國立交通大學

網路工程研究所

碩 士 論 文

3GPP 毫微基地台之負載平衡研究

Study of Load Balance in 3GPP Femto-cell Network

研 究 生：李忠炘

指導教授：趙禧綠　教授

中 華 民 國 九 十 九 年 八 月

3GPP 毫微基地台之負載平衡研究

Study of Load Balance in 3GPP Femto-cell Network

研 究 生：李忠炘 Student：Chung-Shin Li

指導教授：趙禧綠 Advisor：Hsi-Lu Chao

國 立 交 通 大 學

網 路 工 程 研 究 所

碩 士 論 文

A Thesis

Submitted to Institute of Network Engineering

College of Computer Science

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Master

in

Computer Science

August 2010

Hsinchu, Taiwan, Republic of China

中華民國九十九年八月

3GPP 毫微基地台之負載平衡研究

學生：李忠炘　　　　教授：趙禧綠

國立交通大學網路工程研究所碩士班

摘　要

不像大型基地台能夠使用很寬的頻帶，毫微基地台能使用的頻帶有所限制，因此可達到的總傳輸速度會比大型基地台還小；另外，毫微基地台有可能與其他裝置共用後端的實體線路頻寬，例如桌上型電腦。因此，相較於大型基地台，毫微基地台可同時服務的行動裝置數量將會有所限制。

大多數負載平衡的方法只考慮單一參數。雖然考慮的參數所影響的效能能夠表現的不錯，但是卻造成其它沒考慮到的效能有較為低落的情形。為了克服此項缺點，我們提出的負載平衡方法考慮了行動裝置的移動、服務品質的要求以及毫微基地台間的負載係數，藉此讓更多的使用者能夠同時進行資料傳輸，並且維持較高的傳輸速度和毫微基地台間的平衡。

假設鄰近毫微基地台間的涵蓋範圍有部分重疊，並且有許多的行動裝置位在重疊範圍內。針對那些行動裝置，首先我們利用訊雜比的大小和變化來估計各個行動裝置停留在涵蓋範圍內的時間係數。接著利用毫微基地台剩餘的資源，我們計算出毫微基地台能提供的頻寬，並藉此得知毫微基地台能夠提供的服務品質係數。再來，我們根據毫微基地台的負載狀況來計算它們之間的平衡係數。我們利用以上三個係數進行行動裝置的挑選並將之換手到其他毫微基地台，以達到負載平衡的目的。而我們也提出了兩種方法，以便用來挑選被換手的行動裝置。最後我們列出模擬結果並且說明之。

關鍵詞：3GPP、毫微基地台、服務品質、負載平衡。

Study of Load Balance in 3GPP Femto-cell Network

Student: Chung-Shin Li Advisor: Hsi-Lu Chao

Institute of Network Engineering College of Computer Science

National Chiao Tung University

Abstract

The usable frequency bandwidth of femto-cells is narrower than macro BSs'. Hence, the total data rate of femto-cells is lower than macro BSs'; besides, the backhaul of femto-cells may be shared by other devices, such as PC. The number of concurrently connecting users is restricted by those reasons mentioned above.

Many load balance methods consider one parameter only. Although the evaluated performance affected by that parameter works well, other performance affected by non-considered parameters may not work well. In order to overcome the weakness mentioned above, we propose a load balance method that considers the movement of user equipment (UE), QoS requirement and the load balance index between femto-cells.

Assume that the coverage area of a femto-cell is partially overlapped with another femto-cell, and there are many UEs within the partially overlapping area. Firstly, we use the magnitude and variation of signal to noise ratio (SNR) to estimate the remaining time index of a UE within the overlapping area. Secondly, we make use of the remaining resource of a femto-cell to calculate the remaining bandwidth, and then we use the acquired remaining bandwidth to calculate the satisfaction index of a UE. Thirdly, we use the loading of those femto-cells to calculate the load balance index. Finally, based on those three indexes mentioned above, we choose a UE to handover for the purpose of load balance, and we propose two strategies to select the UE. We describe the performance evaluation and explain it in chapter 5.

Keyword: 3GPP, femto-cell, QoS, load balancing

Table of Contents

List of Figures

List of Tables

Chapter 1 Introduction

Both 3rd Generation Partnership Project (3GPP) and Worldwide Interoperability for Microwave Access (WiMAX) forum are developing the next generation communication systems that are called Long Term Evolution (LTE) and WiMAX respectively. Both 3GPP and WiMAX forum also propose their own femto-cell system recently. And Femto-cells are named as Home eNode Bs (HeNBs) in 3GPP.

Telecom service providers must keep installing macro base stations (BS) in order to deal with the increasing number of 3G and 4G users. There are many costs if the operator needs to setup a new macro BS, such as BS itself, housing, electricity and backhaul to the core network. Although there are many macro BSs, users still get poor signal quality inside buildings since macro BSs must be placed outside buildings. Adopting femto-cells can reduce costs and improve signal quality inside buildings.

Femto-cells are very small and low cost BSs, but femto-cells still need to provide the same functionalities as macro BSs do. Note that femto-cells are approximately of the same size as the current Wi-Fi access points. Femto-cells are deployed at customers' premises, such as home and office. In addition, femto-cells are configured by the core network automatically, so customers can install femto-cells by themselves without professional knowledge. Femto-cells are powered from the customers' electricity sockets and the customers' internet connections are used as backhaul connections.

For operators, deploying femto-cells can reduce costs such as housing and electricity bill that are necessary for macro BSs. Furthermore, many value-added services will be integrated into femto-cells, and customers that have installed femto-cells are unlikely to change their subscription if they are satisfied with the services. Femto-cells also enhance the indoor coverage. As mentioned above, lack of indoor coverage is the weakness when deploying macro BSs only.

For customers, in addition to enhanced indoor coverage, short communication rage between femto-cells and UEs means that using time of mobile devices can be prolonged. And operators may give discounts to those who installed femto-cells.

Although femto-cells, or named as HeNBs for LTE in 3GPP, have many benefits for operators and users, the development of femto-cells still encounter many challenges. The basic requirement of HeNBs is that users can place HeNBs in anywhere without the aid of operators, and it may raise many problems.

Because that HeNBs could be placed in anywhere, HeNBs need to negotiate with the core network to setup all necessary parameters. After the installation of HeNBs, owners of the HeNBs may want to restrict unknown users from accessing their own HeNBs, so 3GPP defines the Closed Subscriber Group (CSG). Obviously, HeNBs do not have the information of the CSG after installation; HeNBs must negotiate with the core network to acquire the necessary information.

The paging message is needed when a connection request to the mobile station is raised. In the cellular environment, the paging message is sent to a group of macro BSs. The set of those macro BSs are called a paging group. Because the positions of those macro BSs are known by the operator, the design of paging group can be done easily. In the HeNB environment, the locations of HeNBs cannot be known in advance, the setup of paging group must be dynamic. And the paging group must be carefully designed in order to avoid the flooding of paging messages.

The traffic between a HeNB and the core network goes through the internet, so the security of both data and control flow are important. There must be a security tunnel between a HeNB and the core network to protect the messages from eavesdropping and altering.

Different kinds of HeNBs may have different functionalities in the future. Thus, the authentication between a HeNB and the core network is very important. In addition, each operator has their own operation frequency. HeNBs must acquire the frequency parameter from the core network and operate in the pre-defined frequency. Optionally, if a connection is established between two HeNBs, data packets are not necessarily transmitted from the source HeNB to the core network and then from the core network to the destination HeNB. If the data packets can be transmitted directly from one HeNB to another without going through the core network, the loading in core network can be released.

The motivation and objectives of this thesis are described as follows. As mentioned before, the number of concurrently connecting users is restricted by the limited data rate of femto-cells, and there will be many HeNB users in buildings. This means that UEs' connections would be blocked because of the lack of resources. In the mean time, the number of connecting users would be further restricted if HeNBs provide QoS support. If connections of the overloaded HeNB can be transferred to the non-overloaded one by handover, the number of concurrently connecting users can be increased.

We consider QoS requirement and load balance index between HeNBs to achieve our goal. In addition to QoS requirement and load balance between HeNBs, we also consider the movement of UEs. In other words, we would choose the UE to the appropriate HeNB based on the movement and the UE can acquire the minimum QoS requirement at least. Hence, the proposed method can balance the loading between femto-cells while guarantee the minimum QoS requirement.

The reset of this thesis is organized as follows. We give an overview of 3GPP HeNB system in chapter 2. And we introduce some related works about load balance in chapter 3. Then the proposed load balance method is discussed in chapter 4. And we give the performance evaluations in chapter 5.

Chapter 2 3GPP HeNB System

3GPP proposes the Evolved Universal Terrestrial Radio Access Network (E-UTRAN) since release 8[1]. Fig. 1 is the overall E-UTRAN architecture. In 3GPP E-UTRAN, macro BSs are called eNBs (evolved NodeB) and femto-cells are called HeNBs. As depicted in Fig. 1, every eNB connects to each other using the X2 interface, and each eNB connects to the MME/S-GW using the S1 interface. As we can see in Fig. 1, there is no X2 interface between two HeNBs, so control messages between two HeNBs must go through the HeNB GW. And each HeNB connects to the HeNB GW or the MME/S-GW using the S1 interface.

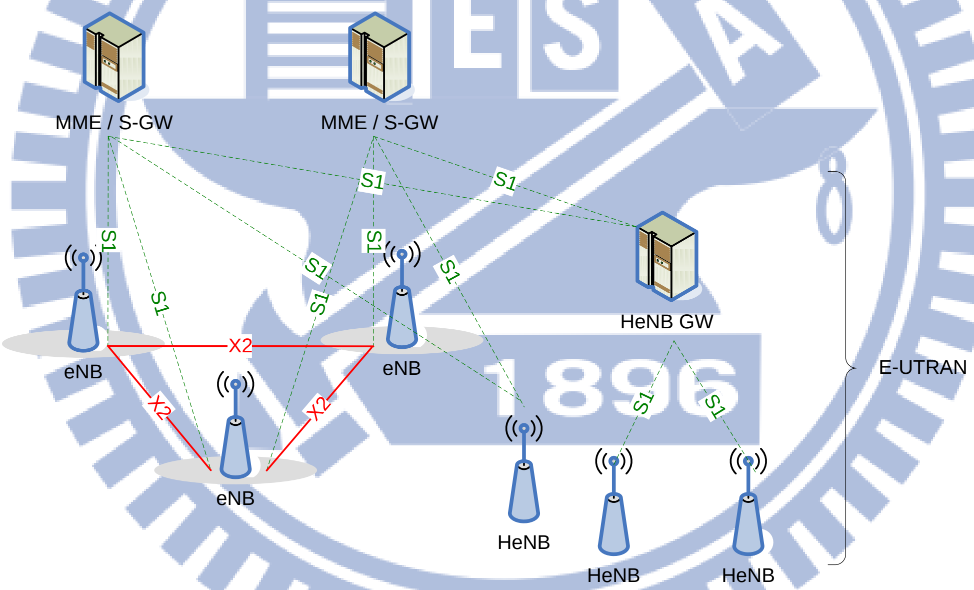

Figure 1 : Overall E-UTRAN architecture with deployed HeNB GW

2.1 HeNB Architecture

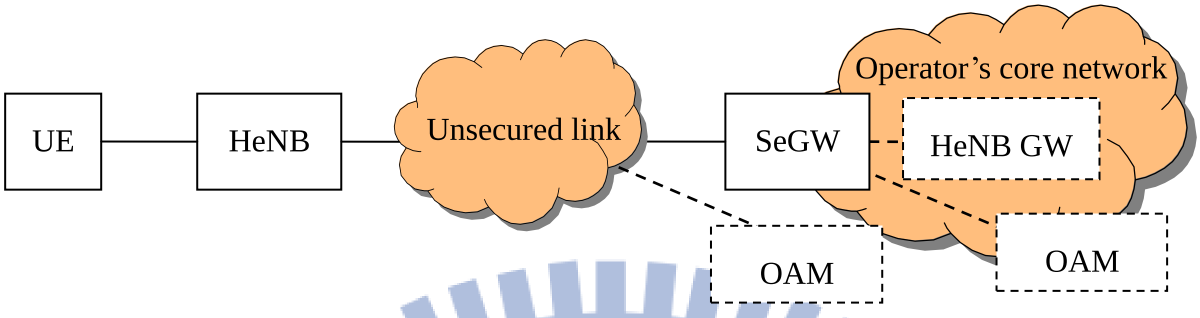

Figure 2 : System architecture of HeNB environment

The system architecture of HeNBs is illustrated in Fig. 2[2]. Security Gateway (SeGW) performs the mutual authentication between the HeNB and the operator's core network. SeGW and HeNB gateway (HeNB GW) are logically separate entities within operator's network, and SeGW may be integrated into HeNB GW. SeGW also plays the role of firewall to separate the core network from the public Internet. The connections between HeNBs and the core network must go through the HeNB GW, and the HeNB GW concentrates connections from numerous HeNBs.

The Operations, Administration and Maintenance (OAM) node is also called the Self-Organized Network (SON) server which provides the auto-configuration and auto-optimization functions for HeNBs. SON server can be used to initiate and verify software updates of HeNBs. Note that SON server uses the TR-069 management protocol to configure HeNBs, and TR-069 is specified by the DSL forum.

As we can see in Fig. 2, data and control traffic between the HeNB and the SeGW is transmitted through the unsecured link. In other words, security tunnel is needed between the HeNB and the SeGW to protect the information transmitted in backhaul link. There are many approaches to protect information, and IPSec may be adopted for integrity and confidentiality protection. In addition, Certificate-based authentication or EAP-AKA-based authentication provides the mutual authentication between the HeNB and the SeGW.

2.2 Frame Structure and Scheduling

2.2.1 Frame Structure

LTE supports both FDD and TDD. Therefore, there are two types of frame structure supported in LTE:

- Type 1: designed for FDD
- Type 2: designed for TDD

Frame structure type 1 is illustrated in Fig. 3. Each radio frame contains 10 sub-frames. The length of each sub-frame is 1ms. And each sub-frame contains two 0.5 ms slots. Because uplink and downlink are separated in different frequency domain, one radio frame is available for both uplink and downlink in a 10 ms interval,

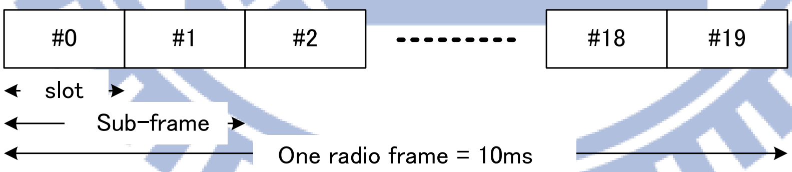

Figure 3 : Frame structure type 1

Frame structure type 2 is illustrated in Fig. 4. Each 10 ms radio frame is divided into 2 half frames. And each half frame contains eight 0.5 ms slots and three special fields: DwPTS, GP and UpPTS. The length of DwPTS and UpPTS is configurable, and the total length of DwPTS, GP and UpPTS equals to 1ms. GP is used for downlink to uplink transition. DwPTS, UpPTS and other slots are used for either uplink or downlink transmission. In TDD mode, there are seven configurations, and according to the configuration, the percentage for both uplink and downlink is configurable. Those configurations are described in Table 1. In table 1, subframes with notation “D” and “U” is used for downlink and uplink respectively, and “S” represents those special fields. Take configuration 0 as an example, subframe 0 and 5 are used for downlink transmission, and subframe 1 and 6 are used for the downlink to uplink transition. Other sub frames are used for uplink transmission.

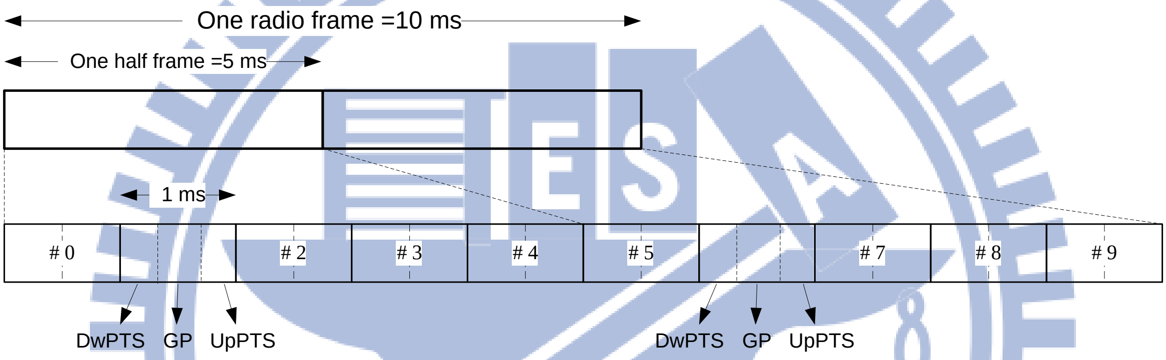

Figure 4 : Frame structure type 2

Table 1: Uplink and downlink allocations for type 2 frame structure

Configuration	**Switch-point periodicity**	**Subframe number**									
		0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
0	5 ms	D	S	U	U	U	D	S	U	U	U
1	5 ms	D	S	U	U	D	D	S	U	U	D
2	5 ms	D	S	U	D	D	D	S	U	D	D
3	10 ms	D	S	U	U	U	D	D	D	D	D
4	10 ms	D	S	U	U	D	D	D	D	D	D
5	10 ms	D	S	U	D	D	D	D	D	D	D
6	5 ms	D	S	U	U	U	D	S	U	U	D

2.2.2 Physical Resource and Scheduling

The physical resource structure of LTE for downlink and uplink are illustrated in Fig. 5 and 6[3]. Each physical resource contains N_{symb}^{UL} (N_{symb}^{DL}) consecutive symbols in time domain, and N_{sc}^{RB} consecutive subcarriers in frequency domain. The allocated frequency bandwidth of the HeNBs determines the total number of physical resources in frequency domain.

Notice that Fig. 5 and 6 only illustrate one slot in the time domain, and one sub-frame contains two slots. When scheduling, the basic assignment unit of physical resources is called Physical Resource Block (PRB). A PRB is composed of one sub-frame in time domain, and N_{sc}^{RB} consecutive subcarriers in frequency domain. In other words, one PRB contains $2\times N_{symb}^{DL}\times N_{sc}^{RB}$ (or $2\times N_{symb}^{UL}\times N_{sc}^{RB}$) resource elements.

The scheduler in HeNB determines how many PRBs should be assigned to each ongoing call at each Transmission Time Interval (TTI), and each TTI is defined as 1ms. Take Fig. 7 as an example, the scheduler executes at every TTI. At the first TTI, the scheduler assigns two PRBs for user A, and one PRB for user B. At the second TTI, the scheduler assigns one PRB for user C、D and E respectively. Note again that PRB is the basic resource unit for scheduling.

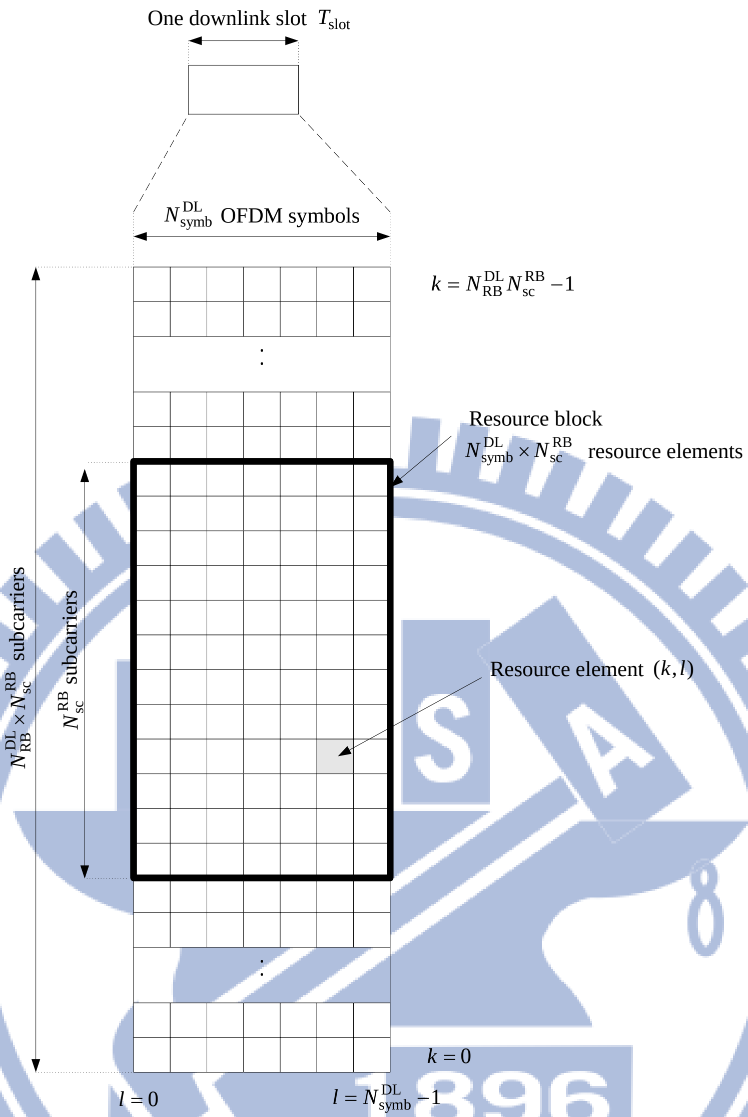

Figure 5 : Physical structure of downlink

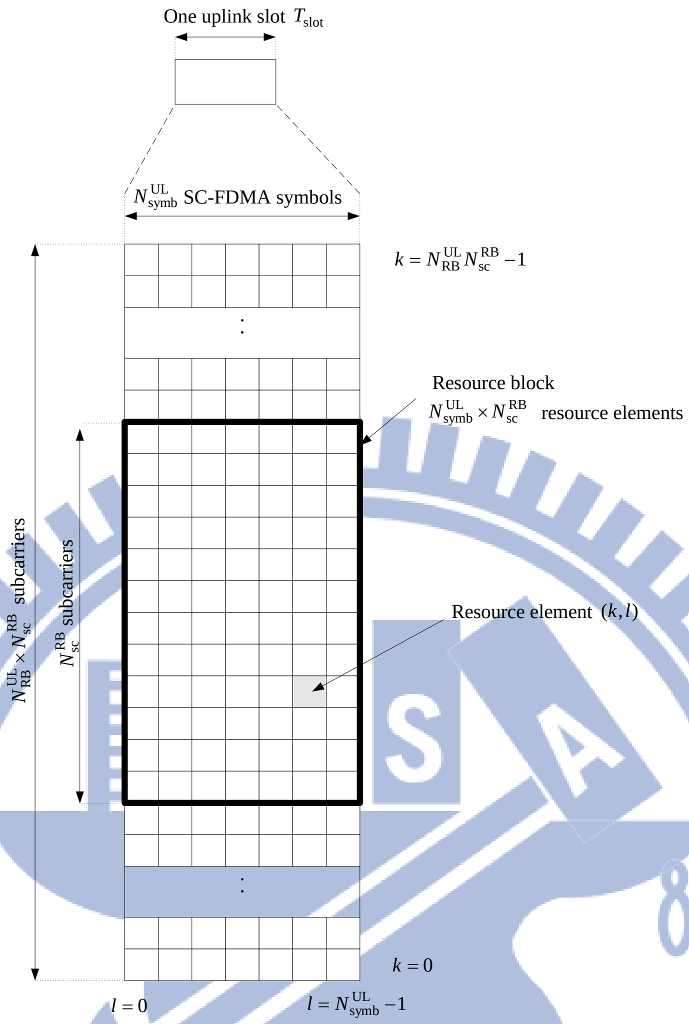

Figure 6 : Physical structure of uplink

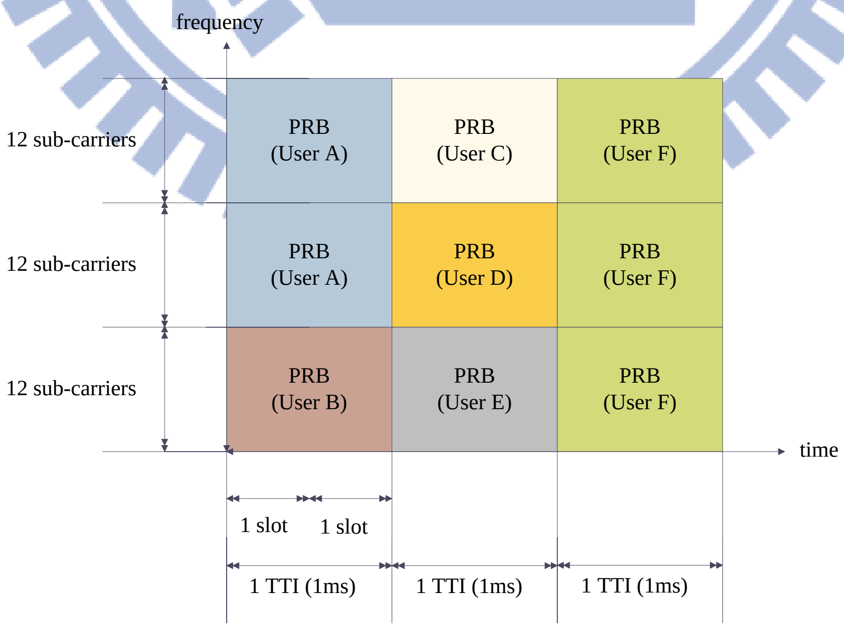

Figure 7 : Example of PRB allocation

Chapter 3 Related Work

Many methods have been proposed for load balance. There are two categories to achieve load balance: power control based and handover based. In the meantime, there are several ways to trigger load balance methods.

3.1 Categories to Achieve Load Balance

One of the categories to achieve load balance is to control the pilot power of BSs. In general, if the BS overloads, it decreases its pilot power in order to prevent incoming connections. The other category to achieve load balance is handover. When the BS overloads, it can force one of its serving UEs to handover to the neighbor BS, and loading of the BS can be released.

3.1.1 Power Control based Category

In [4], authors calculate the relative load factor of each BS initially. The pilot power is reduced if the relative load factor was more than the threshold, and the pilot power is increased if the relative load factor was less than the threshold. However, decreasing the pilot power causes blind spots which are defined as the position where E_c/N_0 of the received pilot signal is lower than a predefined threshold. After the initial adjustment, authors execute the second step based on the blind spot ratio. Authors assume that mobile stations can report the quality of received pilot signal. And authors estimate the blind spot ratio according to the reported quality. In the second step, with the increase in blind spot ratio, the probability to strengthen the pilot signal increases. On the contrary, the lower the ratio, the higher the probability is to reduce the power of pilot signal. Finally, the proposed method adjusts the pilot power again according to the probability acquired in the second step. Load balance through pilot power adjustment can control the number of connecting users efficiently. However, UEs that are in idle mode must receive the pilot signal periodically. Adjusting the power of pilot signal means that some UEs would not be capable of camping on BSs. So it is not practical.

In [5], each local coverage area is treated as a bubble. The air within each bubble can be analogous to the traffic served by each cell. Temporary vacuum is treated as an un-served traffic. Bubbles must oscillate to fulfill the vacuum, and it alters the size of each coverage area. Altering the size of coverage area means to control the power of pilot signal. Treating each coverage area as a bubble is novel idea, but the proposed method needs the angles between UEs and BSs, and it is very difficult to acquire those angles. So the proposed method is difficult to implement.

3.1.2 Handover based Category

Most of the balancing techniques utilize handover to achieve load balance. In [6], both WLAN APs or Universal Mobile Telecommunication Service (UMTS) BSs calculate their utilities and broadcast it. Each mobile station receives the broadcasted utilities. If the utility of the serving AP (BS) is less than the received utility, the mobile station handovers to the network that has a higher utility. By introducing network utility, the proposed method is capable of being used in WLAN/UMTS interworking system. However, simulation scenarios seem to be having great impact on the performance of the proposed method, so it may be not suitable to deploy the proposed method in the real world.

In [7], there is an entity called AP Resource Advertisement Server (ARAS). Firstly, ARAS collects the bandwidth information of each UE. Secondly, ARAS calculates the available bandwidth of each AP based on the collected information, and then ARAS transmits the available bandwidth information to each “AP to AP Side Advertisement

Translator" (ASAT). And ASAT forwards that information to UEs by multicast. Finally, UEs periodically select the best AP by the received available bandwidth information. Obviously, it is clever to transmit the bandwidth information using multicast because it can reduce message traffic. However, there is a drawback if all the UEs receive the same bandwidth information. The AP with maximum available bandwidth would be selected by all the UEs that receive the same message, and then the AP would be congested.

3.2 Trigger Events for Load Balance

Load balance methods can be triggered by several events. One of the trigger events is that cell overloads. Obviously, it's very reasonable to perform load balance method when cell overloads. Another trigger event is that the balance index among cells is less than the predefined threshold, and the balance index is calculated by the loading of those cells. If loading of those cells are the same, the balance index is 1. On the other hand, the more uneven the loading distributes, the lower the index is.

3.2.1 Overload based Event

In [8], authors sum up the load indexes of cells first. The index is defined as Eq. (1).

$$\sum_{i=1}^{k}(\rho_i - \delta)^+ \qquad (1)$$

In Eq. (1), ρ_i is the loading of cell *i*, and δ is the predefined threshold. If the loading of one of those cells exceeds the predefined threshold, load balancing is triggered. And the proposed balance method only stops when the acquired value of Eq. (1) is less than zero. The proposed method does not mention how to select the UE for handover if there are multiple UEs that all of them can make Eq. (1) zero. In reality, users would move around. UE handovered by load balance may be handovered back because of the movement of the UE, so the proposed method seems to be inappropriate if all the UEs are movable, and this is why selecting the UE for handover from multiple UEs is so important.

3.2.2 Balance Index based Event

In [9], the balance index β is calculated first, and it is defined as Eq. (2).

$$\beta = \frac{(\sum_{i=1}^{k} \rho_i)^2}{k \sum_{i=1}^{k} \rho_i^2} \tag{2}$$

If β is less than 1, the proposed method in [9] is executed. The method sorts those cells into three groups: under-loaded, balanced and overloaded. Those under-loaded cells allow new or handover connections. Balanced cells allow new connections only. And over-loaded cells deny any new or handover connections. And the proposed method selects the candidate mobile station for load transfer. The transferred mobile station is handovered from the overloaded cell to the under-loaded or balanced cell. Again, the proposed method does not consider the scenario that UEs are movable. When overloaded, the proposed method may choose the UE that would handover back to the source cell.

Chapter 4 Proposed method

This chapter presents the proposed method. The system architecture is described in chapter 4.1. And details of the proposed method are presented in chapter 4.2. In chapter 4.2, we give an example scenario initially to help reader to understand the concept of the proposed method. In chapter 4.3, we present how to acquire the available resources in the target HeNB. Finally, we depicted the handover procedure in chapter 4.4.

4.1 System Architecture

Fig. 8 presents the system architecture of the proposed method. As depicted in Fig. 8, Radio Resource Control (RRC) layer is used for radio resource control between UEs and HeNBs and S1 interfaces are used to transmit control messages between the SON server and HeNBs. SON server is used to control the behaviors of those HeNBs, and the most important factor deciding whether to perform handover or not is the received signal quality in UE. Hence, UE must constantly transmit the measured signal quality to the serving HeNB by RRC layer in order to assist the handover procedure. In addition to the signal quality of serving HeNB, UEs also transmit the measured signal quality of neighbor HeNBs in the proposed system architecture. Each HeNB sends those signal quality information to the SON server through the S1 interface in order to assist the SON server to carry out the load balance procedure.

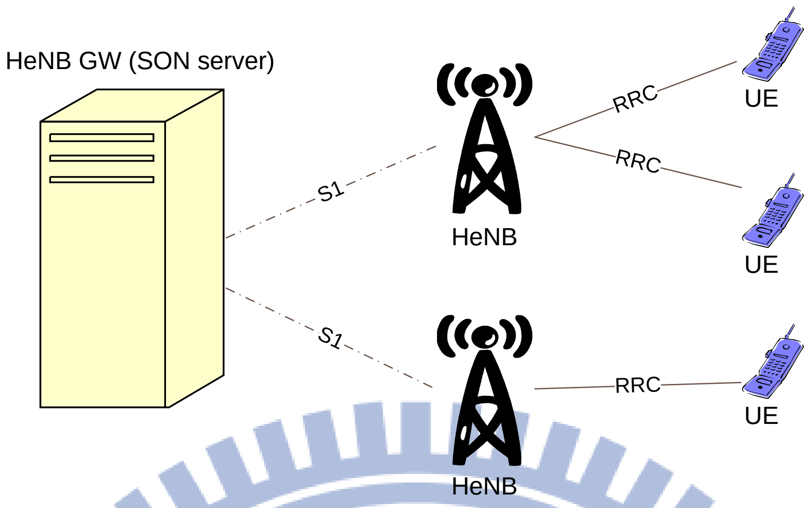

Figure 8 : System architecture of the proposed method

4.2 Proposed Load Balance Method

This chapter describes the details of the proposed method. We assume that there are many HeNBs and UEs in the environment, and there is a SON server located on the HeNB GW that is responsible to control all the HeNBs.

Take Fig. 9 as an example, if there are too many UEs that connect to the same HeNB, the HeNB would be overloaded. And if we handover one of the UEs, that is located on the overlapping area between two HeNB, to the non-overloading HeNB, the loading of the overloaded one can be released, and it is capable of accepting new calls.

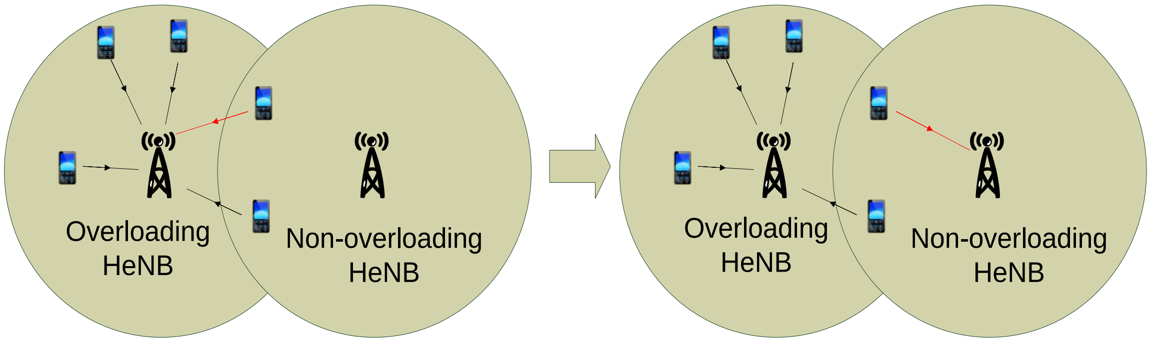

Figure 9 : Example scenario

One of the important factors that we must consider is the timing to perform the proposed load balance method. Obviously, if the loading of a HeNB is close to the critical level, those following new connections will be blocked because of the lack of resources. So the loading of HeNBs is used to decide whether to perform the load balance procedure or not. Because PRB is the basic transmission unit in LTE physical layer, it is a perfect unit to define the critical loading. Assume that a HeNB has K usable PRBs in the frequency domain. If the unused PRBs in the HeNB are less than 20% of K, then the proposed method will be executed in the SON server.

Fig. 10 is the flowchart of the proposed method. Firstly, we try to find an overloaded HeNB. Secondly, if one of the HeNBs within the system overloads, we would execute the matrix calculation which is described in chapter 4.2.1. Thirdly, based on the V matrixes obtained in chapter 4.2.1, we can acquire the target HeNB and the UE that is going to be handovered. Finally, we handover the UE to the target HeNB.

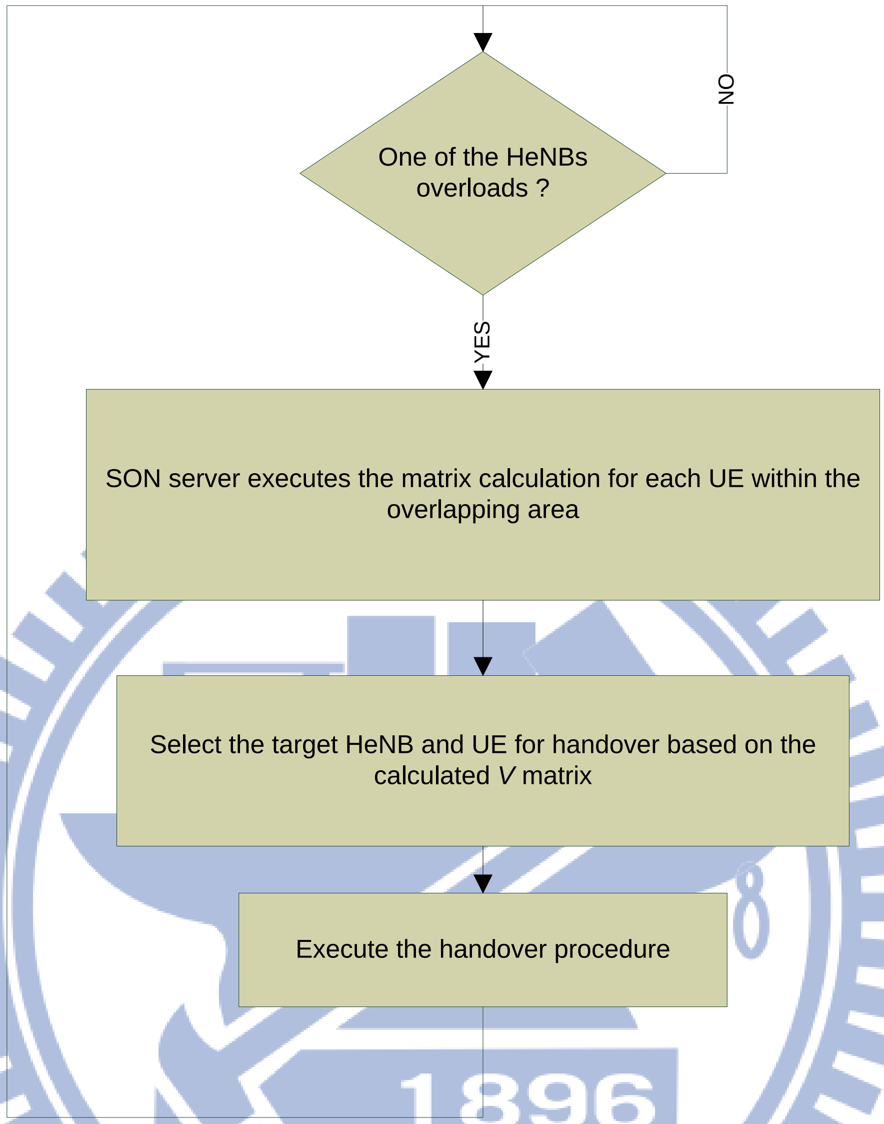

Figure 10 : Flow chart of the proposed method

Following statements define all the variables that are used in the proposed method.

- N^u is the set of HeNBs that includes UE_u's serving HeNB and neighbor HeNBs. Note that UE_u's serving HeNB is the overloaded one.
- U is the set of UEs in the overlapping area, and all the UEs in U are connecting to the overloaded HeNB.
- P^u is the parameter matrix with three columns to be calculated and N^u rows for each HeNB, and $u \in U$.
- W^u is the weight matrix with three rows of weight and N^u columns for each neighbor HeNB, and $u \in U$.
- V^u is the value matrix, where $u \in U$. And there are $N^u * N^u$ elements in V^u. The

diagonal elements within V^u are regarded as the calculated value for each HeNB. In other words, $V^u_{n,n}$ is used to represent the calculated value for UE_u and $HeNB_n$. And $V^u_{i,j}$ is not used in the proposed method, where $i \notin j$.

- B_u is the maximum bandwidth requirement of UE_u.
- The minimum bandwidth requirement of UE_u is regarded as b_u.
- Supportable bandwidth in target HeNB is regarded as ρ_n.
- Supportable bandwidth in serving HeNB is regarded as ρ_s.
- K_i is the number of used PRBs in $HeNB_i$ before the load balance method is executed.
- K^u_{ser} is the number of PRBs that UE_u occupies in the serving HeNB.
- K^u_n is the number of occupied PRBs if UE_u handover to $HeNB_n$.

4.2.1 Matrix Calculation

For each UE_u, the SON server calculates Eq. (3) in the overloaded HeNB, where $u \in U$.

$$\left[P^u\right]_{N^u*3} \cdot \left[W^u\right]_{3*N^u} = \left[V^u\right]_{N^u*N^u} \tag{3}$$

There are three columns in the P^u matrix, and those three columns in the n^{th} row are normalization values of indexes for $HeNB_n$, where $n \in N^u$. The first column represents the *UE remaining time index*. Each moving UE has different dwell time in different HeNBs, so we use this column to represent the relatively remaining time in each HeNB for the moving UE_u. The n^{th} element in this column is the relatively remaining time index for UE_u staying in $HeNB_n$. With the help of *UE remaining time index*, we know the HeNB on which UE_u has the longest sojourn time.

The second column contains the normalized values of the *UE satisfaction indexes*, and the n^{th} element in the second column is the normalized value of *UE satisfaction index* for UE_u if UE_u was handovered to $HeNB_n$. *UE satisfaction index* helps us to find the HeNB on which UE_u can acquire the maximum satisfaction.

The third column contains the normalized values of the *HeNB load balance indexes*, and the n^{th} element in the third column is the normalized value of *load balance index* if UE_u was handovered to $HeNB_n$. We can estimate the balance index among those HeNBs after handover, and it helps us to choose the target HeNB with a good balance index.

W^u matrix contains the weight of each element in the P^u matrix. And V^u matrix contains all the calculated values for each $HeNB_n$, $n \in N^u$. In simple terms, if $HeNB_n$ has a relatively large value in the V^u matrix, then we can acquire better system performance if UE_u connects to $HeNB_n$.

We give an example in Fig. 11, and there are three HeNBs. If a HeNB overloads, each UE_u has one P^u, W^u and V^u matrix in the SON server. Take the first row as an example, $P^1_{1,1}$, $P^1_{1,2}$ and $P^1_{1,3}$ represent three normalized values for $HeNB_1$ and UE_1. After the calculation, we can acquire $V^1_{1,1}$, $V^1_{2,2}$ and $V^1_{3,3}$ in the V^1 matrix and those values are used for the handover UE selection in chapter 4.2.2.

$$\begin{pmatrix} P^1_{1,1} & P^1_{1,2} & P^1_{1,3} \\ P^1_{2,1} & P^1_{2,2} & P^1_{2,3} \\ P^1_{3,1} & P^1_{3,2} & P^1_{3,3} \end{pmatrix} \begin{pmatrix} W^1_{1,1} & W^1_{1,2} & W^1_{1,3} \\ W^1_{2,1} & W^1_{2,2} & W^1_{2,3} \\ W^1_{3,1} & W^1_{3,2} & W^1_{3,3} \end{pmatrix} = \begin{pmatrix} V^1_{1,1} & V^1_{1,2} & V^1_{1,3} \\ V^1_{2,1} & V^1_{2,2} & V^1_{2,3} \\ V^1_{3,1} & V^1_{3,2} & V^1_{3,3} \end{pmatrix}$$

Figure 11: Example of the matrix calculation

Following paragraphs describe how to calculate those elements in the P^u matrix.

A. **UE remaining time element ($P^u_{n,1}$)**

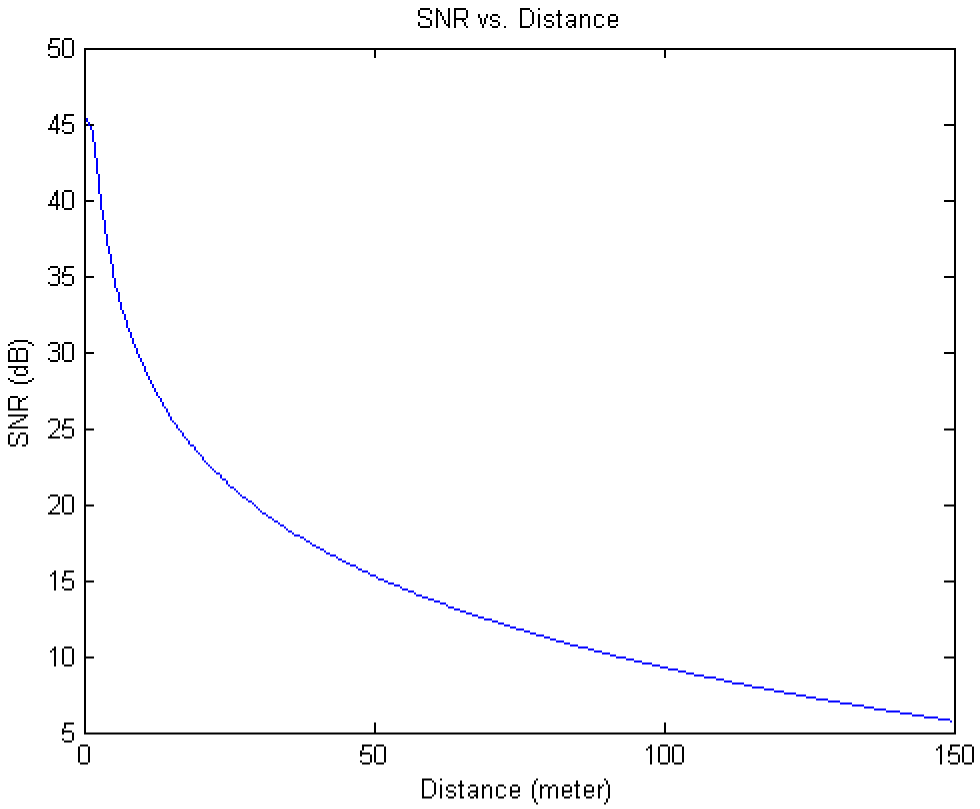

Figure 12: SNR vs. Distance

$$Time = \frac{Distance}{Speed} \tag{4}$$

Firstly, we use the two-ray ground reflection model[10] to construct the mapping between SNR and distance, and it is depicted in Fig. 12. Note that the relative speed between the UE and each neighbor HeNB is the same. And according to Eq. (4), we know that the ratio of time is the ratio of distance.

Assume that the distance between UE_u and the coverage boundary of each neighbor $HeNB_n$ is RD^u_n (Remaining Distance), $n \in N^u$. Derived from Eq. (4), we calculate $P^u_{n,1}$ in

Eq. (5), and $P^u_{n,1}$ is one of the elements in P^u.

$$P^u_{n,1} = \frac{RD^u_n}{\sum_{i \in N^u} RD^u_i} \qquad (5)$$

UEs continue to monitor the SNR values of neighbor HeNBs. If the UE is approaching a HeNB, then the monitored SNR values are going larger; else the monitored SNR values are going smaller. If the UE does not move or circle around the HeNB with a constant radius, then the monitored SNR values are the same. Assume that the coverage radius of each HeNB is R and the distance between UE$_u$ and HeNB$_n$ is r^u_n, and then RD^u_n is calculated as Eq. (6). Note that r^u_n is acquired using the received SNR and two-ray ground reflection model. The longer the UE stays in HeNB$_n$, the larger the RD^u_n is. That is, if $P^u_{i,1}$ is larger than $P^u_{j,1}$, the time that UE$_u$ stays in HeNB$_i$ is longer than the time that UE$_u$ stays in HeNB$_j$. The example of Eq. (6) is depicted in Fig. 13 and 14.

$$\begin{cases} RD^u_n = r^u_n & \text{, if UE}_u \text{ does not move or circle around the HeNB} \\ RD^u_n = R - r^u_n & \text{, if UE}_u \text{ is leaving the HeNB}_n \\ RD^u_n = (R^2 + (r^u_n)^2)^{\frac{1}{2}} & \text{, if UE}_u \text{ is approaching the HeNB}_n \end{cases} \qquad (6)$$

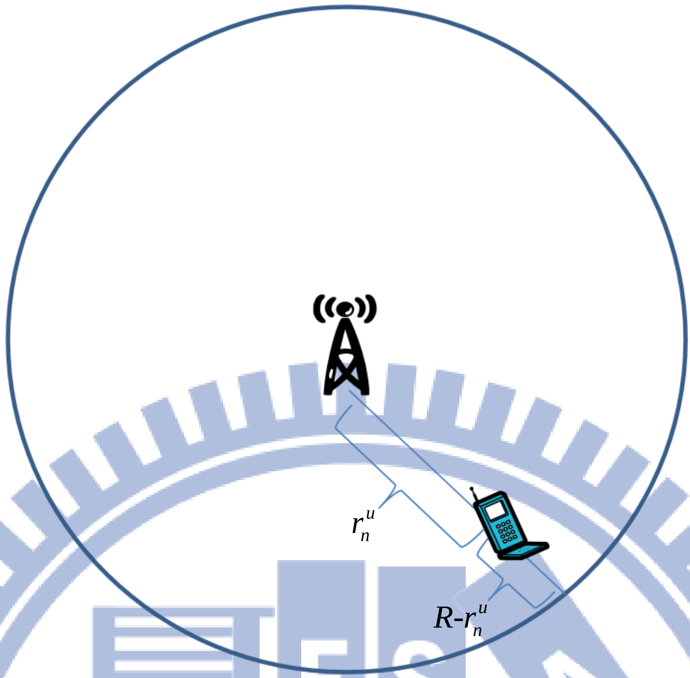

Figure 13: Example of Eq. (6) if UE is leaving the HeNB$_i$

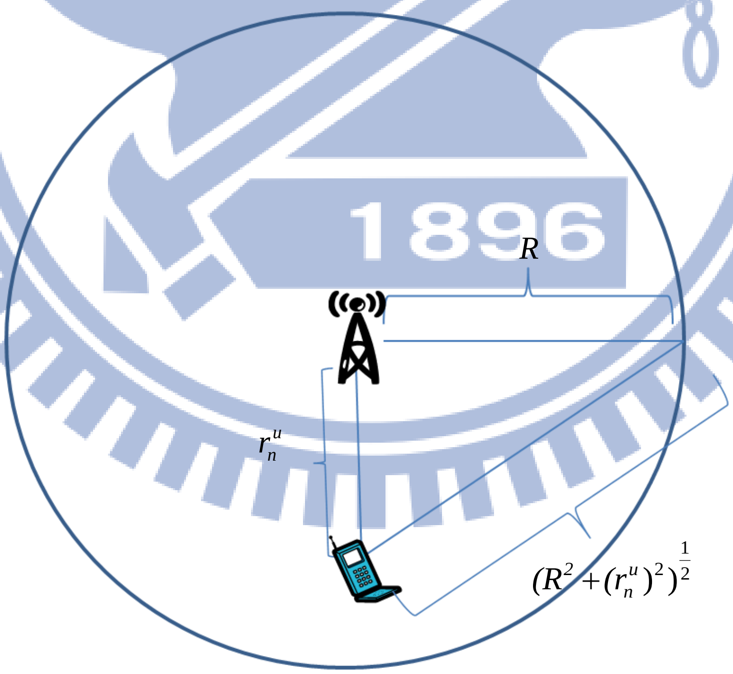

Figure 14 Example of Eq. (6) if UE is approaching the HeNB$_i$

B. **UE satisfaction element ($P^u_{n,2}$)**

The UE in the overlapping area may have multiple neighbor HeNBs, and not all the neighbor HeNBs can satisfy the QoS requirement of the UE, so UE satisfaction must be taken into account.

[11] shows a way to calculate the satisfaction index of the UE, and we propose the modified formula as Eq. (8). Then we normalize satisfaction indexes into elements in P^u, and the normalization formula is presented in Eq. (7).

$$P^u_{n,2} = \frac{S^u_n}{\sum_{i \in N^u} S^u_i} \tag{7}$$

$$S^u_n = \frac{1}{1+e^{-\alpha(K*G-\beta)}} \tag{8}$$

Where,

$$K = \frac{\rho_n - b_u}{B_u - b_u} \tag{9}$$

$$G = \frac{\rho_n}{\rho_s} \tag{10}$$

B_u is the maximum bandwidth requirement of the UE and b_u is the minimum bandwidth requirement of the UE. ρ_n is the acquired bandwidth in target HeNB, and ρ_s is the acquired bandwidth in source HeNB. Hence, G is the bandwidth gain between the source HeNB and the target HeNB. And α and β are constants.

If we are not aware of ρ_n, the calculation of Eq. (8) is impossible. So we use the remaining PRBs in $HeNB_n$ and the SNR between $HeNB_n$ and UE_u to calculate ρ_n. And the calculation steps are explained in chapter 4.4.1.

C. **HeNB load balance element ($P^u_{n,3}$)**

One of the goals in the proposed method is to balance the loading of HeNBs. [9] proposes a formula to calculate the load balance index of Wi-Fi APs. And we modify it as Eq. (12). Then we normalize balance indexes into elements in P^u, and the normalization formula is presented in Eq. (11).

$$P^u_{n,3} = \frac{L^u_n}{\sum_{i \in N^u} L^u_i} \tag{11}$$

$$L^u_n = \frac{(\sum_{i \in N^u} O^u_{i(n)})^2}{num(N^u)\sum_{i \in N^u}(O^u_{i(n)})^2} \tag{12}$$

Assume that L^u_n is calculated for UE$_u$ and HeNB$_n$. $O^u_{i(n)}$ is calculated as Eq. (13), and $num(N^u)$ means the number of elements within the set N^u.

$$\begin{cases} O^u_{i(n)} = K_i + K^u_n & \text{, if } i == n \\ O^u_{i(n)} = K_i - K^u_{ser} & \text{, if HeNB}_i \text{ is the serving HeNB} \\ O^u_{i(n)} = K_i & \text{, others} \end{cases} \tag{13}$$

K_i is the number of occupied PRBs in HeNB$_i$ before handover, and K^u_n is the number of occupied PRBs if UE$_u$ handover to HeNB$_n$. K^u_{ser} is the number of PRBs that UE$_u$ occupied in the serving HeNB. How do we calculate K^u_n is explained in chapter 4.4.2.

4.2.2 Handover UE Selection

After the matrix calculation, we can acquire the V^u matrix. Assume that $V^u_{s,s}$ is the calculated value for UE_u and serving (overloading) HeNB, and $V^u_{n,n}$ is the calculated value for UE_u and $HeNB_n$. The handover UE can be chosen according to the following two strategies.

A. Maximum difference

According to Eq. (14), max_N(u) is the index of neighbor HeNB, and $V^u_{\max\_N(u),\ \max\_N(u)}$ is the maximum value in V^u.

According to Eq. (15), we know that $(V^{uM}_{\max\_N(uM),\max\_N(uM)} - V^{uM}_{s,s}) > (V^{u}_{\max\_N(u),\max\_N(u)} - V^{u}_{s,s})$, where u, $uM \in U$, $uM \neq u$. And we handover UE_{uM} to $HeNB_{\max\_N(uM)}$. After handover, UE_{uM} has the maximum increment of value V among all the UEs in the overlapping area, and we can acquire a better enhancement of the system performance.

$$\max\_N(u) = arg\ max_{i \in N^u} (V^u_{i,i}) \tag{14}$$

$$uM = arg\ max_{u \in U} (V^u_{\max\_N(u),\ \max\_N(u)} - V^u_{s,s}) \tag{15}$$

B. Minimum V

Using Eq. (16), we know that UE_{um} has the minimum V value among all the UEs in the overlapping area. We choose UE_{um} for handover since it has the minimum V value, and it is handovered to $HeNB_{\max\_N(um)}$.To handover UE_{um} can decrease the loading of the overloaded HeNB, and UE_{um} is the most unsuitable UE staying in the overloaded HeNB among all the UEs.

$$um = arg\ min_{u \in U} V^u_{s,s} \tag{16}$$

4.3 Mapping of Bandwidth and PRB

3GPP gives the table of bandwidth using Modulation and Coding Scheme (MCS) and the number of PRBs[12]. 3GPP also gives the limitation that Block Error Rate (BLER) must be less than 10%. Assume that the SNR is constant, then we know that higher MCS index causes higher BLER[13]. In other words, the MCS index has an upper bound for a particular SNR. The SNR between HeNBs and UEs can be acquired by control signal, and then the suitable MCS indexes are known. Given the SNR, we choose the maximum MCS index that satisfies the BLER constrain. Take Fig. 15 as an example, if the monitored SNR is 5, then the maximum MCS index that satisfies the BLER constrain is 11. Likewise, if the monitored SNR is 10, then the maximum MCS index that satisfies the BLER constrain is 17.

From [12], the MCS indexes can be mapped to Transport Block Size (TBS) indexes. Table 2 is the snapshot of the mapping between TBS indexes and MCS indexes. Assume that the TBS index is acquired from the MCS index. Using the table in [12], the supportable bandwidth can be calculated by the number of unused PRBs. Likewise, SON server can obtain the number of required PRBs for a UE by the bandwidth requirement. And it is described in the following two sections.

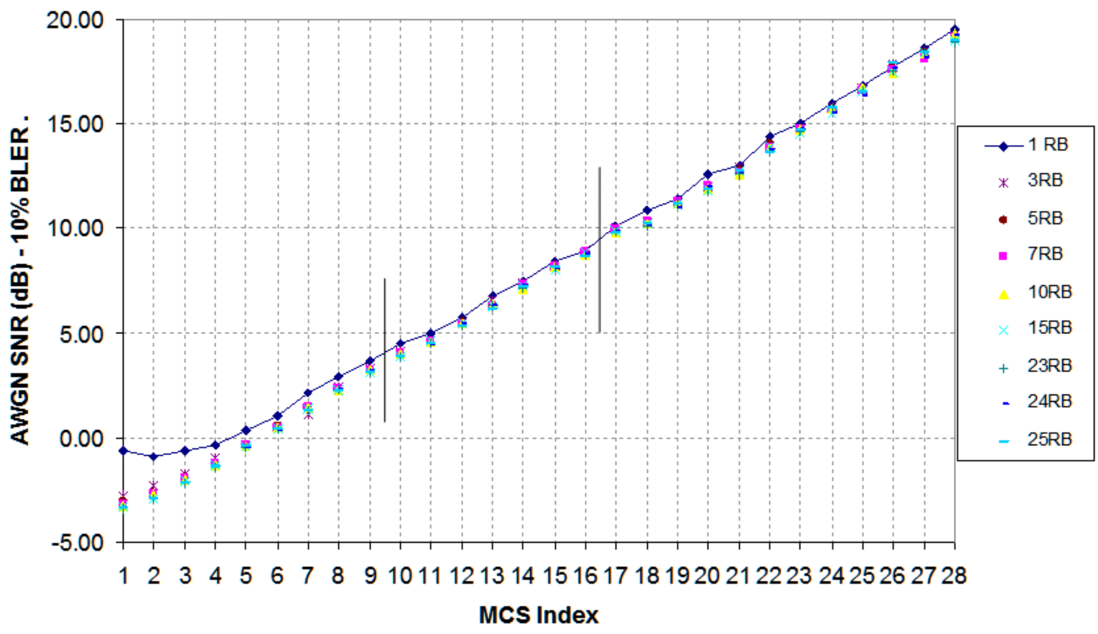

Figure 15: AWGN SNR vs. MCS index for 25RB MCS/TBS table

Table 2: Snapshot of the mapping between MCS indexes and TBS indexes

MCS Index I_{MCS}	Modulation Order Q_m	TBS Index I_{TBS}
0	2	0
1	2	1
2	2	2
3	2	3
4	2	4
5	2	5
6	2	6

4.3.1 Available Bandwidth in Target HeNB

Given the TBS index and number of unused PRBs, we can acquire the supportable bandwidth by table lookup. The table in chapter 7.1.7.2.1 of [12] shows the mapping of TBS indexes, number of PRBs and bandwidth. Using the TBS index and the number of unused PRBs, supportable bandwidth in the HeNB is then acquired. Note that the bandwidth in that table is regarded as bits per microsecond. Take Table 3 as an example, if the TBS index is 3 and there are 5 unused PRBs in the HeNB, then the supportable in target HeNB is 256 bits per microsecond.

Table 3: Snapshot of the mapping among TBS indexes, number of PRBs and bandwidth

I_{TBS}	N_{PRB}									
	1	2	3	4	5	6	7	8	9	10
0	16	32	56	88	120	152	176	208	224	256
1	24	56	88	144	176	208	224	256	328	344
2	32	72	144	176	208	256	296	328	376	424
3	40	104	176	208	256	328	392	440	504	568
4	56	120	208	256	328	408	488	552	632	696
5	72	144	224	328	424	504	600	680	776	872
6	328	176	256	392	504	600	712	808	936	1032

4.3.2 Required Number of PRBs in Target HeNB

Given the TBS index and bandwidth request, we can acquire the number of required PRBs by table lookup. The table in chapter 7.1.7.2.1 of [12] shows the mapping of TBS indexes, number of PRBs, and bandwidth. Using TBS index and the requested bandwidth, the number of required PRBs can be acquired in that table. Take Table 3 as an example, if the TBS index is 4 and the requested bandwidth is 400 bits per microsecond, then the number of required PRBs is 6.

4.4 Handover Procedure

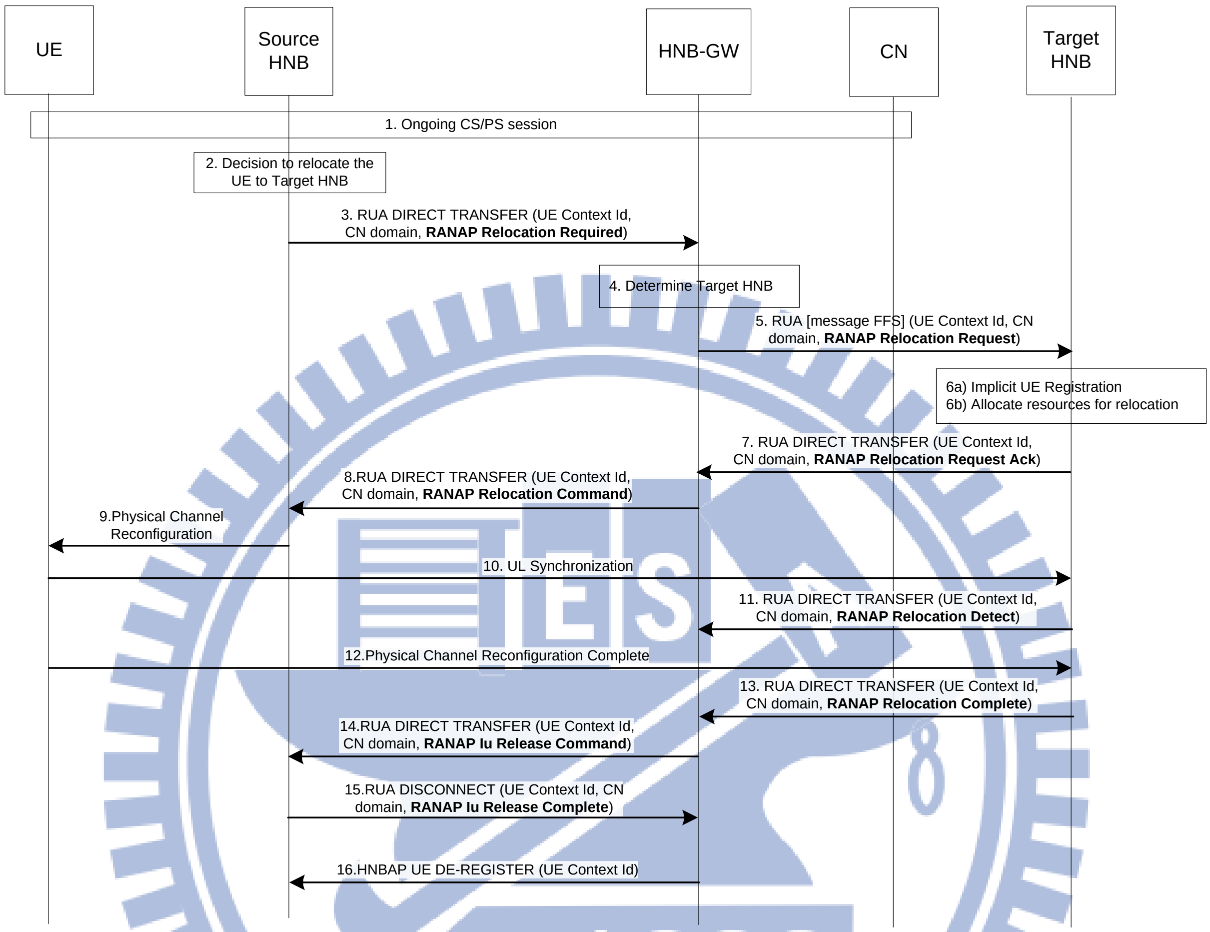

Figure 16: Handover procedure proposed by 3GPP

Fig. 16[14] is the handover procedure proposed by 3GPP.

1. UE has an ongoing session to the core network via source HeNB and HeNB GW.
2. Source HeNB decides to relocate the UE to target HeNB,
3. The source HeNB sends the RANAP *relocation Required*[15] message encapsulated in the RUA *Direct transfer* message to the HeNB GW. The core network domain and UE-id are also included in the RUA *Direct transfer* message.
4. The HeNB GW, where the SON server is located on, determines the target HeNB.
5. HeNB GW sends RANAP *Relocation Request* message encapsulated in the RUA

message to target HeNB to ask for the handover preparation.

6. Target HeNB registers the UE implicitly using the UE-id and core network domain. The target HeNB also prepares appropriate resources for the relocation.
7. The target HeNB sends RANAP *Relocation Request Ack* message encapsulated in RUA *Direct transfer* message back to HeNB GW to inform the completion of the preparation.
8. HeNB GW sends RANAP *Relocation Command* message encapsulated in RUA *Direct Transfer* message to source HeNB, commanding the source HeNB to relocate the UE to the target HeNB.
9. The source HeNB commands the UE to reconfigure physical channel.
10. The UE synchronize its UL channel with target HeNB.
11. The target HeNB detects the synchronization of the UE, and sends RAN *Relocation Detect* message encapsulated in RUA *Direct Transfer* message to HeNB GW.

12-13. After the completion of physical channel reconfiguration, the target HeNB sends the RANAP *Relocation Complete* message to the HeNB GW, meaning that the UE is now attaching to the target HeNB instead of source HeNB.

14. After the relocation is completed, the HeNB GW sends RANAP *lu Release Command* message to inform the source HeNB to release the *lu* channel between the UE and itself.
15. After the release is completed, the source HeNB sends the RANAP *lu Release Complete* message encapsulate in RUA *Disconnect* message to the HeNB GW.
16. Finally, the HeNB GW sends the HNBAP *UE DE-REGISTER* message to source HeNB for the purpose of deregistration of the UE.

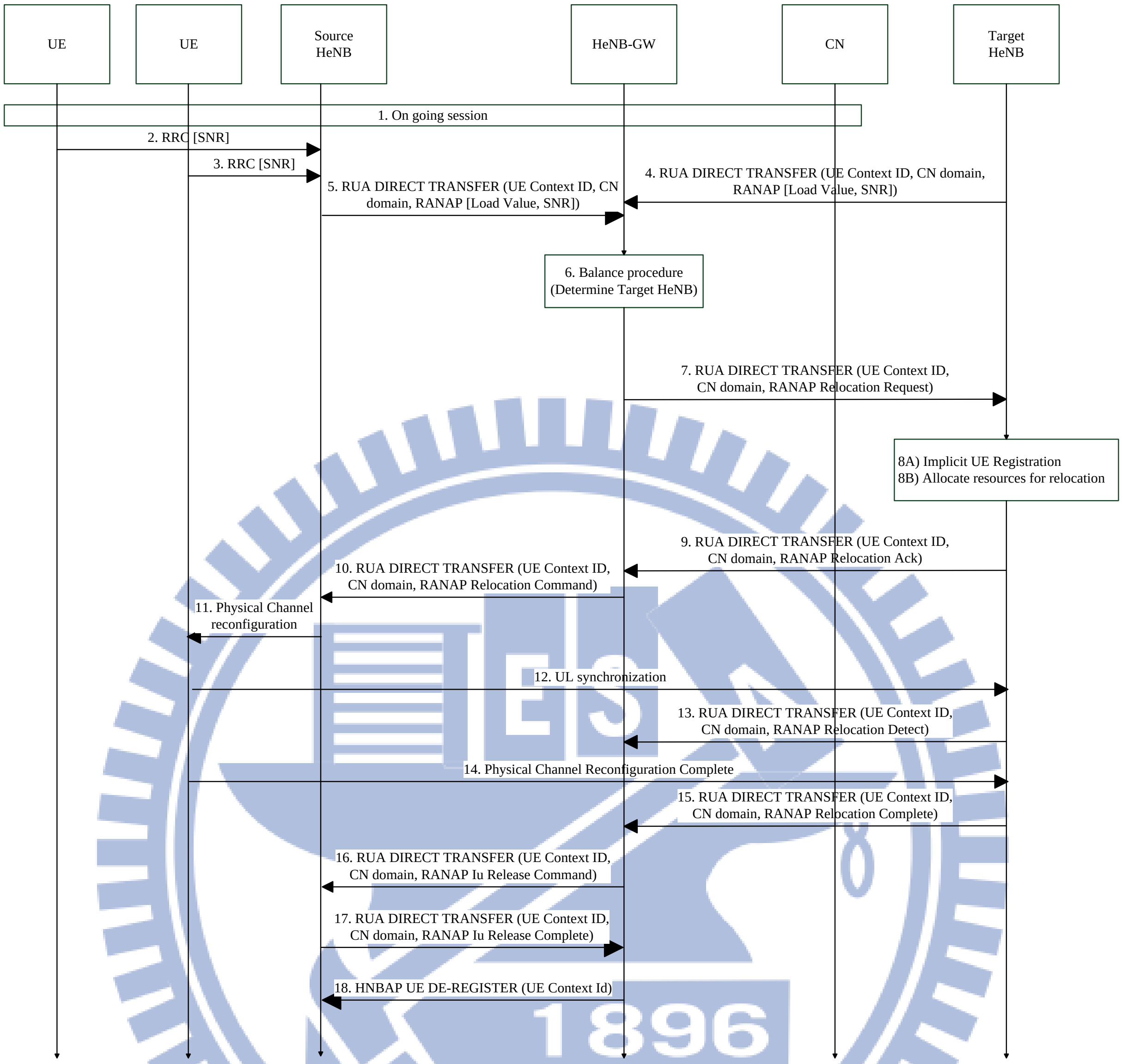

Figure 17: Proposed handover procedure

Fig. 17 is the proposed handover procedure which is modified from Fig 16. Because the original handover procedure cannot support the proposed load balance method, the modification of original procedure is needed.

1. Each UE has an ongoing session to the core network via source HeNB and HeNB GW.

2-3. Each UE constantly transmits the measured signal quality through the RRC (Radio

Resource Control)[16] layer of LTE.

4-5. Each HeNB sends the *Load value* of itself and the collected *SNR* encapsulated in RUA *Direct Transfer* to the HeNB GW, where the SON server is located on. Note that the transmitted *Load value* and *SNR* is used in the proposed method.

6. SON server that is located on HeNB GW performs the balance procedure based on those received *SNR* and *HeNB load value*. Once the balance procedure determines the source HeNB, target HeNB and the UE that is to be handovered, HeNB GW would transmit *RANAP relocation Request* to target HeNB.

7-18. These steps are similar to Fig. 16.

The major difference between the original handover procedure and the modified handover procedure is that: in the original handover procedure, the source HeNB initiates the handover, and it is caused by UE movement most of the time; in the modified handover procedure, the handover procedure is initiated by SON server and it is caused by the lack of resources in the source HeNB. The purpose of these two handover procedures is different, and that is the reason why the original handover procedure needs to be modified.

Chapter 5 Performance Evaluation

In this chapter, the performance of our load balance method and another load balance method is compared. And the results are described and discussed later.

5.1 Simulation environments

The simulation topology is depicted in Fig. 18 and simulation parameters are listed in Table 4. The size of simulation map is 1000*1000 m^2, and there are 13 HeNBs in the map. As depicted in Fig. 18, each circle represents the coverage area of a HeNB, and all the HeNBs are not movable. UEs are placed randomly in the simulation map and all the UEs' speed are distributed in normal distribution.

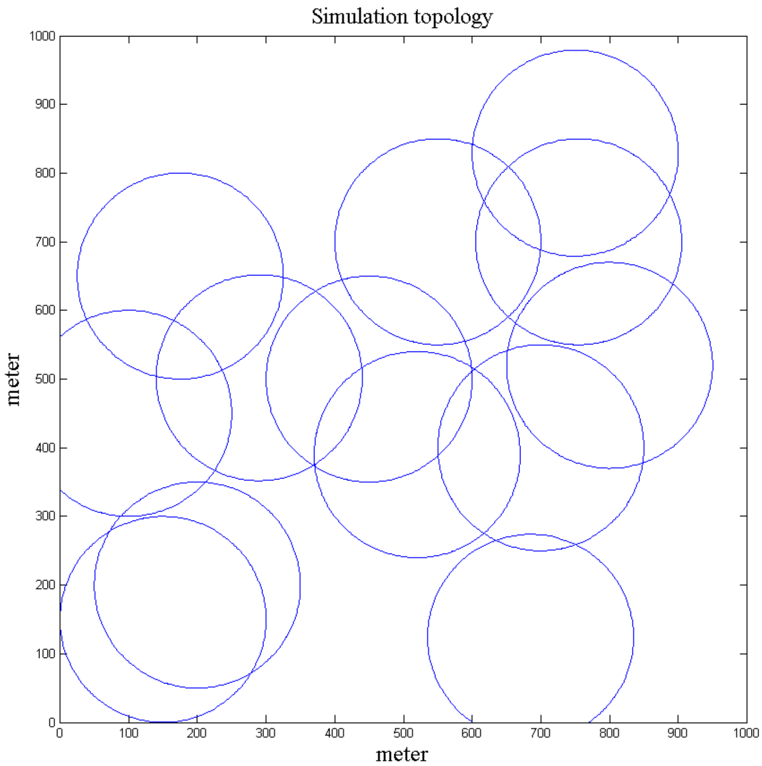

Figure 18: Simulation topology

Table 4: Simulation parameters

Number of PRBs per HeNB	50 (10MHz)
Number of HeNBs	13
UE speed	0 ~ 1 m/s
Transmission range	150m
Service distribution	VoIP(50%), Data(50%)
Call life time	Exponential distribution with average 30 seconds for VoIP user Exponential distribution with average 10 seconds for data user
$W_{i,i}^{u}$ in the W^{u} matrix	1/3
Maximum bandwidth requirement for data user	3Mbps
Minimum bandwidth requirement for data user	1.5Mbps
VoIP bandwidth requirements	64kbps, 32kbps, 24kbps, 16kbps, 8kbps
Simulation topology	1000*1000 m^2
α	30
β	0.3
Critical threshold	80%
Simulation time	100 seconds

5.2 Compared load balance method

The proposed load balance method is compared with the method that is presented in [8]. Firstly, the compared method defines the load of a BS in Eq. (17). And those parameters in Eq. (17) are listed below.

- $R_{mc}(i)$ is the modulation and coding rate (bits/symbol).
- $GR(i)$ is the guaranteed bit rate.
- $SF(i)$ is the spread factor (chips/symbol) of CDMA, UMTS or IEEE 802.11b. If SF does not exist, then we set $SF(i)=1$.
- R_s is the number of data symbols that BS can transmit in one second, i.e. R_s is data symbol rate.
- R_C (chip/second) is defined as Eq. (18). And R_C is considered as the total resources.

$$\rho = \frac{1}{R_c}\sum_{i=1}^{M}\frac{GR(i)SF(i)}{R_{mc}(i)} \tag{17}$$

$$R_c = R_s * SF \tag{18}$$

Secondly, the compared method proposes a load index in Eq. (19), where δ is the load threshold and a^+ is defined in Eq. (20). Obviously, if ξ_2 is equal to 0, then the loading of all the BSs are less than the load threshold.

$$\xi_2 = \sum_{i=1}^{K}(\rho_i - \delta)^+ \tag{19}$$

$$a^+ = \max(a, 0) \tag{20}$$

The compared method proposes Eq. (21) to identify the pair (i, j) for load balance, where i is the index of UE and j is the index of BS. And w_{ij} is the load contribution of UE_i to

BS_j. If the loading of one of the BSs is larger than load threshold δ, then ξ_2 would be larger than zero. And the compared method will try to identify the pair (i, j) that make ξ_2 zero. Finally the load balance is executed by handovering UE_i to BS_j

$$\xi_2(i,j) = (\rho_0 - w_{i0} - \delta)^+ + (\rho_j + w_{ij} - \delta)^+ + \sum_{l \neq \{0,j\}} (\rho_l - \delta)^+ \tag{21}$$

5.3 Simulation results

The performance is evaluated by success probability, balance index, throughput, number of affected HeNBs per balance and one simple scenario. Note that the number of affected HeNBs per balance means that when an UE is handovered to the target HeNB, target HeNB may be overloaded because of the load contribution of that UE. If target HeNB is overloaded because of the load contribution of the handover UE, then the target HeNB is regarded as an affected HeNB.

In those figures of results, the notation "*maximum difference strategy*" means that we use the *maximum difference* selection strategy in the simulation, and the notation "*Minimum V strategy*" means that we use the *minimum V* selection strategy in the simulation. The notation "*NALB*" means that we use the compared method in the simulation and "*No load balance*" means that there is no load balance method during the simulation.

Success probability is compared in Fig. 19. X axis represents the number of UEs and Y axis represents the call success probability. Success probability is defined by Eq. (22). In Fig. 19, when the number of UEs is less than 50, all the success probabilities are higher than 0.96. And it is obviously that with the increase in the number of UEs, the success probability decreases. We can see that the success probability with no load balance method decreases substantially during the simulation. Because of the consideration of user mobility, the success probabilities of the proposed method with different selection strategies are higher than the compared method's. We can see that the success probability with the maximum difference selection strategy is higher than the one with the minimum V selection strategy, and it is because that the maximum difference selection strategy selects the UE that is the most unsuitable one staying in the overloading HeNB. And that's why the success probability of the maximum difference strategy is higher than the minimum V strategy.

$$\textit{Success probability} = \textit{number of success calls / number of total calls} \tag{22}$$

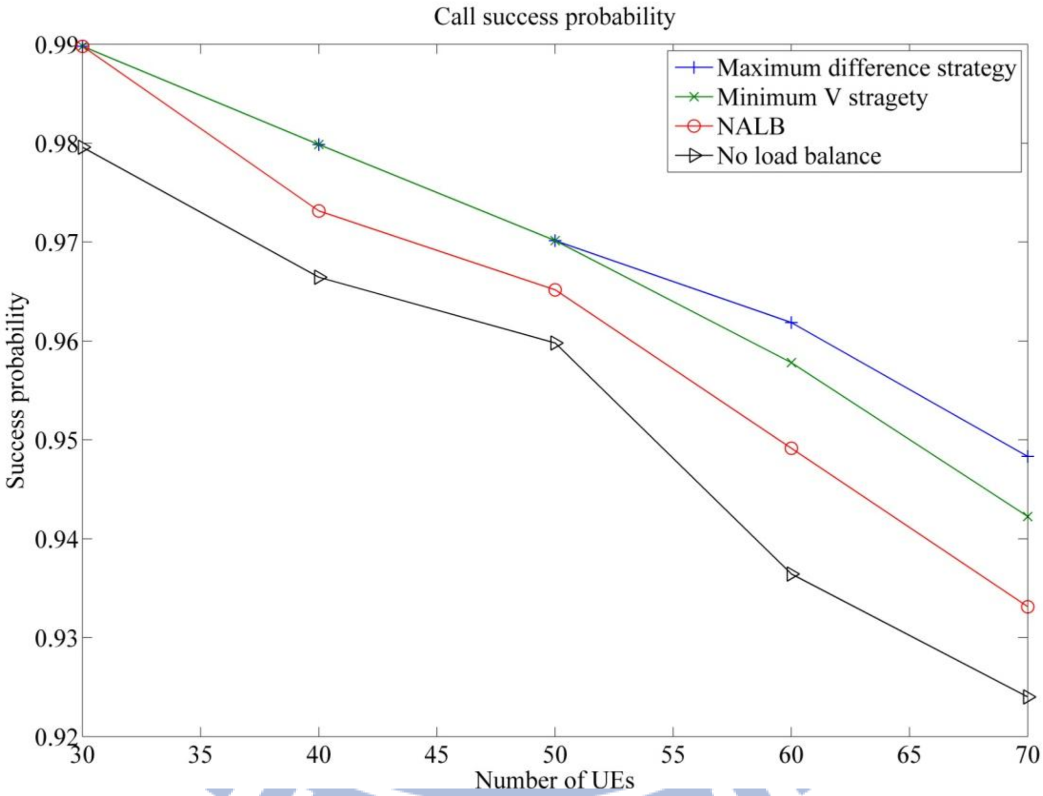

Figure 19: Success probability

The comparison of balance indexes is depicted in Fig. 20. The definition of balance index is slightly different with the one that used in chapter 4.2.1, and it is defined in Eq. (23), where *M* is the set of all the HeNBs and O_i is the average number of used PRBs in $HeNB_i$. The calculation of Eq. (12) in chapter 4.2.1 involves only the overloaded HeNB, the target HeNB and the HeNBs whose coverage area is overlapped with the overloaded HeNB. On the contrary, the balance index in Eq. (24) involves all the HeNBs during the simulation.

We can see that when the number of UEs is small, the balance index is low as well. This is because that some of the HeNBs are connected by UEs and some are not. And that is why the balance index is low. With the increase in number of UEs, each HeNB have one or two UEs connecting to them at least. And the larger the number of UEs, the higher the balance index is. The balance index with no load balance method has the lowest value because there is no handover when a HeNB overloads. We can see that the balance index with the proposed method is higher than the one with the compared method, and it is because the compared method considers loading only, and the proposed method considers not only the loading but also the balance indexes.

$$Balance\ index = \frac{(\sum_{i\in M} O_i)^2}{n(M)\sum_{i\in M} O_i^2} \tag{23}$$

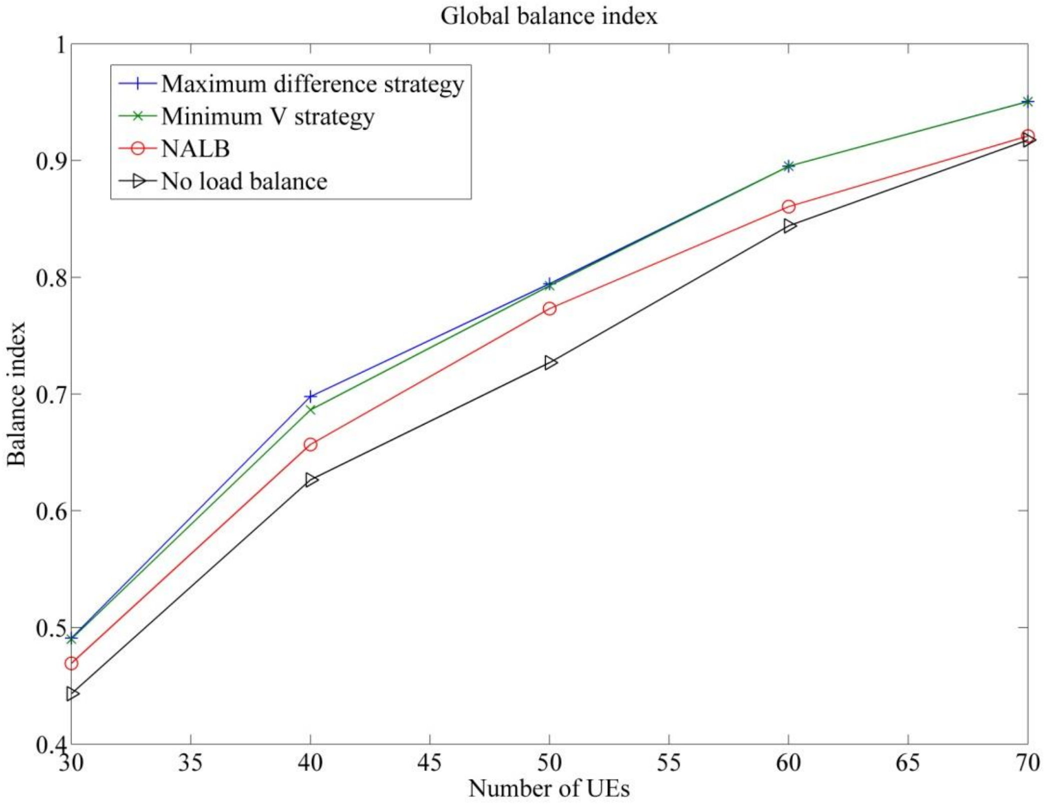

Figure 20: Balance index

The average throughput of data UEs is depicted in Fig. 21. With the increase in number of UEs, the average data throughput decreases, and this is because there are more UEs sharing the resources. Because the maximum bandwidth requirement is 3Mbps and the minimum bandwidth requirement is 1.5 Mbps, average data throughput falls within the region between 1.5 to 3 Mbps. We can see that the curve with no load balance method has the lowest data throughput, and the data throughput of the compared method is lower than the proposed selection strategies'. This is because that the proposed method has the highest balance index. Higher balance index means that the loading is distributed to those HeNBs more evenly, and that is why the proposed method has the largest average data throughput under the same number of UEs.

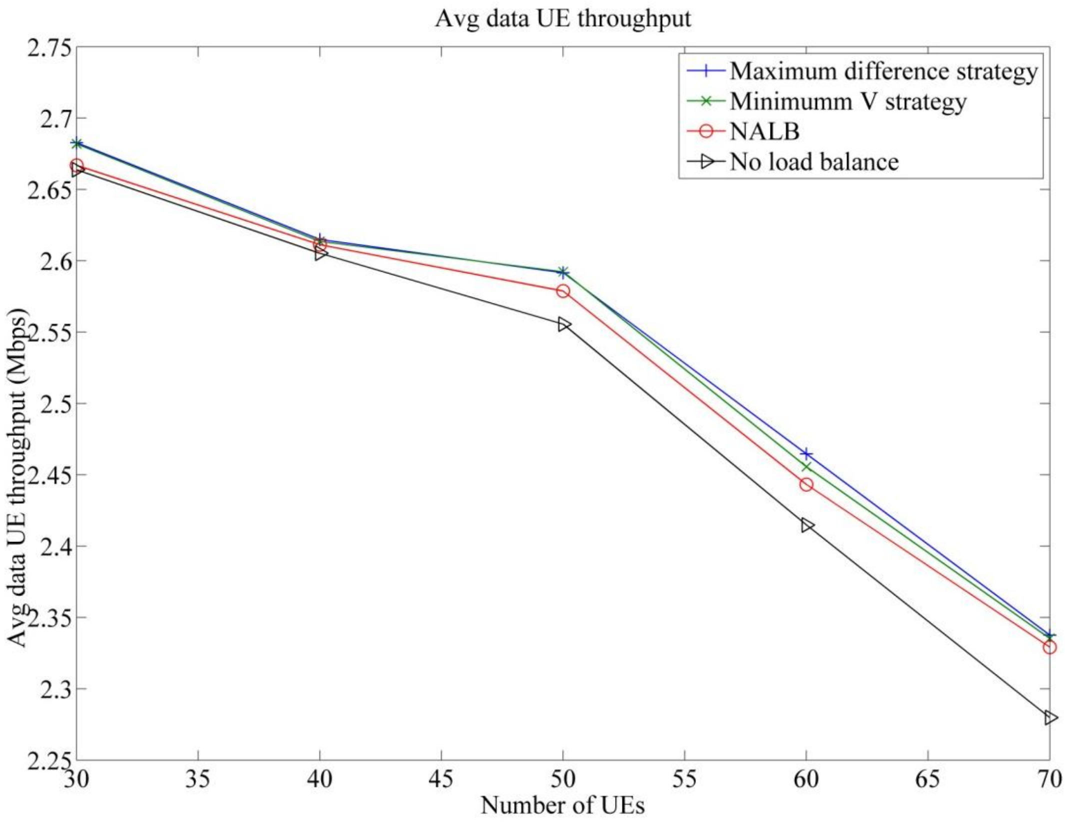

Figure 21: Average data UE throughput

The number of affected HeNBs per balance is shown in Fig. 22. When the HeNB overloads, the compared method does not mention how to select the UE for handover if there are many UEs that can make ξ_2 zero. On the contrary, the proposed method selects the handover UE based on the balance index. Better balance index means that when an UE is handovered to the target HeNB, the loading of target HeNB may be much less than the critical level, so the target HeNB is not likely to be overloaded. Poor balance index means that the loading of some HeNBs may be very close to the critical level, and if the loading of target HeNB is close to the critical level, the target HeNB is very likely to be overloaded if a UE handovered to it. And that is why the proposed method has a smaller number of affected HeNBs.

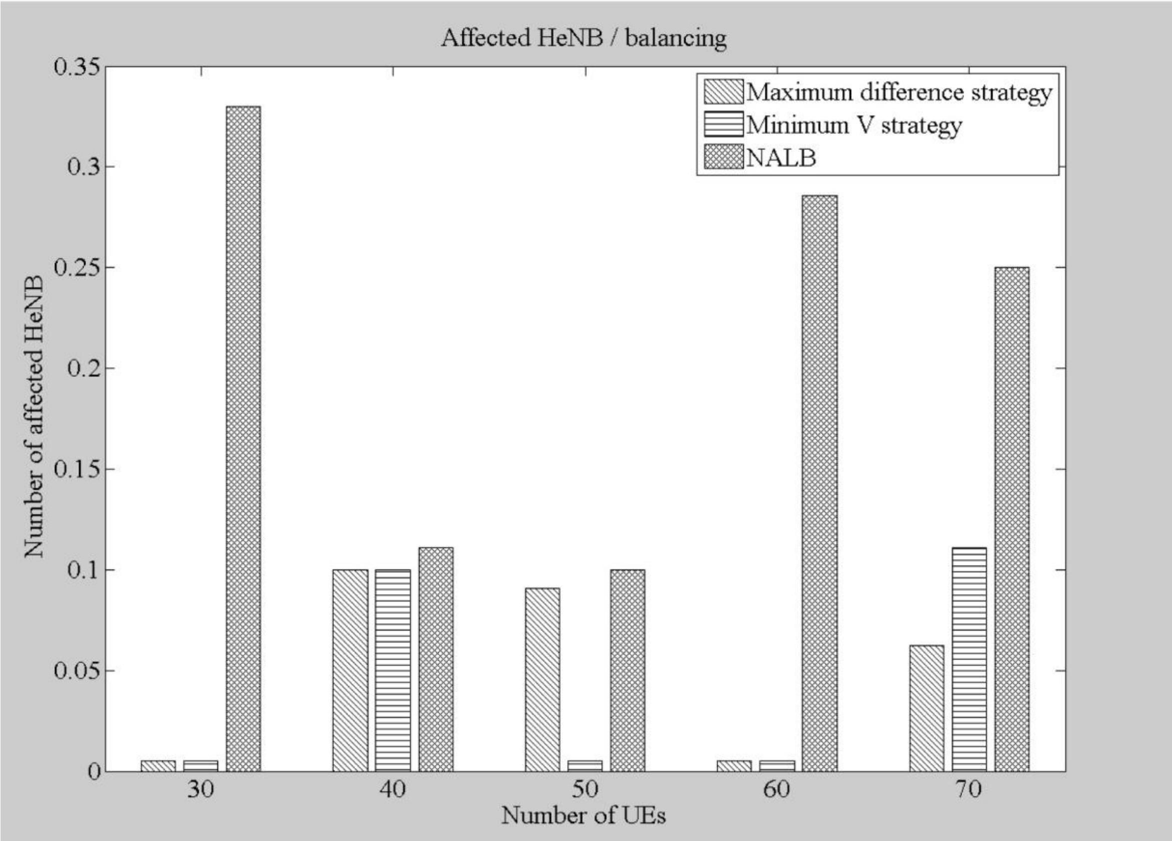

Figure 22: Number of affected HeNB

We present a simple scenario in Fig. 23. In this scenario, there are three HeNBs, i.e.

HeNB 0, HeNB 1 and HeNB 2. There are six non-moving UEs generating traffic during the simulation, and each HeNB covers two UEs respectively. We put a mobile UE in HeNB 0 initially, and the UE moves horizontally through HeNB 1 and then HeNB 2. In this simulation, we want to show that using the proposed method, and we can handover the UE to the correct HeNB according to the direction of UE movement. Table 5 shows the serving HeNB of the mobile UE with different load balance methods. When the mobile UE moves to the position with X axis equal to 387m, the HeNB 0 overloads. It is because that with the increase in distance, the SNR decreases, and in order to maintain the QoS, the number of required PRBs also rises; and that is why HeNB 0 overloads. With the proposed method, the mobile UE handover to HeNB 2 with X axis equal to 387m; on the contrary, the compared method handover the mobile UE to HeNB 1. This is because that the proposed method considers not only satisfaction, but also the UE movement. The reason why the compared method chooses HeNB 1 is that loading in HeNB 2 is higher than that in the HeNB 1. From Fig. 23, we can see that it is better to handover the movable UE to HeNB 2 instead of HeNB 1 according to UE movement. So it is better to use the proposed method instead of the compared one in this scenario.

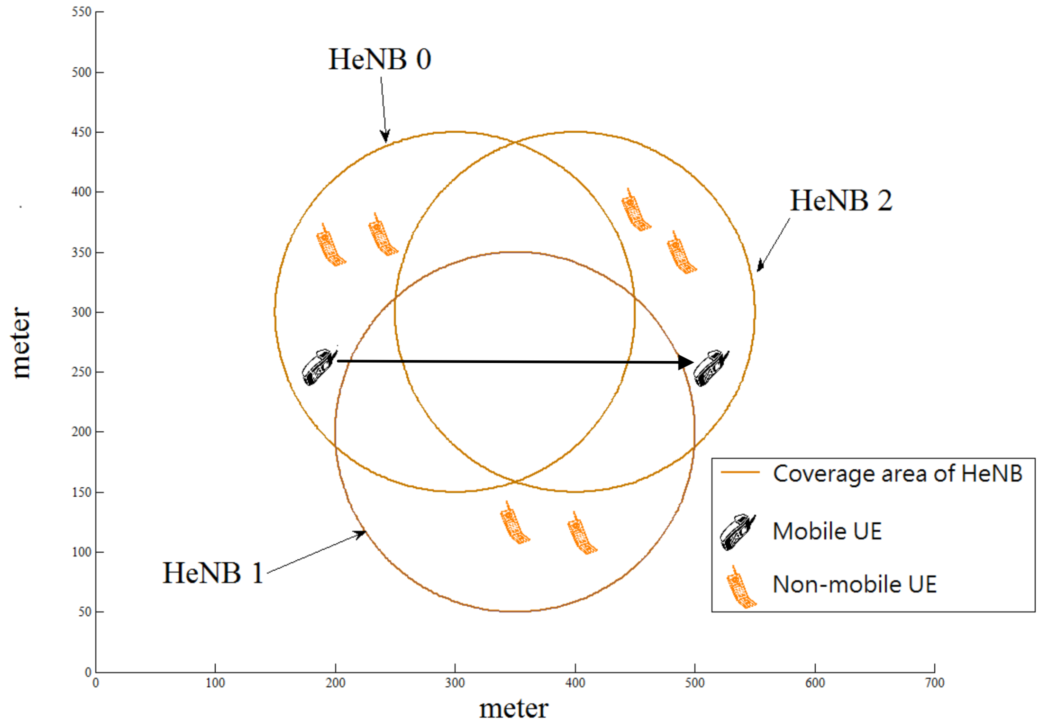

Figure 23: Simple simulation scenario

Table 5: Serving HeNB of the mobile UE

HeNB	**HeNB 0**	**HeNB 1**	**HeNB 2**
Method	**X axis of movable UE**		
Maximum diff.	180 m ~ 387 m		387 m ~ 520 m
Minimum value	180 m ~ 387 m		387 m ~ 520 m
compared	180 m ~ 387 m	387 m ~ 447 m	447 m ~ 520 m

Chapter 6 Conclusion and future work

In this thesis, we propose a load balance method which considers UE movement, QoS satisfaction and load balance between HeNBs in the 3GPP HeNB environment. The consideration of UE movement prevents the moving UE from handovering back to the ex-serving HeNB after the load balance procedure. The QoS satisfaction guarantees that the UE receives the minimum QoS requirement at least after the load balance procedure. The consideration of load balance between HeNBs makes an even distribution of loading between HeNBs, and the simulations show that the consideration of load balance has benefits for call success probability, average data UE throughput, system balance index and the number of affected HeNBs per load balance.

We show that the proposed method works well beyond the compared method through the simulation. However, there are many kinds of QoS requirements, such jitter and delay for VoIP connections and VoIP is an important application for the HeNB environment. We only consider one QoS requirement, which is throughput, in the proposed method. Thus, we will focus on jitter and delay for VoIP connections in our following work.

References

[1] 3GPP TS 36.300 v8.11.0, "Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2," ed, 2010.

[2] 3GPP TR 33.820 v8.3.0, "Security of Home Node B (HNB) / Home evolved Node B (HeNB)," ed, 2009.

[3] 3GPP TS 36.211 v8.9.0, "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical channels and modulation," ed, 2010.

[4] L. Jing, *et al.*, "Improved pilot power adjustment for load balancing in the CDMA system," in *Wireless Communications, Networking and Mobile Computing, 2005. Proceedings. 2005 International Conference on*, 2005, pp. 1020-1023.

[5] D. Lin, *et al.*, "A bubble oscillation algorithm for distributed geographic load balancing in mobile networks," in *INFOCOM 2004. Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies*, 2004, p. 338.

[6] Z. Yu, *et al.*, "Utility-based load balancing in WLAN/UMTS internetworking systems," in *Radio and Wireless Symposium, 2008 IEEE*, 2008, pp. 587-590.

[7] K. Ishizu and H. Harada, "A Load-Balancing Framework for Cognitive Wireless Network to Coexist with Legacy WiFi Systems," in *Vehicular Technology Conference, 2009. VTC Spring 2009. IEEE 69th*, 2009, pp. 1-5.

[8] N.-V. Quoc-Thinh, *et al.*, "Novel Approach for Load Balancing in Heterogeneous Wireless Packet Networks," in *Network Operations and Management Symposium Workshops, 2008. NOMS Workshops 2008. IEEE*, 2008, pp. 26-31.

[9] H. Velayos, *et al.*, "Load balancing in overlapping wireless LAN cells," in *Communications, 2004 IEEE International Conference on*, 2004, pp. 3833-3836 Vol.7.

[10] *NS2 simulator*. Available: http://isi.edu/nsnam/ns

[11] S. Pal, *et al.*, "User-satisfaction based differentiated services for wireless data networks," in *Communications, 2005. ICC 2005. 2005 IEEE International Conference on*, 2005, pp. 1174-1178 Vol. 2.

[12] 3GPP TS 36.213 v8.8.0, "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures," ed, 2009.

[13] Motorola, "TBS and MCS Signaling and Tables, 3GPP R1-081638," ed, Apr. 2008.

[14] 3GPP TS 25.467 v9.0.0, "UTRAN architecture for 3G Home Node B (HNB); Stage 2," ed, 2009.

[15] 3GPP TS 25.413 v8.5.0, "UTRAN Iu interface Radio Access Network Application

Part (RANAP) signalling," ed, 2010.

[16] 3GPP TS 36.331 v8.9.0, "Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC); Protocol specification," ed, 2010.