problem in inter-domain communication, we firstly proposed the self-adaptive route optimization scheme. The proposed scheme consists of two sub-schemes: time-saving sub-scheme and mobility-transparency sub-scheme. The time-saving sub-scheme can effectively minimize the data transmission overhead of inter-domain communication, which is suitable for the low mobility and large communication traffic scenario. While the mobility-transparency sub-scheme can reduce the location update overhead to the greatest extent, which is suitable for the high mobility scenario. A threshold is used to adaptively adopt the most appropriate sub-scheme for a given scenario. Furthermore, the self-adaptive route optimization scheme is extended for intra-domain communication. In the extended scheme, MRs make use of once standard location update process to construct the topologies of their subtrees. When an MR receives data, it will check whether the destination belongs to its subtree or not. Hence, the intra-domain communication can be limited in a pretty small region. Consequently, the communication and location update overheads of intra-domain are reduced.

Chapter 4

A Multi-Objective Distributed Network Selection Scheme for Hybrid 5G Environments

In this chapter, we propose a distributed network selection scheme for hybrid 5G environments. The network selection problem is formulated as a multi-objective optimization problem which maximizes the channel capacity and minimizes the block probability simultaneously. By taking the throughput metric into consideration, the formulated multi-objective optimization problem is transformed into a maximization problem. We solve the transformed maximization problem to calculate the network selection result in a distributed method. The calculated network selection result is proved to be a Pareto Optimal solution of the original multi-objective optimization problem. The proposed scheme guarantees that based on limited local information, each user can select a new network device with high channel capacity and low block probability. Comprehensive experiment results show that the proposed scheme promotes the total throughput and user served ratio significantly.

This chapter is organized as follows. Section 4.1 introduces the motivation of the distributed network selection scheme for hybrid 5G environments. The system description and problem formulation are given in Section 4.2. In Section 4.3, we study the estimation method of block probability. Section 4.4 presents our proposed network selection scheme for hybrid 5G environments. Section 4.5 is the performance evaluation. Finally, we summarize this chapter in Section 4.6.

4.1 Introduction

The emergence of 5G will not replace the existing technologies[1] but be more integrative and hybrid: combining with existing technologies to provide ubiquitous high-rate and seamless communication service [8]. As we move toward 5G era, environment becomes so complex that the handoff problem faces with new challenges. The data rate in 5G is expected to be roughly 1000× compared with current 4G technology [48], hence the handoff problem requires a faster processing [49]. Furthermore, as the number of Base Stations (BSs) and mobile devices dramatically increases, the centralized control may not be efficient. On the contrary, more intelligent mobile devices can play important roles in handoff. Moreover, increasingly serious data security problem reminds users[2] do not share their private information with others. Thus, it is glad to see a fast, distributed, privacy-preservation and user centered handoff scheme in hybrid 5G environments. Motivated by this, we will study the handoff problem for hybrid 5G environments in this chapter.

Consider a scenario as shown in Fig. 4.1 where 3G [50], LTE, WiMAX and 5G BSs construct a hybrid 5G environment. Users in the hybrid 5G environment do not share their private information with others. Moving in this scenario, users may need to transfer their network connections from one BS to another. This kind of transferring operation is called handoff [37]. The handoff problem refers that when a user has several available BSs in a handoff, the user needs to decide to which BS the network connection should be transferred [51]. Take a user for instance. As the user moves far away from 3G BS, the signal strength received from 3G BS gets so weak that the user has to transfer his (or her) network connection to a new BS. This user has three possible choices: LTE, 5G and WiMAX BSs [52]. He (or she) has to decide which BS should be selected. It seems that

---

[1] The existing technologies include 3G, LTE, and so on.

[2] The terms *user* and *mobile device* are interchangeable in the chapter.

the handoff problem is very simple, the user only needs to select the best performance one. However, the user has difficulties to know the network selection behaviors of other users. If there are too many other users making the same selection, this user is possible to be blocked [53], [54]. As a result, the objectives of network selection are to select a high performance BS and avoid being blocked.

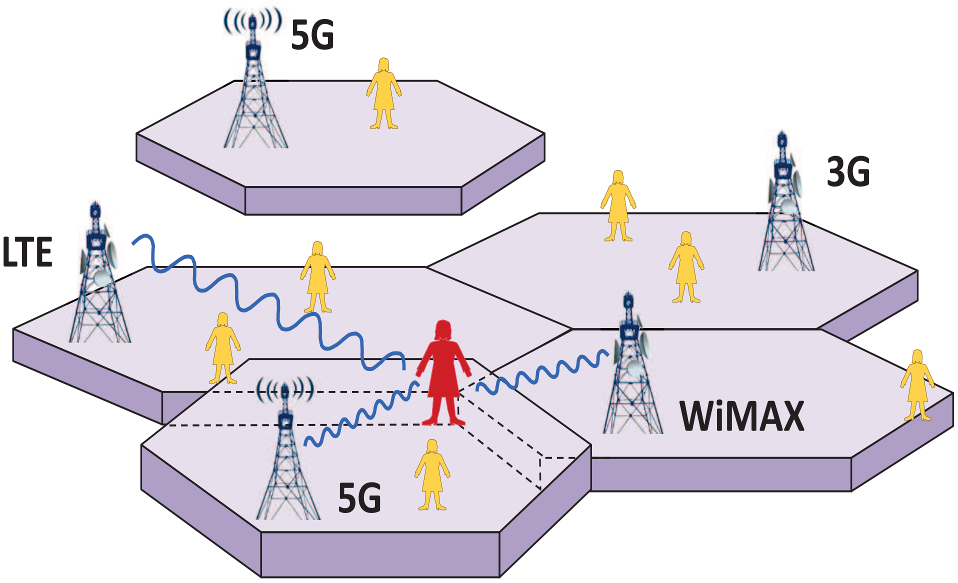


Figure 4.1: Illustrative example for network selection problem.

There are two kinds of approaches to solve the handoff problem in general: the network centered approach and the user centered approach. In the network centered approach, networks are responsible for computing and making the decisions. In the user centered approach, users will be in charge of the network selection. Considering the requirement of privacy-preservation in hybrid 5G environment, users are not suggested to send out their private information (*e.g.*, number of available networks, basic bandwidth requirement and so on) [55]. Under this limitation, networks are unable to obtain adequate information from users for the network selection. As a result, the user centered approach is more suitable for the hybrid 5G environment than the network centered approach.

In this chapter, we propose a user centered multi-objective handoff scheme for hybrid 5G environment. In our proposed scheme, users are divided into two classes: *non-handoff* *users* and *handoff users*. Non-handoff users will stay in the connections with their current BSs. While handoff users will transfer their network connections to new BSs based on limited local information. Local information refers to the private information of the user itself, the parameters of BSs and two pieces of public information (*i.e.*, the total numbers of handoff and non-handoff users inside each available BS). When a user needs to select a new BS in a handoff, it will calculate the achievable data receiving rates of all its available BSs. Furthermore, the user also has to **infer** the network selection behaviors of other users in order to **estimate** its block probability for each available BS. By jointly considering the achievable data receiving rate and block probability, the user can select the most appropriate BS in a handoff. The main contributions of this chapter are summarized as follows:

- We study the relations between two users, and define the *correlation degree*. The correlation degree could efficiently distinguish the categories of relations, and sufficiently reflect the association strength.
- We formulate the handoff problem as a multi-objective optimization problem which maximizes the achievable data receiving rate and minimizes the block probability. Then, we transform the formulated multi-objective optimization problem into an equivalent maximization problem.
- We solve the transformed maximization problem by a distributed method in polynomial time and linear space. We further prove that the solution of the transformed maximization problem is a Pareto Optimal [56] result of the original multi-objective optimization problem.

4.2 System Description and Problem Formulation

In this section, we formulate the handoff problem for hybrid 5G environments. Specifically, we consider a hybrid 5G environment which consists of n BSs. Let $\mathcal{B}$ be the set of BSs, $\mathcal{B} = \{b_1, b_2, \cdots, b_n\}$. These BSs support different wireless technologies. With the support of the Media-Independent Handover (MIH) standard [57], we can focus on the handoff problem from the perspective of algorithm without caring about the differences between communication technologies. Denote the frequency band of BS b_i ($b_i \in \mathcal{B}, i = 1, 2, \cdots, n$) as ω_i in MHz. b_i equally allocates its frequency band among serving users. In order to guarantee the quality of service of each user, b_i will serve at most η_i users at the same time.

Consider that there are m users. Let $\mathcal{U}$ be the set of users, $\mathcal{U} = \{u_1, u_2, \cdots, u_m\}$. If user u_j ($u_j \in \mathcal{U}, j = 1, 2, \cdots, m$) is inside the coverage area of a BS, this BS is called an **available BS** of u_j. An **adjacency matrix** $\delta(t)$ is used to reflect the available relationship between BSs and users at time t as follows. By introducing the time-slotted idea [58], a continuous period of time is divided into discrete time samples. In the rest of chapter, time t is referred to the t th time slot. The system status in a time slot is assumed to be stable.

$$\delta(t) = \begin{array}{c} \\ \boldsymbol{b_1} \\ \boldsymbol{b_2} \\ \vdots \\ \boldsymbol{b_n} \end{array} \begin{array}{c} \begin{array}{cccc} \boldsymbol{u_1} & \boldsymbol{u_2} & \cdots & \boldsymbol{u_m} \end{array} \\ \begin{bmatrix} \delta_{11}(t) & \delta_{12}(t) & \dots & \delta_{1m}(t) \\ \delta_{21}(t) & \delta_{22}(t) & \dots & \delta_{2m}(t) \\ \vdots & & \ddots & \vdots \\ \delta_{n1}(t) & \delta_{n2}(t) & \dots & \delta_{nm}(t) \end{bmatrix} \end{array},$$

where

$$\delta_{ij}(t) = \begin{cases} 1, & \text{BS } b_i \text{ is available to user } u_j \text{ at time } t, \\ 0, & \text{otherwise.} \end{cases} \tag{4.1}$$

For BS b_i, the number of users inside its coverage area is $\sum_{j=1}^{m} \delta_{ij}(t)$ which should satisfy the following constraint.

$$0 \leq \sum_{j=1}^{m} \delta_{ij}(t) \leq m. \tag{4.2}$$

For user u_j, the number of available BSs is $\sum_{i=1}^{n} \delta_{ij}(t)$. In hybrid 5G environment, u_j may have several available BSs. That is the value of $\sum_{i=1}^{n} \delta_{ij}(t)$ should satisfy the

Table 4.1: Notation Summary in Chapter 4

n	Number of base stations
$\mathcal{B}$	Set of base stations $\{b_i\}$, $i = 1, 2, \cdots, n$
ω_i	The frequency band of base station b_i in MHz
ω_i'	The bandwidth that a user can get from base station b_i in MHz
η_i	The maximum users that base station b_i can serve simultaneously
m	Number of users
$\mathcal{U}$	Set of users $\{u_j\}$, $j = 1, 2, \cdots, m$
$\delta(t)$	Adjacency matrix of $\mathcal{B}$ and $\mathcal{U}$ at time t, i.e., $[\delta_{ij}(t)]$
$\theta(t)$	Conjunction matrix of $\mathcal{B}$ and $\mathcal{U}$ at time t, i.e., $[\theta_{ij}(t)]$
γ_{ij}	Basic bandwidth requirement of user u_j for base station b_i in Mbps
$s_{ij}(t)$	Received signal power of user u_j from base station b_i at time t in watts
$d_{ij}(t)$	Euclidean distance between base station b_i and user u_j at time t
ρ_i	Transmission power of base station b_i in watts
h_{ij}	Channel fading gain of channel (b_i, u_j)
λ	Pass loss exponent
ζ^2	Background additive white Gaussian noise in watts
$g_{ij}(t)$	The interference caused by base station b_i to user u_j in watts at time t
$q_{ij}(t)$	The achievable data receiving rate of user u_j from base station b_i at time t in Mbps
$v_j(t)$	Identifier that if user u_j at time t is a handoff user or a non-handoff user
$\mathcal{V}(t)$	$(v_j(t))$, $j = 1, 2, \cdots, m$
$\mathcal{B}_j(t)$	Set of available base stations for user u_j at time t, i.e., $\{b_{j_i}\}$, $\mathcal{B}_j(t) \subseteq \mathcal{B}$
$\mathcal{F}_j(t)$	Network selection result of handoff user u_j at time t, i.e., (f_{j_ij}), $\lvert\mathcal{F}_j(t)\rvert = \lvert\mathcal{B}_j(t)\rvert$
$\mathcal{P}_j(t)$	Block probabilities of user u_j for available base stations at time t, i.e., (p_{j_ij}), $\lvert\mathcal{P}_j(t)\rvert = \lvert\mathcal{B}_j(t)\rvert$
$Q_j(t)$	Achievable data receiving rates provided by available base stations for user u_j at time t, i.e., (q_{j_ij}), $\lvert Q_j(t)\rvert = \lvert\mathcal{B}_j(t)\rvert$
$\Theta_i(t)$	Number of non-handoff users which are connecting to base station b_i at time t
$\Delta_i(t)$	Number of hand-off users inside the coverage area of base station b_i at time t
$\alpha_{ij}(t)$	Probability that base station b_i will be selected by handoff user u_j at time t
$\beta_{ij}(t)$	Probability inferred by u_j that base station b_i will be selected by another handoff user at time t
$\mathbb{P}_j(r_i(t))$	Probability inferred by u_j that there are $r_i(t)$ other handoff users who have selected b_i as their new base station at time t
$\tau_{ij}(t)$	Throughput of channel (b_i, u_j) at time t in Mbps
ε	The maximal moving velocity of user in m/s

following constraint.

$$0 \leq \sum_{i=1}^{n} \delta_{ij}(t) \leq n. \tag{4.3}$$

Although user u_j has several available BSs, it can connect to at most one of its available BSs at any time. The connected available BS is called the **current BS** of user u_j. A **conjunction matrix** $\theta(t)$ is used to reflect the connected relationship between BSs and users at time t as follows.

$$\theta(t) = \begin{array}{c} \\ \boldsymbol{b_1} \\ \boldsymbol{b_2} \\ \vdots \\ \boldsymbol{b_n} \end{array} \begin{array}{c} \begin{array}{cccc} \boldsymbol{u_1} & \boldsymbol{u_2} & \cdots & \boldsymbol{u_m} \end{array} \\ \begin{bmatrix} \theta_{11}(t) & \theta_{12}(t) & \dots & \theta_{1m}(t) \\ \theta_{21}(t) & \theta_{22}(t) & \dots & \theta_{2m}(t) \\ \vdots & & \ddots & \vdots \\ \theta_{n1}(t) & \theta_{n2}(t) & \dots & \theta_{nm}(t) \end{bmatrix} \end{array},$$

where

$$\theta_{ij}(t) = \begin{cases} 1, & \text{current BS } b_i \text{ is connected by user } u_j \text{ at time } t, \\ 0, & \text{otherwise.} \end{cases} \tag{4.4}$$

For BS b_i, the number of serving users is $\sum_{j=1}^{m} \theta_{ij}(t)$ which should satisfy the following constraint.

$$0 \leq \sum_{j=1}^{m} \theta_{ij}(t) \leq min\left(\eta_i, \sum_{j=1}^{m} \delta_{ij}(t)\right). \tag{4.5}$$

Each serving user can get $\omega_i'(t)$ MHz bandwidth from BS b_i at time t. The value of $\omega_i'(t)$ is calculated as follows.

$$\omega_i'(t) = \frac{\omega_i}{\sum_{j=1}^{m} \theta_{ij}(t)}, \forall i, \forall j, 0 \leq \theta_{ij}(t) \leq \delta_{ij}(t). \tag{4.6}$$

Since user u_j can connect to at most one current BS, the number of current BS $\sum_{i=1}^{n} \theta_{ij}(t)$ should satisfy the following constraint.

$$0 \leq \sum_{i=1}^{n} \theta_{ij}(t) \leq min\left(1, \sum_{i=1}^{n} \delta_{ij}(t)\right). \tag{4.7}$$

For each BS-user pair (b_i, u_j), assume that the received signal power of user u_j from available BS b_i at time t is $s_{ij}(t)$ in watts. Let $d_{ij}(t)$ denote the Euclidean distance between BS b_i and user u_j at time t. When b_i transmits a signal for each channel with power ρ_i in watts, $s_{ij}(t)$ is then calculated as follows.

$$s_{ij}(t) = \delta_{ij}(t) \cdot \rho_i \cdot h_{ij} \cdot d_{ij}(t)^{-\lambda}, \tag{4.8}$$

where the channel fading gain h_{ij} follows an exponential distribution with rate μ ($h_{ij} \sim exp(\mu)$), and the pass loss exponent $\lambda \geq 2$ (varies depending on channel conditions).

Since different BSs are assumed to use different frequency bands, there is no interference among BSs. For 5G supported BS which utilizes the Orthogonal Frequency Division Multiple Access (OFDMA, also commonly applied to LTE, WiMAX and IEEE 802.11 b/g supported devices) to avoid the interference among users. For those BSs which do not utilize the OFDMA, some techniques such as Code Division Multiple Access (CDMA, commonly applied to 3G devices) and orthogonal codes are assumed to be used in order to waken the interference among users. Let $g_{xj}(t)$ in watts be the interference caused by BS b_x ($b_x \in \mathcal{B}, x \neq i$) to user u_j at time t, where b_x transmits signal by using the same frequency as user u_j. The value of g_{xj} can be calculated as follows.

$$g_{xj}(t) = \rho_x \cdot h_{xj} \cdot d_{xj}(t)^{-\lambda}, \tag{4.9}$$

where $d_{xj}(t)$ is the Euclidean distance between BS b_x and user u_j at time t. According to the Shannon theorem, the achievable data receiving rate of user u_j from BS b_i at time t denoted by $q_{ij}(t)$ in Mbps is calculated as follows.

$$q_{ij}(t) = \omega_i'(t) \cdot \log\left[1 + \frac{s_{ij}(t)}{\sum_{b_x \in \mathcal{B}, x \neq i} g_{xj}(t) + \zeta^2}\right], \tag{4.10}$$

where ζ^2 is the background additive white Gaussian noise (AWGN).

In hybrid 5G environment, different kinds of BSs are assumed with different basic bandwidth requirements. Let γ_{ij} denote the basic bandwidth requirement of user u_j for BS b_i in Mbps. Suppose that the current BS of user u_j is b_c. If the achievable data receiving rate from b_c cannot meet the basic bandwidth requirement ($q_{cj}(t) < \gamma_{cj}$), user u_j will perform handoff. We call these users who need to perform handoff **handoff users**. If the achievable data receiving rate can satisfy the basic bandwidth requirement ($q_{cj}(t) \geq \gamma_{cj}$), user u_j will stay in the connection with its current BS b_c. We call these users who do not need handoff **non-handoff users**. A vector $\mathcal{V}(t) = (v_1(t), v_2(t), \cdots v_m(t))$ is used to identify the kinds of users. The value of $v_j(t)$ is given as follows, where $j = 1, 2, \cdots, m$.

$$v_j(t) = \begin{cases} 0, & \text{user } u_j \text{ is a handoff user at time } t, \\ 1, & \text{user } u_j \text{ is a non-handoff user at time } t. \end{cases} \tag{4.11}$$

Let $\Delta_i(t)$ be the number of **handoff users** which are **inside** the coverage area of BS b_i at time t. The value of $\Delta_i(t)$ is then calculated as follows.

$$\Delta_i(t) = \sum_{j=1}^{m} \left\{\delta_{ij}(t) \cdot [1 - v_j(t)]\right\}. \tag{4.12}$$

Let $\Theta_i(t)$ be the number of **non-handoff users** which are **connecting** to BS b_i at time t. The value of $\Theta_i(t)$ is then calculated as follows.

$$\Theta_i(t) = \sum_{j=1}^{m} \left[\theta_{ij}(t) \cdot v_j(t)\right]. \tag{4.13}$$

Note that there is no centralized control entity. Users perform network selection in a distributed way. Furthermore, users are assumed do not share their private information (such as the number of available BSs, channel capacities, and so on) for privacy preservation. Therefore, each user has to make its own network selection based on local information. Local information is acquired by a user including the private information of itself, parameters of BSs and two pieces of public information (*i.e.*, $\Delta_i(t)$ and $\Theta_i(t)$).

Users have a lot of ways to obtain the public information, such as BSs periodically broadcast, device-to-device communication and standard location update. At the beginning of each time slot, users can send Hello messages to their available BSs to announce their presences. After collecting these Hello messages, BSs count the number of handoff and non-handoff users, then broadcast the values. This procedure can be enhanced through the device-to-device communications in some special scenarios [59]: those devices which have already known the public information can notify their neighbors about the public information. In order to further reduce the overhead and information refresh time, BSs can make use of the location update processes provided by the communication standards (*e.g.*, GSM 03.12 [60], 3GPP TS 23.012 [61], Mobile IP [62], [63]). By embedding the Hello message and public information into the Channel Request, Immediate Assignment and other control frames, the overhead and refresh time will be reduced to a very low level even can be neglected [64].

Let $\mathcal{B}_j(t) = \{b_{j_1}, b_{j_2}, \cdots, b_{j_k}\}$ be the set of available BSs for user u_j at time t, where $\mathcal{B}_j(t) \subseteq \mathcal{B}$, $k = |\mathcal{B}_j(t)| = \sum_{i=1}^{n} \delta_{ij}(t)$. Since u_j is a handoff user at time t, it has to select **a new** BS from $\mathcal{B}_j(t)$. Let $\mathcal{F}_j(t) = (f_{j_1 j}(t), f_{j_2 j}(t), \cdots, f_{j_k j}(t))$ be the network selection result of user u_j at time t, where $|\mathcal{F}_j(t)| = |\mathcal{B}_j(t)|$. The value of $f_{j_i j}(t)$ is given as follows, where $i = 1, 2, \cdots, k$.

$$f_{j_i j}(t) = \begin{cases} 1, & \text{new BS } b_{j_i} \text{ is selected by } u_j \text{ at time } t, \\ 0, & \text{otherwise.} \end{cases} \tag{4.14}$$

For handoff user u_j, let $Q_j(t) = (q_{j_1 j}(t), q_{j_2 j}(t), \cdots, q_{j_k j}(t))$ be the achievable data receiving rates provided by its available BSs, where $|Q_j(t)| = |\mathcal{B}_j(t)|$. Hence, the achievable data receiving rate that u_j can obtain from its new BS is $\mathcal{F}_j(t) \cdot [Q_j(t)]^T$ in Mbps, where $[Q_j(t)]^T$ is the transposition of $Q_j(t)$.

$$\mathcal{F}_j(t) \cdot [Q_j(t)]^T = \sum_{i=1}^{k} [f_{j_i j}(t) \cdot q_{j_i j}(t)] \,. \tag{4.15}$$

The achievable data receiving rate provided by new BS should satisfy the basic bandwidth requirement of user u_j for the new BS. That is, the value of $\mathcal{F}_j(t) \cdot [Q_j(t)]^T$

should be subject to the following constraint.

$$\mathcal{F}_j(t) \cdot [Q_j(t)]^T \geq \sum_{i=1}^{k} [f_{j_i j}(t) \cdot \gamma_{j_i j}(t)] \,. \tag{4.16}$$

If no available BS can satisfy the above constraint, the handoff of u_j will fail. In order to guarantee the quality of experience of other users, u_j will be discarded by its **current** BS. For handoff user u_j, its available BS b_{j_i} ($b_{j_i} \in \mathcal{B}_j(t)$) can serve at most η_{j_i} users simultaneously. Since there are $\Theta_{j_i}(t)$ non-handoff users connecting to BS b_{j_i} at time t, b_{j_i} can serve at most $\eta_{j_i} - \Theta_{j_i}(t)$ handoff users. Note that time t refers to the t th time slot. Handoff requests will come to a BS successively during a time slot. If there are more than $\eta_{j_i} - \Theta_{j_i}(t)$ handoff users that have chosen b_{j_i} as their new BS at time t, the after coming handoff requests will be blocked. These blocked handoff users will wait in a First-In-First-Out (FIFO) queue.

Let $p_{j_i j}(t)$ be the probability that handoff user u_j is blocked, when it tries to handoff to the BS b_{j_i} at time t. The calculation method of $p_{j_i j}(t)$ will be given in Section 4.3. Let $\mathcal{P}_j(t) = (p_{j_1 j}(t), p_{j_2 j}(t), \cdots, p_{j_k j}(t))$. Then, the block probability of u_j for its new BS is $\mathcal{F}_j(t) \cdot [\mathcal{P}_j(t)]^T$, where $[\mathcal{P}_j(t)]^T$ is the transposition of $\mathcal{P}_j(t)$.

$$\mathcal{F}_j(t) \cdot [\mathcal{P}_j(t)]^T = \sum_{i=1}^{k} [f_{j_i j}(t) \cdot p_{j_i j}(t)] \,. \tag{4.17}$$

For a single handoff user, the objectives of its network selection are to maximize the achievable data receiving rate provided by the new BS, and to minimize the block probability. We theoretically formulate the handoff problem as a multi-objective optimization problem as follows.

$$\begin{aligned} \mathcal{O}_1 &= \text{Maximize } \; \mathcal{F}_j(t) \cdot [Q_j(t)]^T \\ \mathcal{O}_2 &= \text{Minimize } \; \mathcal{F}_j(t) \cdot [\mathcal{P}_j(t)]^T \end{aligned} \tag{4.18}$$

subject to

$$\sum_{i=1}^{k} f_{j_i j}(t) + v_j(t) \leq 1, \; j \in \mathbb{Z}^+, \; 1 \leq j \leq m, \tag{4.19a}$$

$$\sum_{i=1}^{k} [f_{j_i j}(t) \cdot \gamma_{j_i j}(t)] \leq \sum_{i=1}^{k} [f_{j_i j}(t) \cdot q_{j_i j}(t)] \,, \tag{4.19b}$$

$$b_{j_i} \in \mathcal{B}_j(t), \quad \mathcal{B}_j(t) \subseteq \mathcal{B}, \quad k = |\mathcal{B}_j(t)| \,. \tag{4.19c}$$

The first constraint Eqn. (4.19a) indicates that a non-handoff user ($v_j(t) = 1$) does not have any new BS and a handoff user ($v_j(t) = 0$) has at most one new BS. The second constraint Eqn. (4.19b) guarantees that the achievable data receiving rate provided by the new BS can satisfy the basic bandwidth requirement of a handoff user. The last constrain Eqn. (4.19c) reveals that the network selection of a handoff user should be implemented within its available BS set.

4.3 Block Probability Estimation

Based on limited local information, each handoff user tries to select a new BS which can provide the maximal achievable data receiving rate and minimal block probability. The calculation method of achievable data receiving rate has been given in Section 4.2. In this section, we will explain the estimation method of block probability.

4.3.1 Relations Between Users

The block probability of a handoff user relates to the network selection behaviors of other handoff users. However, under the premise of privacy preservation, a user has no idea of other handoff users. The calculation of block probability relies on the **inferences** made by a handoff user to other handoff users. In order to assist a handoff user in inferring the network selection behaviors of other handoff users, we study the relations between handoff users in this subsection.

In hybrid 5G environments, each handoff user has several available BSs. We investigate the relations between any two handoff users based on their available BS sets. In general, the relations of a pair of handoff users can be divided into two categories: independent relation and correlated relation.

Definition 4.3.1 (*The independent relation of a pair of handoff users*). Let (u_i, u_j) denote any pair of handoff users. Their available network sets at time t are $\mathcal{B}_i(t)$ and $\mathcal{B}_j(t)$ respectively. If $|\mathcal{B}_i(t) \cap \mathcal{B}_j(t)| = 0$, u_i and u_j have the independent relation. □

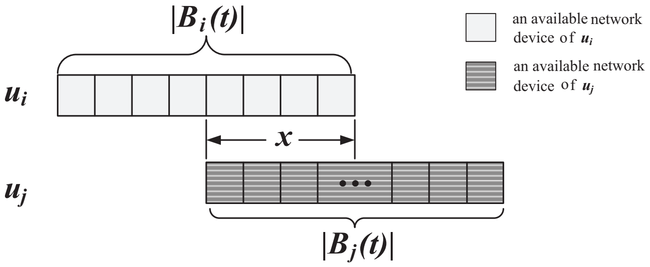


FIGURE 4.2: The correlation degree of (u_i, u_j).

When u_i and u_j are independent, the network selection behavior of u_i has no direct impact on u_j, and vice versa. Hence in the handoff process, a user only needs to consider those users who are in the correlated relations.

Definition 4.3.2 (*The correlated relation of a pair of handoff users*). For any pair of handoff users (u_i, u_j), if there is at least one BS which is available to both of them, then they have the correlated relation. Consequently, u_i and u_j are mutually neighbors. □

In order to reflect the strength of correlated relation, we define the *correlation degree* as follows.

Definition 4.3.3 (*The correlation degree of a pair of handoff users*). (u_i, u_j) is any pair of handoff users, their available BS sets are $\mathcal{B}_i(t)$ and $\mathcal{B}_j(t)$ respectively. The correlation degree of (u_i, u_j) is the probability that when selecting a BS from $\mathcal{B}_i(t) \cup \mathcal{B}_j(t)$, the selected BS is available to both u_i and u_j. □

Let $\mathcal{L}(u_i, u_j)$ denote the correlation degree of (u_i, u_j). Suppose that $| \mathcal{B}_i(t) \cap \mathcal{B}_j(t) | = x$ as shown in Fig. 4.2, then $\mathcal{L}(u_i, u_j)$ can be calculated by the following equation,

$$\mathcal{L}(u_i, u_j) = \frac{x}{|\mathcal{B}_i(t)| + |\mathcal{B}_j(t)| - x},$$
$$\text{where} \quad x \in \mathbb{Z}^+, \ 0 < x \leq min(|\mathcal{B}_i(t)|, |\mathcal{B}_j(t)|). \tag{4.20}$$

Note that if u_i and u_j are independent, the value of x is 0, and the correlation degree $\mathcal{L}(u_i, u_j) = 0$. If u_i and u_j are correlated, $\mathcal{L}(u_i, u_j) \in (0, 1]$. As a result, we can extend the defining field of x in Eqn. (4.20) to $[0, min(|\mathcal{B}_i(t)|, |\mathcal{B}_j(t)|)]$ and use just one metric *correlation degree* to distinguish the categories of relations, and reflect the strength of association. The *correlation degree* metric also has the following attribute.

Theorem 4.3.1: For any pair of handoff users (u_i, u_j), their available BS sets are $\mathcal{B}_i(t)$ and $\mathcal{B}_j(t)$ respectively. If the correlation degree $\mathcal{L}(u_i, u_j) = 1$, $\mathcal{B}_i(t)$ is equal to $\mathcal{B}_j(t)$.

Proof. Based on Eqn. (4.20), if $\mathcal{L}(u_i, u_j) = 1$, then $2x = |\mathcal{B}_i(t)| + |\mathcal{B}_j(t)|$. For the correlation degree $\mathcal{L}(u_i, u_j) = 1$, we have the following cases.

Case 1: $|\mathcal{B}_i(t)| < |\mathcal{B}_j(t)|$, *i.e.*, $min(|\mathcal{B}_i(t)|, |\mathcal{B}_j(t)|) = |\mathcal{B}_i(t)|$. Substitute this equation into the constraint $x \leq min(|\mathcal{B}_i(t)|, |\mathcal{B}_j(t)|)$ of Eqn. (4.20), we can get that $2x \leq |\mathcal{B}_i(t)| + |\mathcal{B}_i(t)|$. Since $2x = |\mathcal{B}_i(t)| + |\mathcal{B}_j(t)|$, $|\mathcal{B}_j(t)|$ should be not bigger than $|\mathcal{B}_i(t)|$ which contradicts with the premise of Case 1. That is Case 1 will not happen when the correlation degree $\mathcal{L}(u_i, u_j) = 1$.

Case 2: $|\mathcal{B}_i(t)| > |\mathcal{B}_j(t)|$, *i.e.*, $min(|\mathcal{B}_i(t)|, |\mathcal{B}_j(t)|) = |\mathcal{B}_j(t)|$. Similar to the previous case, we can get that $2x \leq |\mathcal{B}_j(t)| + |\mathcal{B}_j(t)|$. Since we already know that $2x = |\mathcal{B}_i(t)| + |\mathcal{B}_j(t)|$, then $|\mathcal{B}_i(t)|$ should be smaller than or equal to $|\mathcal{B}_j(t)|$ which contradicts with the premise of Case 2. That is Case 2 will not happen when the correlation degree $\mathcal{L}(u_i, u_j) = 1$.

For the relationship between $\mathcal{B}_i(t)$ and $\mathcal{B}_j(t)$, we have excluded $|\mathcal{B}_i(t)| < |\mathcal{B}_j(t)|$ and $|\mathcal{B}_i(t)| > |\mathcal{B}_j(t)|$ through the above discussions. Therefore, $|\mathcal{B}_i(t)| = |\mathcal{B}_j(t)|$. Furthermore, from the Definition 4.3.3 we observed that $x = | \mathcal{B}_i(t) \cap \mathcal{B}_j(t) | = |\mathcal{B}_i(t)| = |\mathcal{B}_j(t)|$ when the correlation degree $\mathcal{L}(u_i, u_j) = 1$. As a result, the available BS sets $\mathcal{B}_i(t)$ and $\mathcal{B}_j(t)$ are completely overlapping when their correlation degree is equal to 1. Theorem 4.3.1 is proved. $\square$

Here, we want to explain the reason that why we specially proposed and studied the *correlation degree* metric in this subsection. Remember that, we investigate handoff user relations for the purpose of assisting a handoff user to infer the network selection behaviors of other handoff users. It requires a metric which can reflect the relation between handoff users. Thus, we proposed the *correlation degree* metric in Definition 4.3.3. During the behavior inference, since a handoff user does not know any private information of other handoff users, the handoff user will consider the worst case (*i.e.*, the correlation degree is 1) to be on the safe side. Through Theorem 4.3.1 we observed that the available BS sets of two handoff users will be completely overlapping in the worst case. This conclusion is meaningful since a handoff user can infer the network selection behaviors of other handoff users based on its own available BS set.

4.3.2 Behaviors Inference

For a handoff user u_j ($u_j \in \mathcal{U}$), since u_j has no idea of other handoff users, these handoff users are indistinguishable for u_j. We use u to represent an arbitrary one of them. In order to estimate its block probability for each available BS, u_j has to infer the network selection behavior of u [65].

Suppose that BS b_i is available to handoff user u_j at time t (*i.e.*, $\delta_{ij}(t) = 1$). There are two conditions needed to be satisfied simultaneously, if u_j is blocked when it tries to handoff to b_i [25]. These two conditions are: 1) u_j selects b_i as the new BS in a handoff; 2) before u_j tries to handoff to b_i, b_i is already full load.

For the first condition, we assume that u_j selects BSs based on their achievable data receiving rates. The larger achievable data receiving rate, the higher probability to be selected. As a result, the BS b_i will be selected as the new BS by u_j at time t with the probability $\alpha_{ij}(t)$ as follows.

$$\alpha_{ij}(t) = \frac{q_{ij}(t)}{\sum_{k=1}^{n} [\delta_{kj}(t) \cdot q_{kj}(t)]}. \tag{4.21}$$

During the network selection behavior inference, u_j always considers the worst case with the other handoff user u (*i.e.*, $\mathcal{L}(u, u_j) = 1$). According to Theorem 4.3.1 we can get that the available BS sets of u_j and u are completely overlapping in the worst case. Since u_j does not know the private information of u (such as how much achievable data receiving rate that u can obtain from each available BS, the specific location of u, and so on), u_j has no choice but to assume that u selects BS based on the remaining bandwidth. Note that the bandwidth of BS b_i is ω_i MHz. Moreover, there are $\Theta_i(t)$ non-handoff users are connecting to BS b_i at time t. Each non-handoff users will occupy $\omega'_i(t)$ MHz bandwidth of b_i. As a result, the remaining bandwidth of b_i is $\omega_i - \omega'_i(t) \cdot \Theta_i(t)$ MHz. Hence, u_j infers that b_i will be selected as the new BS by u at time t with the probability $\beta_{ij}(t)$ as follows.

$$\beta_{ij}(t) = \frac{\omega_i - \omega'_i(t) \cdot \Theta_i(t)}{\sum_{k=1}^{n} \delta_{kj}(t) \cdot \left[\omega_k - \omega'_k(t) \cdot \Theta_k(t)\right]}. \tag{4.22}$$

Note that there are $\Delta_i(t)$ handoff users (including u_j) inside the coverage area of BS b_i at time t. Let $\mathbb{P}_j(r_i(t))$ denote the probability that before u_j, there are $r_i(t)$ handoff users that have chosen b_i as their new BS at time t. The value of $\mathbb{P}_j(r_i(t))$ is calculated as follows.

$$\mathbb{P}_j(r_i(t)) = \binom{\Delta_i(t)-1}{r_i(t)} \cdot \beta_{ij}(t)^{r_i(t)} \cdot (1-\beta_{ij}(t))^{\Delta_i(t)-1-r_i(t)},$$
$$\text{where } \binom{x}{y} = \frac{x!}{y! \cdot (x-y)!}. \tag{4.23}$$

If u_j is blocked when it tries to handoff to the new BS b_i, that means b_i has been full load. As a result, the value of $r_i(t)$ should satisfy the following constraint.

$$\eta_i - \Theta_i(t) \leq r_i(t) \leq \Delta_i(t) - 1. \tag{4.24}$$

Based on its private information, parameters of BSs and two pieces of public information (*i.e.*, the number of non-handoff users $\Theta_i(t)$ and the number of handoff users $\Delta_i(t)$), handoff user u_j estimates its block probability for BS b_i at time t denoted by $p_{ij}(t)$ as follows.

$$p_{ij}(t) = \alpha_{ij}(t) \cdot \sum_{r_i(t)=\eta_i-\Theta_i(t)}^{\Delta_i(t)-1} \mathbb{P}_j(r_i(t)). \tag{4.25}$$

4.4 Proposed Network Selection Scheme

4.4.1 Scheme Detail

The network selection problem is formulated as a multi-objective optimization problem. For most of multi-objective optimization problems, there dose not exist a solution which simultaneously optimizes each objective. In our scheme, a handoff user is unable to find a network device which exactly provides maximal channel capacity and minimal block probability simultaneously either. However, our proposed scheme is able to find a Pareto Optimal [56] network selection result for the formulated multi-objective optimization problem.

Suppose that the solution space of our multi-objective optimization problem is the shaded area in Fig. 4.3. Two different points A and B satisfy our two objectives $\mathcal{O}_1$ and $\mathcal{O}_2$ respectively. The line between points A and B indicates the Pareto Optimal solutions of our multi-objective optimization problem. A network selection result is Pareto Optimal if and only if there does not exist another network selection result which promotes at least one objective without demoting any one objective. In this subsection, we will explain how to solve the formulated multi-objective optimization problem and find a Pareto Optimal network selection result.

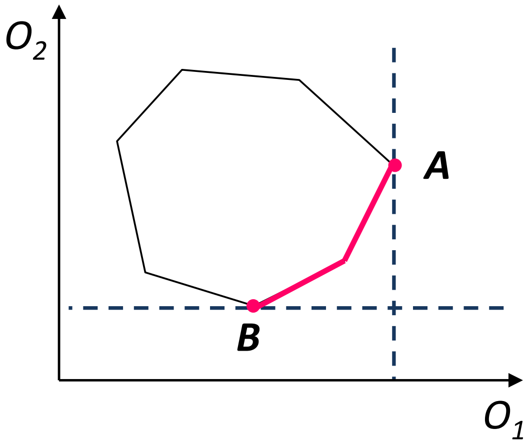


Figure 4.3: Solution space of a multi-objective optimization problem.

By taking the *throughput* metric into consideration, we firstly transform the original multi-objective optimization problem into a maximization problem. As an available network device of handoff user u_j, network device a_{j_i} is tagged with two attributes: channel capacity denoted by $q_{j_ij}(t)$ and block probability denoted by $p_{j_ij}(t)$. Let $\tau_{j_ij}(t)$ be the throughput of channel (a_{j_i}, u_j) at time t in Mbps. If a_{j_i} is not selected as the new network device by handoff user u_j at time t (i.e., $f_{j_ij}(t) = 0$), $\tau_{j_ij}(t) = 0$. If a_{j_i} is selected as the new network device (i.e., $f_{j_ij}(t) = 1$) but handoff user u_j is blocked in a_{j_i}, $\tau_{j_ij}(t) = 0$. If a_{j_i} is selected as the new network device and u_j successfully get the network service, $\tau_{j_ij}(t) = q_{j_ij}(t)$. In summary, the value of throughput $\tau_{j_ij}(t)$ is calculated as follows.

$$\tau_{j_ij}(t) = f_{j_ij}(t) \cdot q_{j_ij}(t) \cdot [1 - p_{j_ij}(t)] \,. \tag{4.26}$$

Note that, the *throughput* metric involves both of two attributes (i.e., channel capacity and block probability) what a handoff user is concerned. Furthermore, the throughput metric is proportional to the channel capacity attribute and inversely proportional to the block probability attribute. Thus, it is reasonable to substitute the following objective $\mathcal{O}_3$ for the original multiple objectives $\mathcal{O}_1$ and $\mathcal{O}_2$ under the same constraints listed in Eqn. (4.19).

$$\mathcal{O}_3 = \text{Maximize} \sum_{i=1}^{k} \{f_{j_ij}(t) \cdot q_{j_ij}(t) \cdot [1 - p_{j_ij}(t)]\} \,. \tag{4.27}$$

By solving the maximization problem (i.e., $\mathcal{O}_3$), a handoff user can select a new network device. We will prove that this selected new network device is a Pareto Optimal solution of the original multi-objective optimization problem (i.e., $\mathcal{O}_1$ and $\mathcal{O}_2$) through the following theorem.

Theorem 4.4.1: The solution of the transformed maximization problem is a Pareto Optimal result of the original multi-objective optimization problem.

Proof. (*Reductio ad absurdum*) Denote the network selection result of handoff user u_j at time t by solving the transformed maximization problem $\mathcal{O}_3$ as $\mathcal{F}^{\circ}_j(t) = (f^{\circ}_{j_1j}(t), f^{\circ}_{j_2j}(t), \cdots, f^{\circ}_{j_kj}(t))$, where k is the number of available network devices. Network device a_{j_x} is the new network device in ${\mathcal{F}^{\circ}}_j(t)$. That is, ${f^{\circ}}_{j_xj}(t) = 1$ and other elements in ${\mathcal{F}^{\circ}}_j(t)$ are 0.

Assume that ${\mathcal{F}^{\circ}}_j(t)$ is not a Pareto Optimal solution of the original multi-objective optimization problem $\mathcal{O}_1$ and $\mathcal{O}_2$. There exists another network selection result ${\mathcal{F}^{*}}_j(t) = \Big({f^{*}}_{j_1j}(t), {f^{*}}_{j_2j}(t), \cdots, {f^{*}}_{j_kj}(t)\Big)$ which promotes at least one objective without demoting any one objective. Moreover, ${\mathcal{F}^{*}}_j(t)$ cannot be worked out by solving $\mathcal{O}_3$, and a_{j_y} is the new network device in ${\mathcal{F}^{*}}_j(t)$ $(y \neq x)$. That is, ${f^{*}}_{j_yj}(t) = 1$ and other elements are 0. Consider the following cases.

Case 1: ${\mathcal{F}^{*}}_j(t)$ promotes $\mathcal{O}_1$ and maintains $\mathcal{O}_2$, i.e.,

$$\begin{cases} \sum_{i=1}^{k} \Big[{f^{*}}_{j_ij}(t) \cdot q_{j_ij}(t)\Big] > \sum_{i=1}^{k} \Big[{f^{\circ}}_{j_ij}(t) \cdot q_{j_ij}(t)\Big], \\ \sum_{i=1}^{k} \Big[{f^{*}}_{j_ij}(t) \cdot p_{j_ij}(t)\Big] = \sum_{i=1}^{k} \Big[{f^{\circ}}_{j_ij}(t) \cdot p_{j_ij}(t)\Big]. \end{cases}$$

Note that except for ${f^{*}}_{j_yj}(t)$ and ${f^{\circ}}_{j_xj}(t)$, other elements in ${\mathcal{F}^{*}}_j(t)$ and ${\mathcal{F}^{\circ}}_j(t)$ are 0. Hence we can get that,

$$\begin{cases} {f^{*}}_{j_yj}(t) \cdot q_{j_yj}(t) > {f^{\circ}}_{j_xj}(t) \cdot q_{j_xj}(t) \geq 0, \\ 1 \geq {f^{*}}_{j_yj}(t) \cdot p_{j_yj}(t) = {f^{\circ}}_{j_xj}(t) \cdot p_{j_xj}(t) \geq 0. \end{cases}$$

Furthermore since ${f^{*}}_{j_yj}(t) = {f^{\circ}}_{j_xj}(t) = 1$, ${f^{*}}_{j_yj}(t) \cdot q_{j_yj}(t) \cdot \big[1 - p_{j_yj}(t)\big] \geq {f^{\circ}}_{j_xj}(t) \cdot q_{j_xj}(t) \cdot [1 - p_{j_xj}(t)]$. Add the remaining terms whose values are 0 to this inequality, we can get that $\sum_{i=1}^{k} \Big\{{f^{*}}_{j_ij}(t) \cdot q_{j_ij}(t) \cdot [1 - p_{j_ij}(t)]\Big\} \geq \sum_{i=1}^{k} \Big\{{f^{\circ}}_{j_ij}(t) \cdot q_{j_ij}(t) \cdot [1 - p_{j_ij}(t)]\Big\}$.

Case 2: ${\mathcal{F}^{*}}_j(t)$ maintains $\mathcal{O}_1$ and promotes $\mathcal{O}_2$, i.e.,

$$\begin{cases} \sum_{i=1}^{k} \Big[{f^{*}}_{j_ij}(t) \cdot q_{j_ij}(t)\Big] = \sum_{i=1}^{k} \Big[{f^{\circ}}_{j_ij}(t) \cdot q_{j_ij}(t)\Big], \\ \sum_{i=1}^{k} \Big[{f^{*}}_{j_ij}(t) \cdot p_{j_ij}(t)\Big] < \sum_{i=1}^{k} \Big[{f^{\circ}}_{j_ij}(t) \cdot p_{j_ij}(t)\Big]. \end{cases}$$

Note that except for ${f^*}_{j_yj}(t)$ and ${f^\circ}_{j_xj}(t)$, other elements in ${\mathcal{F}^*}_j(t)$ and ${\mathcal{F}^\circ}_j(t)$ are 0. Hence we can get that,

$$\begin{cases} 0 \leq {f^*}_{j_yj}(t) \cdot q_{j_yj}(t) = {f^\circ}_{j_xj}(t) \cdot q_{j_xj}(t), \\ 0 \leq {f^*}_{j_yj}(t) \cdot p_{j_yj}(t) < {f^\circ}_{j_xj}(t) \cdot p_{j_xj}(t) \leq 1. \end{cases}$$

Furthermore since ${f^*}_{j_yj}(t) = {f^\circ}_{j_xj}(t) = 1$, ${f^*}_{j_yj}(t) \cdot q_{j_yj}(t) \cdot \left[1 - p_{j_yj}(t)\right] \geq {f^\circ}_{j_xj}(t) \cdot q_{j_xj}(t) \cdot [1 - p_{j_xj}(t)]$. Similar to Case 1 we can get that $\sum_{i=1}^{k} \left\{ {f^*}_{j_ij}(t) \cdot q_{j_ij}(t) \cdot [1 - p_{j_ij}(t)] \right\} \geq \sum_{i=1}^{k} \left\{ {f^\circ}_{j_ij}(t) \cdot q_{j_ij}(t) \cdot [1 - p_{j_ij}(t)] \right\}$.

Case 3: ${\mathcal{F}^*}_j(t)$ promotes both $\mathcal{O}_1$ and $\mathcal{O}_2$, i.e.,

$$\begin{cases} \sum_{i=1}^{k} \left[{f^*}_{j_ij}(t) \cdot q_{j_ij}(t) \right] > \sum_{i=1}^{k} \left[{f^\circ}_{j_ij}(t) \cdot q_{j_ij}(t) \right], \\ \sum_{i=1}^{k} \left[{f^*}_{j_ij}(t) \cdot p_{j_ij}(t) \right] < \sum_{i=1}^{k} \left[{f^\circ}_{j_ij}(t) \cdot p_{j_ij}(t) \right]. \end{cases}$$

Similar to Case 1 and Case 2 we can get that ${f^*}_{j_yj}(t) \cdot q_{j_yj}(t) \cdot \left[1 - p_{j_yj}(t)\right] > {f^\circ}_{j_xj}(t) \cdot q_{j_xj}(t) \cdot [1 - p_{j_xj}(t)]$. Add the remaining terms whose values are 0 to this inequality, we can get that $\sum_{i=1}^{k} \left\{ {f^*}_{j_ij}(t) \cdot q_{j_ij}(t) \cdot [1 - p_{j_ij}(t)] \right\} > \sum_{i=1}^{k} \left\{ {f^\circ}_{j_ij}(t) \cdot q_{j_ij}(t) \cdot [1 - p_{j_ij}(t)] \right\}$.

We consider all possible cases and get a conclusion that is, $\sum_{i=1}^{k} \left\{ {f^*}_{j_ij}(t) \cdot q_{j_ij}(t) \cdot \left[1 - p_{j_ij}(t)\right] \right\}$ is bigger than or equal to $\sum_{i=1}^{k} \left\{ {f^\circ}_{j_ij}(t) \cdot q_{j_ij}(t) \cdot \left[1 - p_{j_ij}(t)\right] \right\}$. However since ${\mathcal{F}^\circ}_j(t)$ is the solution of $\mathcal{O}_3$ and ${\mathcal{F}^*}_j(t)$ is not, the former summation should be smaller than the latter summation. As a result, the assumption that ${\mathcal{F}^\circ}_j(t)$ is not Pareto Optimal is failed. Theorem 4.4.1 is proved. $\square$

At the beginning of each time slot, users compare the channel capacities of their current network devices with the corresponding basic bandwidth requirements, and decide whether to implement the vertical handoffs or not. If a user does not need vertical handoff (non-handoff user), it will stay in the connection with its current network device. Otherwise, the user (handoff user) will select a new network device through Algorithm 1. In a real system, Algorithm 1 will be executed on mobile terminals. It is necessary to consider the general limitations of mobile terminals, such as small storage space and low processing capacity. Hence, we will analyze the computation and memory complexities of Algorithm 1 through Theorem 4.4.2.

Algorithm 1: Steps of the Proposed Network Selection Scheme for handoff User u_j

```
Input: available network device set at time $t$ $\mathcal{A}_j(t)$; for $\forall a_{j_i} \in \mathcal{A}_j(t)$: number of
       channels $l_{j_i}$, bandwidth per channel $b_{j_i}$, number of handoff and non-handoff
       users $\Gamma_{j_i}(t)$ and $\Theta_{j_i}(t)$, received signal power $s_{j_ij}(t)$, noise interference power
       $n_{j_ij}(t)$ and basic bandwidth requirement $\gamma_{j_ij}$.
Output: network selection result $\mathcal{F}_j(t)$.
1  $max = 0, index = 0$;
2  for $\forall a_{j_i} \in \mathcal{A}_j(t)$ do
3      Calculate the channel capacity $q_{j_ij}(t)$ by Eqn. (4.10);
4      Estimate the block probability $p_{j_ij}(t)$ by Eqn. (4.25);
5      if $q_{j_ij}(t) \geq \gamma_{j_ij}$ then
6          Calculate the throughput $\tau_{j_ij}(t)$ by Eqn. (4.26);
7          if $\tau_{j_ij}(t) \geq max$ then
8              $\tau_{j_ij}(t) \rightarrow max$;
9              the index of the selected network device $index = i$;
10         end
11     end
12 end
13 for $i = 1; i \leq |\mathcal{A}_j(t)|; i{+}{+}$ do
14     if $i == index$ then
15         the selected network device is $a_{j_i}$, $f_{j_ij}(t) = 1$;
16     end
17     else
18         $f_{j_ij}(t) = 0$;
19     end
20     return $\mathcal{F}_j(t)$;
21 end
```

Theorem 4.4.2: The computation complexity of the proposed scheme is $O(mn)$, the memory complexity of the proposed scheme is $O(n)$.

Proof. The major computational work of Algorithm 1 consists of three parts: calculate the channel capacities of available network devices (Line 3); estimate the block probabilities for available network devices (Line 4); scan the available network devices and calculate their throughput (Line 6), then determine the new network device (from Line 5 to 15).

Consider a scenario which has m users and n network devices. The first part is just a numerical calculation, its computation complexity is $O(n)$. For the second part, the computation complexity of $\mathbb{P}_j(r_i(t))$ is $O(1)$ (Eqn. (4.23)). In order to estimate the block probability for an available network device, a handoff user has to perform at most $m-1$ times calculations of $\mathbb{P}_j(r_i(t))$ (Eqn. (4.25)). Therefore, the computation complexity of the second part is $O(mn)$. Since the calculation of throughput is also a simple numerical calculation, the computation complexity of the third part is $O(n)$. As a result, the computation complexity of our proposed scheme is $O(mn)$.

Since the first part is just a numerical calculation, the memory complexity of this part is $O(n)$. For the second part, we can make use of the recurrence relation as shown in Eqn. (4.28) during the calculation process of Eqn. (4.23). Therefore, the memory complexity of Eqn. (4.23) is $O(1)$. Consequently, the memory complexity of the second part is $O(n)$. For the third part, since we only need to store the information of the current optimal network device, the memory complexity of the third part is $O(1)$. As a result, the memory complexity of our proposed scheme is $O(n)$. □

$$\mathbb{P}_j(r_i(t)+1) = \mathbb{P}_j(r_i(t)) \cdot \frac{\beta_{ij}(t) \cdot [\Gamma_i(t) - 1 - r_i(t)]}{[1-\beta_{ij}(t)] \cdot [r_i(t)+1]}. \tag{4.28}$$

Above discussions illustrate that our scheme can be solved in polynomial time and linear space. The proposed network selection scheme is suitable for ordinary mobile terminals.

4.4.2 An Example

In this subsection, we illustrate the proposed scheme by using an example. Consider a heterogeneous wireless network environment which consists of six network devices, i.e., a_i, $i = 1, 2, \cdots, 6$ as shown in Fig. 4.4. These network devices support different wireless techniques. In the coverage area of six network devices, there are several users. These users make the network selection in a distributed way. We take a single user u_1 as an example. Other users use the same method. Since u_1 has no idea of other users, other users are indistinguishable for u_1. We use u to represent an arbitrary one of them. The available network device set of u_1 at time t is $\mathcal{A}_1(t)$, where $\mathcal{A}_1(t) = \{a_2, a_3, a_4, a_5, a_6\}$. For convenience, we suppose that the basic bandwidth requirements of user u_1 for its available network devices are all 3 Mbps. That is, $\gamma_{i1} = 3$ Mbps, for $i = 2, 3, 4, 5, 6$. The current network device of user u_1 at time t is a_5. Since the channel capacity provided by the current network device $q_{51}(t)$ is 0.2 Mbps, which is smaller than the corresponding basic bandwidth requirement γ_{51}, user u_1 is a handoff user. Handoff user u_1 performs the following steps to select a new network device.

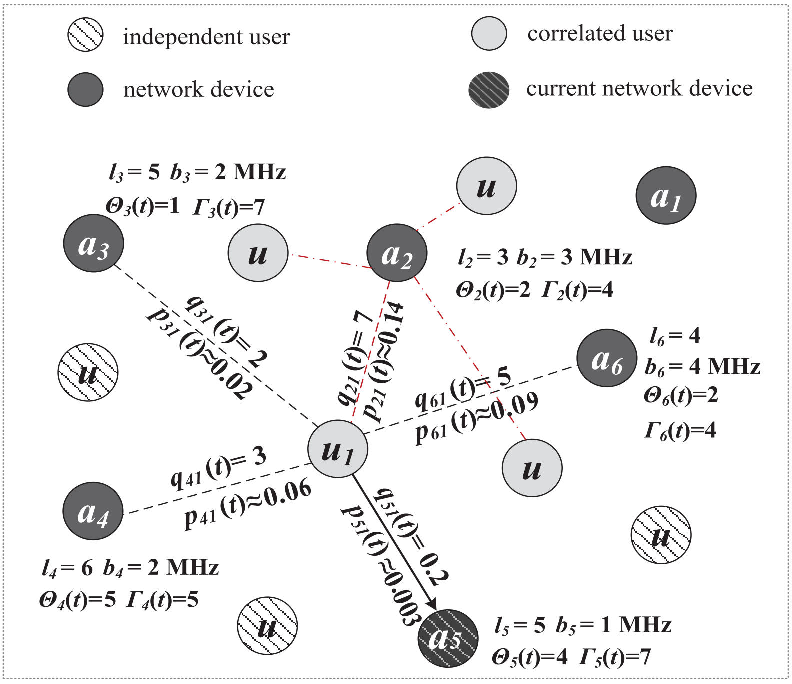


Figure 4.4: An example of the proposed scheme.

Step 1. Handoff user u_1 calculates the channel capacity of each available network device. The channel capacity of network device a_2 denoted by $q_{21}(t)$ is 7 Mbps. Similarly, u_1 calculates that the channel capacities of network devices a_3, a_4, a_5 and a_6 denoted by $q_{31}(t)$, $q_{41}(t)$, $q_{51}(t)$ and $q_{61}(t)$ are 2 Mbps, 3 Mbps, 0.2 Mbps and 5 Mbps respectively. These calculated channel capacities are shown in Fig. 4.4.

Step 2. Handoff user u_1 estimates its block probability for each available network device.

Take the block probability in network device a_2 at time t which is denoted by $p_{21}(t)$ for instance. Following the former discussion we know that handoff user u_1 selects network devices based on their channel capacities. Let $\alpha_{21}(t)$ denote the probability that network device a_2 will be selected by handoff user u_1 at time t. The value of $\alpha_{21}(t)$ is then calculated as follows.

$$\alpha_{21}(t) = \frac{q_{21}(t)}{\sum_{i=2}^{6} q_{i1}(t)} = \frac{7}{17.2}.$$

In order to estimate the block probability $p_{21}(t)$, handoff user u_1 also has to infer the network selection behaviors of other handoff users. Under the premise of privacy preservation, handoff user u_1 makes inference based on local information. The local information refers to the number of channels in network device a_i denoted by l_i, the bandwidth per channel in network device a_i denoted by b_i, the number of non-handoff users which are connecting to network device a_i at time t denoted by $\Theta_i(t)$, and the number of handoff users which are inside the coverage area of network device a_i at time t denoted by $\Gamma_i(t)$. Since handoff user u_1 does not know more detailed information about other handoff users, u_1 has no choice but to assume that other handoff users select network devices based on the remaining bandwidth. Note that network device a_2 has 3 channels, and 2 of them are occupied by two non-handoff users at time t. Therefore, the remaining bandwidth of network device a_2 at time t is equal to $[l_2 - \Theta_2(t)] \cdot b_2 = 3$ MHz. Let $\beta_{21}(t)$ be the probability that network device a_2 will be selected by another handoff user at time t. The value of $\beta_{21}(t)$ is then calculated as follows.

$$\beta_{21}(t) = \frac{[l_2 - \Theta_2(t)] \cdot b_2}{\sum_{i=2}^{6} \{[l_i - \Theta_i(t)] \cdot b_i\}} = \frac{3}{22}.$$

Note that network device a_2 can serve $l_2 - \Theta_2(t) = 1$ handoff user at time t, and there are $\Gamma_2(t) = 4$ handoff users (including u_1) inside the coverage area of network device a_2. Handoff user u_1 is blocked in network device a_2 means that before u_1, there are one or more than one other handoff users that have chosen a_2 as their new network device at time t. Let $r_2(t)$ be the number of handoff users who have selected a_2 as their new network device before u_1 at time t. The value of $r_2(t)$ should satisfy the following relation.

$$l_2 - \Theta_2(t) \leq r_2(t) \leq \Gamma_2(t) - 1.$$

The block probability $p_{21}(t)$ is then calculated as follows.

$$p_{21}(t) = \alpha_{21}(t) \cdot \sum_{r_2(t)=1}^{3} \left\{ \binom{\Gamma_2(t)-1}{r_2(t)} \cdot \beta_{21}(t)^{r_2(t)} \cdot \Big[1 - \beta_{21}(t)\Big]^{\Gamma_2(t)-1-r_2(t)} \right\} \approx 0.14.$$

Similarly, handoff user u_2 estimates its block probabilities for network devices a_3, a_4, a_5 and a_6 denoted by $p_{31}(t)$, $p_{41}(t)$, $p_{51}(t)$ and $p_{61}(t)$ respectively. The estimated block probabilities are shown in Fig. 4.4.

Step 3. Handoff user u_1 scans its available network devices and selects the new network device.

At first, network devices a_3 and a_5 will be eliminated since their channel capacities cannot satisfy the basic bandwidth requirements of handoff user u_1. After this, handoff user u_1 calculates the throughput provided by network devices a_2, a_4 and a_6. The throughput provided by a_2 is equal to $q_{21}(t) \cdot [1 - p_{21}(t)] \approx 6.02$ Mbps. While the throughput provided by a_4 and a_6 are approximately equal to 2.82 Mbps and 4.55 Mbps respectively. Since network device a_2 can provide the highest throughput, a_2 will be selected as the new network device by handoff user u_1 at time t.

4.5 Performance Evaluation

We compare the proposed scheme with two recent typical distributed handoff schemes: the multiplicative scheme [21] and the two-step scheme [22] under various network conditions. Over a 500m × 500m rectangular flat space, we randomly place 3 BSs and several users. A BS is available to a user when the distance between them is smaller than the coverage radius of this BS. In order to simulate a small hybrid 5G environment, we set the parameters of these 3 BSs refer to 3G, 4G and 5G techniques respectively. According to the 3G (W-CDMA/HSDPA) standard [66], we set the coverage radius of 3G BS to be

Table 4.2: Experimental Parameters for Distributed Network Selection Scheme

Parameter	**Value**
Number of BSs	3
Coverage radii of BSs	7km, 50km, 25km
Maximum number of serving users in BSs	10, 20, 15
Bandwidths of BSs	5MHz, 20MHz, 40MHz
Transmission powers of BSs	10 watts, 20 watts, 40 watts
Basic bandwidth requirements of users for BSs	2 Mbps, 4 Mbps, 6 Mbps
Time slot	1 second
Channel fading gain h	$h \sim exp(1)$
Additive white Gaussian noise power ζ^2	$\zeta^2 \sim N(0, 1)$ watts
Moving velocities of users	$0 \sim 5$ m/s

7 km, set the bandwidths and transmission power to be 5 MHz and 10 watts. According to the 4G (802.16a) standard [67], we set the coverage radius, bandwidth and transmission power to be 50 km, 20 MHz and 20 watts respectively. So far the 5G standard is still being figured out. However, Andrews *et al.* [8] pointed out the 5G BS will have higher bandwidth, higher transmission power, smaller cell size and ever-smaller serving users compared with 4G BS. Thus, we set the coverage radius, bandwidth and transmission power of 5G BS to be 25 km, 40 MHz and 40 watts accordingly. Users are moving around inside the hybrid 5G environment. If the current location of a user is denoted by a two-dimensional coordinate (x, y), this user will be inside $(x \pm \triangle t \cdot \varepsilon, y \pm \triangle t \cdot \varepsilon)$ after a period of time $\triangle t$, where ε is the maximal moving velocity of the user [68], [69]. For convenience we assume that users have the same basic bandwidth requirements for a single BS. We set the basic bandwidth requirements of users to be equal to or greater than 2 Mbps, which corresponds to the video conference demanding. Some important experimental parameters are presented in Table 4.2 [67]. The concerned performance metrics are total throughput and ratio of users served. Simulation experiments are repeated one thousand times and the results are presented with 95% confidence interval.

4.5.1 Total Throughput

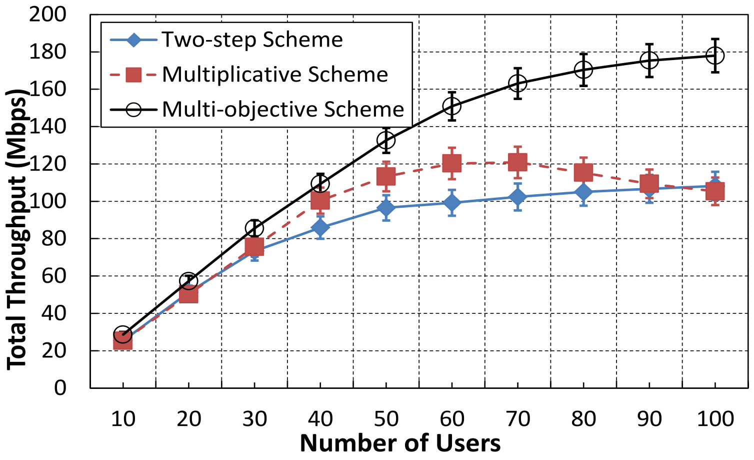
Total throughput is defined as the sum of throughput that handoff and non-handoff users can obtain. According to the analysis and discussion in Section 4.4, the throughput metric can reflect two performance attributes that users care about: achievable data receiving rate and block probability. We study the total throughput when the number of users varies under free space propagation ($\lambda = 2$), flat-earth reflection ($\lambda = 3$) and diffraction losses ($\lambda = 4$) environment conditions in Fig. 4.5.

The general trend is that the total throughput will be higher as more users join in. For the same scenario, the proposed multi-objective scheme always has the highest total throughput. From the crosswise comparison we observe that the total throughput in three schemes declines in tougher environments. Another interesting observation is that the total throughput in multiplicative scheme slightly reduces when the number of users is bigger than around 50. After careful deliberation, we consider that the reason behind this phenomenon is network congestion.

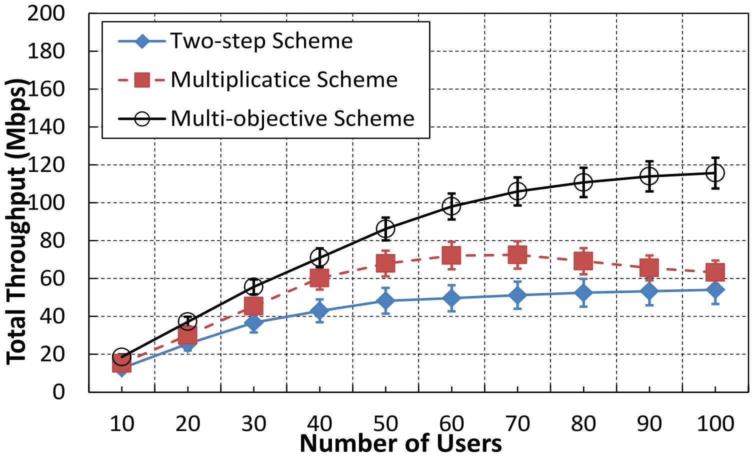
4.5.2 Ratio of Users Served

Ratio of users served refers to the ratio of users who have the network service. Following the notations made in problem formulation, the ratio of users served is equal to $\frac{\sum_{i=1}^{n}\Theta_i(t)+\sum_{i=1}^{n}\sum_{j=1}^{m}\{f_{ij}(t)\cdot[1-p_{ij}(t)]\}}{m}$, where m is the number of uses, $\sum_{i=1}^{n}\Theta_i(t)$ is the number of non-handoff users and $\sum_{i=1}^{n}\sum_{j=1}^{m}\{f_{ij}(t)\cdot[1-p_{ij}(t)]\}$ is the number of handoff successful users. The ratio of users served metric is used to reflect the fairness in three handoff schemes.

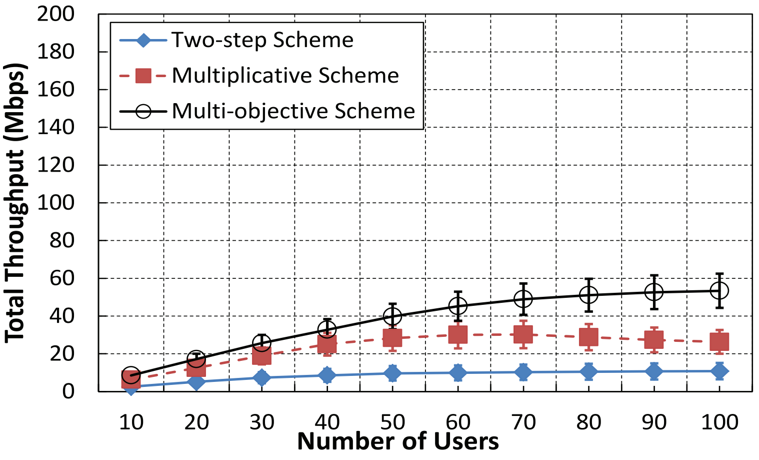
The experiment results shown in Fig. 4.6 revel that there are more users can get service in our proposed scheme. Furthermore, the ratio of users served in our scheme will maintain stable then decline as the number of users increases. Comparatively, the ratios of users served in two contrast schemes will slightly increase then decrease. Moreover, there is an obvious downtrend in multiplicative scheme when the number of users is around 50. It will not be difficult to find that the inflection point of multiplicative scheme in Fig. 4.6 is very close to that in Fig. 4.5. This is another proof of network congestion.



(a) pass loss exponent $\lambda = 2$

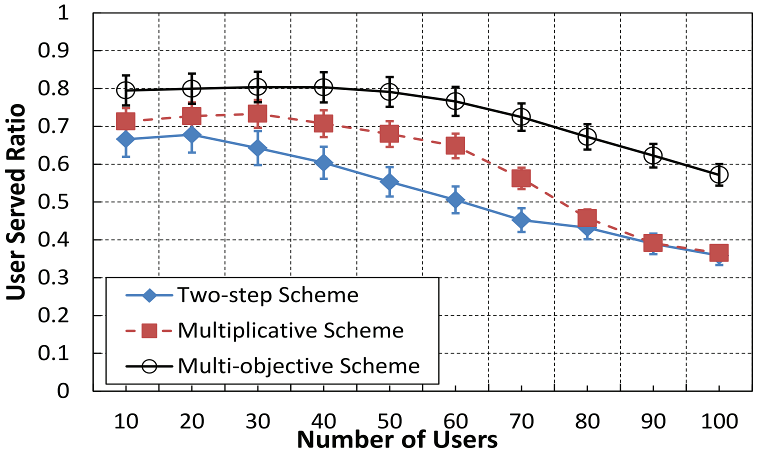


(b) pass loss exponent $\lambda = 3$

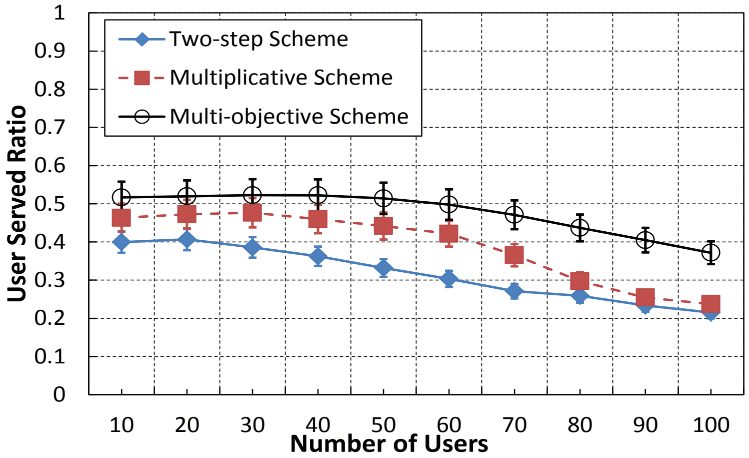


(c) pass loss exponent $\lambda = 4$

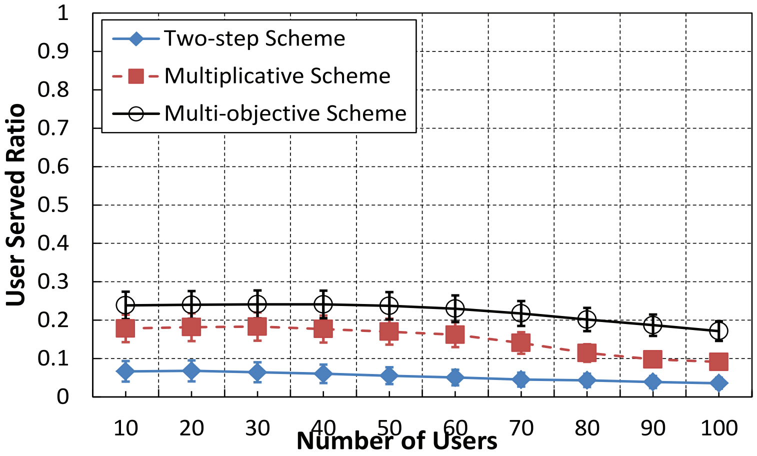
FIGURE 4.5: Number of users vs. throughput.



(a) pass loss exponent $\lambda = 2$



(b) pass loss exponent $\lambda = 3$



(c) pass loss exponent $\lambda = 4$

Figure 4.6: Number of users vs. ratio of users served.

4.6 Summary

In this chapter, we proposed a user centered handoff scheme fulfilling multiple objectives for hybrid 5G environments. We consider the general limitations in hybrid 5G environments that users are unwilling to share their private information and centralized control usually is inefficient in large scare scenario. Based on limited local information, a user has to make the network selection by itself. We exploited two performance attributes to evaluate BSs: achievable data receiving rate and blocking probability. When a handoff user needs to select a new BS, it will calculate the achievable data receiving rates of available BSs. Then the user has to infer the network selection behaviors of other users in order to estimate its blocking probability for each available BS. By jointly considering these two attributes, a user can select the most appropriate one as its new BS for a handoff.

Chapter 5

A Software-Defined Networking based Centralized Vertical Handoff for Heterogeneous Wireless Network Environments

In this chapter, we propose a novel vertical handoff scheme with the support of the Software-Defined Networking (SDN) technique for heterogeneous wireless networks. The proposed scheme solves two important issues in vertical handoff: *network selection* and *handoff timing*. In this chapter, the network selection is formulated as a 0-1 integer programming problem which maximizes the sum of channel capacities that handoff users can obtain from their new access points. After the network selection process is finished, a user will wait for a time period. Only if the new access point is consistently more appropriate than the current access point during this time period, will the user transfer its inter-network connection to the new access point. Our proposed scheme ensures that a user will transfer to the most appropriate access point at the most appropriate time. Comprehensive simulation has been conducted. It is shown that the proposed scheme reduces the number of vertical handoffs, maximizes the total throughput and user served ratio significantly.

This chapter is organized as follows. Section 5.1 introduces the heterogeneous wireless networks and the handoff problem. Section 5.2 is the system description and problem formulation. Section 5.3 and Section 5.4 present the proposed scheme for network selection and handoff timing issues respectively. Section 5.5 is the performance evaluation. Section 5.6 summarizes this chapter.

5.1 Introduction

Heterogeneous wireless networks integrate a variety of wireless techniques to provide ubiquitous services [70]. In heterogeneous wireless networks, users may need to transfer their inter-network connections from one access point to another. The transferring operation among different kinds of access points is called vertical handoff [71]. There are two important issues [72] needed to be solved in vertical handoff: network selection and handoff timing. The network selection issue is to select an access point, to which the inter-network connection should be transferred. The handoff timing issue is to determine when the inter-network connection transferring should be implemented. The emergence of Software-Defined Networking (SDN) technique [73] makes it possible to solve these two issues of vertical handoff in a novel perspective. SDN is a new networking paradigm, which provides a global centralized control of access points. In this chapter, we make use of this feature of SDN and study the vertical handoff problem for heterogeneous wireless networks.

Consider the scenario shown in Fig. 5.1. There is a general heterogeneous wireless network environment that consists of Wi-Fi, LTE, WiMAX and 3G. A user walks to the company from home. When the user is at home, his (or her) smart phone connects to the Wi-Fi in the house. After the user goes out of home, his (or her) smart phone may connect to LTE, or WiMAX of the public library, or 3G. There are three available networks. This user needs to know which one should be selected (network selection issue) [74]. After a new access point is selected, this user also needs to know when the inter-network connection should be transferred to the new access point (handoff timing issue) [10]. Solutions to the network selection and handoff timing issues compose the vertical handoff scheme [75].

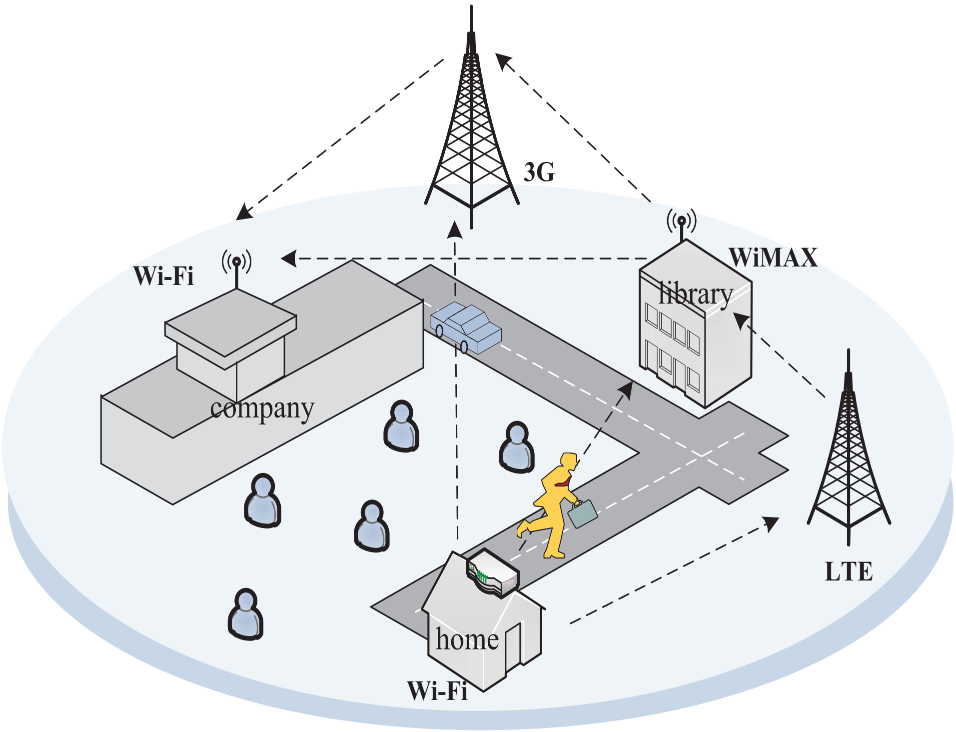


FIGURE 5.1: Illustration of vertical handoff.

Due to lack of the global view, most of existing vertical handoff schemes failed to be global optimal. The emergence of software-defined networking [76] technique provides a chance to break this limitation. An SDN controller has an abstracted centralized control of access points. We make use of this feature of SDN, and propose a novel vertical handoff scheme named S-DNVH. In S-DNVH, users are divided into two classes: non-handoff users and handoff users. Non-handoff users will stay in the connections with their current access points. While handoff users will send their handoff requests to an SDN controller. The network selection is formulated as a 0-1 integer programming problem [77] by the SDN controller, with the objective of maximizing the sum of channel capacities that handoff users can obtain from their new access points. After the network selection process is finished, handoff users have to wait for a certain time period [23]. After the time period, the SDN controller evaluates the performances of current access points and new access points for users. Only if the new access points are consistently more appropriate than the current access points, will users transfer their inter-network connections to the new access points. The main contributions of this chapter are summarized as follows:

- To the best of our knowledge, we are the *first* to apply the SDN technique in the study of vertical handoff problem for heterogeneous wireless networks. We investigate the architecture of SDN, and design a compatible vertical handoff procedure.
- We formulate the network selection issue as a 0-1 integer programming problem based on our previous work [78], and propose a network selection algorithm to solve it. In the proposed network selection algorithm, an SDN controller will allocate the most appropriate access point for each user.
- We propose a handoff timing algorithm to determine the time when the network selection results should be implemented. Based on limited information and simple calculation, the proposed handoff timing algorithm can predict the movement directions of users. Only if users are certain to move away from their current access points, will the network selection results be implemented.
- Through comprehensive experiments, we validate that our proposed scheme significantly reduces the number of vertical handoffs, maximizes the total throughput and user served ratio.

5.2 System Model and Preliminaries

5.2.1 Network Architecture

For the upcoming problem formulation, we first exploit the special network architecture of SDN. Traditionally, each access point contains both a control plane and a data plane [76]. The control plane decides whether a traffic flow is admissible or not, and the route that the traffic flow should traverse. The data plane forwards the traffic flow according to the decision made by the control plane. In the SDN architecture, an SDN controller separates control planes from data planes of access points, and provides a centralized control of these access points. The SDN controller communicates with access points via OpenFlow [79], and has a global view of the network environment. This feature of SDN gives us an opportunity to design a global optimal vertical handoff scheme.

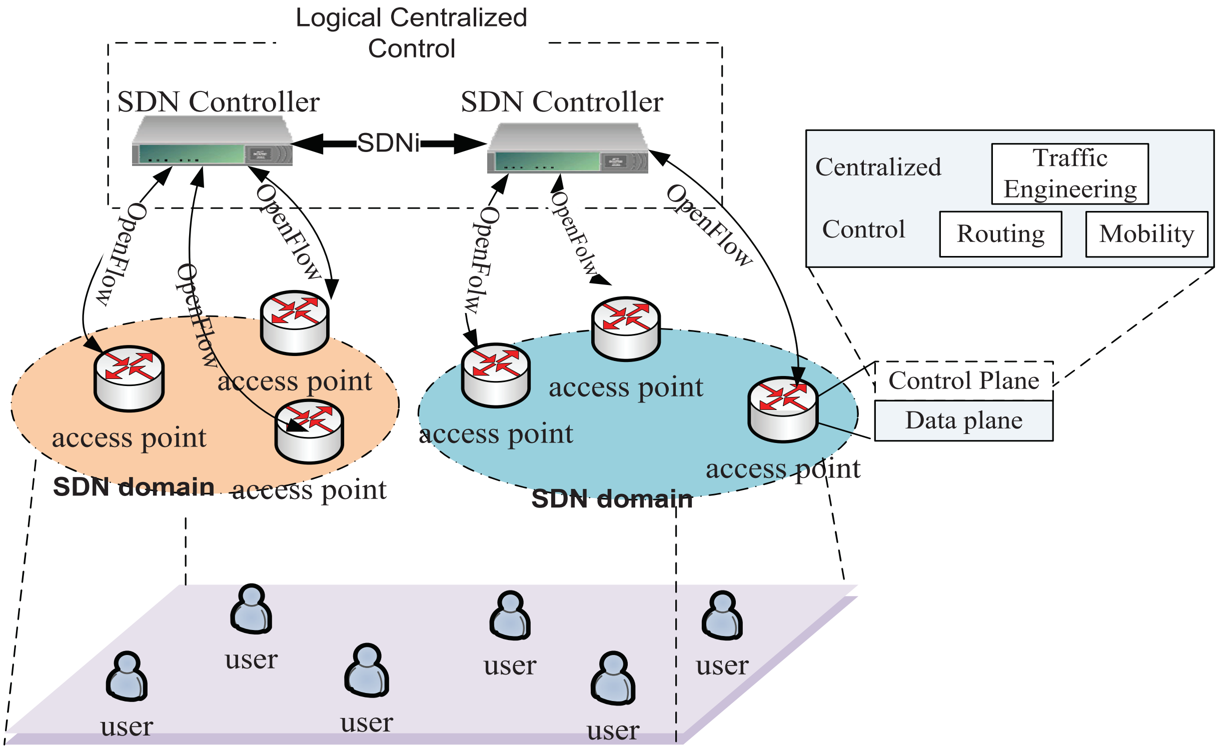


FIGURE 5.2: A network architecture of SDN.

Normally, an SDN controller can manage thousands of access points at the same time. When there are a large number of access points in an area, several SDN controllers can be deployed. Besides of this, in order to group manage the access points, multiple SDN controllers are also required, so that each SDN controller can install a custom policy in its own domain. Stallings [76] pointed out the most common scenario is the numerous and nonoverlapping SDN domains scenario as shown in Fig. 5.2. As we know, a SDN controller centrally controls the access points in its domain. In fact, logical centralized control also exists in the multiple SDN controllers scenario. The IETF is currently working on the SDNi protocol [80] which helps the SDN controllers to exchange information and gain the global information. Thus, if a optimal vertical handoff scheme is proposed for the single SDN controller scenario, the proposed scheme also works for the multiple SDN controllers scenario. The extension work is just to add an "information exchange" stage at the beginning of proposed scheme. In this chapter, we will consider the scenario with single SDN controller.

5.2.2 Problem Formulation

In this subsection, we formulate the vertical handoff problem for heterogeneous wireless network environment by using the software-defined networking technique [81]. Specifically, we consider a heterogeneous wireless network environment which consists of m access points. Let $\mathcal{A}$ be the set of access points, $\mathcal{A} = \{a_1, a_2, \cdots, a_m\}$. These access points support different wireless technologies. With the support of the Media-Independent Handover (MIH) standard [57], all of these access points can be centrally controlled by a single SDN controller. Access point a_i ($a_i \in \mathcal{A}$, $i = 1, 2, \cdots, m$) has l_i channels, a_i equally divides its frequency band among these channels. The bandwidth of each channel in access point a_i is denoted by b_i in MHz. If the access point a_i is connected by a user, this user will occupy one channel of a_i [25]. Hence, the access point a_i can serve at most l_i users simultaneously.

Consider that there are n users. Let $\mathcal{U}$ be the set of users, $\mathcal{U} = \{u_1, u_2, \cdots, u_n\}$. If user u_j ($u_j \in \mathcal{U}$, $j = 1, 2, \cdots, n$) is inside the coverage area of an access point, this access point is called an **available access point** of user u_j. In heterogeneous wireless network environment, u_j may have several available access points. However, u_j can connect to at most one of its available access points at anytime. The connected available access point is called the **current access point** of user u_j. An adjacency matrix $\delta(t)$ is used to reflect the relationship between access points and users at time t as follows.

$$\delta(t) = \begin{array}{c} \\ \boldsymbol{a_1} \\ \boldsymbol{a_2} \\ \vdots \\ \boldsymbol{a_m} \end{array} \begin{array}{c} \begin{array}{cccc} \boldsymbol{u_1} & \boldsymbol{u_2} & \cdots & \boldsymbol{u_n} \end{array} \\ \begin{bmatrix} \delta_{11}(t) & \delta_{12}(t) & \dots & \delta_{1n}(t) \\ \delta_{21}(t) & \delta_{22}(t) & \dots & \delta_{2n}(t) \\ \vdots & & \ddots & \vdots \\ \delta_{m1}(t) & \delta_{m2}(t) & \dots & \delta_{mn}(t) \end{bmatrix} \end{array},$$

where

$$\delta_{ij}(t) = \begin{cases} 1, & \text{current access point } a_i \text{ is connected by user } u_j \text{ at time } t, \\ 0, & \text{otherwise.} \end{cases} \tag{5.1}$$

Since most of the applications (e.g., music, video and game) require higher download and lower upload rates, users are more concerned about the quality of downlink (from access point a_i to user u_j). All links refer to downlink unless otherwise specified from now on. For convenience, the notations used in this chapter are summarized in Table 5.1.

For each access point-user pair (a_i, u_j), assume that the received signal power of user u_j from available access point a_i at time t is $s_{ij}(t)$ in watts. Let $d_{ij}(t)$ denote the Euclidean distance between access point a_i and user u_j at time t. Let p_i denote the transmission power of access point a_i in watts. Then the value of $s_{ij}(t)$ can be calculated as follows [82].

$$s_{ij}(t) = p_i \cdot h_{ij} \cdot d_{ij}(t)^{-\alpha}, \tag{5.2}$$

where the channel fading gain h_{ij} follows an exponential distribution with rate μ ($h_{ij} \sim exp(\mu)$), and the pass loss exponent $\alpha > 2$ (varies depending on the channel conditions).

With the help of some techniques such as Orthogonal Frequency Division Multiplexing (OFDM, commonly applied to LTE, WiMAX and IEEE 802.11 b/g supported devices) [83] and Code Division Multiple Access (CDMA, commonly applied to 3G devices), the interference among users which belong to the same access point can be controlled in a low level. In order to further weaken this kind of interference, we assume that the orthogonal codes are used in CDMA.

For the interference from other access points, let $g_{xj}(t)$ in watts be the interference caused by access point a_x ($a_x \in \mathcal{A}, x \neq i$) to user u_j at time t, where a_x transmits signal by using the same frequency band as user u_j. Let $d_{xj}(t)$ denote the Euclidean distance between access point a_x and user u_j at time t. Let p_x and h_{xj} denote the transmission power of access point a_x in watts and channel fading gain respectively. Then the value of $g_{xj}(t)$ can be calculated as follows.

$$g_{xj}(t) = p_x \cdot h_{xj} \cdot d_{xj}(t)^{-\alpha}. \tag{5.3}$$

Since the transmission powers of users are relative small and the locations of users erratically change all the time, the interference from other users is usually negligible [84]. If the access point a_i is connected by a user, this user will occupy one channel of

TABLE 5.1: Notation Summary in Chapter 5

m	Number of access points
$\mathcal{A}$	Set of access points $\{a_i\}$, $i = 1, 2, \cdots, m$
$\mathcal{A}'_j$	Set of available access points for user u_j, $\mathcal{A}'_j \subseteq \mathcal{A}$
$\mathcal{X}_j$	Set of access points which transmit signal by using the same frequency band as user u_j, $\mathcal{X}_j \subseteq \mathcal{A}$
l_i	Number of channels in access point a_i
b_i	Bandwidth of each channel in access point a_i in MHz
n	Number of users
$\mathcal{U}$	Set of users $\{u_j\}$, $j = 1, 2, \cdots, n$
$\delta(t)$	Adjacency matrix of $\mathcal{A}$ and $\mathcal{U}$ at time t, i.e., $[\delta_{ij}(t)]$
γ_j	Basic bandwidth requirement of user u_j in Mbps
$s_{ij}(t)$	Received signal power of user u_j from access point a_i at time t in watts
$d_{ij}(t)$	Euclidean distance between access point a_i and user u_j at time t
p_i	Transmission power of access point a_i for each channel in watts
h_{ij}	Channel fading gain of channel (a_i, u_j)
α	Pass loss exponent
$n_{ij}(t)$	Additive White Gaussian Noise (AWGN) power of channel (a_i, u_j) at time t in watts
$q_{ij}(t)$	Channel capacity of channel (a_i, u_j) at time t in Mbps
$\mathcal{V}(t)$	Identifier vector of user set $\mathcal{U}$ at time t, i.e., $(v_j(t))$, $j = 1, 2, \cdots, n$
$\mathcal{R}_j(t)$	Channel capacities provided for user u_j at time t, i.e., $(r_{ij}(t))$, $i = 1, 2, \cdots, m$
$\mathcal{R}(t)$	Channel capacities provided for user set $\mathcal{U}$, i.e., $\left[[\mathcal{R}_j(t)]^{\mathrm{T}}\right]$, $j = 1, 2, \cdots, n$
$\mathcal{F}_j(t)$	Network selection result of user u_j at time t, i.e., $(f_{ij}(t))$, $i = 1, 2, \cdots, m$
$\mathcal{F}(t)$	Network selection results of user set $\mathcal{U}$, i.e., $\left[[\mathcal{F}_j(t)]^{\mathrm{T}}\right]$, $j = 1, 2, \cdots, n$

a_i. Hence, the access point a_i can serve at most l_i users simultaneously. Note that the bandwidth of each channel in access point a_i is b_i MHz. Let $n_{ij}(t)$ denote the Additive White Gaussian Noise (AWGN) [85] power of the channel (a_i, u_j) at time t in watts. Let $\mathcal{X}_j$ ($\mathcal{X}_j \subseteq \mathcal{A}$) be the set of access points which transmit signal by using the same frequency band as the user u_j. According to the Shannon equation [86], the channel capacity denoted by $q_{ij}(t)$ in Mbps is calculated as follows.

$$q_{ij}(t) = b_i \cdot \log_2 \left[1 + \frac{s_{ij}(t)}{\sum_{a_x \in \mathcal{X}_j} g_{xj}(t) + n_{ij}(t)} \right]. \tag{5.4}$$

Assume that all of the transmitted data can be received correctly. Consider that a user may carry out multiple tasks at the same time, there are infinite data needed to be received by each user. Thus, we can make an assumption that the data receiving rates of users are equal to the channel capacities. Furthermore, in order to guarantee the quality of experience, user u_j has the basic bandwidth requirement denoted by γ_j in Mbps. Suppose that the current access point of user u_j is a_c. If the channel capacity of current access point cannot meet the corresponding basic bandwidth requirement ($q_{cj}(t) < \gamma_j$), user u_j will perform the vertical handoff. We call these users who need to perform handoff **handoff users**. If the channel capacity of current access point can satisfy the basic bandwidth requirement ($q_{cj}(t) \geq \gamma_j$), user u_j will stay in the connection with its current access point a_c. We call these users who do not need handoff **non-handoff users**. A vector $\mathcal{V}(t) = (v_1(t), v_2(t), \cdots, v_n(t))$ is used to identify the kinds of users at time t. The value of $v_j(t)$ is given as follows, where $j = 1, 2, \cdots, n$.

$$v_j(t) = \begin{cases} 0, & \text{user } u_j \text{ is a handoff user at time } t, \\ 1, & \text{user } u_j \text{ is a non-handoff user at time } t. \end{cases} \tag{5.5}$$

Let $r_{ij}(t)$ denote the channel capacity that access point a_i can provide for user u_j at time t, if a_i is selected as the **new access point**. The value of $r_{ij}(t)$ is given in Eqn. (5.6). If u_j is a handoff user ($v_j(t) = 0$), $r_{ij}(t)$ is equal to $q_{ij}(t)$. If u_j is a non-handoff user ($v_j(t) = 1$), $r_{ij}(t)$ is 0.

$$r_{ij}(t) = \begin{cases} q_{ij}(t), & v_j(t) = 0, \\ 0, & v_j(t) = 1. \end{cases} \tag{5.6}$$

Let $\mathcal{R}_j(t) = (r_{1j}(t), r_{2j}(t), \cdots, r_{mj}(t))$. Based on $\mathcal{R}_j(t)$, the SDN controller selects a **new** access point for u_j. The network selection result of u_j at time t is denoted by $\mathcal{F}_j(t) = (f_{1j}(t), f_{2j}(t), \cdots, f_{mj}(t))$. The value of $f_{ij}(t)$ is given as follows, where $i = 1, 2, \cdots, m$.

$$f_{ij}(t) = \begin{cases} 1, & \text{new access point } a_i \text{ is selected by user } u_j \text{ at time } t, \\ 0, & \text{otherwise.} \end{cases} \tag{5.7}$$

The channel capacity that handoff user u_j can obtain from the **new** access point is $\mathcal{F}_j(t) \cdot [\mathcal{R}_j(t)]^{\mathrm{T}}$ in Mbps, where $[\mathcal{R}_j(t)]^{\mathrm{T}}$ is the transposition of $\mathcal{R}_j(t)$.

$$\mathcal{F}_j(t) \cdot [\mathcal{R}_j(t)]^{\mathrm{T}} = \sum_{i=1}^{m} [f_{ij}(t) \cdot r_{ij}(t)] \tag{5.8}$$

The SDN controller has to ensure that if an access point a_i is selected as the **new** access point of user u_j, the channel capacity provided by a_i should satisfy the basic bandwidth requirement of u_j. That is, $\mathcal{F}_j(t) \cdot [\mathcal{R}_j(t)]^{\mathrm{T}}$ should subject to the following constrain.

$$\mathcal{F}_j(t) \cdot [\mathcal{R}_j(t)]^{\mathrm{T}} \geq \gamma_j \tag{5.9}$$

If there is no access point can satisfy the basic bandwidth requirement of user u_j, the vertical handoff of u_j is failed. Moreover, u_j will be discarded by its **current** access point.

Note that if user u_j is a non-handoff user ($v_j(t) = 1$), it does not have any new access point ($\sum_{i=1}^{m} f_{ij}(t) = 0$). Thus, we can get a constraint for non-handoff users that is $\sum_{i=1}^{m} f_{ij}(t) + v_j(t) = 1$. If user u_j is a handoff user ($v_j(t) = 0$), when the vertical handoff of u_j is failed, u_j will not have any new access point neither ($\sum_{i=1}^{m} f_{ij}(t) = 0$). In this case, we can get a relationship that is $\sum_{i=1}^{m} f_{ij}(t) + v_j(t) = 0$. If user u_j is a handoff user ($v_j(t) = 0$) and the vertical handoff of u_j is successful, the SDN controller will allocate a new access point to u_j ($\sum_{i=1}^{m} f_{ij}(t) = 1$). In this case, we can get a relationship that is $\sum_{i=1}^{m} f_{ij}(t) + v_j(t) = 1$. In summary, the network selection result of user u_j should satisfy the following constraint.

$$\sum_{i=1}^{m} f_{ij}(t) + v_j(t) \leq 1. \tag{5.10}$$

Given a set of users $\mathcal{U}$, their network selection results are denoted by $\mathcal{F}(t) = [[\mathcal{F}_1(t)]^{\mathrm{T}}, [\mathcal{F}_2(t)]^{\mathrm{T}}, \cdots, [\mathcal{F}_n(t)]^{\mathrm{T}}]$. Note that access point a_i has l_i channels, since each user will occupy one channel, a_i can server at most l_i users simultaneously. There are $\sum_{j=1}^{n} [v_j(t) \cdot \delta_{ij}(t)]$ non-handoff users are connecting to access point a_i at time t. Therefore, the number of handoff users assigned to a_i ($\sum_{j=1}^{n} f_{ij}(t)$) should satisfy the following constraint.

$$\sum_{j=1}^{n} [v_j(t) \cdot \delta_{ij}(t)] + \sum_{j=1}^{n} f_{ij}(t) \leq l_i. \tag{5.11}$$

We express $\mathbb{T}(\mathcal{F}(t))$ as the sum of the channel capacities that handoff users can obtain from their **new** access points. $\mathbb{T}(\mathcal{F}(t))$ is then calculated as follows.

$$\mathbb{T}(\mathcal{F}(t)) = \sum_{j=1}^{n} \left\{ \mathcal{F}_j(t) \cdot [\mathcal{R}_j(t)]^{\mathrm{T}} \right\}. \tag{5.12}$$

The goal in the chapter is to maximize the sum of channel capacities that handoff users can obtain from their new access points $\mathbb{T}(\mathcal{F}(t))$. Non-handoff users will stay in the connections with their current access points. The optimization problem of maximizing $\mathbb{T}(\mathcal{F}(t))$ is theoretically formulated as follows.

$$\text{Maximize} \quad \mathbb{T}(\mathcal{F}(t)) = \sum_{i=1}^{m} \sum_{j=1}^{n} [f_{ij}(t) \cdot r_{ij}(t)] \tag{5.13}$$

subject to

$$\sum_{i=1}^{m} f_{ij}(t) + v_j(t) \leq 1, j = 1, 2, \cdots, n, \tag{5.14a}$$

$$\sum_{i=1}^{m}[f_{ij}(t)\cdot r_{ij}(t)] \geq \gamma_j, j = 1, 2, \cdots, n, \tag{5.14b}$$

$$\sum_{j=1}^{n}[v_j(t)\cdot \delta_{ij}(t)] + \sum_{j=1}^{n} f_{ij}(t) \leq l_i, i = 1, 2, \cdots, m. \tag{5.14c}$$

The first constraint Eqn. (5.14a) indicates that each user has at most one **new** access point. The second constraint Eqn. (5.14b) guarantees that the channel capacity provided by the **new** access point can satisfy the basic bandwidth requirement of a handoff user. The last constraint Eqn. (5.14c) ensures that the number of users connected to an access point is smaller than the number of channels in this access point.

The vertical handoff problem is formulated as a 0-1 integer programming problem. Although the formulated problem is an NP-hard problem, the computation complexity is acceptable in its particular application context. The latest OpenFlow version 1.4 supported access points provide 40 GbE services [87], and the computing capability of SDN controller is considered to be infinite [88]. Compared to the powerful SDN devices, most of vertical handoff scenarios involve limited number of users. Thus, we think that the network selection process is completed in a very short time interval which can be neglected. Moreover, if we eliminate non-handoff users from consideration, the computation complexity can be further reduced.

5.3 Network Selection Algorithm

Based on previous formulations, we study the network selection issue of the vertical handoff in this section. We assume that users are selfish, they will select access points in the Always Best Connected (ABC) way [9] if allowed. That is, users always choose the access points which have the best performance as their new access points. As an example shown in Fig. 5.3, there are three network access points (*i.e.*, a_1, a_2, a_3). These access points support different wireless technologies. In the coverage area of three access points, there are four users (*i.e.*, u_1, u_2, u_3, u_4). At first, user u_1 was connecting to the access point a_1. When the channel capacity of (a_1, u_1) cannot meet its basic bandwidth requirement, u_1 performed the vertical handoff. Suppose that the performance of a_2 is better than a_3, thus u_1 will choose a_2 as its new access point. At the same time, other users (u_2, u_3 and u_4) may also choose a_2 as their new access points. Since the resources are limited, a_2 will be exhausted. While there is no user connects to a_3. System resources are unreasonably utilized.

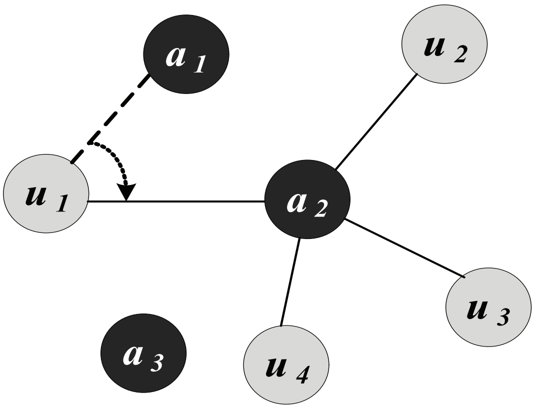


FIGURE 5.3: An example of the network selection issue.

In order to efficiently utilize the system resources and construct a good access point associating strategy for each user, we propose a network selection algorithm for vertical handoff. An SDN controller selects access points for users in three phases: initialization, request matrix construction and network selection (Fig. 5.4).

5.3.1 Algorithm Details

Phase 1. Initialization

At the beginning of each time slot, users evaluate their current access points and determine whether to perform vertical handoffs or not. If a user needs vertical handoff (handoff user), it will send request frames to its available access points. The values of request frames are equal to the channel capacities. If a user does not need vertical handoff (non-handoff user), it also sends request frames to its available access points. In this case, the request frames are just like Hello messages, and their values are 0.

Phase 2. Request matrix construction

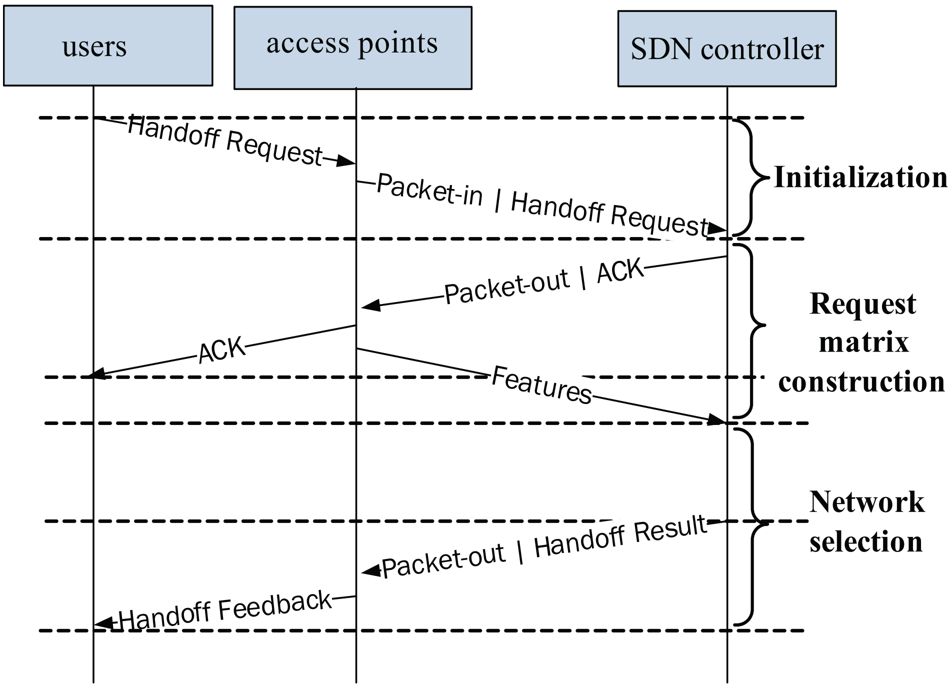


FIGURE 5.4: The process of network selection.

The request frames sent from users will be converged in the SDN controller. According to these received frames, the SDN controller constructs a request matrix $\mathcal{R}(t) = \left[[\mathcal{R}_1(t)]^{\mathrm{T}},\ [\mathcal{R}_2(t)]^{\mathrm{T}},\ \cdots,\ [\mathcal{R}_n(t)]^{\mathrm{T}}\right]$. At first, the request matrix $\mathcal{R}(t)$ is incomplete because users may not be able to reach all access points. There are some elements, whose values are unknown. These elements are defined as the unassigned elements.

Definition 3.4.1 (*Unassigned element*). The unassigned element is the element of a request matrix, whose value is unknown due to the corresponding access point and user cannot communicate directly. □

Since the SDN controller has a global view, it can complete the request matrix $\mathcal{R}(t)$ after a simple calculation in which the unassigned elements will be set to be 0.

Theorem 3.4.1 The calculation rule of unassigned element maintains the consistency of vertical handoff request matrix, and will not affect the network selection results.

Proof. Each column of the request matrix $\mathcal{R}(t)$ corresponds to a vector $\mathcal{R}_j(t)$, where $j = 1, 2, \cdots, n$. For non-handoff users, all elements in $\mathcal{R}_j(t)$ are 0, and the value of corresponding column in $\mathcal{R}(t)$ will be set to be 0. During the network selection process, there is no new access point will be assigned to these users. For handoff users, the unassigned elements in $\mathcal{R}(t)$ which correspond to their unavailable access points are set to be 0. During the network selection process, some other access points with positive evaluations will be selected. □

Phase 3. Network selection

After constructing the vertical handoff request matrix $\mathcal{R}(t)$, the SDN controller selects new access points for handoff users. The network selection is formulated as a 0-1 programming problem (Eqn. (5.13)). The SDN controller calculates the network selection results by solving this 0-1 programming problem. There are many tools can be used in solving linear problems like LINGO and CVX on MATLAB. The network selection results are presented as a $m \times n$ matrix $\mathcal{F}(t)$. According to $\mathcal{F}(t)$, the SDN controller sends out feedback frames. If the element $f_{ij}(t)$ is 1, that means a_i is the new access point of user u_j. Therefore, the SDN controller sends an OpenFlow message [89] to access point a_i. Then, access point a_i sends a feedback frame to user u_j to notify this result.

5.3.2 An Example

In this subsection, we will use an example to explain the network selection algorithm in detail. Specifically, we consider a scenario shown in Fig. 5.5. There are three access points (*i.e.*, a_1, a_2, a_3). These access points support different wireless technologies. An SDN controller centralized controls these access points. In the coverage area of three access points, there are five users (*i.e.*, u_1, u_2, u_3, u_4, u_5). The basic bandwidth requirements of these five users are 2 Mbps (*i.e.*, $\gamma_1 = 2$ Mbps), 3 Mbps (*i.e.*, $\gamma_2 = 3$ Mbps), 4 Mbps (*i.e.*, $\gamma_3 = 4$ Mbps), 4.5 Mbps (*i.e.*, $\gamma_4 = 4.5$ Mbps) and 4 Mbps (*i.e.*, $\gamma_5 = 4$ Mbps) respectively. Suppose that the channel capacity of (a_1, u_1) is 4 Mbps, and u_1 is a non-handoff user. Meanwhile, other four users (u_2, u_3, u_4 and u_5) are handoff users.

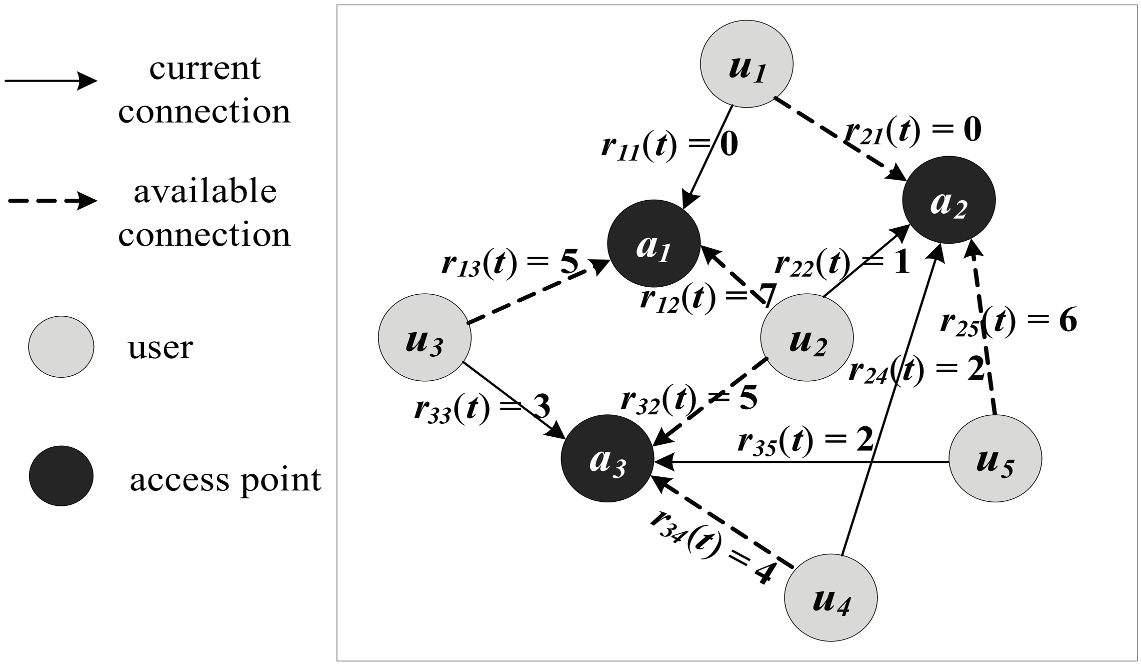


FIGURE 5.5: An example of the proposed network selection algorithm.

Phase 1. Initialization

Since u_1 is a non-handoff user, it will send request frames $r_{11}(t)$ and $r_{21}(t)$ to the available access points a_1 and a_2 respectively. The values of $r_{11}(t)$ and $r_{21}(t)$ are 0. Meanwhile, other four users u_2, u_3, u_4 and u_5 are handoff users, they also have to send out request frames. Take u_2 for instance, a_1, a_2 and a_3 are available to u_2. Let $r_{12}(t)$ denote the handoff request frame sent from u_2 to a_1. In our example $r_{12}(t)$ is 7, which means if u_2 selects a_1 as its new access point, the channel capacity provided by a_1 is 7 Mbps. Similarly, other users send the vertical handoff request frames to their available access points.

Phase 2. Request matrix construction

After receiving the request frames, the SDN controller constructs a request matrix $\mathcal{R}(t)$ as shown in Eqn. (5.15). At first, there are four unassigned elements (*i.e.*, $r_{14}(t)$, $r_{15}(t)$, $r_{23}(t)$ and $r_{31}(t)$) in the matrix $\mathcal{R}(t)$. Take the element $r_{14}(t)$ for instance, since the access point a_1 is unavailable to user u_4, u_4 will not send a request frame to a_1.

Therefore, the SDN controller cannot determine the value of $r_{14}(t)$ in the beginning.

$$\mathcal{R}(t) = \begin{array}{c} \\ \boldsymbol{a_1} \\ \boldsymbol{a_2} \\ \boldsymbol{a_3} \end{array} \begin{array}{c} \begin{array}{ccccc} \boldsymbol{u_1} & \boldsymbol{u_2} & \boldsymbol{u_3} & \boldsymbol{u_4} & \boldsymbol{u_5} \end{array} \\ \begin{bmatrix} 0 & 7 & 5 & r_{14}(t) & r_{15}(t) \\ 0 & 1 & r_{23}(t) & 2 & 6 \\ r_{31}(t) & 5 & 3 & 4 & 2 \end{bmatrix} \end{array}. \tag{5.15}$$

Based on this primary request matrix, the SDN controller calculates the values of unassigned elements. All of the unassigned elements are set to be 0, and the completed request matrix $\mathcal{R}(t)$ is as follows.

$$\mathcal{R}(t) = \begin{array}{c} \\ \boldsymbol{a_1} \\ \boldsymbol{a_2} \\ \boldsymbol{a_3} \end{array} \begin{array}{c} \begin{array}{ccccc} \boldsymbol{u_1} & \boldsymbol{u_2} & \boldsymbol{u_3} & \boldsymbol{u_4} & \boldsymbol{u_5} \end{array} \\ \begin{bmatrix} 0 & 7 & 5 & \mathbf{0} & \mathbf{0} \\ 0 & 1 & \mathbf{0} & 2 & 6 \\ \mathbf{0} & 5 & 3 & 4 & 2 \end{bmatrix} \end{array}. \tag{5.16}$$

Phase 3. Network selection

Based on the request matrix $\mathcal{R}(t)$, the SDN controller formulates the network selection as a 0-1 programming problem (Eqn. (5.13)). The solution of this 0-1 programming problem is the network selection result, which is presented as a matrix $\mathcal{F}(t)$. Assume that the number of channels in access points a_1, a_2 and a_3 are 2, 3, 4 respectively. After some calculations, the SDN controller can get a 3×5 matrix $\mathcal{F}(t)$ as follows.

$$\mathcal{F}(t) = \begin{array}{c} \\ \boldsymbol{a_1} \\ \boldsymbol{a_2} \\ \boldsymbol{a_3} \end{array} \begin{array}{c} \begin{array}{ccccc} \boldsymbol{u_1} & \boldsymbol{u_2} & \boldsymbol{u_3} & \boldsymbol{u_4} & \boldsymbol{u_5} \end{array} \\ \begin{bmatrix} f_{11}(t) & f_{12}(t) & f_{13}(t) & f_{14}(t) & f_{15}(t) \\ f_{21}(t) & f_{22}(t) & f_{23}(t) & f_{24}(t) & f_{25}(t) \\ f_{31}(t) & f_{32}(t) & f_{33}(t) & f_{34}(t) & f_{35}(t) \end{bmatrix} \end{array}$$

$$
= \begin{array}{c c}
 & \begin{array}{ccccc} \boldsymbol{u_1} & \boldsymbol{u_2} & \boldsymbol{u_3} & \boldsymbol{u_4} & \boldsymbol{u_5} \end{array} \\
\begin{array}{c} \boldsymbol{a_1} \\ \boldsymbol{a_2} \\ \boldsymbol{a_3} \end{array} & \begin{bmatrix} 0 & 0 & \mathbf{1} & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathbf{1} \\ 0 & \mathbf{1} & 0 & 0 & 0 \end{bmatrix}
\end{array}. \tag{5.17}
$$

As a non-handoff user, u_1 will not get any new access point. User u_1 stays in the connect with its current access point a_1. Consequently, access point a_1 can only serve one more user. In order to maximize the sum of channel capacities that handoff users can obtain from their new access points, the SDN controller allocates access point a_1 to user u_3 instead of user u_2. For user u_4, since the channel capacities provided by its available access points a_2 and a_3 cannot satisfy the basic bandwidth requirements, the vertical handoff of u_4 is failed. User u_4 does not have any new access point, and will be discarded by its current access point a_2. According to the network selection result $\mathcal{F}(t)$, the SDN controller sends out feedback frames. The value of $f_{13}(t)$ is 1 means the new access point of user u_3 is a_1. That is, u_3 should transfer its inter-network connection from the current access point a_3 to a_1. Therefore, a_1 will send a feedback frame to u_3 to notify this selection result. Similarly, a_2 sends a feedback frame to u_5. a_3 sends a feedback frame to u_2.

5.4 Handoff Timing Algorithm

Since users are always moving around in wireless environment, the network selection results should not be implemented immediately. For example as shown in Fig. 5.6, user u_1 is moving back and forth. When user u_1 leaves the access point a_1 and closes to the access point a_2, its inter-network connection will be transferred from a_1 to a_2. When u_1 leaves a_2 and backs to a_1, the inter-network connection will be transferred from a_2 to a_1 again. The inter-network connection of user u_1 is switched between access points a_1 and a_2 times and times again [90]. This common example reveals that inappropriate handoff timing will incur numerous unnecessary handoffs.

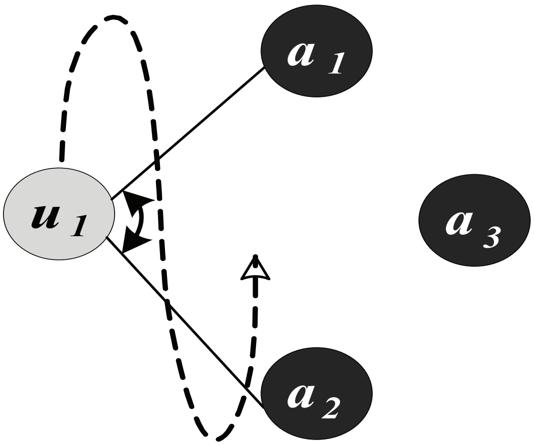


Figure 5.6: An example of the handoff timing issue.

5.4.1 Discussions

In our proposed scheme, users will wait for a stability period τ [23] after their network selection processes are finished. During the stability period, if the new access points are consistently more appropriate than their current access points, users will handoff to their new access points. Before we start introducing our proposed handoff timing algorithm, there are three things needed to be explained specially.

The first thing is about the meaning of *appropriate*. In this paper, we use “appropriate” instead of “better”. The reason is even the new access point is better than the current one, if the current access point can satisfy the basic bandwidth requirement of a user, this user should not perform the vertical handoff.

The second thing is about the handoff overhead. When we are judging whether a new access point is appropriate or not, we need to take both the performance gain and the handoff overhead into account. According to the IEEE 802.11 standard handoff procedure [91], users will experience the Authentication phase and the Reassociation phase before transferring to their selected new access points. Anshul *et al.* [92] studied various scenarios and observed that the average Authentication delay and the average Reassociation delay are 1.3 ms and 2.3 ms respectively. During the Authentication phase, two Authentication frames will be exchanged between the handoff user and the new access point. Normally, an Authentication frame contains 24 Bytes header, at least

6 Bytes body and 4 Bytes Frame Check Sequence (FCS). Therefore, the control frames will cost o_a Mbps bandwidth during the Authentication phase as the following equation shows.

$$o_a = \frac{2 \cdot (3.4e-5) \text{ Mb}}{1.3e-3 \text{ s}} \doteq 0.0052 \text{ Mbps}. \tag{5.18}$$

During the Reassociation phase, the handoff user will send out an Reassociation-Request frame and the new access point will reply an Reassociation-Response frame. An Reassociation-Request frame consists of 24 Bytes header, at least 10 Bytes body and 4 Bytes FCS. While an Reassociation-Response frame has 24 Bytes header, at least 6 Bytes body and 4 Bytes FCS. Therefore, the control frames will cost o_r Mbps bandwidth during the Reassociation phase as the following equation shows.

$$o_r = \frac{(3.8e-5+3.4e-5) \text{ Mb}}{2.3e-3 \text{ s}} \doteq 0.0031 \text{ Mbps}. \tag{5.19}$$

The last thing is about the length of a *stability period*. For a handoff user u_j, its available access points set is denoted by $\mathcal{A}'_j$, and the size of $\mathcal{A}'_j$ is $|\mathcal{A}'_j|$. Let τ_x be the length of stability period in the x th round, where $x = 1, 2, \cdots$. For an available access point a_i ($a_i \in \mathcal{A}'_j$), its channel capacity at time t is denoted by $q_{ij}(t)$. After waiting for a stability period τ_x, the channel capacity of a_i becomes $q_{ij}(t+\tau_x)$. Thus, the change rate of channel capacity during the stability period τ_x can be represented by $\frac{q_{ij}(t+\tau_x)}{q_{ij}(t)}$. We make use of the average change rate of channel capacities for all available access points to adjust the length of stability period. If u_j is suggested to wait for another stability period, the value of τ_{x+1} can be calculated as follows.

$$\tau_{x+1} = \frac{\sum_{a_i \in \mathcal{A}'_j} \dfrac{q_{ij}(t+\tau_x)}{q_{ij}(t)}}{|\mathcal{A}'_j|} \tau_x. \tag{5.20}$$

Eq.(20) indicates that if the average change rate of channel capacities during τ_x is bigger than 1 (*i.e.*, $\tau_{x+1} > \tau_x$), that is the connection quality of user u_j is getting better, u_j will wait for a longer stability period in the next round. Otherwise, u_j will wait for a shorter stability period in the next round. H.J. Wang *et al.* [23] pointed out that the stability period is proportional to the handoff latency. Moreover, S. Sharma *et al.* [93]

proved by experiment that the handoff latency is about 0.1 s. Thus, we accordingly set the initial value of stability period τ_1 to be 0.1 s in this chapter.

5.4.2 Algorithm Details

In this subsection, we take a handoff user u_i for instance to explain our proposed handoff timing algorithm. For the user u_1, suppose that its current access point is a_1, and its new access point which has been determined by the SDN controller is a_2 as shown in Fig.5.7. Following the previous definitions, the channel capacities of a_1 and a_2 at time t were $q_{11}(t)$ and $q_{21}(t)$ respectively. After waiting for a stability period τ, the channel capacities of a_1 and a_2 become $q_{11}(t+\tau)$ and $q_{21}(t+\tau)$ respectively. Based on the channel capacities, the Authentication overhead o_a and the Reassociation overhead o_r, the SDN controller makes a judgement and notifies u_i whether it should transfer the inter-network connection from a_1 to a_2 or not.

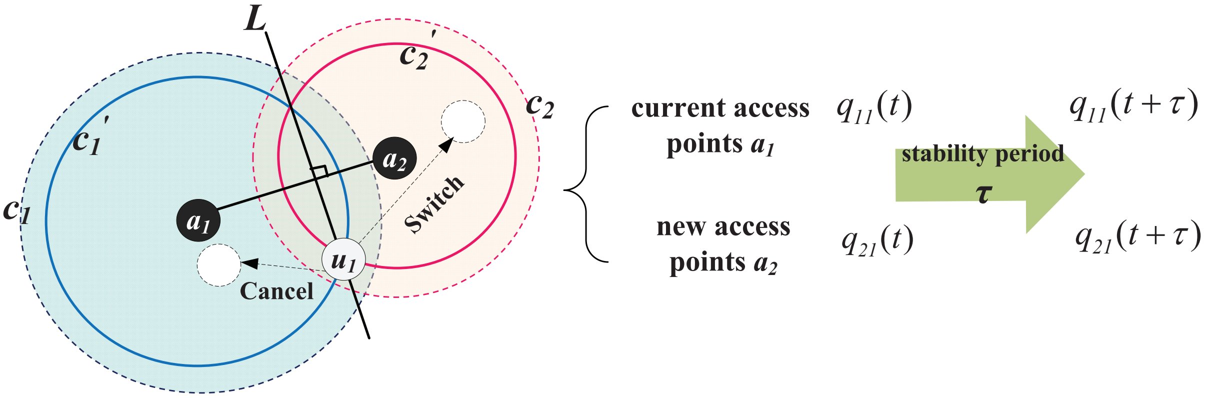


FIGURE 5.7: An example of the proposed handoff timing algorithm.

The network performance can be regarded as stable when considering the user mobility, which is a widely accepted assumption [94], [95]. Therefore, we are working under the assumptions as shown in Fig. 5.7. The coverage area of a_1 is the circular region insides c_1. The closer to a_1, the higher channel capacity that u_1 can get from a_1. c_1 and c_1' are concentric circles. If u_1 moves along c_1', the channel capacity of (a_1, u_1) will not change [96]. Similarly, the coverage area of a_2 is the circular region insides c_2. If u_1 moves along c_2', the channel capacity of (a_2, u_1) will not change. There is a line $\mathcal{L}$ goes through the intersections of c_1' and c_2'. $\mathcal{L}$ is perpendicular to the line between a_1 and a_2. We consider the following two situations.

- If u_1 moves to the left of $\mathcal{L}$, which means u_1 has the tendency of moving closer to a_1. Since user was moving back to its current access point during the stability period, it does not need vertical handoff anymore, and the network selection result is cancelled.
- If u_1 moves to the right of $\mathcal{L}$, which means u_1 has the tendency of moving closer to a_2. Since user was moving away from its current access point during the stability period, it has to transfer the inter-network connection to the new access point at once.

Since the movement trend of user is important for the vertical handoff, some related work tried to predict the movement trend of a user. The existing work is based on location information [97], context [98] or historical record [99] and so on. Each of them requires a large amount of storage space. In this chapter, we predict the movement trend of a user just based on the channel capacities of its current access point and new access point. Discussions are provided for the following nine cases.

1) $q_{11}(t) < q_{11}(t+\tau)$ and $q_{21}(t) < q_{21}(t+\tau)$. Since the channel capacity of a_1 increases, u_1 must be inside c_1'. For the same reason, u_1 is also inside c_2'. That is to say, after a stability period, u_1 locates at the domain d_1 shown in Fig. 5.8 (a). Since the line $\mathcal{L}$ passing through d_1, we cannot determine the movement trend of u_1. As a result, u_1 should wait for another stability period, and then analyze the situation again.

2) $q_{11}(t) < q_{11}(t+\tau)$ and $q_{21}(t) = q_{21}(t+\tau)$. Since the channel capacity of a_2 does not change, u_1 must locate at c_2'. Furthermore, u_1 is inside c_1'. That is to say, after a stability period, u_1 locates at the line segment l_1 shown in Fig. 5.8 (b). l_1 is on the left of $\mathcal{L}$, which means u_1 moves back. As a result, u_1 does not need vertical handoff anymore.

3) $q_{11}(t) < q_{11}(t+\tau)$ and $q_{21}(t) > q_{21}(t+\tau)$. Since the channel capacity of a_2 decreases, u_1 must be outside c_2'. Furthermore, u_1 is inside c_1'. That is to say, after a

stability period, u_1 locates at the domain d_2 shown in Fig. 5.8 (a). d_2 is on the left of $\mathcal{L}$, which means u_1 moves back. As a result, u_1 does not need vertical handoff anymore.

4) $q_{11}(t) = q_{11}(t+\tau)$ and $q_{21}(t) < q_{21}(t+\tau)$. Since the channel capacity of a_1 does not change, u_1 must locate at c'_1. Furthermore, u_1 is inside c'_2. That is to say, after a stability period, u_1 locates at the line segment l_2 shown in Fig. 5.8 (b). l_2 is on the right of $\mathcal{L}$, which means u_1 moves away. For this case, if the handoff overhead is less than the performance gain (*i.e.* $o_a + o_r < q_{21}(t+\tau) - q_{11}(t+\tau)$), u_1 should handoff to the new access point a_2 at once. Otherwise, u_1 will initialize another network selection.

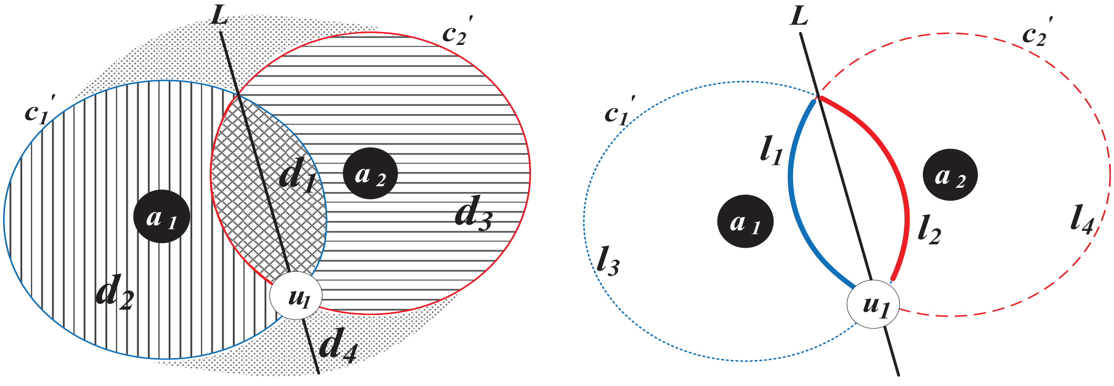
5) $q_{11}(t) = q_{11}(t+\tau)$ and $q_{21}(t) = q_{21}(t+\tau)$. Since the channel capacities of a_1 and a_2 have no change, u_1 still locates at the original point after a stability period. We cannot determine the movement trend of u_1. As a result, u_1 should wait for another stability period, and then analyze the situation again.

6) $q_{11}(t) = q_{11}(t+\tau)$ and $q_{21}(t) > q_{21}(t+\tau)$. Since the channel capacity of a_2 decreases, u_1 must be outside c'_2. Furthermore, u_1 locates at c'_1. That is to say, after a stability period, u_1 locates at the line segment l_3 shown in Fig. 5.8 (b). l_3 is on the left of $\mathcal{L}$, so u_1 does not need vertical handoff anymore.

7) $q_{11}(t) > q_{11}(t+\tau)$ and $q_{21}(t) < q_{21}(t+\tau)$. Since the channel capacity of a_1 decreases, u_1 must be outside c'_1. Furthermore, the channel capacity of a_2 increases, u_1 must be inside c'_2. That is to say, after a stability period, u_1 locates at the domain d_3 shown in Fig. 5.8 (a). Similar to case 4, if the handoff overhead is less than the performance gain (*i.e.* $o_a + o_r < q_{21}(t+\tau) - q_{11}(t+\tau)$), u_1 should handoff to the new access point a_2 at once. Otherwise, u_1 will initialize another network selection.

8) $q_{11}(t) > q_{11}(t+\tau)$ and $q_{21}(t) = q_{21}(t+\tau)$. Since the channel capacity of a_2 does not change, u_1 must locate at c'_2. Furthermore, the channel capacity of a_1 decreases, u_1 is outside c'_1. That is to say, after a stability period, u_1 locates at the line segment l_4 shown in Fig. 5.8 (b). l_4 is on the right of $\mathcal{L}$, so u_1 takes the same measures as case 4 and case 7.

9) $q_{11}(t) > q_{11}(t+\tau)$ and $q_{21}(t) > q_{21}(t+\tau)$. Since the channel capacity of a_2 decreases, u_1 must be outside c'_2. Furthermore, u_1 is outside c'_1. That is to say, after a stability period, u_1 locates at the domain d_4 shown in Fig. 5.8 (a). The line $\mathcal{L}$ passing



(a) user moves into a region (b) user moves on a line

FIGURE 5.8: The movement direction of a user.

through d_4, so we cannot determine the movement trend of u_1. As a result, u_1 should wait for another stability period, then analyze the situation again.

As we have explained that after the new access point is selected, a user should transfer its inter-network connection to the new access point at an appropriate time. For this purpose, we proposed a handoff timing algorithm. The SDN controller only needs to know the channel capacities of current access point and new access point for each handoff user. Based on limited information, the SDN controller can determine the time when a handoff user should implement its network selection result. Following the above approach we can see that the network selection result will be implemented only if the user is certain to move away from its current access point.

5.5 Performance Evaluation

In this section, we provide the performance evaluation of our Software-Defined Networking based Vertical Handoff (S-DNVH) scheme. We compare the proposed scheme with two typical existing schemes: the Always Best Connected (ABC) scheme [9] and the Smooth Adaptive Soft Handover Algorithm (SASHA) [26] under various network conditions. The concerned performance metrics are the number of handoffs, total throughput and the user served ratio. As we know, the SDN technique originated in and is commonly used in the campus networks. Furthermore, most of access points in the campus networks are the IEEE 802.11 standard [100] supported devices. Therefore, we will refer to the IEEE 802.11 standard to set the parameters of access points during the experiment. If other standards are required in the practical applications, this experiment procedure can be repeated by using the corresponding parameters. Simulation experiments are repeated one thousand times and the results are presented with 95% confidence interval.

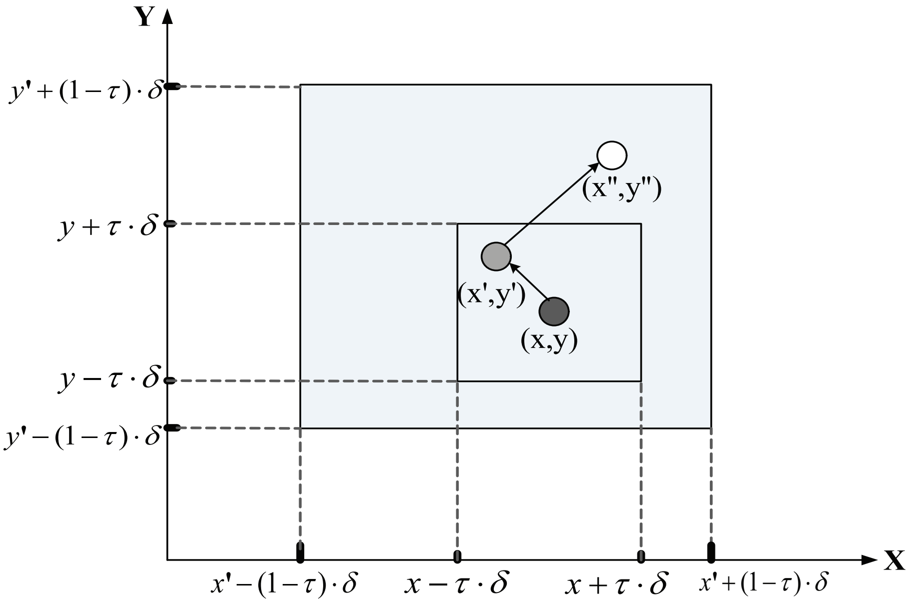


FIGURE 5.9: Movement model of users.

5.5.1 Experiment Setup

Over a 500m × 500m rectangular flat space, there are 9 access points and n users. These 9 access points belong to 3 different types, and each type has 3 entities. Different types of access points have different coverage radii, number of channels, bandwidths and other different attributes. Furthermore, different types of access points are assumed to use different frequency bands, and same type of access points use the same frequency band. An access point is available to a user when the distance between them is smaller than the coverage radius of this access point. Users are moving around inside the considered area. If the current location of a user is denoted by a two-dimensional coordinate (x, y), this user will be inside $(x\pm \triangle t \cdot \delta, y\pm \triangle t \cdot \delta)$ after a period of time $\triangle t$, where δ is the maximal moving velocity of the user. Following this rule, the movement model of users used in our experiments can be illustrated as Fig. 5.9, where (x, y) is the location of a user at the beginning of a time slot, (x', y') is the location of the user after a stability period, and (x'', y'') is the location of the user at the end of the time slot. We set the basic bandwidth requirements of users to be 2 Mbps, which corresponds to the video conference demanding. Based on references [101] [102], we have the main experimental parameters which are listed in Table 5.2.

In order to construct the SDN scenario and evaluate the performance of proposed S-DNVH scheme, we simulate the OpenFlow switches by using the Mininet VM 2.2.1. The SDN controller is implemented through the Floodlight V 1.2. The SDN controller and OpenFlow switches exchange information via OpenFlow 1.3. In order to avoid too much change to the SDN controller, the calculation work is assigned to MATLAB. Hence, a logic complete SDN controller is composed by Floodlight and MATLAB. At first, we write a Python file which is able to insert the features of users into the Data fields of Packet-In messages. We make use of the Packet-In message to send information from OpenFlow switches to the SDN controller. Based on Eclipse, we encapsulate an I/O API for the SDN controller. This API is used for SDN controller to generate a file which contains the information received from switches, and read a file which contains the network selection results calculated by the CVX on MATLAB. Once the SDN controller gets the network selection results, it will insert the network selection results into the Data fields of Packet-Out messages, and send the Packet-Out messages to OpenFlow switches. The architecture of our simulation system is shown in Fig. 5.10.

5.5.2 Experiment Results

5.5.2.1 Number of Vertical Handoffs

We study the number of vertical handoffs when the number of users varies under free space propagation ($\alpha = 2$), flat-earth reflection ($\alpha = 3$), and diffraction losses ($\alpha = 4$) environment conditions in Fig. 5.11. The general trend is that there will be more vertical handoffs in each time slot as the number of users increases. For the same

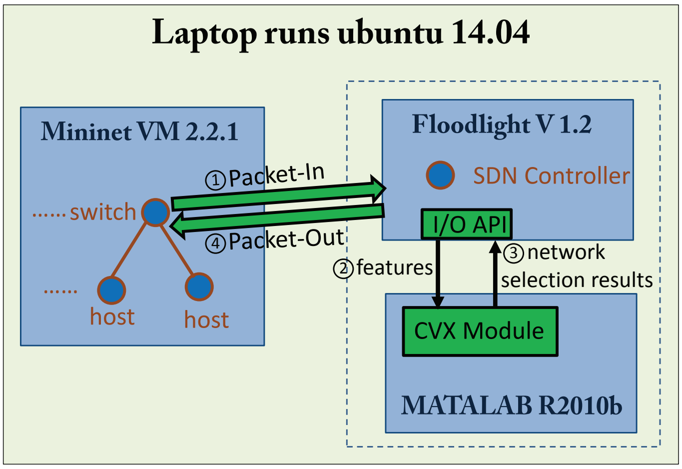


Figure 5.10: Structure of simulation system.

Table 5.2: Experimental Parameters for SDN based Vertical Handoff Scheme

Parameter	**Value**
Simulator	MATLAB, Flooding, MATLAB
Number of access points	9 (each kind has 3 devices)
Coverage radii of access points	30 m, 50 m, 100 m
Transmission powers of access points	0.002 watts, 0.005 watts, 0.02 watts
Maximal number of serving users in access points	10, 20, 30
Basic bandwidth requirements of users	2 Mbps
Bandwidths of access points	20 MHz, 40 MHz, 40 MHz
Initial value of stability period τ_1	0.1 second
Time slot	1 second
Channel fading gain h	$h \sim exp(1)$
Noise power per channel n	$n \sim N(0,1)$ watts
Moving velocities of users	$0 \sim 5$ m/s
Confidence interval δ	95 %

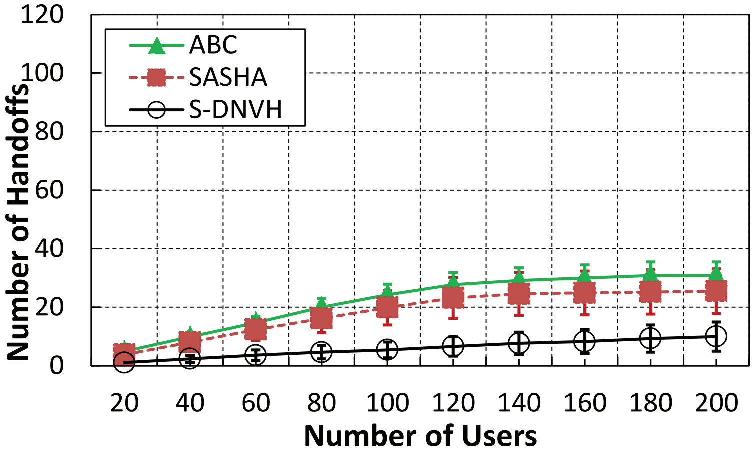
number of users, larger pass loss exponent environments always have more handoffs. From the longitudinal comparison we observe that the proposed S-DNVH scheme can reduce the number of handoffs significantly. This advantage is more remarkable in tough environments. For example in Fig. 5.11 (c), the number of handoffs in S-DNVH scheme is nearly 76% less than ABC scheme in the worst case.

Another interesting phenomenon is that the number of vertical handoffs will increase as more users join in. After the number of users is bigger than around 120, the number of vertical handoffs in the two contrast schemes becomes relative stable. However, this phenomenon does not exist in our proposed scheme. We also observe that the ABC scheme is particularly sensitive to environment condition. The number of handoffs in ABC varies dramatically as the pass loss exponent changes. Compared to ABC, situations in other two schemes are much more peaceful. Furthermore, the number of handoffs in SASHA is getting closer to that in our proposed S-DNVH when the pass loss exponent increases.

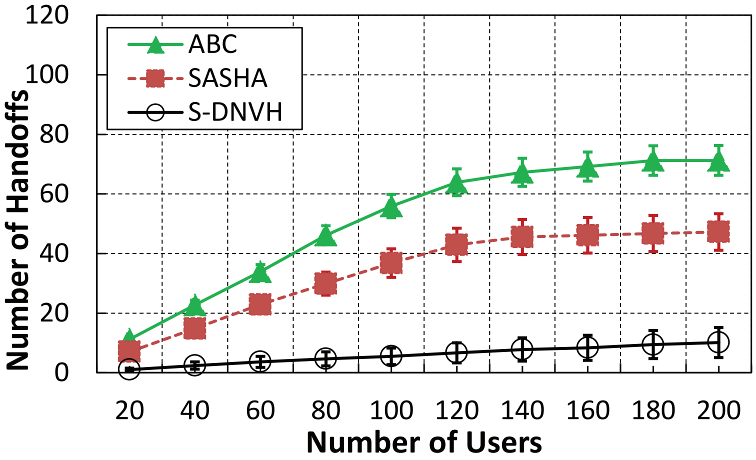
5.5.2.2 Throughput

We compare the total throughput under different network settings in Fig. 5.12. The total throughput is defined as the sum of channel capacities that non-handoff users and handoff users can obtain. Since each user is assumed to occupy at most one channel of access point, higher total throughput will be achieved as more users join in at first. Then, the value of total throughput has slower growth when the number of users is larger than a certain value. From Fig. 5.12 we find that, this special value is around 160. Note that, the experimental 9 access points theoretically can support maximum 180 users at the same time. Our experimental result is very close to this theoretical value.

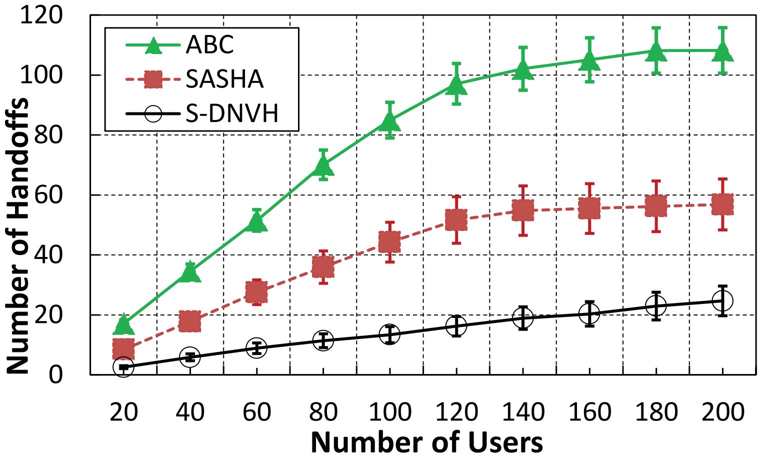
From any sub-figure of Fig. 5.12 we can find that S-DNVH scheme always has the highest total throughput. Fig. 5.12 also reveals that an inverse proportion operates between the total throughput and the pass loss exponent. Furthermore, from the crosswise comparison we can observe that the throughput in our proposed S-DNVH is getting closer to that in SASHA when the pass loss exponent increases. A similar tendency is also observed in Fig. 5.12.



(a) pass loss exponent $\alpha = 2$

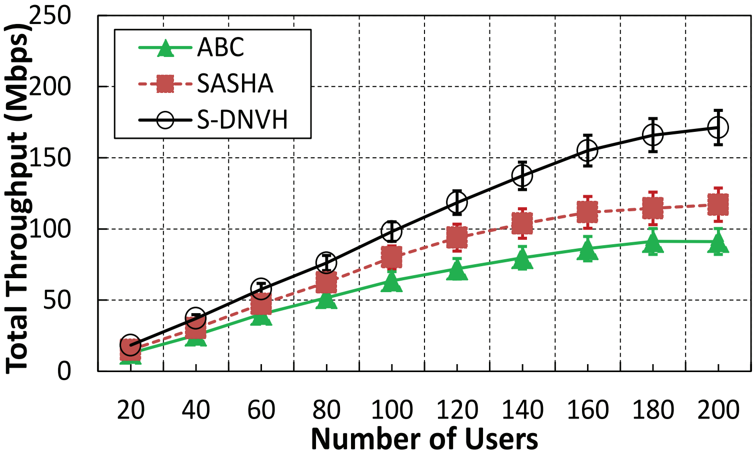


(b) pass loss exponent $\alpha = 3$

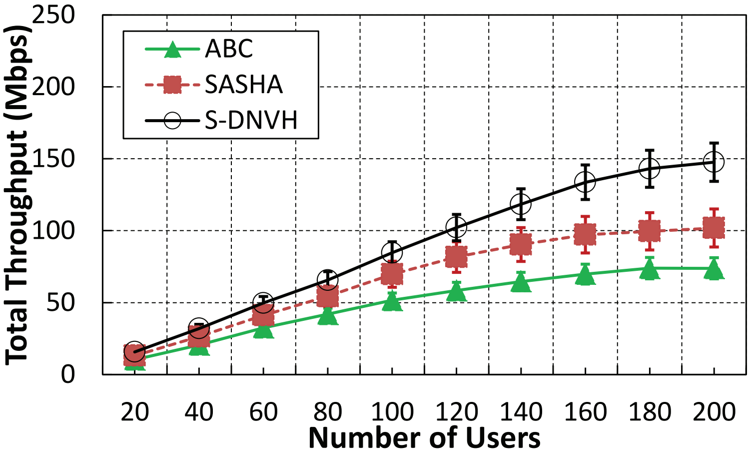


(c) pass loss exponent $\alpha = 4$

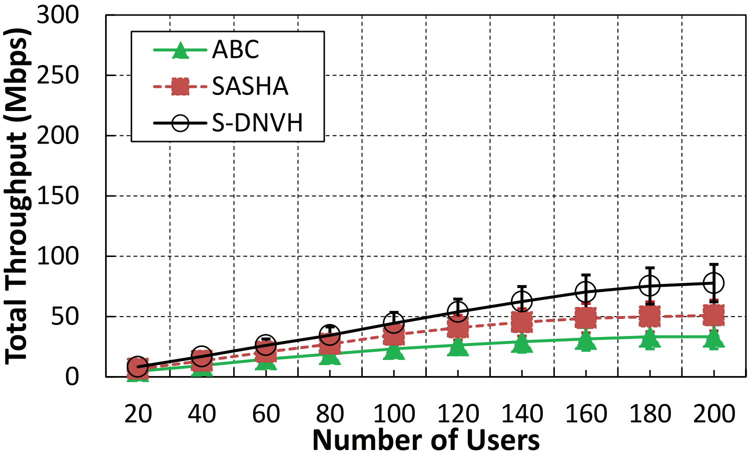
Figure 5.11: Number of users vs. number of handoffs.



(a) pass loss exponent $\alpha = 2$



(b) pass loss exponent $\alpha = 3$



(c) pass loss exponent $\alpha = 4$

Figure 5.12: Number of users vs. total throughput.

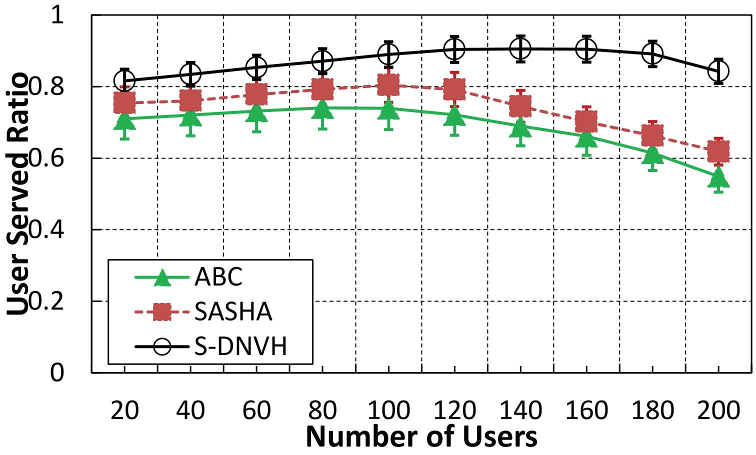
5.5.2.3 User served ratio

The user served ratio is the ratio of users who have the network service. Following the definitions and assumptions made in Section 3, the user served ratio is equal to $\frac{\sum_{j=1}^{n} v_j(t)+\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij}(t)}{n}$, where n is the number of users, $\sum_{j=1}^{n} v_j(t)$ is the number of non-handoff users and $\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij}(t)$ is the number of handoff users whose vertical handoffs are successful. We compare the user served ratio under different network settings in Fig. 5.13. It is shows that the proposed S-DNVH scheme always has the highest user served ratio in different scenarios. As the number of users increases, the user served ratio sightly increases then decreases. Moreover, user served ratio will be less in higher pass loss exponent environments. As the pass loss exponent increases, the user served ratios in SASHA and the proposed S-DNVH get closer to each other.

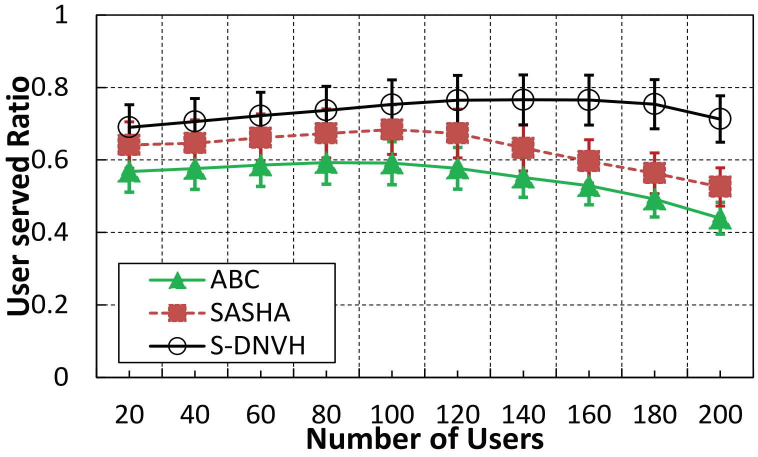
Another interesting observation is that the turning points in the two contrast schemes are around 120. This special value has also been found in Fig. 5.11. Thus, we can infer that the maximal number of served users in the two contrast schemes should be 120. If there are more than 120 users are added to the scenario, the performance of two contrast schemes will degrade. However, our proposed scheme does not be affected by this limitation. Compared with these two schemes, our proposed scheme S-DNVH can make fuller use of the system resources.

5.6 Summary

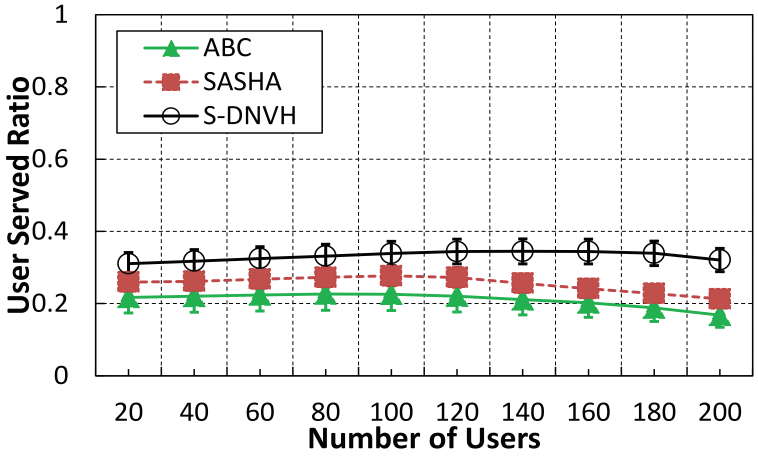
In this chapter, we proposed a novel vertical handoff scheme with the support of SDN technique. The proposed scheme ensures that a user will transfer to the most appropriate new access point at the most appropriate time. From the standpoint of users, we choose the channel capacity as the performance metric. When the channel capacities cannot meet their basic bandwidth requirements, users will initialize vertical handoffs. We formulated the network selection process as a 0-1 integer programming problem, with the objective of maximizing the sum of channel capacities that handoff users can obtain from their new access points. After the network selection process is finished, users have to wait for a stability period. Only if the new access points are consistently more appropriate than their current access points, will users transfer their inter-network connections to



(a) pass loss exponent $\alpha = 2$



(b) pass loss exponent $\alpha = 3$



(c) pass loss exponent $\alpha = 4$

FIGURE 5.13: Number of users vs. user served ratio.

the new access points. We carried out comparison experiments under different network settings. Comparison results demonstrate that the proposed scheme reduces the number of vertical handoffs, maximizes the total throughput and user served ratio significantly.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

With the advancement of telecommunication technology, mobile devices with wireless functional are ubiquitous nowadays. An efficient and effective mobility management scheme for heterogeneous wireless networks becomes more and more important. Motivates by this, we carry out our research on mobility management which has been performed in the dissertation. We focus on three open problems of mobility management: pinball routing problem, network selection problem and handoff timing problem.

In Chapter 3, we investigated the pinball routing problem for nested mobile networks. In order to alleviate the heavy data transmission overhead and location update overhead caused by the pinball routing problem in inter-domain communication, we firstly proposed the self-adaptive route optimization scheme. The proposed self-adaptive route optimization scheme guarantees that the times of encapsulation is always smaller than 2 no matter how deep the destination locates at. Furthermore, we extend the self-adaptive route optimization scheme in order to optimize the routing process of intra-domain communication. In the extended scheme, packets of the intra-domain communication will be limited in a pretty small region. Consequently, the communication and location update overheads of inter-domain and intra-domain communications are reduced.

In Chapter 4, we investigated the network selection problem for distributed scenario, in which there is no centralized control entity. We fully considered the privacy-preservation of users and assumed that users will not share their private information with each others. Under this limitation, we proposed a distributed network selection scheme fulling multiple objectives for heterogeneous wireless network environment. In the proposed distributed network selection scheme, each user will perform its own network selection based on limited local information. The network selection problem is formulated as a multi-objective optimization problem which maximizes the channel capacity and minimizes the blocking probability as the same time. Moreover, the computation and memory complexities of the proposed scheme are relatively small, which is a competitive advantage for practical applications.

In Chapter 5, we studied the network selection problem and the handoff timing problem. We proposed a vertical handoff scheme with the support of SDN technique for heterogeneous wireless network environment. The proposed scheme ensures that a user will transfer to the most appropriate new network devices at the most appropriate time. The network selection problem is formulated as a 0-1 integer programming problem which maximizes the sum of channel capacities of handoff users. After the network selection process is finished, we further propose an algorithm to determine the most appropriate handoff timing. Through intensive comparison experiments, we verify that our proposed vertical handoff scheme reduces the number of vertical handoffs, maximizes the total throughput and user served ratio significantly.

6.2 Future Work

As we move toward 5G era, environment becomes so complex that the mobility management facing with new challenges. The SDN technique is an effective tool to deal with networking problems. As the future work, we want to study the mobility management in 5G and SDN combined environments. Furthermore, the increasingly serious data security problem requires the mobility management strategy to consider the security and privacy-preservation.

Bibliography

[1] Ian F Akyildiz, Jiang Xie, and Shantidev Mohanty. A survey of mobility management in next-generation all-ip-based wireless systems. *IEEE Wireless Communications*, 11(4):16–28, 2004.

[2] Debashis Saha, Amitava Mukherjee, Iti Saha Misra, and Mohuya Chakraborty. Mobility support in ip: a survey of related protocols. *IEEE Network*, 18(6):34–40, 2004.

[3] Vincent WS Wong and Victor Leung. Location management for next-generation personal communications networks. *IEEE Network*, 14(5):18–24, 2000.

[4] Kun Zhu, Dusit Niyato, Ping Wang, Ekram Hossain, and Dong In Kim. Mobility and handoff management in vehicular networks: a survey. *Wireless communications and mobile computing*, 11(4):459–476, 2011.

[5] Mussa Bshara, Umut Orguner, Fredrik Gustafsson, and Leo Van Biesen. Fingerprinting localization in wireless networks based on received-signal-strength measurements: A case study on wimax networks. *IEEE Transactions on Vehicular Technology*, 59(1):283–294, 2010.

[6] Charles Perkins, David Johnson, and Jari Arkko. Mobility support in ipv6. Technical report, 2011.

[7] Vijay Devarapalli, Ryuji Wakikawa, Alexandru Petrescu, and Pascal Thubert. Network mobility (nemo) basic support protocol. Technical report, 2004.

[8] Jeffrey G Andrews, Stefano Buzzi, Wan Choi, Stephen V Hanly, Aurelie Lozano, Anthony CK Soong, and Jianzhong Charlie Zhang. What will 5g be? *IEEE Journal on Selected Areas in Communications*, 32(6):1065–1082, 2014.

[9] Malamati Louta and Paolo Bellavista. Bringing always best connectivity vision a step closer: challenges and perspectives. *IEEE Communications Magazine*, 51(2):158–166, 2013.

[10] Johann Márquez-Barja, Carlos T Calafate, Juan-Carlos Cano, and Pietro Manzoni. An overview of vertical handover techniques: Algorithms, protocols and tools. *Computer Communications*, 34(8):985–997, 2011.

[11] Jasone Astorga, Marina Aguado, Nerea Toledo, and Marivi Higuero. A high performance link layer mobility management strategy for professional private broadband networks. *Journal of Network and Computer Applications*, 36(4):1152–1163, 2013.

[12] Thienne Johnson, Eduardo Zagari, Eleri Cardozo, Rodrigo Prado, and Tomas Badan. *Mobility in IP networks: from link layer to application layer protocols and architectures.* INTECH Open Access Publisher, 2010.

[13] Mohammed Atiquzzaman and Abu S Reaz. Survey and classification of transport layer mobility management schemes. In *Proceedings of the IEEE 16th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2005)*, volume 4, pages 2109–2115. IEEE, 2005.

[14] Sangheon Pack, Taekyoung Kwon, Yanghee Choi, and Eun Kyoung Paik. An adaptive network mobility support protocol in hierarchical mobile ipv6 networks. *IEEE Transactions on Vehicular Technology*, 58(7):3627–3639, 2009.

[15] Hanlim Kim, Geunhyung Kim, and Cheeha Kim. S-ro: Simple route optimization scheme with nemo transparency. In *Information Networking. Convergence in Broadband and Mobile Networking*, pages 401–411. Springer, 2005.

[16] Hosik Cho, Taekyoung Kwon, and Yanghee Choi. Route optimization using tree information option for nested mobile networks. *IEEE Journal on Selected Areas in Communications*, 24(9):1717–1724, 2006.

[17] Md Humayun Kabir, Mohammad Mukhtaruzzaman, and Mohammed Atiquzzaman. Efficient route optimization scheme for nested-nemo. *Journal of Network and Computer Applications*, 36(3):1039–1049, 2013.

[18] Geng-Sheng GS Kuo and Kang Ji. Nxg04-6: Novel hierarchical network mobility support protocol with bidirectional end-to-end route optimization solution for nested mobile networks. In *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM 2006)*, pages 1–6. IEEE, 2006.

[19] Jindong Hou and D O'brien. Vertical handover-decision-making algorithm using fuzzy logic for the integrated radio-and-ow system. *IEEE Transactions on Wireless Communications*, 5(1):176–185, 2006.

[20] Abhishek Roy, Jitae Shin, and Navrati Saxena. Multi-objective handover in lte macro/femto-cell networks. *Journal of Communications and Networks*, 14(5):578–587, 2012.

[21] Quoc-Thinh Nguyen-Vuong, Nazim Agoulmine, El Hadi Cherkaoui, and Laura Toni. Multi-criteria optimization of access selection to improve the quality of experience in heterogeneous wireless access networks. *IEEE transactions on vehicular technology*, 62(EPFL-ARTICLE-182868):1785–1800, 2013.

[22] Chao Liu, Yong Sun, Peng Yang, Zhen Liu, Haijun Zhang, and Xiangming Wen. A two-step vertical handoff decision algorithm based on dynamic weight compensation. In *Proceedings of The 2013 IEEE International Conference on Communications Workshops (IEEE ICC 2013)*, pages 1031–1035. IEEE, 2013.

[23] Harry Jiannan Wang, Randy H Katz, and Jochen Giese. Policy-enabled handoffs across heterogeneous wireless networks. In *Proceedings of the Second IEEE Workshop on Mobile Computing Systems and Applications (WMCSA 1999)*, pages 51–60. IEEE, 1999.

[24] Wen-Tsuen Chen, Jen-Chu Liu, and Hsieh-Kuan Huang. An adaptive scheme for vertical handoff in wireless overlay networks. In *Proceedings of the Tenth International Conference on Parallel and Distributed Systems (ICPADS 2004)*, pages 541–548. IEEE, 2004.

[25] Woongsup Lee and Dong-Ho Cho. Enhanced group handover scheme in multiaccess networks. *IEEE Transactions on Vehicular Technology*, 60(5):2389–2395, 2011.

[26] Bogdan Ciubotaru and G Muntean. Sashaa quality-oriented handover algorithm for multimedia content delivery to mobile users. *IEEE Transactions on Broadcasting*, 55(2):437–450, 2009.

[27] Alpesh Patel, Kent Leung, Mohamed Khalil, Haseeb Akhtar, and Kuntal Chowdhury. Mobile node identifier option for mobile ipv6 (mipv6). Technical report, 2005.

[28] Faqir Zarrar Yousaf and Christian Wietfeld. Solving pinball routing, race condition and loop formation issues in nested mobile networks. *Computer Networks*, 56(4):1357–1375, 2012.

[29] Li Qiang, Jie Li, Mohsen Guizani, and Yusheng Ji. An adaptive route optimization scheme for nested mobile ipv6 nemo environment. In *Proceedings of the 12th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt 2014)*, pages 373–380. IEEE, 2014.

[30] Abu Zafar M Shahriar, Mohammed Atiquzzaman, and William Ivancic. Route optimization in network mobility: Solutions, classification, comparison, and future research directions. *IEEE Communications Surveys & Tutorials*, 12(1):24–38, 2010.

[31] Ahmed A Mosa, Aisha H Abdalla, and Rashid A Saeed. Evaluation of manemo route optimization schemes. *Journal of Network and Computer Applications*, 35(5):1454–1472, 2012.

[32] Kenneth P Laberteaux, Tony K Mak, and Raja Sengupta. System and method of vehicular wireless communication using an integrated coordinating access point, June 28 2011. US Patent 7,970,540.

[33] Jarno Rajahalme, Shane Amante, Sheng Jiang, and Brian Carpenter. Ipv6 flow label specification. 2011.

[34] David Johnson, Charles Perkins, and Jari Arkko. Mobility support in ipv6. Technical report, 2004.

[35] Samer S Hassan and Rosilah Hassan. Ipv6 network mobility route optimization survey. *American Journal of Applied Sciences*, 8(6):579, 2011.

[36] Pascal Thubert, C Bontoux, N Montavont, and B McCarthy. Nested nemo tree discovery. *IETF, draft-thubert-tree-discovery-01. txt (work-in-progress)*, 2004.

[37] Jie Li, Hisao Kameda, and Keqin Li. Optimal dynamic mobility management for pcs networks. *IEEE/ACM Transactions on Networking (TON)*, 8(3):319–327, 2000.

[38] Theresa CY Wang and Steven H Moritz. Mobile unit tracking system, November 15 1994. US Patent 5,365,451.

[39] Jong-Hyouk Lee, Sri Gundavelli, and Tai-Myoung Chung. A performance analysis on route optimization for proxy mobile ipv6. In *Proceedings of the IEEE International Conference on Communications (ICC 2009)*, pages 1–6. IEEE, 2009.

[40] Yi-Bing Lin. Reducing location update cost in a pcs network. *IEEE/ACM Transactions on Networking (TON)*, 5(1):25–33, 1997.

[41] Dae-Man Han and Jae-Hyun Lim. Smart home energy management system using ieee 802.15. 4 and zigbee. *IEEE Transactions on Consumer Electronics*, 56(3):1403–1410, 2010.

[42] Kyunghan Lee, Joohyun Lee, Yung Yi, Injong Rhee, and Song Chong. Mobile data offloading: how much can wifi deliver? In *Proceedings of the 6th International Conference*, page 26. ACM, 2010.

[43] Yunbo Wang, Mehmet C Vuran, and Steve Goddard. Cross-layer analysis of the end-to-end delay distribution in wireless sensor networks. *IEEE/ACM Transactions on Networking (TON)*, 20(1):305–318, 2012.

[44] Pascal Thubert, Marco Molteni, C Ng, H Ohnishi, and E Paik. Taxonomy of route optimization models in the nemo context. *(work in progress). Internet Draft (draft-thubert-nemo-ro-taxonomy-02), Internet Engineering Task Force*, 2004.

[45] Robert J Donaghey. Personal area network with automatic attachment and detachment, May 15 2007. US Patent 7,218,633.

[46] Shr-Shiuan Tzeng, Hsiao-Chen Kuo, Long-Sheng Li, and Yong-Yu Yang. An efficient multicast scheme in the nested nemo. In *Proceedings of the International Symposium on Computer Communication Control and Automation (3CA 2010)*, volume 2, pages 95–98. IEEE, 2010.

[47] Ming Li, Wenjing Lou, and Kui Ren. Data security and privacy in wireless body area networks. *IEEE Wireless Communications*, 17(1):51–58, 2010.

[48] Cheng-Xiang Wang, Fourat Haider, Xiqi Gao, Xiao-Hu You, Yang Yang, Dongfeng Yuan, Hadi Aggoune, Harald Haas, Sam Fletcher, and Erol Hepsaydir. Cellular architecture and key technologies for 5g wireless communication networks. *IEEE Communications Magazine*, 52(2):122–130, 2014.

[49] Ruidong Li, Jie Li, Kui Wu, Yang Xiao, and Jiang Xie. An enhanced fast handover with low latency for mobile ipv6. *IEEE Transactions on Wireless Communications*, 7(1):334–342, 2008.

[50] Yang Xiao, Yi Pan, and Jie Li. Design and analysis of location management for 3g cellular networks. *IEEE Transactions on Parallel and Distributed Systems*, 15(4):339–349, 2004.

[51] Khoriba Ghada, Jie Li, and Yusheng Ji. Energy and mobility-aware topology control in heterogeneous mobile ad hoc networks. *International Journal of Computational Science and Engineering*, 5(2):147–153, 2010.

[52] Jie Li, Yi Pan, and Yang Xiao. Performance study of multiple route dynamic source routing protocols for mobile ad hoc networks. *Journal of parallel and distributed computing*, 65(2):169–177, 2005.

[53] Toshikazu Sakano, Zubair Md Fadlullah, Thuan Ngo, Hiroki Nishiyama, Masataka Nakazawa, Fumiyuki Adachi, Nei Kato, Atsushi Takahara, Tomoaki Kumagai, Hironori Kasahara, et al. Disaster-resilient networking: a new vision based on movable and deployable resource units. *IEEE Network*, 27(4):40–46, 2013.

[54] Yang Xiao, Haizhon Li, Yi Pan, Kui Wu, and Jie Li. On optimizing energy consumption for mobile handsets. *IEEE Transactions on Vehicular Technology*, 53(6):1927–1941, 2004.

[55] Pyo ChangWoo, Li Jie, and Hisao Kameda. An efficient caching scheme for personal communication service networks. *IEICE transactions on communications*, 87(12):3603–3610, 2004.

[56] Kalyanmoy Deb. Multi-objective optimization. In *Search methodologies*, pages 403–449. Springer, 2014.

[57] Kenichi Taniuchi, Yoshihiro Ohba, Victor Fajardo, Subir Das, Miriam Tauil, Yuu-Heng Cheng, Ashutosh Dutta, Donald Baker, Maya Yajnik, and David Famolari. Ieee 802.21: Media independent handover: Features, applicability, and realization. *IEEE Communications Magazine*, 47(1):112–120, 2009.

[58] Meng Li, Hiroki Nishiyama, Nei Kato, Yasunori Owada, and Kiyoshi Hamaguchi. On the energy-efficient of throughput-based scheme using renewable energy for wireless mesh networks in disaster area. *IEEE Transactions on Emerging Topics in Computing*, 3(3):420–431, 2015.

[59] Jiajia Liu, Shangwei Zhang, Nei Kato, Hirotaka Ujikawa, and Kenichi Suzuki. Device-to-device communications for enhancing quality of experience in software defined multi-tier lte-a networks. *IEEE Network*, 29(4):46–52, 2015.

[60] ETSI TC-SMG. Digital cellular telecommunications system - location registration procedures. [Online] http://www.etsi.org/deliver/etsi_gts/03/0312/05.00.00_60/gsmts_0312v050000p.pdf, 1996.

[61] 3GPP Support Office. 3rd generation partnership project - technical specification group core network - location management procedures. [Online] http://www.qtc.jp/3GPP/Specs/23012-520.pdf, 2003.

[62] Jie Li and Hsiao-Hwa Chen. Mobility support for ip-based networks. *IEEE Communications Magazine*, 43(10):127–132, 2005.

[63] Adrian Vasilache, Jie Li, and Hisao Kameda. Threshold-based load balancing for multiple home agents in mobile ip networks. *Telecommunication Systems*, 22(1-4):11–31, 2003.

[64] Jie Li, Yi Pan, and Xiaohua Jia. Analysis of dynamic location management for pcs networks. *IEEE Transactions on Vehicular Technology*, 51(5):1109–1119, 2002.

[65] Jie Li and Yongping Pan. A dynamic hlr location management scheme for pcs networks. In *Proceedings of The 2014 IEEE Conference on Computer Communications (IEEE INFOCOM 2004)*, volume 1. IEEE, 2004.

[66] Chris W Liu. Modeling 3g/wcdma/hsdpa handset transmit system. *Microwave Product Digest*, 2007.

[67] Trichy Venkataraman Krishnamurthy and Rajaneesh Shetty. *4G: Deployment S-trategies and Operational Implications*. Apress, 2014.

[68] Jie Li, Atsushi Kubota, and Hisao Kameda. Location management for pcs networks with consideration of mobility patterns. In *Proceedings of The 2005 IEEE Conference on Computer Communications (IEEE INFOCOM 2005)*, volume 2, pages 1207–1216. IEEE, 2005.

[69] A Kubota, Jie Li, and Hisao Kameda. Optimal location management of mobile terminals with consideration of moving patterns. *(in Japanese), IPSJ (Information Processing Society of Japan) Journal*, 43(12):3922–3930, 2002.

[70] Xuanyu Cao, Xinbing Wang, and Songwu Lu. Function computation over heterogeneous wireless sensor networks. *IEEE Transactions on Parallel and Distributed Systems*, 25(7):1756–1766, 2014.

[71] Stenio Fernandes and Ahmed Karmouch. Vertical mobility management architectures in wireless networks: A comprehensive survey and future directions. *IEEE Communications Surveys & Tutorials*, 14(1):45–63, 2012.

[72] Mark Stemm and Randy H Katz. Vertical handoffs in wireless overlay networks. *Mobile Networks and applications*, 3(4):335–350, 1998.

[73] Suneth Namal, Ishtiaq Ahmad, Andrei Gurtov, and Mika Ylianttila. Sdn based inter-technology load balancing leveraged by flow admission control. In *Proceedings of the IEEE SDN for Future Networks and Services (SDN4FNS 2013)*, pages 1–5. IEEE, 2013.

[74] Qingyang Song and Abbas Jamalipour. A network selection mechanism for next generation networks. In *Proceedings of the IEEE International Conference on Communications (ICC 2005)*, volume 2, pages 1418–1422. IEEE, 2005.

[75] Xiaohuan Yan, Y Ahmet Şekercioğlu, and Sathya Narayanan. A survey of vertical handover decision algorithms in fourth generation heterogeneous wireless networks. *Computer Networks*, 54(11):1848–1863, 2010.

[76] William Stallings. Software-defined networks and openflow. *The Internet Protocol Journal*, 16(1):2–14, 2013.

[77] Thomas H Cormen. *Introduction to algorithms*. MIT press, 2009.

[78] Li Qiang, Jie Li, and Changcheng Huang. A software-defined network based vertical handoff scheme for heterogeneous wireless networks. In *Proceedings of The 2014 IEEE Global Communications Conference (IEEE GLOBECOM 2014)*, pages 4671–4676. IEEE, 2014.

[79] Sankalp Agarwal, Murali Kodialam, and TV Lakshman. Traffic engineering in software defined networks. In *Proceedings of the IEEE INFOCOM 2013*, pages 2211–2219. IEEE, 2013.

[80] H Yin, H Xie, T Tsou, D Lopez, P Aranda, and R Sidi. Sdni: A message exchange protocol for software defined networks (sdns) across multiple domains. *Work in Progress of IETF draft*, 2012.

[81] Sushant Jain, Alok Kumar, Subhasree Mandal, Joon Ong, Leon Poutievski, Arjun Singh, Subbaiah Venkata, Jim Wanderer, Junlan Zhou, Min Zhu, et al. B4: Experience with a globally-deployed software defined wan. In *ACM SIGCOMM Computer Communication Review*, volume 43, pages 3–14. ACM, 2013.

[82] Jeffrey G Andrews, François Baccelli, and Radha Krishna Ganti. A tractable approach to coverage and rate in cellular networks. *IEEE Transactions on Communications*, 59(11):3122–3134, 2011.

[83] Thomas David Novlan, Radha Krishna Ganti, Arunabha Ghosh, and Jeffrey G Andrews. Analytical evaluation of fractional frequency reuse for ofdma cellular

networks. *IEEE Transactions on Wireless Communications*, 10(12):4294–4305, 2011.

[84] Kemeng Yang, Iqbal Gondal, and et al. Combined sinr based vertical handoff algorithm for next generation heterogeneous wireless networks. In *Proceedings of The 2007 Global Telecommunications Conference (IEEE GLOBECOM 2007)*, pages 4483–4487. IEEE, 2007.

[85] Hiroshi Sato. The capacity of the gaussian interference channel under strong interference (corresp.). *IEEE Transactions on Information Theory*, 27(6):786–788, 1981.

[86] Claude Elwood Shannon. A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, 5(1):3–55, 2001.

[87] Pre-loaded switches. [Online] http://www.pica8.com/open-switching/1gbe-10gbe-40gbe-openswitches.php.

[88] Inter-Datacenter WAN with centralized TE using SDN and OpenFlow. [Online] https://www.opennetworking.org/images/stories/downloads/sdn-resources/customer-case-studies/cs-googlesdn.pdf.

[89] Nettle OpenFlow Messages. [Online] https://hackage.haskell.org/package/nettle-openflow-0.2.0/docs/Nettle-OpenFlow-Messages.html.

[90] Janise McNair and Fang Zhu. Vertical handoffs in fourth-generation multinetwork environments. *IEEE Wireless Communications*, 11(3):8–15, 2004.

[91] Arunesh Mishra, Minho Shin, and William Arbaugh. An empirical analysis of the ieee 802.11 mac layer handoff process. *ACM SIGCOMM Computer Communication Review*, 33(2):93–102, 2003.

[92] Handoff delay for 802.11 b wireless lans. [Online] http://www.cs.columbia.edu/∼hgs/papers/2003/AnshulJain80211.ppt.

[93] Srikant Sharma, Ningning Zhu, and Tzi-cker Chiueh. Low-latency mobile ip handoff for infrastructure-mode wireless lans. *IEEE Journal on Selected Areas in Communications*, 22(4):643–652, 2004.

[94] Kaveh Shafiee, Alireza Attar, and Victor Leung. Optimal distributed vertical handoff strategies in vehicular heterogeneous networks. *IEEE Journal on Selected Areas in Communications*, 29(3):534–544, 2011.

[95] Dong Ma and Maode Ma. A qos oriented vertical handoff scheme for wimax/wlan overlay networks. *IEEE Transactions on Parallel and Distributed Systems*, 23(4):598–606, 2012.

[96] Georgios Aristomenopoulos, Timotheos Kastrinogiannis, Zhaojun Li, and Symeon Papavassiliou. An autonomic qos-centric architecture for integrated heterogeneous wireless networks. *Mobile Networks and Applications*, 16(4):490–504, 2011.

[97] Sami Mumtaz, Valdemar Monteiro, Jose Rodriguez, and Christos Politis. Positioning for performance enhancement in shared lte and hspa networks. In *Proceedings of the International Conference on Telecommunications and Multimedia (TEMU 2012)*, pages 169–174. IEEE, 2012.

[98] Tansir Ahmed, Kyandoghere Kyamakya, Markus Ludwig, KR Anne, J Schroeder, S Galler, K Kyamakya, K Jobmann, Dietmar Jannach, Klaus Leopold, et al. *A context-aware vertical handover decision algorithm for multimode mobile terminals and its performance.* na, 2006.

[99] Taehoon Kim, Sang-wook Han, and Youngnam Han. A qos-aware vertical handoff algorithm based on service history information. *IEEE Communications Letters*, 14(6):527–529, 2010.

[100] Guido R Hiertz, Dee Denteneer, Philips Lothar Stibor, Yunpeng Zang, X-avier Pérez Costa, and Bernhard Walke. The ieee 802.11 universe. *IEEE Communications Magazine*, 48(1):62–70, 2010.

[101] Effect of Transmit Power Changes on AP Cell Sizing. [Online] http://wifinigel.blogspot.jp/2014/11/effect-of-transmit-power-changes-on-ap.html.

[102] Best Practices for High Density Wireless Network Design In Education and Small-/Medium Businesses. [Online] http://www.netgear.com/images/pdf/High Density Best Practices.pdf.

List of Publications

1. **Li Qiang**, Jie Li, and Eitan Altman, “A Novel Distributed Network Selection Scheme for Heterogeneous Wireless Network Environments”, *IEEE Transactions on Control of Network Systems (TCNS)*, accepted.
2. **Li Qiang**, Jie Li, and Corinne Touati, “A User Centered Multi-Objective Handoff Scheme for Hybrid 5G Environments”, *IEEE Transactions on Emerging Topics in Computing (TETC)*, accepted.
3. **Li Qiang**, Jie Li, Yusheng Ji, and Changcheng Huang, “A Novel Software-Defined Networking Approach for Vertical Handoff in Heterogeneous Wireless Networks”, *Wireless Communications and Mobile Computing (WCMC)*, accepted.
4. **Li Qiang**, Jie Li, and Changcheng Huang, “A Software-Defined Network based Vertical Handoff Scheme for Heterogeneous Wireless Networks”, *Proceedings of the IEEE Global Communications Conference (IEEE GLOBECOM 2014)*, pp. 4671-4676, December, 2014.
5. **Li Qiang**, Jie Li, and Mohsen Guizani, and Yusheng Ji, “An Adaptive Route Optimization Scheme for Nested Mobile IPv6 NEMO Environments”, *Proceedings of the 12th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt 2014)*, pp. 373-380, May, 2014.