國 立 交 通 大 學

電信工程學系

碩 士 論 文

# 異質網路中結合利益函數與賽局理論之網路接取機制

# Utility Function and Game-Theory Based Network Selection Scheme in Heterogeneous Wireless Networks

研 究 生：蔡宗利

指導教授：張仲儒　博士

中 華 民 國 九 十 七 年 七 月

# 異質網路中結合利益函數與賽局理論之網路接取機制

# Utility Function and Game-Theory Based Network Selection Scheme in Heterogeneous Wireless Networks

研　究　生：蔡宗利　　　　Student：Tsung-Li Tsai

指導教授：張仲儒　博士　　Advisor：Dr. Chung-Ju Chang

國立交通大學

電信工程學系

碩士論文

A Thesis

Submitted to Department of Communication Engineering

College of Electrical and Computer Engineering

National Chiao Tung University

in Partial Fulfillment of the Requirements

for the Degree of Master of Science

in

Communication Engineering

July 2008

Hsinchu, Taiwan

中華民國九十七年七月

# 異質網路中結合利益函數與賽局理論之網路接取機制

研究生：蔡宗利　　　　　指導教授：張仲儒

國立交通大學電信工程學系碩士班

Mandarin Abstract

## 摘要

為了提供行動用戶無縫(seamless)的網路接取，連接到多元異質網路的能力是不可或缺的。賦予行動用戶多模(multi-mode)的能力，能使他們根據通道狀況或自身的服務品質需求(QoS requirement)選擇更適宜的網路。為了支援高傳輸速率的多媒體服務以及高速的行動用戶，一個 WCDMA/WMAN/WLAN 的異質網路被提出。

在 WCDMA/WMAN/WLAN 的異質網路系統中，為了增加可容納的使用者個數、減少換手(handoff)的頻率，並且同時保證使用者的服務品質需求，在本篇論文中我們提出了一個結合利益函數和賽局理論的網路接取機制。利益函數式是用來得到行動用戶對服務品質需求的滿足程度，而賽局理論則是用來達到負載平衡(load balance)和減少換手頻率。藉由利益函數和賽局理論得到的結果，我們最後會決定一個最合適的接取網路。模擬中顯示我們提出的方法可以減少新使用者被拒絕進入系統中的機率，同時能降低已存在的使用者被強迫中止(forced terminated)的機率。除此之外，換手的頻率也大幅度的下降，並且能符合服務品質需求。

# Utility Function and Game-Theory Based Network Selection Scheme in Heterogeneous Wireless Networks

Student: Tsung-Li Tsai Advisor: Chung-Ju Chang

Department of Communication Engineering
National Chiao Tung University

English Abstract

## Abstract

To maximize accommodated call numbers, minimize handoff rate, and support QoS requirements at the same time in the heterogeneous wireless network, a utility function and game theory (UGT) based network selection scheme is proposed in this thesis. When a new call or a handoff call arrives, UGT will find which networks are usable for the call request first. After getting the candidate networks, UGT will compute the utility value from the satisfaction of QoS requirements of the call request and the preference value from predefined cooperative game for each candidate network. The main goal of the cooperative game is to decrease the number of handoff and achieve load balance for high system utilization. Finally, by choosing the maximum linear combination of utility values and preference values from all candidate networks, the most suitable network for the call request can be obtained. The simulation results show that the proposed scheme can reduce the new call blocking rate, the number of forced terminated calls, and the handoff occurrence frequency. Besides, the QoS requirements are satisfied no matter in low traffic load or high traffic load.

Acknowledgement

# 誌 謝

兩年的碩士生涯很快的就過去了。能順利的完成這篇碩士論文，必須感謝非常多人的鼓勵與幫忙。首先要感謝張仲儒教授這兩年來在專業知識和待人處世方面的指導，讓我在學業和生活上都獲益良多。除此之外，還要特別感謝抽空回來指導我們的詠翰和芳慶學長，提供我們最新的資訊和專業上的看法。感謝立峰、志明、耀興、文祥、和振宇學長們，每在我遇到困難的時候，能夠提供我寶貴的意見。還有在進實驗室時熱情歡迎我們的學長姐建安、建興、正昕、世宏、佳泓、佳璇。當然，還要謝謝和我一起努力的浩翔、英奇、維謙、邱胤、巧瑩，大家一起打球真的很快樂。還有努力帶給我們實驗室歡樂時光的學弟妹志遠、欣毅、和儁、盈伃，以及辛苦的助理惟媜和玉棋。有你們的陪伴，研究的路途一點也不寂寞。

最後，我要感謝我的父母、哥哥、女朋友和我的愛犬，有了他們對我的支持和期待，我才有動力度過每一個難關，順利完成階段的目標。

蔡宗利 謹誌

民國九十七年七月

# Contents

# List of Figures

# List of Tables

# Chapter 1
# Introduction

Broadband, multimedia capability, and mobility are usually major concerns in modern communication technologies. Several broadband wireless network standards, such as wideband code division multiple access (WCDMA) wireless cellular network, IEEE 802.16 wireless metropolitan area network (WMAN), and IEEE 802.11 wireless local area network (WLAN), have been developed and are amending to seek for better efficiency and functionality of Internet applications. In order to provide mobile stations (MSs) with seamless Internet access in the heterogeneous wireless networks, the connection capability to various radio access networks (RANs) is also necessary. Equipping the MSs with multi-mode ability can enable the MSs to select proper RANs according to the channel state and the service type. However, each RAN has different features. In order to make appropriate selection for each radio access in the heterogeneous networks, it is necessary to consider some essential information from each RAN when performing the network selection.

Mobility support means that the system has to face the challenge of handoff in the heterogeneous network. There are two types of handoff in the heterogeneous wireless networks. First is the horizontal handoff, which means that the connection is handed over between service areas within same types of RAN. Second is the vertical handoff, which means that the connection is handed over with different types of RAN. It has been a popular research issue in recent years to support service continuity for multi-mode MSs. Stevens-Navarro, Lin, and Wong proposed a Markov-decision-process-based (MDP-based) vertical handoff decision algorithm for heterogeneous wireless networks [1]. They formulated the vertical handoff problem as an MDP with an objective of maximizing an expected total reward of connections. A reward function for a connection is used to model the QoS of the mobile connection. A signaling cost function is used to model the switching and re-routing operations when a vertical handoff occurs. Results show that the MDP-based algorithm gives a high expected total reward and low expected number of vertical handoffs per connection.

A higher mobility MS has larger possibility to experience more handoffs during its call holding time. Usually, more handoffs imply more overhead and higher risk of call dropping. In the cellular/WLAN interworking networks, some literatures proposed that when a real-time service request arrivals, it would be better to select cellular networks first. If there is no free bandwidth available for this real-time service, then depending on whether it is a new call or handoff call, the system will reject it or try WLAN [2]. However, it may be inappropriate that the selection of RANs is just according to the service type. Defining a cost function, which is composed of mobility and QoS requirements such as user's minimum data rate, maximum tolerable delay, bit error rate (BER), and so on, to decide how to select for a radio access in the heterogeneous networks would be a better approach [3]. Furthermore, since there are multi-parameters which will influence the final decision, the multi attribute decision making (MADM) method can also be adopted as the network selection scheme [4].

Lots of literatures have been proposed to select the RAN in the heterogeneous networks. Yilmaz, et al. [5] compared the performance of five simple access selection principles. Nevertheless, they did not consider some important parameters, such as service types, QoS, and mobility. To combine different view point in the access selection may be a good idea. In [6], Chen, et al. proposed a scheme consisting of two algorithms, which are the access selection algorithm on the user side and the price control algorithm on the network side.

Also, in order to increase the overall system utilization for the heterogeneous network, consideration of load balance is very important. Ning, Zhu, Peng, and Lu proposed a load balance algorithm in heterogeneous networks by assigning new calls to under-loaded networks and allowing the MSs with non-real time traffic in overloaded networks to handoff to under-loaded heterogeneous networks at any time [7]. A dynamic load balance algorithm based on sojourn time was proposed for the heterogeneous network too [8]. With this algorithm, the total network utilization can be increased and the blocking and dropping probabilities can be decreased but at the cost of increasing number of handoff.

Selecting a suitable RAN to provide services for an MS can be regarded as a kind of competition behavior. It would be a good approach to adopt game theory [9] to design the access selection algorithm for the heterogeneous networks. Antoniou and Pitsillides model the network selection as a game [10]. They assume there are lots call requests as a set of strategy in a game. However, a call request should be served as soon as it comes. In [11], Niyato and Hossain used a *bankruptcy game* [12] formulation which is a special type of N-person cooperative game to find the solution of the bandwidth allocation problem in a heterogeneous wireless access network. But they only considered non-real time service and did not take the user mobility into consideration. They also proposed a non-cooperative game-theoretic framework in heterogeneous wireless access networks [13]. Nevertheless, above papers of [11] and [13] assumed that a call request can use various RANs simultaneously. For examples, a call request can transmit 30% data through WLAN and 70% through cellular at the same time. This will cause some problems, such as synchronization and extra overhead, in real world situation.

In this thesis, a utility function and game theory (UGT) based network selection scheme is proposed for heterogeneous wireless access networks, where multiple classes of traffic are considered. The UGT scheme intends to maximize accommodated call numbers, minimize handoff rate, and support QoS. First, a utility function is defined to represent the degree of fulfillment of QoS requirements for a call request. Then the game theory is adopted to define a cooperative game in order to achieve load balance among RANs and to decrease the number of handoff. The cooperative game means the players choose their strategy jointly in order to get the best outcomes for them. By solving the cooperative game, the compromising solutions (strategies) for these two different purposes can be obtained. Finally, a linear combination of above two outcomes (results from utility function and cooperative game) is used to decide an appropriate RAN for a call request.

The rest of the thesis is organized as follows. Chapter 2 describes the system model. Chapter 3 introduces game theory. Chapter 4 is the proposed utility function and game theory (UGT) based network selection scheme. Chapter 5 shows simulation results and discussions. Finally, Chapter 6 gives the conclusions and future works.

# Chapter 2
# System Model

## 2.1 WCDMA/WMAN/WLAN Interworking Architecture

A heterogeneous access network containing a WCDMA cellular system, an IEEE 802.16 WMAN system, and an IEEE 802.11 WLAN system is shown in Fig. 2.1. WCDMA services are available at any place, while WMAN and WLAN services are only available regionally. It is assumed that WLANs are deployed only at some places for high-speed data services in the urban area.

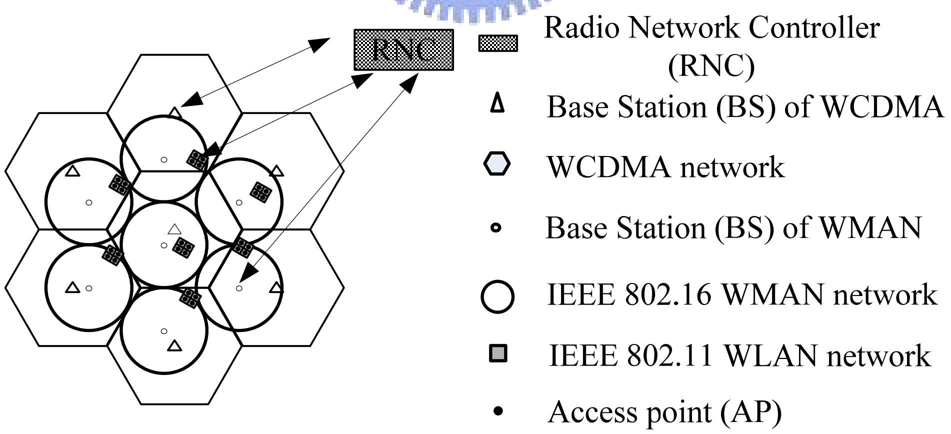



Fig. 2.1 : The network topology of WMAN, WCDMA and WLAN systems

Base stations (BSs) of WMAN and WCDMA networks can collect the information of MSs, including channel quality, velocity, position, direction of motion, and traffic class of a call request [13]. An access point (AP) of the WLAN acts as the

BS in WCDMA network. The proposed utility and game-theory based network selection scheme is designed in a radio network controller (RNC), which gathers information from BSs for selection. These three different wireless access networks are described as follows.

### 2.1.1 WCDMA Cellular Network

For the interference-limited WCDMA networks, the BS needs to control interference in the cell. In this thesis, only the uplink direction is considered, and it is assumed that whenever the uplink channel is assigned, the downlink is established. Also, the transmitted signal power for each MS is assumed to can be adaptively controlled in order to achieve the target received signal power in the BS. Then the achievable bit rate for $MS_j$, denoted as $AR_j$, can be obtained by [15]

$$AR_j = \frac{W}{v_j \cdot (E_b / N_0)_j} \times \frac{P_j}{I_{total} - P_j}, \tag{2.1}$$

where $W$ is the chip rate, $v_j$ is the activity factor of $MS_j$, $(E_b / N_0)_j$ is the signal energy per bit divided by noise spectral density that is required to meet a predefined QoS of $MS_j$, $P_j$ is the signal power from $MS_j$ received at BS, and $I_{total}$ is the total power including thermal noise power received at BS. Note that the $(E_b / N_0)_j$ requirement of $MS_j$ is determined from the bit error rate requirement, service type, velocity, and so on of the $MS_j$.

### 2.1.2 IEEE 802.16 WMAN

Orthogonal frequency division multiple access (OFDMA) has been adopted for IEEE 802.16 WMAN [16]. Suppose there are $K$ sub-channels in the OFDMA system, and each sub-channel consists of $q$ spread out sub-carriers. Thus the channel condition of each sub-channel can be regarded as the same, and the frequency selective condition can be compensated. Assume that each frame includes $L$ OFDMA symbols, and the duration for each frame is $T$. The total number of resource allocation unit, defined as one sub-channel and one OFDMA symbol, in a frame will be $K \times L$. Moreover, equal power control, which means the same allocated power to each service request, is adopted. From [17], an approximation of modulation order, denoted as $M$, when the required BER is given can be obtained by

$$M = \frac{1.5 \times SINR_{a,k}^{(\ell)}}{-\ln(5 \times BER_a^*)} + 1, \tag{2.2}$$

where $SINR_{a,k}^{(\ell)}$ is the received signal to interference and noise ratio (SINR) of service request $a$ on sub-channel $k$ at the $\ell$th OFDMA symbol and $BER^*_a$ is the required bit error rate of service request $a$. However, there are 4 types of modulation: no transmitted, QPSK, 16-QAM, and 64-QAM. So the usable modulation order of service request $a$ on sub-channel $k$ for the $\ell$th OFDMA symbol, denoted by $m_{a,k}^{(\ell)}$, is given as

$$m_{a,k}^{(\ell)} = \begin{cases} 0, & \text{if } M < 4, \\ 2, & \text{if } 4 \le M < 16, \\ 4, & \text{if } 16 \le M < 64, \\ 6, & \text{if } 64 \le M. \end{cases} \tag{2.3}$$

Finally, the total allocated bits $B_a$ to service request $a$ in the current frame can be obtained

$$B_a = \sum_{\ell}^{L} \sum_{k=1}^{K} q \cdot c_{a,k}^{(\ell)} \cdot m_{a,k}^{(\ell)}, \tag{2.4}$$

where $c_{a,k}^{(\ell)}$ is the allocation indicator. The value of $c_{a,k}^{(\ell)}$ equals to 1, if the scheduler allocates the resource on sub-channel $k$ at the $\ell$th OFDMA symbol to the service request $a$. On the contrary, it will be 0.

### 2.1.3 IEEE 802.11 WLAN

The WLAN system supports distributed coordinator function (DCF) mode and point coordinator function (PCF) mode for media access. DCF adopts carrier sense multiple access with collision avoidance (CSMA/CA) protocol with a slotted binary exponential backoff scheme. PCF is a centralized polling protocol controlled by the AP. In order to support service differentiation, IEEE 802.11e proposes the enhanced DCF (EDCF) mode, which allows the AP to initiate a duration of transmission opportunity in the contention period [18]. This standard amendment also defines four differentiated priorities to support QoS. MSs using EDCF mode to transmit data are assumed in this thesis.

Under the EDCF mode, a MS cannot transmit packets until the channel is sensed idle for a time period equal to the arbitration inter-frame space duration (AIFSD). When an MS senses the channel busy during the AIFSD, the backoff time counter is randomly selected from the range $[0, CW-1]$, where $CW$ is the contention window. The value of $CW$ is increased from $CW_{min}$ to $CW_{max}$ if consecutive fail transmissions occur, where $CW_{min}$ is the initial value of $CW$, $CW_{\max} = 2^m CW_{\min}$ is the maximum value of $CW$ and $m$ is called the maximum backoff stage.

The EDCF introduces the concept of access categories (ACs). Different ACs has different $AIFSD[AC]$, $CW_{min}[AC]$, and $CW_{max}[AC]$. Traffic classes with smaller values of $CW_{min}$ and $CW_{max}$ represent higher priorities. $AIFSD[AC]$ for different ACs can be given by

$$AIFSD[AC] = SIFS + AIFS[AC] \times SlotTime, \tag{2.5}$$

where $SIFS$ is the duration of short inter-frame space, $AIFS[AC]$ is a positive integer , $SlotTime$ is the duration for a slot.

## 2.2 Channel Model

The wireless fading channel is composed of large-scale fading and small-scale fading. The large-scale fading comes from path loss and shadowing effect, while the small-scale fading is caused by multipath reflection. The pass loss is modeled as [19]

$$L_{pathloss} = 128.1 + 37.61 \times \log d_{bm} (dB), \tag{2.6}$$

where $d_{bm}$ is the distance between the BS and the MS in kilometers. Assume the log-normal shadowing is with zero mean and standard deviation of 8 dB. For small-scale fading, the Jakes model [20] is used to simulate the fading channel. Furthermore, the channel is assumed to be fixed within a frame and varies independently from frame to frame.

## 2.3 Mobility Model

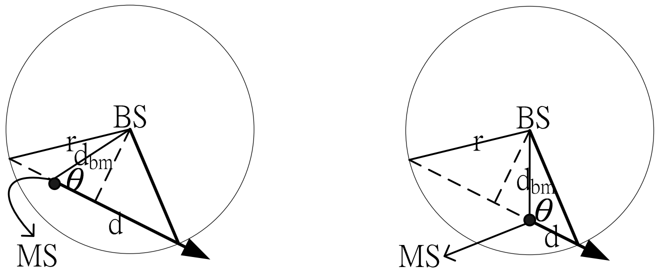
For high mobility MS (30, 50, or 80 km/hr), assume that its speed $v$ and direction of motion are never changed in the network. As shown in Fig. 2.2, $r$ is the radius of network coverage, $\theta$ is the angle between BS and the moving direction of MS, where $0 \le \theta \le \pi$, and $d_{bm}$ is the distance between BS and MS, where $0 \le d_{bm} \le r$. Then the total travel distance in the network, denoted by $d$, can be obtained by

$$d = \sqrt{r^2 - (d_{bm} \cdot \sin\theta)^2} + d_{bm} \cdot \cos\theta, \text{ where } 0 \le d \le 2r. \tag{2.7}$$

So we can get the *dwell time* of the MS in this network, denoted by $T_{dwell}$, by

$$T_{dwell} = d / v. \tag{2.8}$$

For low mobility MS (3 km/hr), its speed is also assumed to be unchanged. But the direction will be changed randomly every certain fixed duration.



(a) when $\theta < 90^0$ (b) when $\theta > 90^0$

Fig. 2.2 : High mobility MS model

## 2.4 Traffic Class

There are four traffic classes considered [21]: conversational class, streaming class, interactive class and background class. The conversational class represents real-time multi-media applications such as telephony (voice). The streaming class includes streaming type of applications, like video on demand (VoD). The interactive class is composed of applications for Web-browsing, chat room, etc. Finally, the background class is the service using best effort transmission, such as file transfer protocol (FTP). It can be found that the first two classes are delay-sensitive (real time), and the last two classes are delay-tolerant (non-real time).

Each service request has different QoS parameters according to their service types. Intuitively, the real-time traffic requests low delay, low jitter, and the number of handoff must keep as low as possible. But they are tolerant of certain level of packet loss. On the other hand, non-real time traffic may request high bandwidth, and low packet loss rate, etc. However, variation of transmission rates is acceptable to them.

## 2.5 Source Model

The conversational class traffic is modeled as the ON-OFF model [22] shown in Fig. 2.3. During ON period, voice packets are generated with rate $D_v$ bps. During OFF period, there is no packet generated. This model has a transition rate with value $y$ in the ON state and a transition rate with value $z$ in the OFF state.

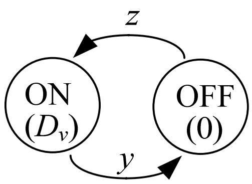


Fig. 2.3 : Voice source model

Fig. 2.4 depicts the source model of streaming class, which is composed of a sequence of video frames generated regularly with a constant interval $T_f$ [19]. Each video frame consists of a fixed number of slices $N_s$, where each slice corresponds to a single packet. The size of packet is denoted by $P_s$, and the inter-arrival time between each packet is $T_p$.

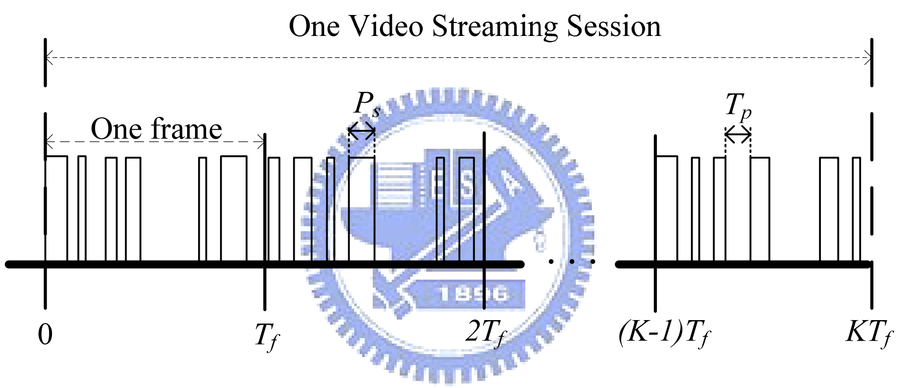


Fig. 2.4 : Video streaming source model

Fig. 2.5 shows the source model of HTTP interactive class. The interactive class traffic can be modeled as a sequence of packet calls (pages), and each packet call consists of a sequence of packet arrivals, which is composed of a main object and several embedded objects [19]. Four parameters, including the inter-arrival time $T_{reading}$ (reading time), main object size $S_m$, embedded object size $S_e$, the number of embedded objects per packet call $N_e$, and the packet inter-arrival time $T_p$ are used in this model.

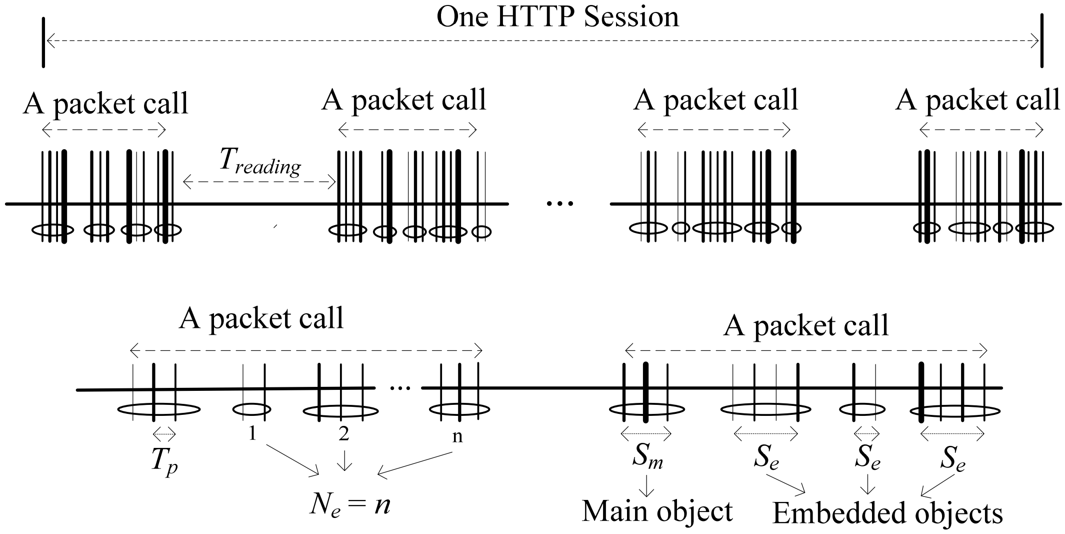



Fig. 2.5 : HTTP source model

The background class traffic is modeled as a sequence of file downloads [19] and is shown in Fig. 2.6. Denote the size of each file by $S_f$, and the inter-arrival time between each file by $T_f$.

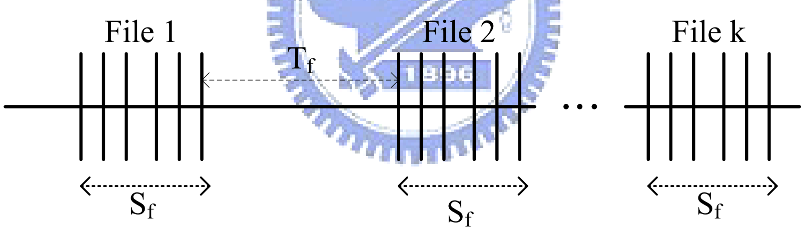



Fig. 2.6 : FTP source model

# Chapter 3
# Game Theory

Game theory can study strategic interactions between agents, where the agent is an actor or player in a model that solves an optimization problem [9]. In strategic games, agents choose strategies which will maximize their payoff, given the strategies of the other agents. There are cooperative and non-cooperative games. In a cooperative game, the players choose their strategy jointly. In the non-cooperative game, each player selects his strategy individually, without any joint-action agreements among players. If a configuration of strategies (one for each player) such that each player's strategy is best for him, given those of the other players, the set of strategies is called be in Nash equilibrium.

A game consists of a set of players, a set of strategies available to these players, and a specification of payoffs for each combination of strategies. From the definition of the payoff function, there are zero-sum games and non-zero-sum games [9]. In zero-sum games the total payoffs to all players always adds to zero. Poker is an example of zero-sum game, because one wins and the other loses. On the contrary, in non-zero-sum games, the summation of payoffs for all players may be larger or less than zero according to different combination of strategies. This means a set of strategies may cause the result of that all win or even all lose!

As shown at table 3.1, the typical example, which is called prisoner's dilemma, is used to explain what the difference between cooperative and non-cooperative game is. For non-cooperative game, prisoner A and B all do not know whether the other side will choose ‘stays silent’ or ‘betrays’. At this situation, they will find ‘betrays’ is the best strategy no matter what the other side’s strategy is. If these two guys are smart, they all need serve 5 years finally. For cooperative game, however, they can know the other side’s strategy. At this situation, they both will choose ‘stays silent’ to get a compromise. That is they just need to serve 6 months respectively.

Table 3.1 : Prisoner’s dilemma

| | Prisoner B Stays Silent | Prisoner B Betrays |
|---|---|---|
| Prisoner A Stays Silent | Each serves 6 months | Prisoner A: 10 years<br>Prisoner B: goes free |
| Prisoner A Betrays | Prisoner A: goes free<br>Prisoner B: 10 years | Each serves 5 years |

For the network access selection problem in heterogeneous networks, a cooperative and non-zero-sum game is defined. Based on the defined game, the goal is to find a set of strategies which satisfy the Nash equilibrium to help the access selection scheme.

# Chapter 4
# Utility Function and Game-Theory Based Network Selection Scheme

When a new call or a handoff call arrives, the system must determine the candidate networks first. The *candidate networks selection* will find which RANs are usable for the call request by checking some thresholds. Note that those thresholds are just used to check the 'available' networks for the call request. After getting the set of candidate networks, whose number is n, the proposed scheme will find the values of $NU_i$ and $NP_i$, for $i = 1,2,\ldots,n$. Note that $NU_i$ and $NP_i$ are gained from *utility function for QoS satisfaction* and *cooperative game for network preference*, respectively. For each candidate network, *Utility function for QoS satisfaction* will compute the utility value from the satisfaction of QoS requirement of the call request for each candidate network. Then, for each candidate networks, the cooperative *game for network preference* will compute the preference value from the network point of view. The main goal of this part is to decrease the number of handoff and achieve load balance for high system utilization. Finally, by chosen the maximum linear combination of utility values and preference values, the most suitable RAN for the call request can be obtained.

## 4.1 Candidate Networks Selection

Two constraints are proposed to select the candidate networks: the signal strength constraint and network loading constraint. An access network must fulfill these constraints and then becomes the candidate network for a call request.

### 4.1.1 Signal Strength Constrain

Define the pilot signal strength from access network $i$ received at the MS as $PW_i$. If the value of $PW_i$ exceeds a given power threshold $PW_{th}$, that is

$$PW_i \geq PW_{th}, \tag{4.1}$$

then the network $i$ will be classified as a candidate network. Otherwise, the network will be neglected from the candidate networks. Notice that the predefined signal strength threshold may be different for different access networks.

### 4.1.2 Network Loading Constraint

The constraint is used to guarantee that the admittance of a call request will not affect the quality of the ongoing connections. Assume that a call request is required to report its traffic characteristic parameters when it asks to access the network. The traffic characteristic parameters of a call include peak rate, utilization (fraction of time the source is active), and mean peak rate duration of the packets. Then the equivalent capacity for the call request, say $a$, denoted by $C_a$, can be obtained [23]. If infinite buffer size is assumed, the equivalent capacity can be derived to be equal to mean rate. This thesis will use the mean rate of a new call request as its equivalent capacity to do the network loading increment/decrement estimation when the call enters/leaves the network.

Define the current existing network loading intensity before accepting the new call is $\mathrm{H}_E$ ($0 \le \mathrm{H}_E \le 1$), and the new loading intensity increment for the call request *a* is $\eta_a$. Then the loading intensity of a candidate network after accepting the call request must be under a predefined threshold loading $\eta_{th}$. That is

$$\mathrm{H}_E + \eta_a \le \eta_{th}. \tag{4.2}$$

On the contrary, this network will not be considered as the candidate network.

In the WCDMA network, the loading intensity increment for a call request *a*, can be estimated as [15]

$$\eta_a = (1+f)\frac{1}{1+W/(C_a \cdot (E_b/N_0)_a)}, \tag{4.3}$$

where *f* is the factor representing for interference from other cells and is defined as the ratio of inter-cell interference to the total interference in the referenced cell, *W* is the chip rate of the WCDMA system, and $(E_b/N_0)_a$ is the required bit-energy to noise-density figure corresponding to the desired link quality of the call request *a*. Clearly, $\mathrm{H}_E = \sum_{e \in E} \eta_e$, where *E* is the set of existing calls in the WCDMA network.

The IEEE 802.16 WMAN uses the OFDMA technique. From chapter 2, the mean capacity of WMAN can be estimated as $4 \times K \times L \times q / T$ (bps). Then the loading intensity increment for a call request *a* can be estimated as

$$\eta_a = C_a / (4 \times K \times L \times q / T), \tag{4.4}$$

The same, $\mathrm{H}_E = \sum_{e \in E} \eta_e$, where *E* is the set of existing calls in the WMAN network.

In WLAN network, the measurement-based network load intensity estimation is used. Assume $T_s$ is the total busy occupation transmission time, including successful transmission time and collision time, in the latest observation duration $T_d$. Define the loading intensity as $H_E = T_s / T_d$. Only when following equation is satisfied, then this network will be still in the set of candidate networks.

$$\mathrm{H}_E \le \eta_{th,WLAN}, \tag{4.5}$$

where $\eta_{th,WLAN}$ is predefined load intensity threshold for WLAN.

## 4.2 Utility Function for QoS Satisfaction

A utility function $U_i$ for each candidate network $i$ is defined to represent the preference of call request. It is a product of three QoS-related evaluation functions, which is given by

$$U_i = f_{B,i} \times f_{D,i} \times f_{R,i}, \tag{4.6}$$

where $f_{B,i}$, $f_{D,i}$, and $f_{R,i}$ are the evaluation functions of data rate, packet delay, and packet dropping rate for access network $i$, respectively. Note that these three evaluation functions of QoS provisioning are designed from the preference of users. For each RAN, if QoS measures that the network can provide is better than the QoS requirements of the call request, then the evaluation functions will get higher evaluation value, denoting more preference of users. As shown in Fig. 4.1, if the QoS measure value for the network can satisfy the QoS requirement of call request (in the QoS satisfaction region), then the evaluation value will increase gently (linearly). On the contrary, if the QoS measure value is in the QoS violation region, the evaluation value will decrease sharply (exponentially).

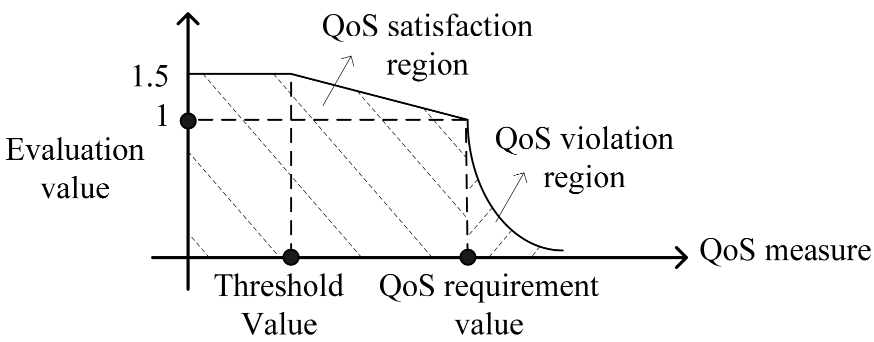



Fig. 4.1 : Evaluation function

Therefore, $f_{B,i}$ is defined as

$$f_{B,i} = \begin{cases} 1.5, & if\ B_{th} \le B_i \\ 1 + 0.5 \times \dfrac{B_i - B_{req}}{B_{th} - B_{req}}, & if\ B_{req} \le B_i < B_{th}, \\ \exp[-\dfrac{(B_i - B_{req})^2}{4 \times B_{req}{}^2}], & if\ B_i < B_{req} \end{cases} \tag{4.7}$$

where $B_i$ is the measured allowed data rate in access network $i$, $B_{req}$ is the data rate requirement of the call request, and $B_{th}$ is a threshold used to represent whether the QoS is highly satisfied. Note that the measured allowed data rate in WCDMA can be obtained by (2.1), and the achievable modulation order in WMAN can be estimated by (2.2), (2.3). Finally, the measured allowed data rate in WLAN is gotten by measurement-based network loading intensity estimation.

Moreover, $f_{D,i}$ is defined as

$$f_{D,i} = \begin{cases} 1.5, & if\ D_i \le D_{th} \\ 1 + 0.5 \times \dfrac{D_{req} - D_i}{D_{req} - D_{th}}, & if\ D_{th} < D_i \le D_{req}, \\ \exp[-\dfrac{(D_i - D_{req})^2}{4 \times D_{req}{}^2}], & if\ D_{req} < D_i \end{cases} \tag{4.8}$$

where $D_i$ is the measured average packet delay in access network $i$, $D_{req}$ is the maximum delay tolerance of the call request, and $D_{th}$ is a threshold used to represent whether the QoS is highly satisfied. Note that the values of average packet delay for different traffic classes are computed separately. Different traffic class has different average packet delay in the same access network.

Similarly, the evaluation functions of packet dropping rate are formulated by

$$f_{R,i} = \begin{cases} 1.5, & if\ R_i \le R_{th} \\ 1 + 0.5 \times \dfrac{R_{req} - R_i}{R_{req} - R_{th}}, & if\ R_{th} < R_i \le R_{req} \\ \exp[-\dfrac{(R_i - R_{req})^2}{4 \times R_{req}{}^2}], & if\ R_{req} < R_i \end{cases}, \tag{4.9}$$

where $R_i$ is the measured average packet dropping rate in access network *i*, $R_{req}$ is the maximum allowable dropping rate of the call request, and $R_{th}$ is a threshold used to represent whether the QoS is highly satisfied. Similar to above case, the values of average packet dropping rate for different traffic classes are computed separately.

Noted that only real-time traffic classes have delay bound, so the non-real time traffic classes need not take $f_{D,i}$ and $f_{R,i}$ into consideration. After getting $U_i$ for each candidate network *I*, the normalized utility value for each candidate access network *i*, denoted by $NU_i$, is defined as

$$NU_i = U_i / \sum_{j=1}^{n} U_j, \tag{4.10}$$

where *n* is the number of candidate network. Clearly, $\sum_{i=1}^{n} NU_i = 1$ and $NU_i \ge 0,\ \forall i$.

## 4.3 Cooperative Game for Network Preference

Consider the load balance among the RANs. It is more preferable to choose the network with low load to serve the call request. The load balance can help to achieve the goal of maximizing the total system utilization. Moreover, consider to decrease the number of handoff, including horizontal and vertical handoff, it is more preferable to decrease the number of handoff to reduce the forced termination probability of call request and signaling overhand for handoffs.

Game theory is here adopted to solve the problem. A network preference cooperative game is defined as follows

**Players**: The players of this game are the candidate access networks. Assume there are ***n*** players : $\{N_1, N_2, \ldots, N_n\}$.

**Strategies**: ***n*** strategies : $\{NP_1, NP_2, \ldots, NP_n\}$. $NP_i$ is the preference value for $N_i$ from network provider view point. Note that $\sum_{i=1}^{n} NP_i = 1$.

**Payoffs**: The payoff for the total candidate networks is defined as

$$PO_{total}(NP_1, NP_2, ..., NP_n) = \sum_{i=1}^{n} A_i \times (NP_i - w_i \times NP_i^2), \tag{4.11}$$

where $A_i = (1 - \mathrm{H}_{E,i} / \eta_{i,th})$, $\mathrm{H}_{E,i}$ is the current loading intensity of network *i* before accepting the call request, $\eta_{i,th}$ is predefined threshold load intensity of network $i$, and $w_i$ is penalty weight of network *i*. The meaning of the payoff function and how to define the penalty weight are described as follows.

### 4.3.1 Meaning of the Payoff Function

Finding the best set of strategies for each candidate network to maximize the payoff function is the main goal of the cooperative game. First, consider the load balance in the heterogeneous networks system. It is more suitable for the call request to choose the access network with low traffic load. The value $A_i$ represents the remaining resource (in the ratio form) available before allocating resource to network $N_i$ for the call request. The more remaining resource of one RAN, the more likely that the call request will choose this RAN. However, assigning more resource (preference value) to a network means other networks will get less resource. If taking the suitability in each access network for a call request into consideration, the situation that lots of resources are allocated to some unsuitable networks must be avoided. The penalty weight $w_i$ is used to achieve this goal. When $w_i$ is less, this means that network $N_i$ is more suitable to the call request.

Generally, if the remaining available resource of one network is higher, then it will get higher preference value (strategy). On the other hand, if the penalty weight of one network is higher, it will obtain lower preference value. That is, each network can get compromising strategy (preference value) in order to maximize the payoff function. The one with the highest preference value means that it is most suitable.

### 4.3.2 Determining the Penalty Weight

As known, if an access network $i$ is more suitable for the call request, the corresponding penalty weight $w_i$ would be lower. Here two factors are considered: dwell time $T_{dwell,i}$ in the network $i$ and relative position in the network $i$ for high mobility MSs and low mobility MSs, respectively.

Assume that the estimated holding time, which is gotten from statistics of call request, is $T_{holding}$. Furthermore, for high mobility MSs, $T_{dwell,i}$ can be obtained from the information of radius of network coverage, velocity, position, and direction of motion of MSs in (2.7) and (2.8). Define $x = T_{holding} / T_{dwell,i}$. When $x > 1$, this means that the call request has high probability to handoff if it chooses the candidate network $i$. Therefore, this network is considered as an unsuitable candidate for the call request in order to decrease the handoff rate. In this situation, the penalty weight $w_i$ is large. On the contrary, when $x < 1$, the call request has high probability to finish the transmission of data in the network $i$. The handoff can be avoided in this case. Then, the penalty weight $w_i$ would be low. Therefore, $w_i$ can be defined as

$$w_i = \begin{cases} 0, & if \quad x \le 0.75 \\ (x - 0.75), & if \quad 0.75 < x \le 1 \\ 0.25 + 3 \times (x\text{-}1), & if \quad 1 < x \le 1.25 \\ 1, & if \quad 1.25 < x \end{cases}. \tag{4.12}$$

However, for low mobility MSs, if an MS is closer to the base station of candidate network $i$, the network $i$ will be more suitable for the call request coming from the MS. The penalty weight $w_i$ will be smaller. On the other hand, if it is far from the BS and at the edge of one cell, the ping-pong effect will happen with large probability. In this situation, the penalty weight for the candidate network will get larger. By this way, the number of handoff can be decreased and the ping-pong effect can be avoided. Define $d_{bm}$ is the distance between BS of network $i$ and MS, $cr_i$ is the radius of network $i$'s coverage, and $cr_{i,th}$ is a predefined value. Then $w_i$ is defined as follow

$$w_i = \begin{cases} 0, & if \quad d_{bm} \le cr_{i,th} \\ (d_{bm} - cr_{th})/(cr_i - cr_{i,th}), & if \quad cr_{i,th} < d_{bm} \le cr_i \\ 1, & if \quad cr_i \ < \ d_{bm} \end{cases}. \tag{4.13}$$

4.3.3 Nash Equilibrium and Optimization problem

After getting $w_i$, the goal is to find the set of strategy which satisfies the Nash equilibrium for the above network preference game. From the definition of Nash equilibrium, the pure strategy $\{NP_1^*, NP_2^*, ..., NP_n^*\}$ is in a Nash equilibrium if

$$PO_{\text{total}}(NP^*_1, NP^*_2, ..., NP^*_n) \ge PO_{\text{total}}(NP_1, NP_2, ..., NP_n) \quad \forall\ NP_1, ..., NP_n. \tag{4.14}$$

In fact, the above game can be formulated as an optimization problem expressed as

$$\begin{aligned} &Maximize \quad && PO_{total}(NP_1, NP_2, \cdots, NP_n) \\ &subject \text{ to} && \sum_{\text{i=1}}^{\text{n}} NP_i = 1, \\ & && NP_1 \ge 0, NP_2 \ge 0, \cdots, NP_n \ge 0. \end{aligned} \tag{4.15}$$

where $PO_{total}$ is a quadratic function. For the above problem which subjects to equality and inequality constraints, the KKT condition [24] can be used to find the solution. With the KKT condition, the solution of (4.15) can be obtained efficiently. See appendix A.

## 4.4 Candidate Networks Decision

Finally, an access network with the maximum compromised evaluative value is expected to obtain. This network decision issue is formulated as an optimization problem given by

$$i^* = \text{Arg } \underset{i}{Max} \ [\alpha NU_i + (1-\alpha) NP_i], \tag{4.16}$$

where $i$ is the $i$th candidate network, $\alpha$ is a constant whose value is between 0 and 1, $NU_i$ is the normalized utility value of candidate network $i$, $NP_i$ is the normalized network preference value of candidate network $i$, and $i^*$ is the chosen access network for the call request.

# Chapter 5
# Simulation Results and Discussions

## 5.1 Simulation Environment

As shown in Fig. 2.1, there are 7 WCDMA cells, 7 WMAN networks, and 28 WLAN networks in the simulation environment. The system parameters in the heterogeneous network are listed in Table 5.1. The channel model and the characteristic of MSs have been introduced in chapter 2.

Table 5.1: System parameters for WCDMA, WMAN, and WLAN

| Parameters | WCDMA | WMAN | WLAN |
|---|---|---|---|
| Cell radius | 1.5 Km | 2 Km | 0.1 Km |
| Frame duration (time slot duration) | 10 ms | 5 ms | 9 us |
| Carrier frequency | 2 GHz | 2.5 GHz | 2.4 GHz |
| load intensity threshold $\eta_{th}$ | 0.75 | 1 | 0.75 |
| Number of cells | 7 | 7 | 28 |
| Chip rate ($W$) | 3.84M bps | | |
| Ratio of inter-cell interference to the total interference in the referenced cell ($f$) | 0.55 | | |
| Number of subchannels ($K$) | | 4 | |
| Number of data subcarriers per subchannel ($q$) | | 48 | |
| Number of slots per frame ($L$) | | 16 | |
| Capacity | | | 2 M bps |

## 5.2 Source Model and QoS Requirements

As described at chapter 2, there are four traffic classes considered. The source model parameters for conversational, streaming, interactive, and background traffic classes are shown in Table 5.2, 5.3, 5.4, and 5.5, respectively.

Table 5.2: Source model parameters for conversational class traffic

| Component | Distribution | Parameters |
|---|---|---|
| ON time | Exponential | Mean=1 sec |
| OFF time | Exponential | Mean=1.35 sec |
| Packets per second | Deterministic | 50 |
| Packet size | Deterministic | 28 bytes |
| Call holding time | Normal | Mean=90 sec, variance=20 sec |
| Data rate during active period | | 11.2 Kbps |
| Active rate | | 0.426 |
| Mean data rate | | 4.77 Kbps |

Table 5.3: Source model parameters for streaming class traffic

| Component | Distribution | Parameters |
|---|---|---|
| Inter-arrival time between each video frame ($T_f$) | Deterministic | 100 ms |
| Number of packets in each video frame ($N_s$) | Deterministic | 8 |
| Packet size ($P_s$) | Truncated Pareto | Min.=40 bytes, Max.=250 bytes Mean=100 bytes, $\alpha$=1.2 |
| Inter-arrival time between packets in a frame ($T_p$) | Truncated Pareto | Min.=2.5 ms, Max.=12.5ms Mean=6 ms, $\alpha$=1.2 |
| Call holding time | Normal | Mean =120 sec, variance =30 sec |
| Data rate during active period | | 133.33 Kbps |
| Active rate | | 0.48 |
| Mean data rate | | 64 Kbps |

Table 5.4: Source model parameters for interactive class traffic

| Component | Distribution | Parameters |
|---|---|---|
| Main object size ($S_m$) | Truncated Lognormal | Min.=100 bytes, Max.=2 Mbytes Mean=10710 bytes, std. dev.=25032bytes |
| Embedded object size ($S_e$) | Truncated Lognormal | Min.=50 bytes, Max.=2 Mbytes Mean=7758 bytes, std. dev.=126168 bytes |
| Number of embedded objects per page ($N_e$) | Truncated Pareto | Mean=5.64, Max.=53 |
| Inter-arrival time between each page ($T_{reading}$) | Exponential | Mean=30 sec |
| Packet size | Deterministic | Chop from objects with size 1500 bytes |
| Packet inter-arrival time ($T_p$) | Exponential | Mean=0.13 sec |
| Call holding time | Normal | Mean =120 sec, variance=30 sec |
| Data rate during active period | | 92.3 Kbps |
| Active rate | | 0.136 |
| Mean data rate | | 12.55 Kbps |

Table 5.5: Source model parameters for background class traffic

| Component | Distribution | Parameters |
|---|---|---|
| file size ($S_f$) | Truncated Lognormal | Min.=50 bytes, Max.=5 Mbytes Mean=2 Mbytes, std. dev.=722 Kbytes |
| Inter-arrival time between each file ($T_f$) | Exponential | Mean = 180 sec |
| Packet size | Deterministic | 3000 bytes |
| Call holding time | Normal | Mean =180 sec, variance =40 sec |
| Data rate during active period | | 88.9 Kbps |
| Active rate | | 1 |
| Mean data rate | | 88.9 Kbps |

As mentioned, the calls with different traffic classes have different QoS requirements. The QoS requirements of each traffic class call are listed in Table 5.6.

Table 5.6: The QoS Requirements of each traffic class

| Traffic class | Requirement | Value |
|---|---|---|
| Conversational (voice) | Required BER | $10^{-3}$ |
| | Required $E_b/N_o$ | 4 dB |
| | Max. delay tolerance | 40 ms |
| | Max. allowable packet dropping rate | 1% |
| Streaming (video) | Required BER | $10^{-4}$ |
| | Required $E_b/N_o$ | 3 dB |
| | Max. delay tolerance | 100 ms |
| | Max. allowable packet dropping rate | 1% |
| Interactive (HTTP) | Required BER | $10^{-6}$ |
| | Required $E_b/N_o$ | 2 dB |
| Background (FTP) | Required BER | $10^{-6}$ |
| | Required $E_b/N_o$ | 1.5 dB |

## 5.3 Iterative TOPSIS Algorithm

The proposed UGT algorithm is compared with the iterative TOPSIS algorithm [4]. Suppose the multi attribute decision making (MADM) method uses the following set of attributes: total capacity, allowed data rate, utilization, packet delay, and packet dropping rate. Then the iterative TOPSIS algorithm is used to solve the MADM method. The iterative TOPSIS algorithm is described as follows:

i. Normalize the value for each of the attributes.

ii. Decide the relative importance of each of the attributes, and weight this value to the corresponding attribute.

iii. Find the best and worst values for each of the new attributes.

iv. Measure the separation (distance) for the best and worst cases for each candidate network $i$, which are denoted by $S_{b,i}$ and $S_{w,i}$, respectively.

v. Measure the preference level $P_i$ for each candidate network $i$. Define $P_i = S_{w,i}/(S_{b,i}+S_{w,i})$.

vi. Remove the candidate network with lowest preference level, and repeat step i. ~ v. until only one network is left. The finally survivor will be the chosen network.

## 5.4 Simulation Results

Suppose that one call request can only connect to one access network at a time here. For each cell, assume the new call arrival rate of conversational, streaming, interactive, and background traffic class calls in the heterogeneous network are $AR \times 1/40$, $AR \times 1/120$, $AR \times 1/120$, and $AR \times 1/240$ (users/second), respectively, where $AR$ is the equivalent arrival rate. In the simulation, $AR$ is chosen from 1, 3, 5, 7, and 9.

Fig. 5.1 shows the new call blocking rate. It can be found that UGT has lower new call blocking rate. That is because UGT chooses lower traffic load network with higher probability than iterative TOPSIS in order to achieve load balance. Iterative TOPSIS also takes the loading intensity (utilization) into consideration, but the final decision is influenced by other attributes. The result shows UGT has a little better performance in the new call blocking rate, generally. However, in the high traffic load, the performance is almost the same.

The handoff call blocking rate is illustrated in Fig. 5.2. It seems that UGT has higher handoff blocking rate than iterative TOPSIS. However, Fig. 5.3 (a) and (b), which depict the number of total handoff calls and the number of failed handoff calls, respectively, show that UGT not only has fewer total handoff calls, but also fewer failed handoff calls. This means UGT has lower number of forced terminated calls. So, in fact, UGT is not worse than iterative TOPSIS.

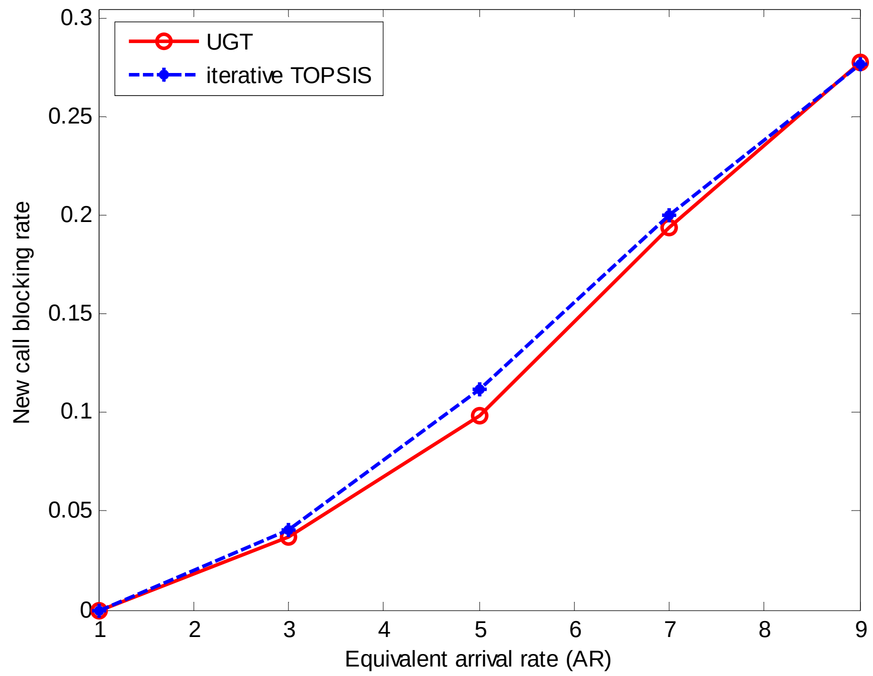



Fig. 5.1 : New call blocking rate

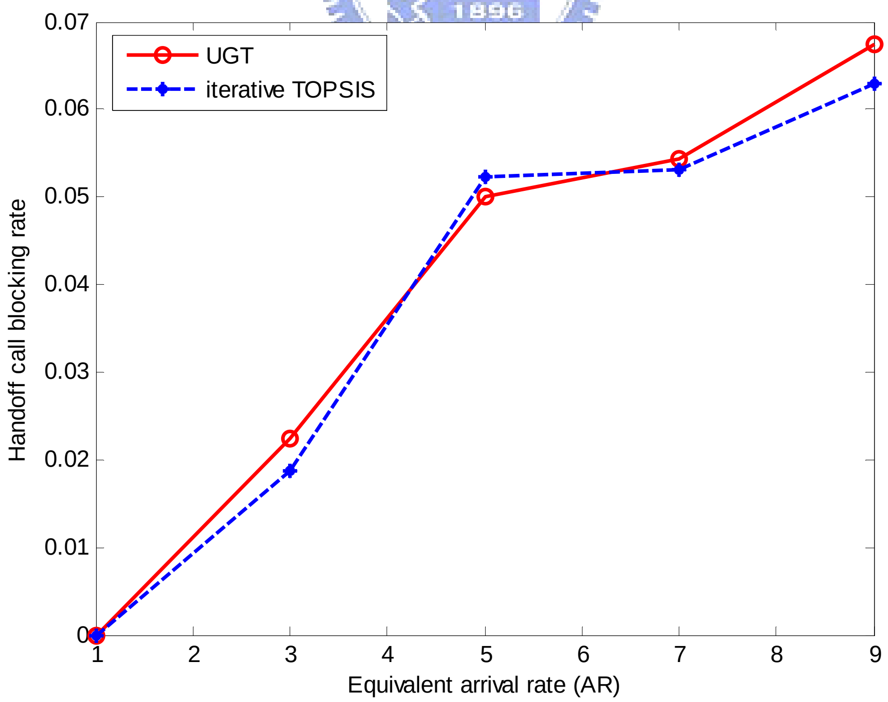
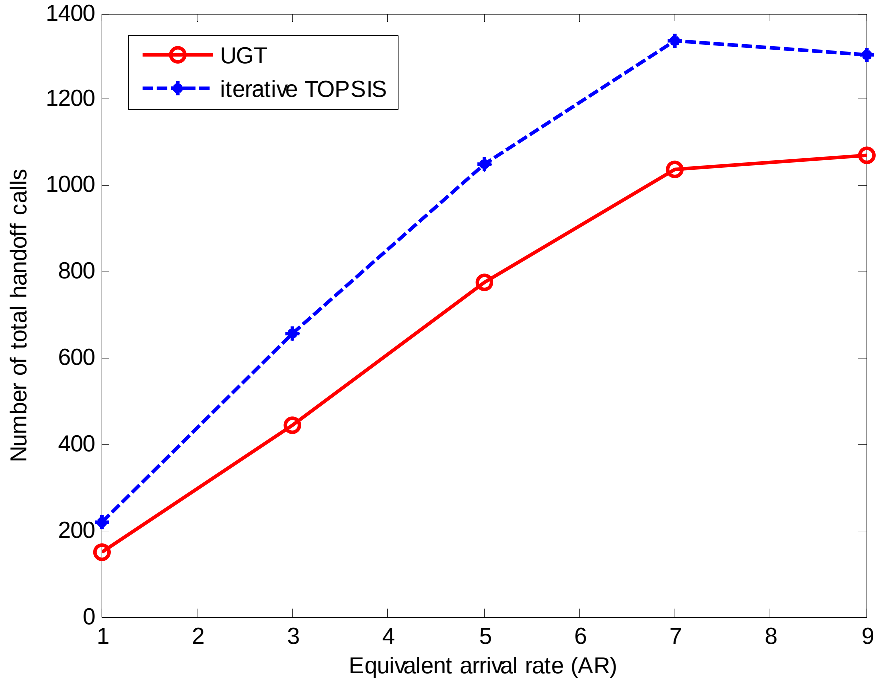
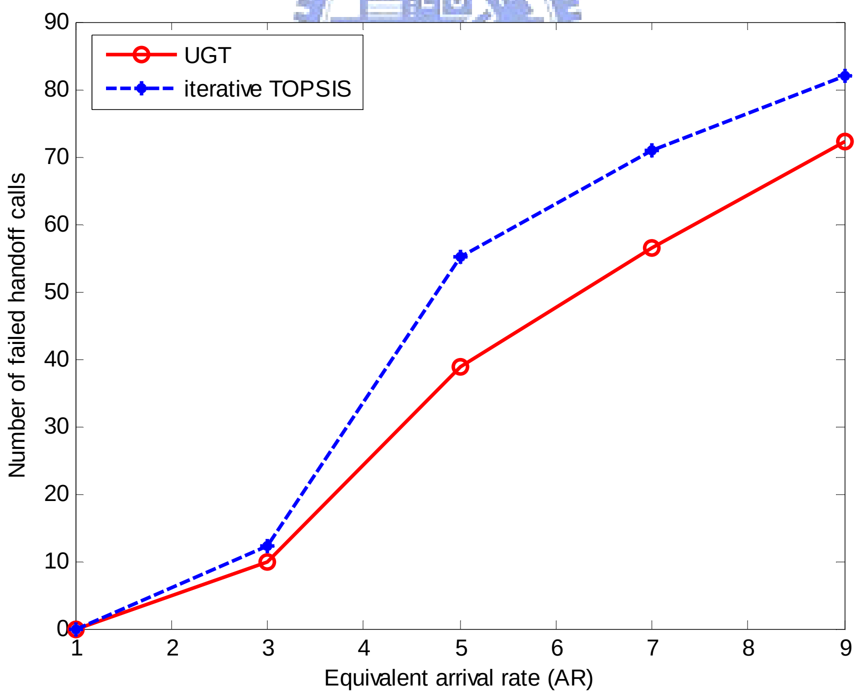



Fig. 5.2 : Handoff call blocking rate




(a) Number of total handoff calls




(b) Number of failed handoff calls

Fig. 5.3 : (a) Number of total handoff calls (b) Number of failed handoff calls

Moreover, it can be found that the trends of new call blocking rate and handoff call blocking rate are very different. That is because the system always reserves 5% resource for handoff calls. When the normalized loading intensity of one network exceeds 95%, a new call will be blocked immediately. On the contrary, a handoff call will not be blocked until the normalized loading intensity reaches 100%. This causes the new call blocking rate will rise exponentially, but handoff call blocking rate will close to the saturated line when the arrival rate gets high gradually.

Handoff occurrence frequency, defined as the number of handoffs per call, is shown at Fig. 5.4. Generally, UGT has lower handoff occurrence frequency than iterative TOPSIS. The result comes from that UGT takes the mobility into consideration, iterative TOPSIS does not. It can be found that the non-real time call has higher handoff frequency in UGT than that in iterative TOPSIS. Since the real time call is more sensitive to the occurrence of handoff, UGT will do it best to avoid handoff for real time call. That is the penalty weight $w$ has higher influence for real time call than non-real time call in UGT.

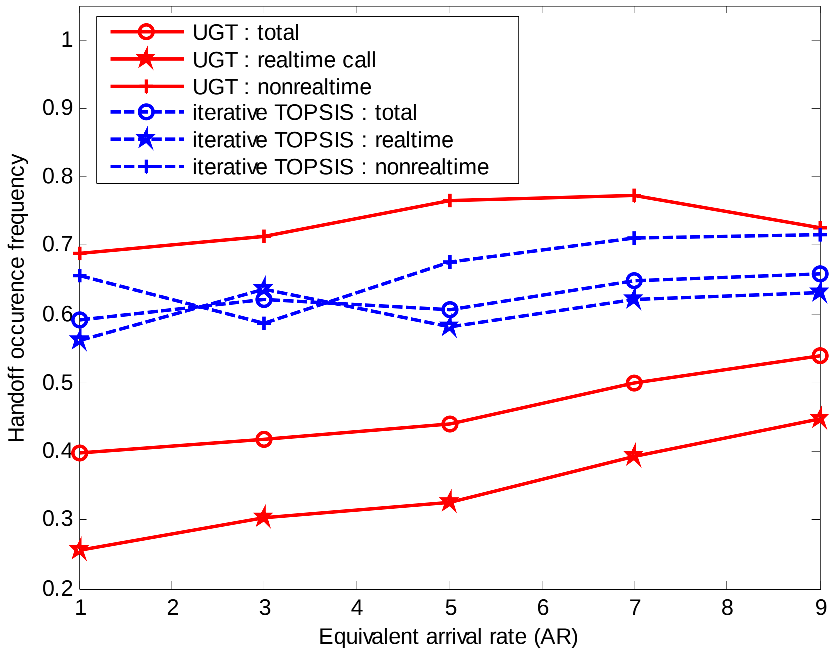



Fig. 5.4 : Handoff occurrence frequency

Fig. 5.5 shows the total throughput and throughput of each network. It can be found that iterative TOPSIS has higher throughput than UGT, and the main difference comes from the throughput in WCDMA and WLAN. The phenomenon can be explained by observing Fig. 5.6 (a) and (b), which plot the number of calls and the number of non-real time calls, respectively. First, the number of calls in WCDMA is analyzed. It can be found that iterative TOPSIS has fewer calls in WCDMA. Moreover, they are almost non-real time calls. On the contrary, UGT has more number of calls in WCDMA, and they are almost real time calls. In the low traffic load, the allowed data rate exceeds the calls' requirement a lot in WCDMA. Since iterative TOPSIS has more non-real time calls in WCDMA and the FTP calls always come with burst, the throughput will get higher obviously. When it comes to the calls in WLAN, it can be found there are more WLAN calls for iterative TOPSIS than that for UGT. Since the number of calls has not achieved its capacity, iterative TOPSIS will have higher throughput clearly.

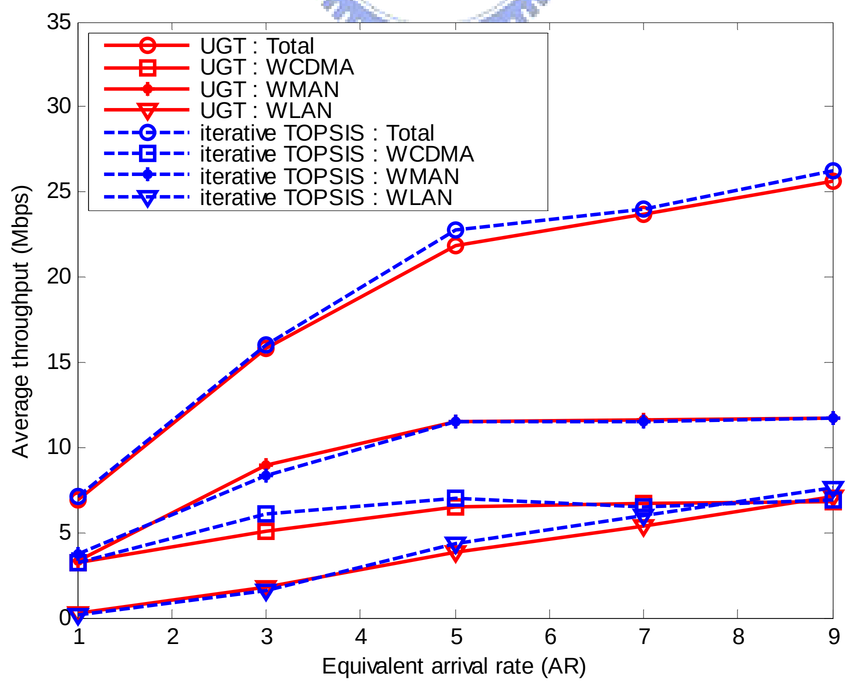
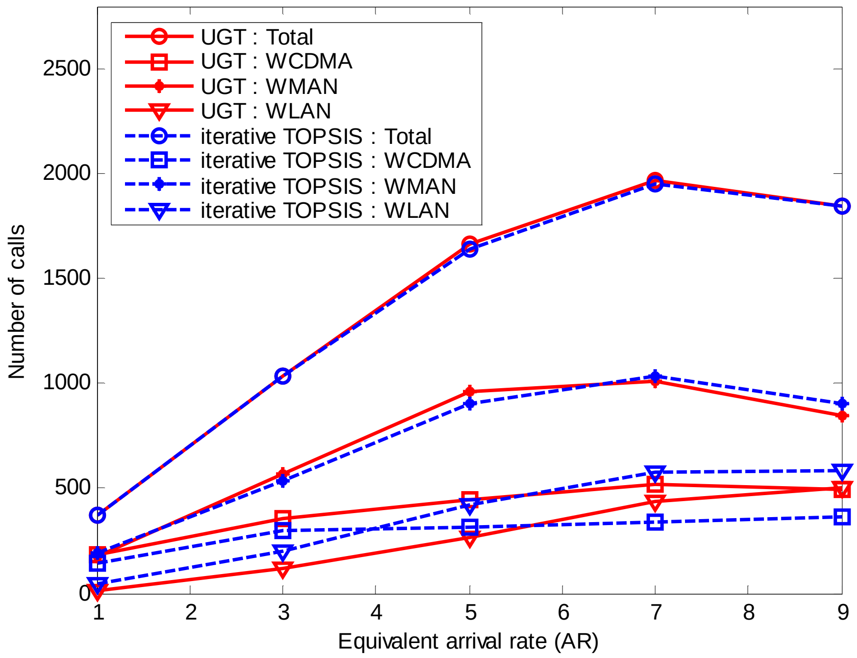
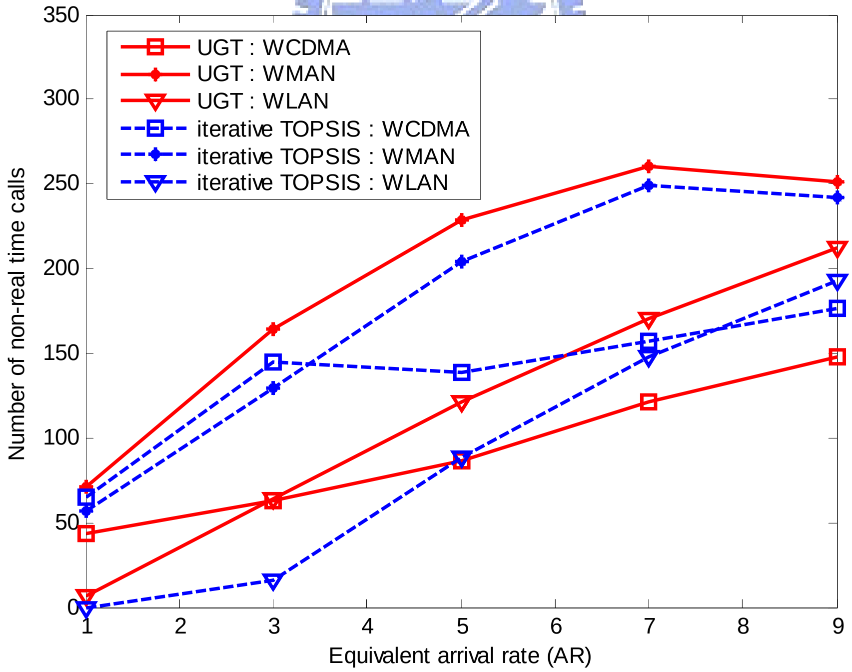



Fig. 5.5 : Total throughput and throughput of each network




(a) Number of calls




(b) Number of non-real time calls

Fig. 5.6 : (a) Number of calls (b) Number of non-real time calls

The average delay for voice and video call in the heterogeneous network are shown in Fig 5.7 and 5.8, respectively. It can be found that the average delay for voice call is almost the same. That is because this traffic class call is highest priority. However, the average delay for video call in high traffic load is higher for UGT than that for iterative TOPSIS in WMAN. This is because there are more video calls for UGT than that for iterative TOPSIS in high traffic load in WMAN. In this situation, WMAN may not have enough resource when a burst comes for video streaming, and then their delay will get high.

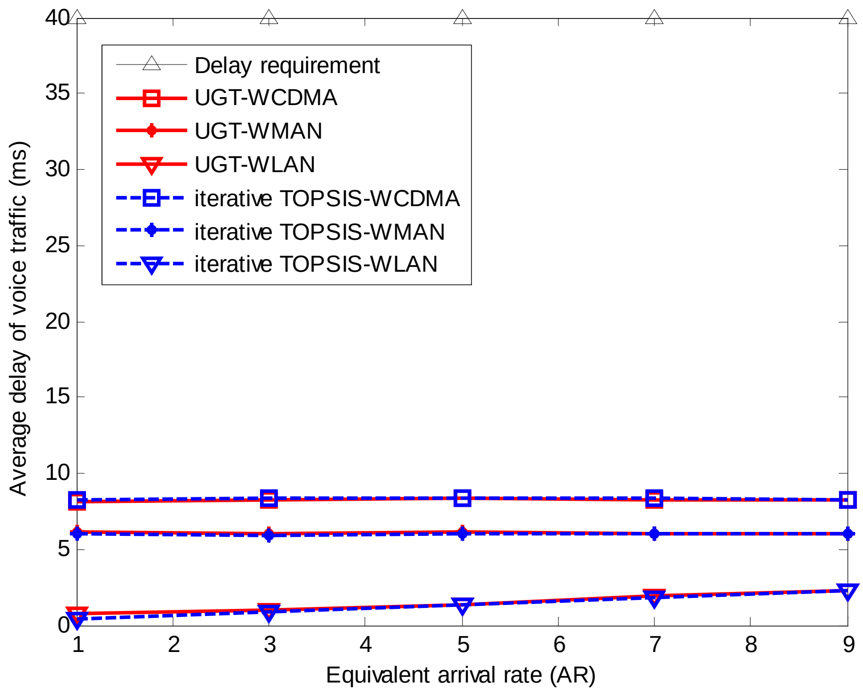



Fig. 5.7 : Average delay of voice traffic

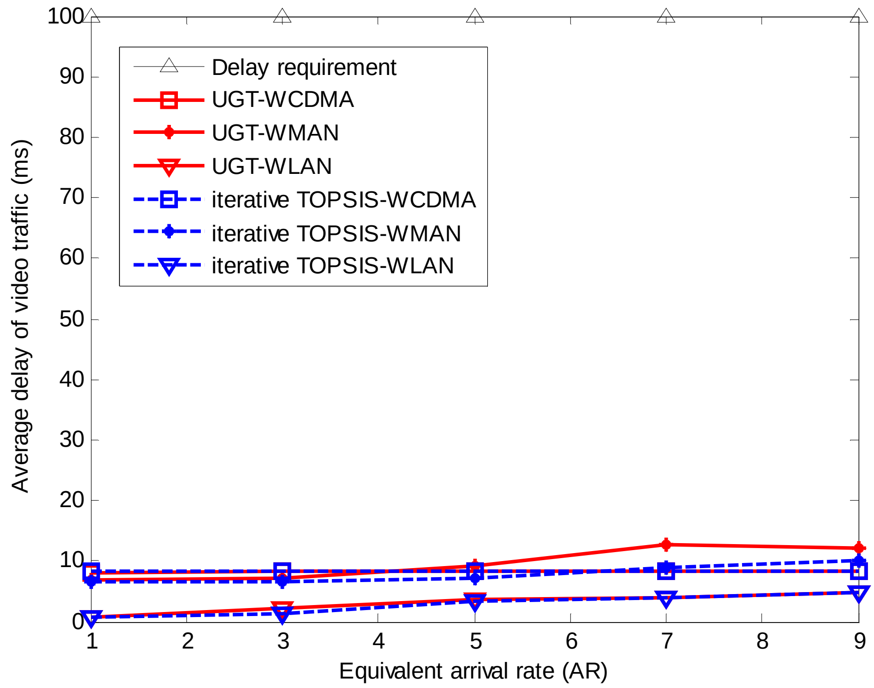



Fig. 5.8 : Average delay of video traffic

The average dropping rate for voice and video call are shown in Fig. 5.9 and Fig. 5.10, respectively. It can be found that the maximum packet dropping rate requirement is satisfied for each scheme. However, the dropping rate is higher in UGT than that in iterative TOPSIS. This is because UGT sees those networks as the same if they can provide enough good QoS requirement, just as shown in Fig. 4.1. On the contrary, iterative TOPSIS see the network as the best if it can provide best QoS requirement for it. Moreover, it can be found that UGT has fewer number of calls in WLAN than iterative TOPSIS has, but the dropping rate is higher in UGT. This is because the calls are almost non-real time calls in UGT. In the design of WLAN in this thesis, it is assumed that when a FTP call gets the right of channel usage, it will transmit 3000 bytes. That is it will occupy at least 12 ms! On the contrary, real time calls transmit much fewer bits than non-real time calls. So the system with more non-real time calls will have higher delay variance. This situation will cause higher dropping rate.

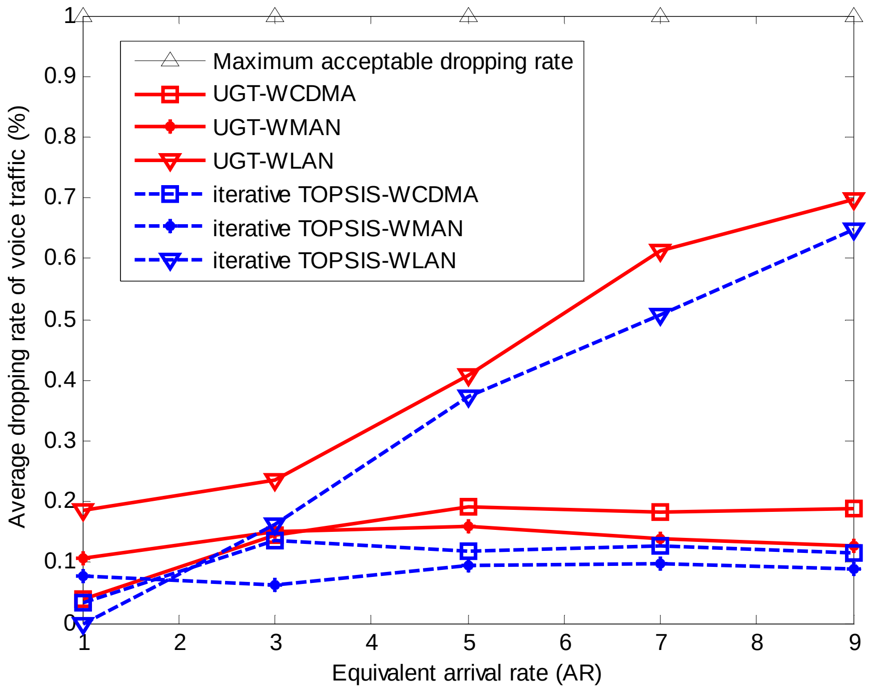



Fig. 5.9 : Average dropping rate of voice traffic

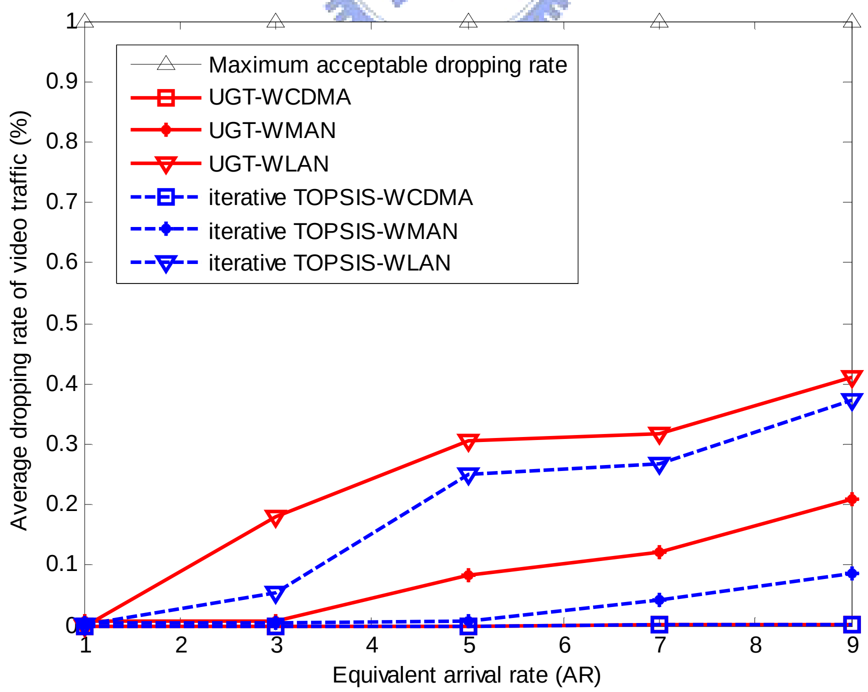



Fig. 5.10 : Average dropping rate of video traffic

# Chapter 6
# Conclusions and Future Works

In this thesis, a utility and game-theory (UGT) based network selection scheme is proposed for heterogeneous wireless access network. By considering four multimedia services, including conversational, streaming, interactive, and background, a call admission control is performed first to find which network can be used when a call request comes. After getting the set of candidate networks, a utility value is obtained to represent the satisfaction degree of QoS requirement. Moreover, in order to achieve load balance and consider mobility factor, a cooperative game is defined to get the preference value for each network. Finally, the most suitable network for the call request can be decided by linear combination of above set of values.

Simulation results show that UGT has lower total throughput than iterative TOPSIS while satisfying the QoS requirements of each traffic class. As known, the difference mainly comes from the non-real time calls. By sacrificing little throughput of non-real time calls, UGT can obtain lower new call blocking rate, fewer forced terminated calls, and fewer handoff occurrence frequency. Lower new call blocking rate and fewer forced terminated calls mean that the heterogeneous system can accommodate more calls. Besides, UGT reduces the handoff occurrence frequency about 30% than iterative TOPSIS generally. However, this value even exceeds 50% for real time calls! With lower handoff occurrence frequency, some problems, happening during the processing of handoff calls, can be avoided substantially. In this aspect, iterative TOPSIS is overwhelmed by UGT. When it comes to the dropping rate, UGT is higher than iterative TOPSIS obviously. But they are all under the maximum acceptable dropping rate. Allowing a little higher dropping rate to exchange for other better performance, which is more critical, may be very worthful. The interesting phenomenon can be observed in our simulation results.

The work can be extended to vertical handoff problem. In this thesis, it is assumed that the handoff occurs only when the call is out of the coverage the original network. However, the handoff can be performed in advance to get better system performance, just like [7]. To make the handoff decision, UGT can be used. At each observation period, an existing call must to decide whether it needs to hand off or not. Some modification of UGT may be very suitable for this problem.

# Appendix A
# KKT Conditions

Consider the following maximizing problem which subjects to equality and inequality constraints:

$$\begin{aligned} &\textit{Maximize} && f(\mathbf{NP}) = \sum_{i=1}^{n} A_i \times (NP_i - w_i \cdot NP_i^2) \\ &\textit{subject}\ \text{to} && h(\mathbf{NP}) = \sum_{i=1}^{n} NP_i - 1 = 0, \\ & && \mathbf{g}(\mathbf{NP}) = \begin{bmatrix} NP_1 & NP_2 & \cdots & NP_n \end{bmatrix}^T \geq \mathbf{0}, \end{aligned} \tag{A.1}$$

where $\mathbf{NP} \in R^n$, $f : R^n \to R$, $h : R^n \to R$, $\mathbf{g} : R^n \to R^n$, and $\mathbf{0}^T = [0 \quad 0 \quad \cdots \quad 0]_{1\times n}$. Note that $n \in \mathbb{N}$, and $\mathbb{N}$ is the set of positive integers. The Karush-Kuhn-Tucker (KKT) condition [24] can be used to find the solution of above problem. Define $\lambda$ as the Lagrange multiplier vector, and $\boldsymbol{\mu} \in R^n$ as the KKT multiplier vector. The KKT condition consists of five parts (three equality and two inequality equations), and is given below

1) $\boldsymbol{\mu} \geq \mathbf{0}$,

2) $Df(\mathbf{NP}) + \lambda \cdot Dh(\mathbf{NP}) + \boldsymbol{\mu}^T D\mathbf{g}(\mathbf{NP}) = \mathbf{0}$, where D is the derivative operator.

3) $\boldsymbol{\mu}^T \mathbf{g}(\mathbf{NP}) = 0$,

4) $h(\mathbf{NP}) = 0$,

5) $\mathbf{g}(\mathbf{NP}) \geq \mathbf{0}$,

Put (A.1) into the KKT condition, then the following results are obtained

$$[\mu_1 \quad \mu_2 \quad \cdots \quad \mu_n]^T \geq \mathbf{0}, \tag{A.2}$$

$$\begin{cases} A_1(1-2w_1NP_1)+\lambda+\mu_1=0 \\ A_2(1-2w_2NP_2)+\lambda+\mu_2=0 \\ \vdots \\ A_n(1-2w_nNP_n)+\lambda+\mu_n=0 \end{cases}, \tag{A.3}$$

$$\mu_1 \cdot NP_1 + \mu_2 \cdot NP_2 + \cdots + \mu_n \cdot NP_n = 0, \tag{A.4}$$

$$\sum_{i=1}^{n} NP_i - 1 = 0, \tag{A.5}$$

$$[NP_1 \quad NP_2 \quad \cdots \quad NP_n]^T \geq \mathbf{0}. \tag{A.6}$$

From (A.2)、(A.4) and (A.6), if $\mu_i > 0$, then $NP_i = 0$. Here, by (A.2), considering three cases to solve the problem.

***Case 1:*** *No value of* $\boldsymbol{\mu}$ *is equal to 0.*

That is $\mu_i > 0$, *for* $i = 1 \sim n$, and $NP_i = 0$, *for* $i = 1 \sim n$. The result conflicts with (A.5), so the set of solution is impossible.

***Case 2:*** *Only one value of* $\boldsymbol{\mu}$ *is equal to 0.*

Assume $\mu_i = 0$, where $i \in \{1, 2, \cdots, n\}$; $\mu_j > 0$, $NP_j = 0$, *for* $j = 1 \sim n$, $j \neq i$. From (A.5), $NP_i = 1$. From (A.3), $\lambda = A_i(2w_i - 1)$, $\mu_j = -\lambda - A_j$, *for* $j = 1 \sim n$, $j \neq i$. The values of $\boldsymbol{\mu}$ must be checked that whether they satisfy (A.2) or not. If satisfied, then this set of solution is valid.

***Case 3:*** *More than one value of* $\boldsymbol{\mu}$ *is equal to 0.*

Assume $\mu_{i_1} = 0, \mu_{i_2} = 0, \cdots, \mu_{i_p} = 0$, where $i_1, i_2, \cdots, i_p \in \{1, 2, \cdots, n\}$;
$\mu_j > 0, NP_j = 0,$ *for* $j = 1 \sim n,\ j \neq i_1, i_2, \cdots, i_p$.

Put the results into (A.3), the following equations are obtained:

$$\begin{cases} NP_{i_2} = a_{i_2} + b_{i_2} NP_{i_1} \\ NP_{i3} = a_{i_3} + b_{i_3} NP_{i_1} \\ \quad\quad \vdots \\ NP_{i_p} = a_{i_p} + b_{i_p} NP_{i_1} \end{cases}, \tag{A.7}$$

where $a_{i_k} = (A_{i_k} - A_{i_1})/(2A_{i_k} \cdot w_{i_k})$, $b_{i_k} = (A_{i_1} \cdot w_{i_1})/(A_{i_k} \cdot w_{i_k})$, for $k = 2 \sim p$. Combine (A.5) and (A.7),

$$NP_{i_1} = (1 - \sum_{k=2}^{p} a_{i_k})/(1 + \sum_{k=2}^{p} b_{i_k}). \tag{A.8}$$

From (A.3), $\lambda = A_{i_1}(2w_{i_1} NP_{i_1} - 1)$ ; $\mu_j = -\lambda - A_j,\ for\ j = 1 \sim n,\ j \neq i_1, i_2, \cdots i_p$. Finally, the values of $\boldsymbol{\mu}$ and **NP** must be checked that whether they satisfy (A.2) and (A.6), respectively. If satisfied, then this set of solution is valid.

In fact, there are total $(2^n - 1)$ situations (excluding the situation which all **NP** equal to 0). For each situation, check whether the solution satisfies the KKT condition or not. Because this function is a quadratic equation, the solution which satisfies the KKT condition must be the optimal solution.

# Bibliography

[1] E. Stevens-Navarro, Y. Lin, and V. W. S. Wong, "An MDP-based vertical handoff decision algorithm for heterogeneous wireless networks," *IEEE Transactions on Vehicular Technology*, Vol. 57, no. 2, pp. 1243-1254, 2007.

[2] W. Song, H. Jiang, W. Zhuang, and X. Shen, "Resource management for QoS support in cellular/WLAN interworking," *IEEE Network*, Vol. 19, no. 5, pp. 12-18, Sept.-Oct. 2005.

[3] Y. H. Chen, N. Y. Yang, C. J. Chang, and F. C. Ren, "A utility function-based access selection method for heterogeneous WCDMA and WLAN networks," *IEEE 18th International Symposium on PIMRC-Sept. 2007*, pp. 1-5.

[4] F. Bari and V. C. M. Leung, "Automated network selection in a heterogeneous wireless network environment," *IEEE Network*, Vol. 21, no. 1, pp. 34-40, Jan.-Feb. 2007.

[5] O. Yilmaz, A. Furuskar, J. Pettersson, and A. Simonsson, "Access selection in WCDMA and WLAN multi-access networks," *IEEE VTC-Spring 2005*, Vol. 4, pp. 2220-2224.

[6] J. Chen, K. Yu, G. Yang, Z. Y. Feng, Y. Ji, P. Zhang, Q. Huang, Y. Bai, L. Chen, and M. Minomo, "An ecology-based adaptive network control scheme for radio resource management in heterogeneous wireless networks," *BIMNICS-Dec. 2006*, pp. 1-5.

[7] G. Ning, G. Zhu, L. Peng, and X. Lu, "Load balancing based on traffic selection in heterogeneous overlapping cellular networks," *The First IEEE and IFIP International Conference in CANET*, Sept. 2005.

[8] G. Ning, G. Zhu, Q. Li, and R. Wu, "Dynamic load balancing based on sojourn time in multitier cellular systems," *IEEE VTC*-Spring 2006, Vol. 1, pp. 111-116.

[9] A. Dixit and S. Skeath, "Games of strategy," *New York:W.W. Norton*, 2004.

[10] J. Antoniou and A. Pitsillides, "4G Converged Environment: Modeling Network Selection as a Game," *IEEE Proceeding on IST Conf.*, pp. 1-5, July 2007.

[11] D. Niyato and E. Hossain, "A cooperative game framework for bandwidth allocation in 4G heterogeneous wireless networks," *IEEE ICC-June 2006*, Vol. 9, pp. 4357-4362.

[12] M. Pulido, J. S. Soriano, and N. Llorca, "Game theory techniques for university management: An extended bankruptcy model," *Operation Research*, Vol. 109, pp. 129-142, 2002.

[13] D. Niyato, and E. Hossain, "A noncooperative game-theoretic framework for radio resource management in 4G heterogeneous wireless access networks," *IEEE Transactions on Mobile Computing*, Vol. 7, no. 3, pp. 332-345, March 2008.

[14] 3GPP TR 23.271, "Functional stage 2 description of Location Services (LCS)," *3rd Generation Partnership Project, Tech. Rep.*, Sep. 2007.

[15] H. Holma and A. Toskala, "WCDMA for UMTS," *John Wiley and Sons, Ltd*, 2002.

[16] IEEE Std. 802.16e, "IEEE standard for local and metropolitan area networks part 16: air interface for fixed broadband Wireless access systems amendment for physical and medium access control layers for combined fixed and mobile operation in licensed bands," Oct. 2005.

[17] A. J. Goldsmith and S. G. Chua, "Variable-rate variable-power MQAM for fading channels," *IEEE Transactions on Communications*, Vol. 45, pp.1218-1230, Oct. 1997.

[18] IEEE Std. 802.11e/D12.0, "IEEE Standard for Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: Medium Access Control (MAC) Enhancements for Quality of Service (QoS)," Nov. 2004.

[19] 3GPP TR 25.892, "Feasibility study for OFDM for UTRAN enhancement," *3rd*

*Generation Partnership Project*, *Tech. Rep.*, 2004-06.

[20] L. Gordon, "Principle of Mobile Communication," *Kluwer Academic*, 1958

[21] 3GPP TR 23.107, "QoS concept and architecture release 6," *3rd Generation Partnership Project, Tech. Rep.*, March 2004.

[22] Universal Mobile Telecommunication System, "Selection procedures for the choice of radio transmission technologies of the UMTS", UMTS Std. 30.03, 1998.

[23] R. Guerin, H. Ahmadi, and M. Naghshineh, "Equivalent capacity and its application to bandwidth allocation in high-speed networks," *IEEE Journal on selected Areas in Communications*, Vol. 9, pp. 968-981, Sep. 1991.

[24] E. K. P. Chong and S. H. Zak, "An introduction to optimization," *John Wiley and Sons, Inc*, 2001.

# Vita

Tsung-Li Tsai was born in 1984 in Changhua, Taiwan. He received the B.E. degree in electrical engineering from National Cheng-Kung University, Tainan, Taiwan, in 2006, and the M.E. degree in the department of communication engineering, college of electrical and computer engineering from National Chiao Tung University, Hsinchu, Taiwan, in 2008. His research interests include radio resource management and wireless communication.