國立交通大學

資訊科學與工程研究所

碩士論文

WiMAX 環境下以停留時間為基礎的
ASN-GW 重新定錨演算法

A Residence-Time-Based ASN-GW Relocation Algorithm for WiMAX Networks

研 究 生：楊子興

指導教授：簡榮宏　教授

中華民國 九十八年六月

WiMAX 環境下以停留時間為基礎的 ASN-GW 重新定錨演算法
A Residence-Time-Based ASN-GW Relocation Algorithm for WiMAX Networks

研究生：楊子興	Student：Tzu-Hsing Yang
指導教授：簡榮宏	Advisor：Rong-Hong Jan

國立交通大學
資訊科學與工程研究所
碩士論文

A Thesis
Submitted to Institute of Computer Science and Engineering
College of Computer Science
National Chiao Tung University
in partial Fulfillment of the Requirements
for the Degree of
Master
in

Computer Science

June 2009

Hsinchu, Taiwan, Republic of China

中華民國九十八年六月

WiMAX 環境下以停留時間為基礎的 ASN-GW 重新定錨演算法

研究生：楊子興　　　　指導教授：簡榮宏 博士

國立交通大學資訊科學與工程研究所

摘　　要

在 WiMAX 點對點網路架構（WiMAX End-to-End Network Architecture）中，為了降低換手（Handover）延遲，當行動裝置（Mobile Station）在存取服務網路（Access Service Network）以及基地台（Base Station）之間移動時，會使用存取服務網路換手（ASN anchored mobility）。然而在行動裝置都不更換定錨點的狀況下，存取服務網路換手會導致存取服務網路閘道器（ASN Gateway）的負載大幅增加。因此，行動裝置需要一個機制來決定何時執行重新定錨（Relocation），以及決定需要重新定錨的行動裝置的數量及對象。在本篇論文中，我們提出了一個以平均停留時間為基礎的存取服務網路重新定錨演算法，在適當的時機針對適合的行動裝置進行重新定錨。模擬的結果顯示出，我們提出的演算法可以有效的降低與穩定存取服務網路閘道器的負載情形，並且降低了整個系統在進行重新定錨的負擔。

A Residence-Time-Based ASN-GW Relocation Algorithm for WiMAX Networks

Student：Tzu-Hsing Yang Advisor：Dr. Rong-Hong Jan

INSTITUTE OF COMPUTER SCIENCE AND ENGINEERING

NATIONAL CHIAO TUNG UNIVERSITY

Abstract

In the WiMAX End-to-End Network Architecture, the mobile station (MS) performs Access Service Network (ASN) anchored mobility to lower handoff latency when it moves between base stations (BSs) and ASNs. However, the ASN anchored mobility will induce a heavy load of ASN gateway (ASN-GW). Thus, relocating anchored ASN-GW for the MS is needed. It is a challenge for system designers to decide when to perform ASN-GW relocation, how many MSs should be relocated and which MSs should perform relocation. In the thesis, we propose an average residence time based (ART-based) ASN-GW relocation algorithm to determine an appropriate time and select a set of MSs to relocate their anchored ASN-GWs. The simulation results show that the ART-based ASN-GW relocation algorithm can reduce and steady the loads of ASN-GWs and decrease the cost for performing relocations.

致謝

在這兩年的研究所生活，首先感謝我的指導教授簡榮宏老師，感謝老師悉心的教導及提供的完善資源，時常的討論並找出論文正確的方向，使我在這兩年中獲益匪淺；感謝老師的大力協助，使得我的論文能夠更完整而嚴謹。

兩年裡的日子中，實驗室裡的學長姐(世昌、安凱、鈺翔、嘉泰、蕙如、奇育、家瑋、文彬)、同學(志賢、淑盈)以及學弟妹們(宇田、嘉瑋)的共同努力以及共同生活的點點滴滴，不管是在學術上的討論或者各種活動的共同參與，都讓我的這兩年的研究生生活變得絢麗多彩。

最後感謝我的家人及朋友，感謝你們在背後強力的支持以及鼓勵我，這是讓我能夠持續努力的最大動力。感謝你們的體諒、包容與關懷，僅以此文獻給我最摯愛的家人和朋友們。

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The IEEE 802.16 [1]-[2], a newly developed broadband wireless communication technology, provides broadband wireless services for last mile access. It is also known as World Interoperability Microwave Access (WiMAX). The IEEE 802.16d [1] defined the interface of air for fixed broadband wireless access systems. The IEEE 802.16e [2] enhances the IEEE 802.16d by supporting the mobility. However these IEEE 802.16 standards only specify physical (PHY) layer and Media Access Control (MAC) layer. To build a complete network system, WiMAX forum [3] proposes an network system architecture, called WiMAX End-to-End Network Systems Architecture, to standardize higher layer specifications for WiMAX [4].

Figure 1 shows the WiMAX End-to-End Network Systems Architecture. The architecture consists of two kinds of networks, *Access Service Network* (ASN) and *Connectivity Service Network* (CSN). Each ASN has one or more ASN gateways (ASN-GWs) and base stations (BSs) which connect with the ASN-GW. Note that the ASN provides radio access to WiMAX subscribers. CSN consist of a set of servers which provide Internet Protocol (IP) connectivity services. These servers include AAA server, DHCP server and home agent (HA). Besides, a reference point is defined as a conceptual point between two network components. The reference point R3 refers to the interface between ASN and CSN and the reference point R4 is between two ASN-GWs.

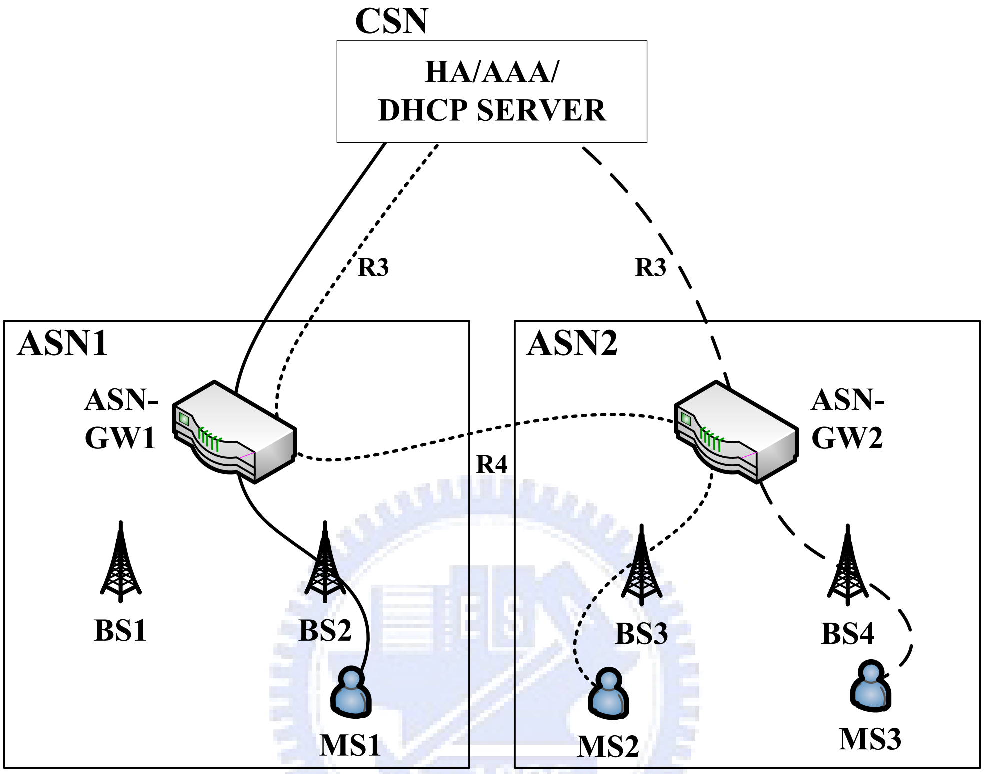


Figure 1. End-to-End Network Systems Architecture

WiMAX Forum adopts Mobile IP [5] to support IP mobility. The home agent (HA) of a mobile station (MS) is located in CSN which is operated by the MS's home network service provider. The functionality of foreign agent (FA) is supported by ASN-GW. WiMAX Forum defines two mobility management methods, *ASN anchored mobility* and *CSN anchored mobility* for the WiMAX End-to-End Network System Architecture. ASN anchored mobility management is defined as mobile station movement between BSs and ASNs without changing the reference point R3. CSN anchored mobility management uses the Mobile IP protocol to construct new reference point R3 between CSN and the target ASN.

For example, suppose that an MS connects to WiMAX network to access a web server, called a correspondent node (CN), in the Internet. A connection is set up between MS, BS2, ASN-GW1, CSN and CN (see Figure 2). We denoted this connection as Flow 1 in Figure 2. Later, the MS may move from BS2 to BS3. An inter-ASN handoff occurs between ASN1 and ASN2. In order to lower handoff latency, the MS performs the ASN anchored mobility instead of CSN anchored mobility. A data tunnel will be established between ASN-GW1 (anchor ASN-GW) and ASN-GW2 (serving ASN-GW). After that the new connection, as Flow 2 in Figure 2, is set up between MS, BS3, ASN-GW2, ASNGW1, CSN and CN. Note that end-to-end delay between CN and MS might be longer and the load of ASN-GW1 will become very heavy if there are many MSs anchored at ASN-GW1.

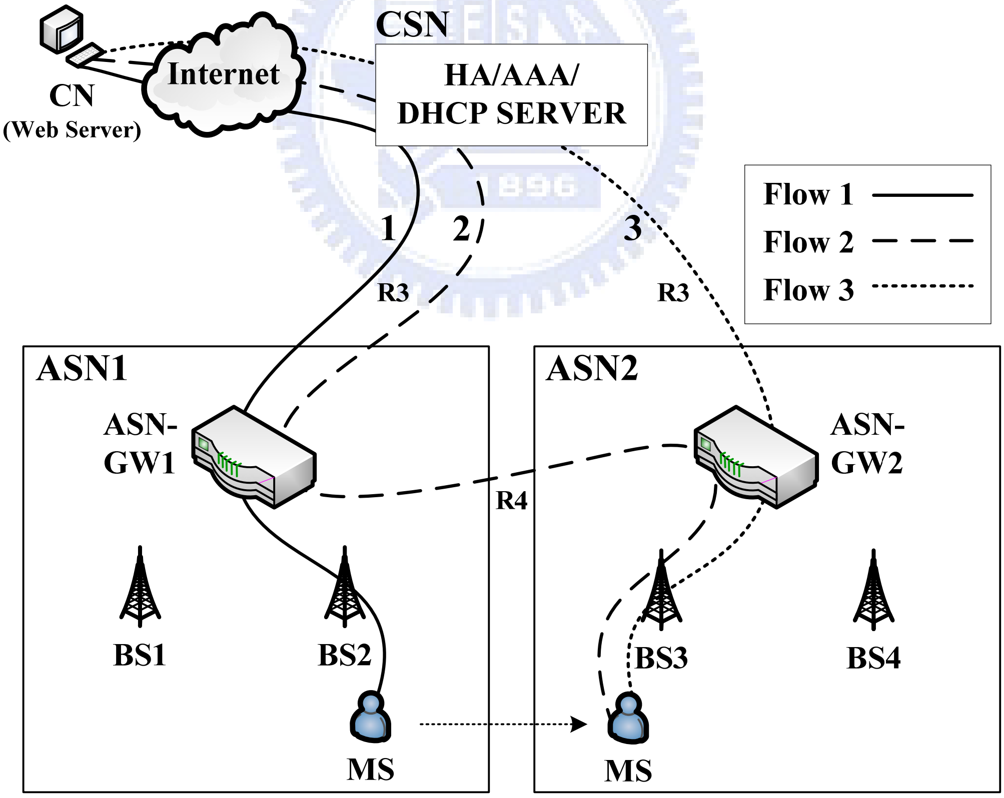


Figure 2. Mobility in WiMAX End-to-End Architecture

The following two methods can deal with the above-mentioned situation: (1) the system asks some MSs to perform CSN anchored mobility if the load of ASN-GW1 becomes heavy. This refers as ASN-GW relocation. A tunnel between CSN and ASN2 (i.e. the tunnel between HA in CSN and FA in ASN-GW2) is established and two tunnels between CSN and ASN1, and between ASN-GW1 and ASN-GW2 are removed. After that a new connection (between MS, BS3, ASN-GW2, CSN and CN, i.e. Flow 3 in Figure 2) is set up. (2) The MS performs CSN anchored mobility instead of ASN anchored mobility whenever the handoff occurs. However, ping-pong phenomenon might occur if performing CSN anchored mobility directly. For example, as shown in Figure 3, suppose that the MS which anchored at ASN-GW1 leaves to ASN2 just for a very short visit, and then the MS moves back to ASN1 for a long stay. If the MS performs CSN anchored mobility directly, then the connection as showing in fig. 3(a) will switch to fig. 3(b) shortly and switch back to fig. 3(c) which is the same connection to fig. 3(a). The CSN anchored mobility is performed twice, but in the end the MS still anchored at ASN-GW1. The tunnel between CSN and ASN-GW2 is shortly used. If the MS performs ASN anchored mobility, as Figure 4 illustrating, the tunnel between CSN and ASN-GW1 is always used. In this case, performing ASN anchored mobility is better because the overhead and handoff delay for ASN anchored mobility are less than that for CSN anchored mobility.

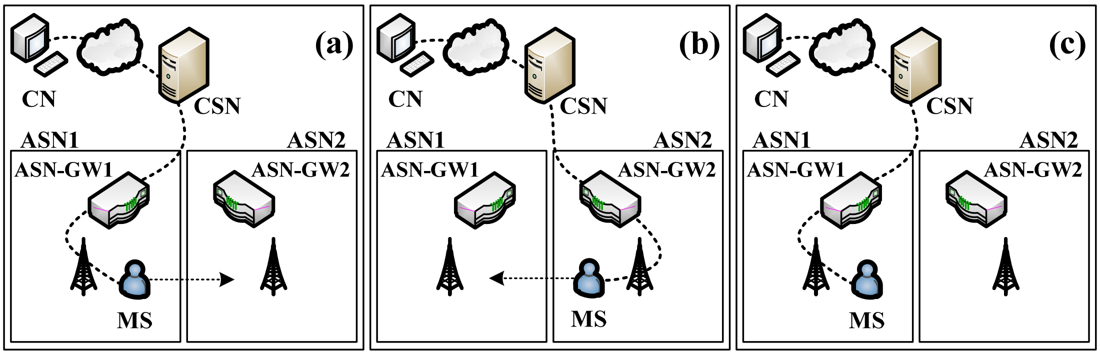


Figure 3. An example with CSN Anchored Mobility

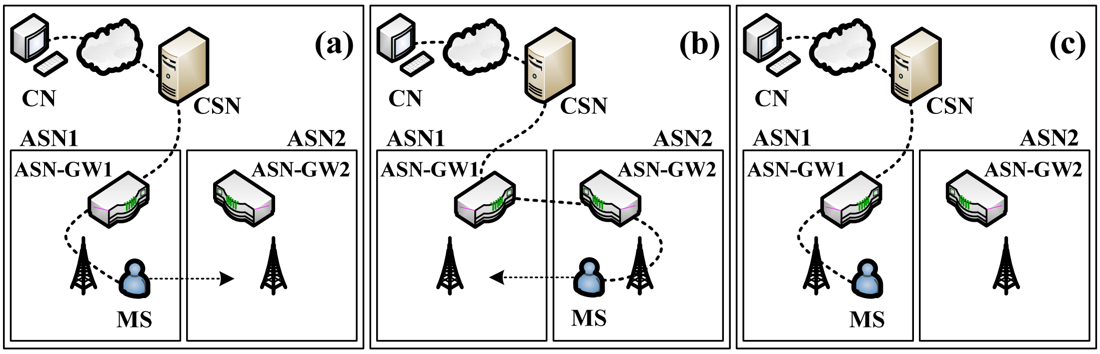


Figure 4. An example with ASN Anchored Mobility

In this thesis, we consider the following two mobility management problems for the WiMAX End-to-End Network Systems:

Problem 1: How to choose a suitable mobility (ASN or CSN anchored mobility) for MS when a handoff occurs.

Problem 2: As the example we mentioned, if the ASN anchored mobility handoff is performed, then the number of MSs anchored with old ASN-GW does not reduce. The load of the old ASN-GW will grow. Thus, the system has to perform ASN-GW relocation to release the load of the old ASN-GW. However, WiMAX standards only define the ASNGW relocation procedure. It does not specify when to perform ASN-GW relocation, how many MSs should be relocated and which MSs should perform relocation.

For Problem 1, this thesis proposes an average residence time based (ART-based) method to determine a suitable mobility (ASN or CSN anchored mobility) for MS. For Problem 2, the ART-based method is applied to determine when to perform ASN-GW relocation, how many MSs should be relocated and which MSs should perform relocation in this thesis. The simulation results show that the ART-based method decreases the load of ASN-GWs, and it has lower number of relocations and lower number of MSs that ASN-GW served.

The rest of this thesis is organized as follows. In chapter 2 the related work is reviewed. The ART-based ASN-GW relocation algorithm is discusses in chapter 3, and chapter 4 presents the simulation results. Finally, chapter 5 gives the conclusion.

Chapter 2

Related Work

Several studies have been proposed for load control or gateway relocation in IP-based mobile networks or cellular networks. In Mobile IP, a bottleneck may occur at HA if HA is serving a large number of MSs. Many load control mechanisms [6]-[9] have been presented to balance the load among all HAs in home network. These load balance methods are helpful for Mobile IP protocol but they can not apply to ASN GW relocation in WiMAX networks. This is because the role of HA is differs greatly from the ASN-GW's.

In [10], a mobility-based load control mechanism is proposed for Mobility Anchor Point (MAP) to avoid getting overloaded in Hierarchical Mobile IP network. If a new MS is going to anchor at a loaded MAP, the MAP will redirect one MS with the Session-to-mobility ratio (SMR) greater than a given threshold to its HA where SMR a ratio of the session arrival rate to the handoff rate. Otherwise, if there is not any MS with SMR greater than the SMR threshold, the new entry MS will be rejected to anchor. Therefore, the mobility-based load control mechanism controls the load of the MAP and relocates MS to HA. The MAP in Hierarchical Mobile IP network is similar ASN-GW in WiMAX networks. However, we cannot relocate the MS to its HA in

WiMAX network.

For Universal Mobile Telecommunications System (UMTS), the serving radio network controller (SRNC) relocation is studied in [11]. When an MS hands off to the BS under a new RNC (target RNC), the source RNC initiates SRNC relocation. Hence, the movement of MS can trigger the source RNC to perform SRNC relocation. However, in WiMAX networks, the ANS-GW relocation cannot be initiated only by depending on the mobility of MS because the ping-pong phenomenon might occur seriously.

Recently, two ASN-GW relocation algorithms, non-predictive and predictive algorithms, for WiMAX networks are proposed in [12]. The algorithms are designed to determine when to perform ASN-GW relocation and how many anchored MSs should be relocated. As the figure 5 illustrating, MS1 is a serving MS of ASN-GW2 and it is an anchored MS of ASN-GW1. That is, an anchored MS of ASN-GW is an MS which is anchored at the ASN-GW but serving by other ASN-GW. In [12], the ASN-GW relocation is initiated by only considering about the load of anchored MSs. If an ASN-GW with high load of serving MSs but with low load of anchored MSs, the algorithms would not perform ASN-GW relocations because the load of anchored MSs is low. Thus, the load of this ASN-GW always keeps high. The other problem is that the algorithms only determine how many anchored MSs should be relocated. They do not determine which anchored MSs should be selected to relocate. As a result, ping-pong phenomenon may occur and the algorithms will perform many unnecessary ASN-GW relocations.

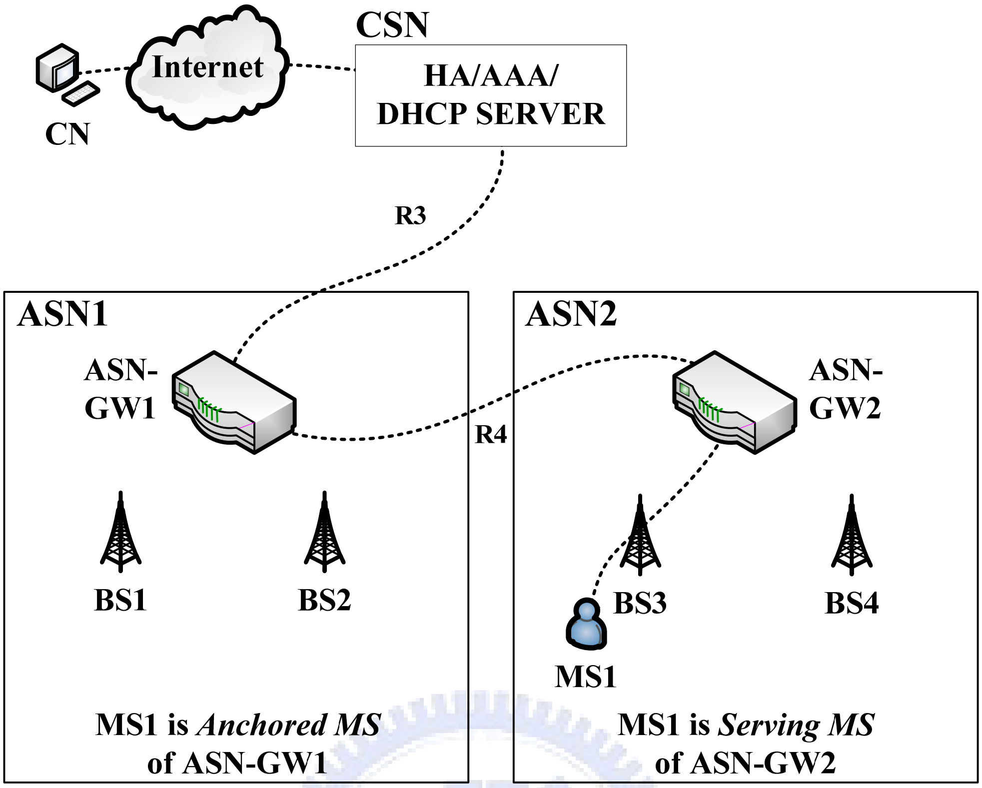


Figure 5. Anchored and Serving MS in the WiMAX End-to-End Architecture

Chapter 3

Residence-Time-Based ASN-GW Relocation Algorithm

In this chapter, we propose an average residence time based (ART-based) method in section 3.1 to determine suitable mobility for MS. It is the solution for the problem 1 mentioned in chapter 1. For problem 2, we propose a relocation strategy in section 3.2 to determine which MS should be relocated, when to perform relocation, and how many MS should change their anchored ASN-GW.

3.1 Mobility Selection

One of the most important ideas in the proposed method is average residence time (ART). ASN-GW records the time at which each serving MS arrives and the time at which the serving MS leaves. From two times, ASN-GW computes an elapsed time called as a sample residence time (SRT). Whenever it obtains a new sample residence time, ASN-GW adjusts the ART for the MS. Usually, ASN-GW stores the ART as a weighted average and uses new sample residence times to change the average slowly. That is, the new ART can be computed by

$$T_i = \alpha \times T_{i-1} + (1-\alpha) \times S_i$$

where T_i is the new ART, T_{i-1} is the old ART, S_i is the new sample residence time, and $0 < a < 1$.

Consider that a WiMAX network having ASN-GWs, $G_1,\ G_2,\ \cdots\ ,G_m$, and MSs $M_1,\ M_2,\ \cdots\ ,M_n$. Let $T(a, b)$ be the current ART for M_b maintained in ASN-GW G_a. We define an ART threshold B_a for each G_a, $a = (1,\ 2,\ \ldots,\ m)$ as

$$B_a = T_{\min}^{(a)} + \beta \times \left| T_{\max}^{(a)} - T_{\min}^{(a)} \right|$$

where $T_{\min}^{(a)} = \min\{\, T(a,\ b) \mid b = 1,\ 2,\ \cdots,\ n \,\}$, $T_{\max}^{(a)} = \max\{\, T(a,\ b) \mid b = 1,\ 2,\ \cdots,\ n \,\}$, and $\beta\ (0 \le \beta \le 1)$ is a constant weighting factor. When the MS M_j performs the inter-ASN handoff from G_s to G_t, the ASN-GW G_t checks to see if M_j's ART $T(t, j)$ is greater than B_t or not. If the answer is yes, then the ASN-GW G_t chooses CSN anchored mobility for MS M_j. Otherwise, the ASN-GW G_t chooses ASN anchored mobility for MS M_j. Choosing a value for weighting factor can be difficult. If the B_a value is close to $T_{\min}^{(a)}$ (i.e. β = 0), then the ASN-GW G_t always chooses CSN anchored mobility for MS M_j. This will cause ping-pong phenomenon which wastes ASN-GWs' resources and network bandwidth. If $B_a = T_{\max}^{(a)}$ (i.e. β = 1) MSs always perform ASN anchored mobility and ASN-GWs' load will grow quickly. This thesis recommends setting $\beta = 0.75$.

There is an example of a WiMAX network with two ASN-GWs, G_s and G_t and two MSs M_1 and M_2 as shown in Figure 6. The current ARTs $T(a, b)$, a = s, t, and b = 1, 2, 3, are showing at Table 1.

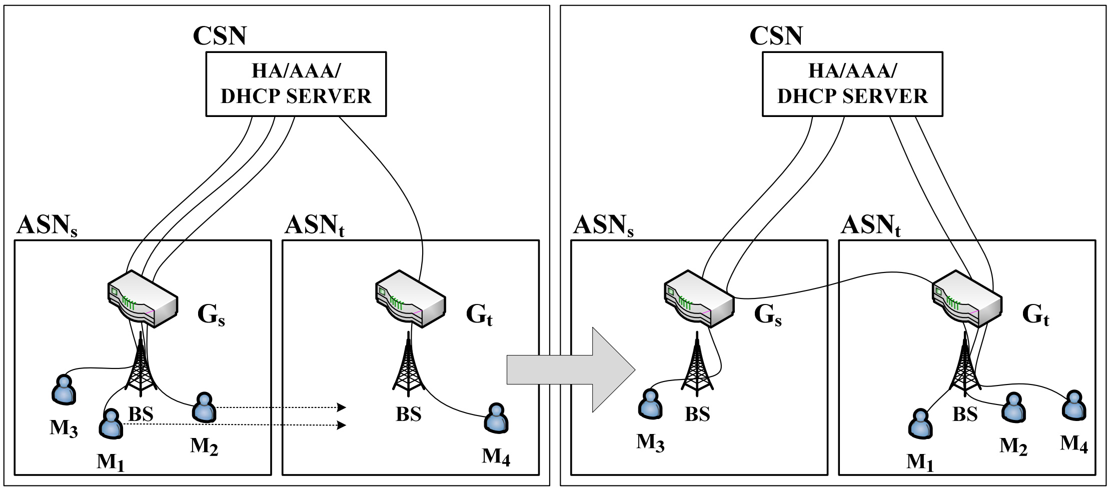


Figure 6. An example for suitable mobility selection (with $\beta = 0.75$)

Table 1. Current ARTs for the example of Fig. 6

T(a, b)	M_1	M_2	M_3	M_4
G_s	80	30	95	5
G_t	40	75	10	90

We give $\beta = 0.75$ and the B_a for G_s and G_t are as follows.

$$B_s = T_{\min}^{(s)} + \beta \times \left| T_{\max}^{(s)} - T_{\min}^{(s)} \right| = 5 + 0.75 \times |95 - 5| = 72.5$$

$$B_t = T_{\min}^{(t)} + \beta \times \left| T_{\max}^{(t)} - T_{\min}^{(t)} \right| = 10 + 0.75 \times |90 - 10| = 70$$

M_1 has the ART at ASN$_\text{t}$ $T(t,1) = 40$ smaller than B_t. When M_1 handoff from ASN$_\text{s}$ to ASN$_\text{t}$, M_1 will use ASN anchored mobility and it still anchored at G_s. On contrary, M_2 has the ART at ASN$_\text{t}$ $T(t,1) = 75$ greater than B_t. Thus, M_2 will use CSN anchored mobility when it moves from ASN$_\text{s}$ to ASN$_\text{t}$. M_2 will anchor at G_t after the movement. However, if we give $\beta = 0.95$ and 0.25 the B_t is re-calculated as follows.

$$(\beta = 0.95)\ \ B_t = T_{\min}^{(t)} + \beta \times \left| T_{\max}^{(t)} - T_{\min}^{(t)} \right| = 10 + 0.95 \times |90 - 10| = 86$$

$$(\beta = 0.25)\ \ B_t = T_{\min}^{(t)} + \beta \times \left| T_{\max}^{(t)} - T_{\min}^{(t)} \right| = 10 + 0.25 \times |90 - 10| = 30$$

If we set β as 0.95, the B_t is close to $T_{\max}^{(t)}$. When M_1 and M_2 moves from ASN_s to ASN_t, both of them will use ASN mobility. On contrary, B_t is close to $T_{\min}^{(t)}$ when β set as 0.25. Both of M_1 and M_2 has ART greater than B_t. After they handoff from ASN_s to ASN_t, they will use CSN anchored mobility and change their anchored ASN-GW from G_s to G_t.

3.2 Relocation Strategy

In this section, we propose a relocation strategy. The MS selection method is defined in section 3.2.1 and when to perform relocation is given in section 3.2.2.

3.2.1 Mobile Station Selection

Consider an example of a WiMAX network with three ASN-GWs, G_1, G_2, G_3 and eight MSs $M_1, M_2, \ldots, M_8$ as shown in Figure 7. The sets of the serving MSs for G_1, G_2 and G_3 are $\{M_1, M_2\}$, $\{M_3, M_4, M_5\}$ and $\{M_6, M_7, M_8\}$, respectively. In addition, G_2 has the anchored MS set $A_2 = \{M_2, M_6, M_7\}$. The current ARTs $T(a, b)$, $a = 1, 2, 3$, $b = 1, 2, \ldots, 8$, are showing at Table 2.

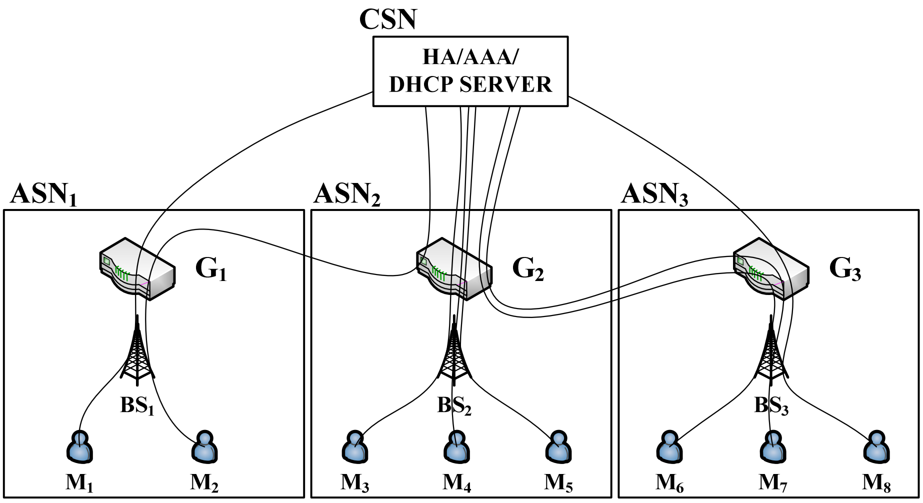


Figure 7. An example of WiMAX network with three ASN-GWs

Table 2. ARTs for the example of Fig. 7

$T(a, b)$	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8
G_1	25	75	5	10	10	15	5	10
G_2	50	15	90	95	60	50	40	15
G_3	35	5	10	10	10	10	80	45

Suppose that the load of G_2 is too heavy. The system has to perform ASN-GW relocation to release the load of the G_2. The problem is which MSs should be selected from the anchored MS set A_2 to perform relocation. In order to avoid the ping-pong phenomenon, we consider the proportion R_{M_j} of ART at serving ASN-GW to the ART at anchored ASN-GW for each $M_j \in A_2$. This is, for the MS M_j anchored at G_s and serving by G_t, the proportion R_{M_j} is defined as

$$R_{M_j} = \frac{T(t,j)}{T(s,j)}$$

where $T(t, j)$ is the ART of M_j for serving ASN-GW G_t and $T(s, j)$ is the ART of M_j for anchored ASN-GW G_s. For example, the proportion values for each $M_j \in A_2$ are

$$R_{M_2} = \frac{T(1,2)}{T(2,2)} = \frac{75}{15} = 5$$

$$R_{M_6} = \frac{T(3,6)}{T(2,6)} = \frac{10}{50} = 0.2$$

$$R_{M_7} = \frac{T(3,7)}{T(2,7)} = \frac{80}{40} = 2$$

The large value of R_{M_j}, e.g. $R_{M_2} = 5$, means that the MS M_j may stay at the serving ASN-GW much longer than stay at the anchored ASN-GW. Thus, it should be selected with high priority to perform relocation. Contrary, if R_{M_j}, e.g. $R_{M_6} = 0.2$, is very small, it represents the MS M_j may move back to its anchored ASN-GW immediately. The ASN-GW should avoid performing relocation for anchored MSs with R_{M_j} is very small. Thus, we use R_{M_j} as a criterion to determine if M_j is selected to relocate or not.

3.2.2 Relocation Timing Selection

In order to reduce the load of ASN-GWs and avoid ASN-GWs performing too many relocations, selecting a proper time to perform relocation is essential for ASN-GWs. We define a measurement for the ASN-GW's load which is modified from [12]. This measurement is based on drop rate. We apply Random Early Detection (RED) [13] to mark and drop packets. If the queue of an ASN-GW is not full but the traffic of the ASN-GW is getting heavy, we mark some arrived packets randomly as same as RED. While the queue is full, we randomly drop those marked packets and new arrived packets. At every time interval i, we find a drop rate $r_i \ (0 \le r_i \le 1)$ as follows

$$r_i = \frac{d_i}{f_i}$$

where d_i is the number of marked and dropped packets during time interval i and f_i is the number of received packets during time interval i. Thus, the load of an ASN-GW at time interval k can be defined as a weighted moving average [14] of $r_i \ (i = k,\ k-1,\ \cdots,\ k-(h-1))$ as follows.

$$L_k = \frac{\sum_{i=k-(h-1)}^{k} w_i r_i}{\sum_{i=k-(h-1)}^{k} w_i}$$

where w_i is a weighting factor of r_i and $w_k > w_{k-1} > \ldots > w_{k-(h-1)}$. According to L_k, the ASN-GWs are able to know the current load condition. Note that the load L_k will fall into interval [0, 1]. The older drop rate records may have effect on the load of ASN-GW. Thus, we define a constraint for L_k, h which is the maximum number of latest drop rate records should be taken to calculate L_k.

As Figure 8 showing, although the ASN-GW already records drop rates of each 6 interval, only 5 records are used during the calculating of the L_k of the 6th interval. The record of first interval is not used because the h of the ASN-GW is given as 5. In this way, the L_k can represent the current load condition more actually.

***h*=5**

k	**0**	**1**	**2**	**3**	**4**	**5**	**6**
r_i	0	0.03	0.05	0.08	0.15	0.18	0.24
w_i	--	--	2^1	2^2	2^3	2^4	2^5

$$L_6 = \frac{2^5 \times 0.24 + 2^4 \times 0.18 + 2^3 \times 0.15 + 2^2 \times 0.08 + 2^1 \times 0.05}{2^5 + 2^4 + 2^3 + 2^2 + 2^1} = 0.1965$$

Figure 8. An example of L_k calculating

This thesis defines three thresholds of load ℓ_L, ℓ_M and ℓ_H ($0 < \ell_L < \ell_M < \ell_H < 1$), and the ASN-GW divides the load condition into four states. The table 3 shows four states for which MSs should be relocated and what kind of mobility will be used by MSs. The operations ASN-GW will perform as follows.

Table 3. Four states relocation strategy

State	Relocation Target	Mobility Used
State 4 $\ell_M \le L_k < \ell_H$	Relocate all of Anchored MS ordered by R_{M_j}	CSN anchored Mobility
State 3 $\ell_M \le L_k < \ell_H$	Relocate Anchored MS with $R_{M_j} > R_{th}$	CSN Anchored Mobility ($L_{k-1} \le L_k$) ASN Anchored Mobility ($L_{k-1} > L_k$)
State 2 $\ell_L \le L_k < \ell_M$		ASN Anchored Mobility
State 1 $0 \le L_k < \ell_L$	No Relocations	

(1) **State 1** ($0 \le L_k < \ell_L$): At this state, the load of ASN-GW is light. No relocation is performed.

(2) **State 2** ($\ell_L \le L_k < \ell_M$): The load is medium. The ASN-GW has enough ability to cope with all of traffic generated by MSs. However, in order to avoid the load getting heavier, the ASN-GW selects the MS M_j with $R_{M_j} \le R_{th}$ from anchored MS set to perform relocation.

(3) **State 3** ($\ell_M \le L_k < \ell_H$): The load of the ASN-GW is heavy. The ASN-GW still selects the MS M_j with $R_{M_j} > R_{th}$ from anchored MS set to perform relocation. If the load of ASN-GW is increasing (i.e. $L_k \ge L_{k-1}$), the ASN-GW asks all MSs use CSN anchored mobility when they handoff to other ASN.

(4) **State 4** ($\ell_M \le L_k < \ell_H$): The load of the ASN-GW is very heavy. The ASN-GW may not be able to handle the traffic from all of mobile stations. In order to prevent the load of the ASN-GW keeping very heavy, the

ASN-GW has to perform relocation M_j ordered by the R_{M_j} from the anchored MS set. The relocation will continually perform until the load of ASN-GW lower than ℓ_M. The number of MSs which will be relocated at the interval k starts at 1. The number will double at every interval until the number reaches the maximum MS number of once relocation. In addition, the ASN-GW asks all MSs use CSN anchored mobility when they handoff to other ASN.

Figure 9 demonstrates the detail flow chart of the ART-Based Algorithm.

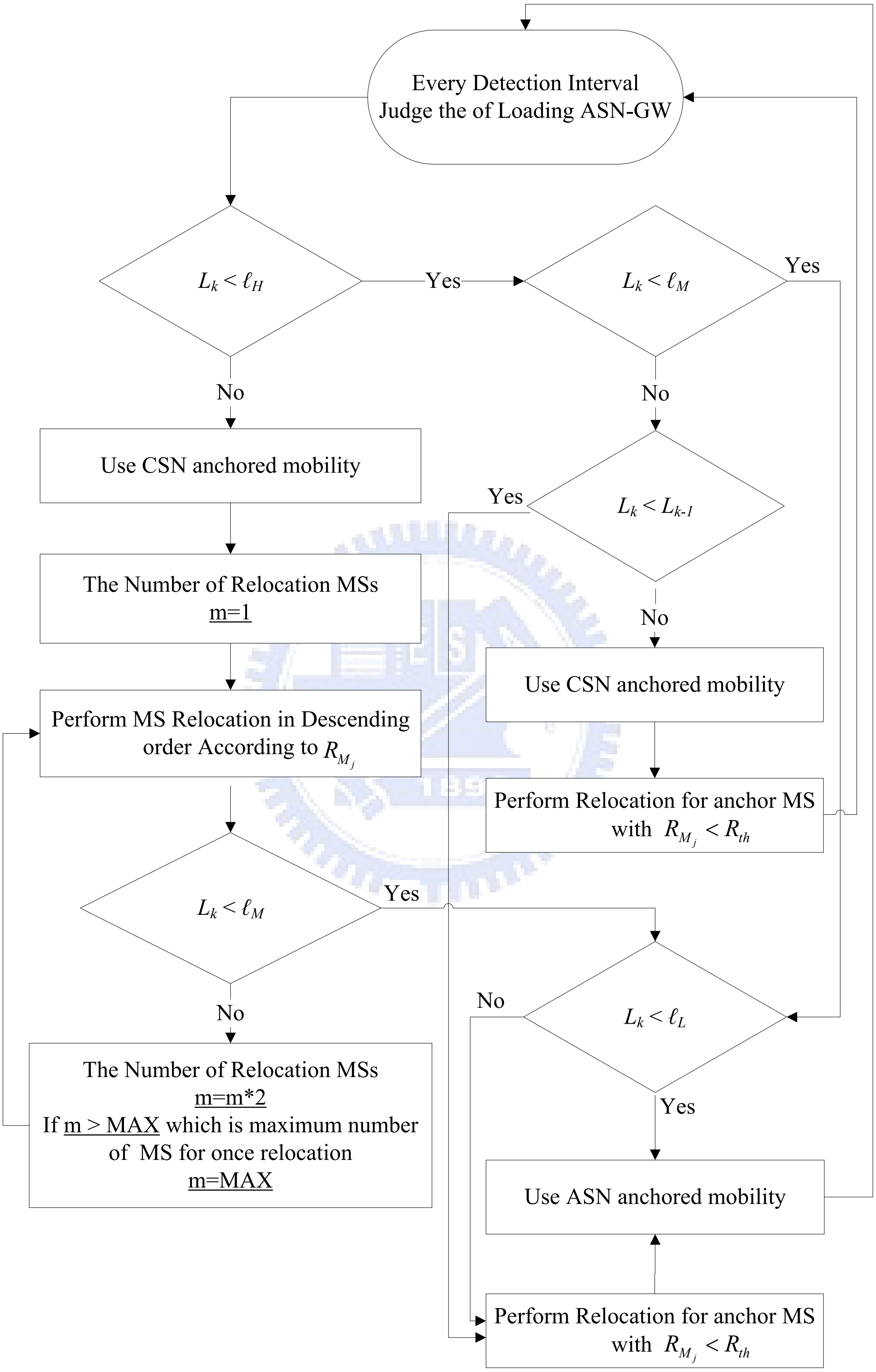


Figure 9. Average Residence Time Based Relocation Algorithm

Consider an example of a WiMAX network with three ASN-GWs, G_1, G_2, G_3 and six MSs $M_1, M_2, \ldots, M_6$ as shown in Figure 10. At the interval k, M_4 will handoff from G_2 to G_3.The ARTs $T(a, b)$ and R_{M_b}, $a = 1, 2, 3$, $b = 1, 2, \ldots, 8$, of time interval k are showing at Table 4.

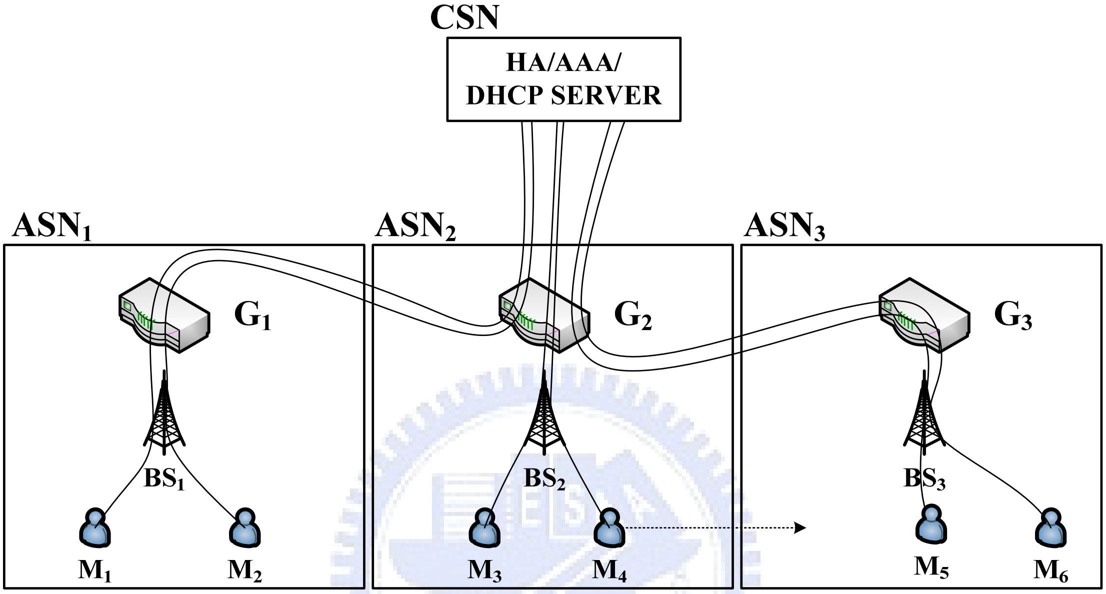


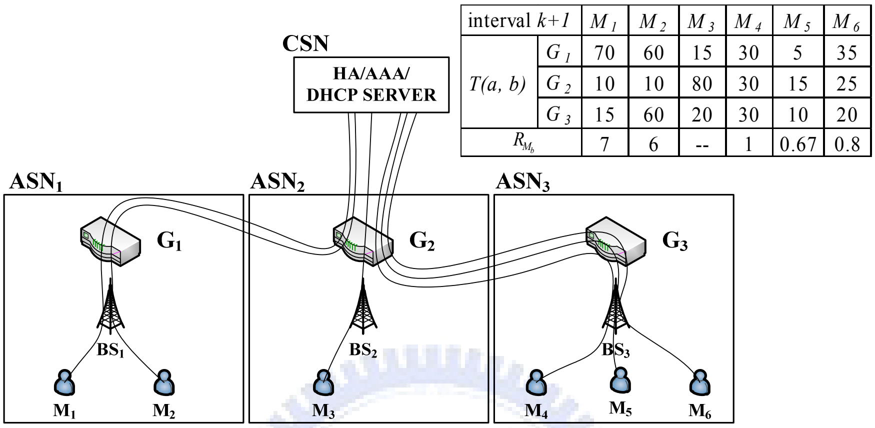
Figure 10. An example for 4 states relocation strategy

Table 4. ARTs and R_{M_b} of time interval k for the example of Fig. 10

interval k		M_1	M_2	M_3	M_4	M_5	M_6
$T(a, b)$	G_1	70	60	15	30	5	35
	G_2	10	10	80	30	15	25
	G_3	15	60	20	30	10	20
R_{M_b}		7	6	--	--	0.67	0.8

Although the network system has the same condition, if the load condition of the anchored ASN-GW is different, there will be different results at the next interval. In the following, we will show four outcomes of the network system after go through each four states of load condition with the same ARTs and movement at time interval *k*.

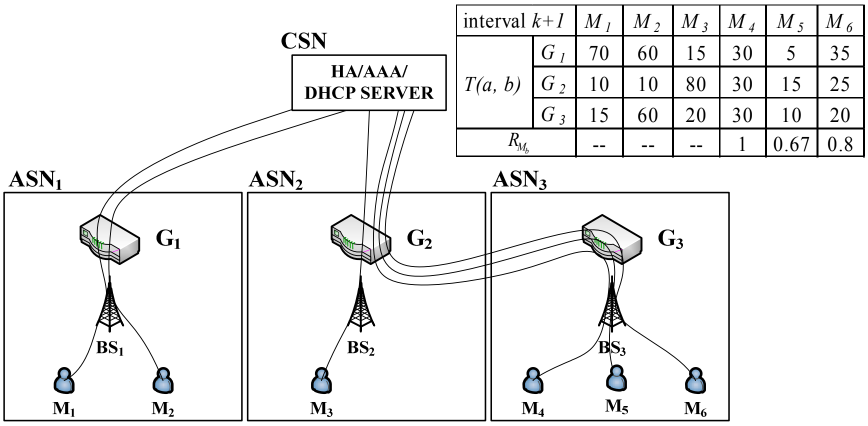
Suppose that the load of G_2 is at state 1 (i.e. $0 \le L_k < \ell_L$) at interval *k*. No relocations will occur and the M_4 uses ASN anchored mobility to handoff from G_2 to G_3. Figure 11 shows the system condition at interval *k+1* after the state 1 operation is performed at interval *k*.



interval *k+1*		M_1	M_2	M_3	M_4	M_5	M_6
T(a, b)	G_1	70	60	15	30	5	35
	G_2	10	10	80	30	15	25
	G_3	15	60	20	30	10	20
R_{M_b}		7	6	--	1	0.67	0.8

Figure 11. System condition at interval *k+1* after the state 1 operation

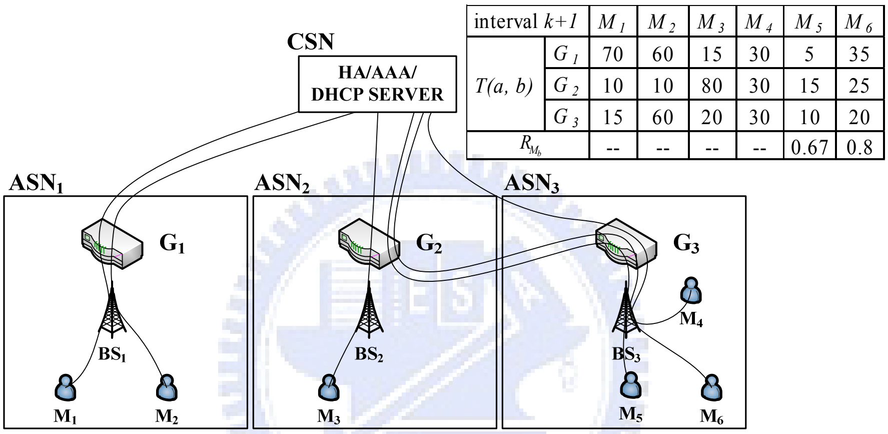
Suppose that the load of G_2 is at state 2 (i.e. $\ell_L \le L_k < \ell_M$) at interval *k*. G_2 selects the MS M_b with $R_{M_b} \le R_{th}$ from anchored MS set (i.e. { M_1, M_2, M_5, M_6}) to perform relocation. Note that we set R_{th} as 5. Thus, M_1 and M_1 has $R_{M_b} \le R_{th}$, and they will be relocated to G_1 at interval *k*. Besides, with load condition at state 2, M_4 uses ASN anchored mobility to handoff from G_2 to G_3. Figure 12 shows the system condition at interval *k+1* after the state 2 operation is performed at interval *k*.



interval *k+1*		M_1	M_2	M_3	M_4	M_5	M_6
T(a, b)	G_1	70	60	15	30	5	35
	G_2	10	10	80	30	15	25
	G_3	15	60	20	30	10	20
R_{M_b}		--	--	--	1	0.67	0.8

Figure 12. System condition at interval *k+1* after the state 2 operation

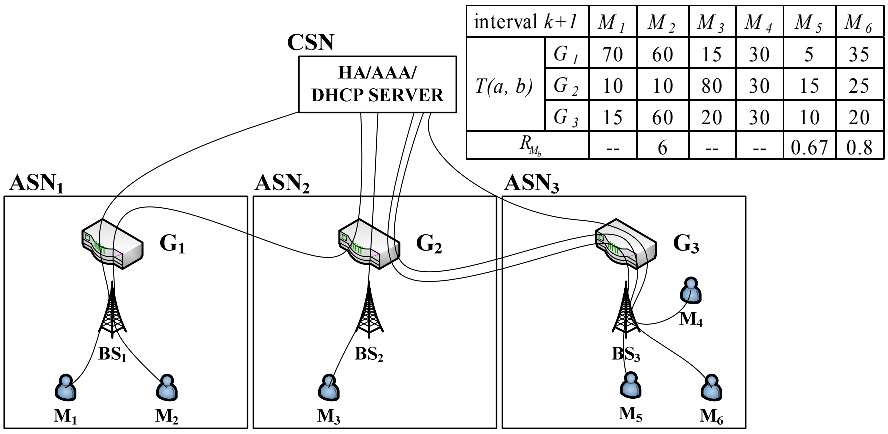
Suppose that the load of G_2 is at state 3 (i.e. $\ell_M \le L_k < \ell_H$) at interval *k* and $L_{k-1} \le L_k$. It is same as state 2 that G_2 selects the MS M_b with $R_{M_b} \le R_{th}$ from anchored MS set, and M_1, M_2 will be relocated to G_1 at interval *k*. Because the load condition at state 3 and it is increasing (i.e. $L_{k-1} \le L_k$), M_4 uses CSN anchored mobility to handoff from G_2 to G_3. Figure 13 shows the system condition at interval *k+1* after the state 3 operation is performed at interval *k*.



interval *k+1*		M_1	M_2	M_3	M_4	M_5	M_6
T(a, b)	G_1	70	60	15	30	5	35
	G_2	10	10	80	30	15	25
	G_3	15	60	20	30	10	20
R_{M_b}		--	--	--	--	0.67	0.8

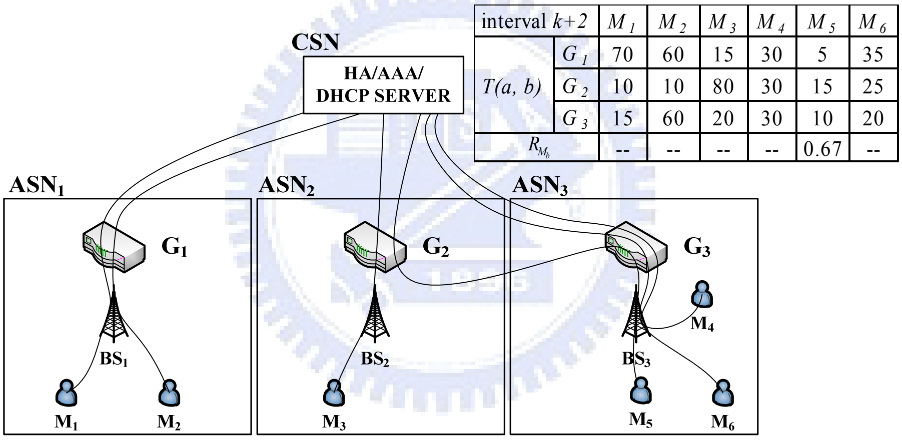
Figure 13. System condition at interval *k+1* after the state 3 operation

Suppose that the load of G_2 is at state 4 (i.e. $\ell_H \le L_k \le 1$) at interval *k*. G_2 selects the MS M_b according to their R_{M_b} from anchored MS set, and the relocation priority is $M_1 > M_2 > M_6 > M_5$ at interval *k*. The number of MS should be relocation starts at 1. Therefore, at interval *k*, only M_1 will be relocated. At this state, M_4 uses CSN anchored mobility to handoff from G_2 to G_3. Figure 14 shows the system condition at interval *k+1* after the state 4 operation is performed at interval *k*. Suppose that the load of G_2 is still at state 4 at interval *k+1*. The relocation priority at interval *k+1* will change as $M_2 > M_6 > M_5$ and the number of relocation will double as 2. Thus, two MSs (i.e. M_2 and M_6) will be relocated at interval *k+1*. Figure 15 shows the system condition at interval *k+2* after the state 4 operation is performed at interval *k+1*.



interval *k+1*		M_1	M_2	M_3	M_4	M_5	M_6
$T(a, b)$	G_1	70	60	15	30	5	35
	G_2	10	10	80	30	15	25
	G_3	15	60	20	30	10	20
R_{M_b}		--	6	--	--	0.67	0.8

Figure 14. System condition at interval *k+1* after the state 4 operation



interval *k+2*		M_1	M_2	M_3	M_4	M_5	M_6
$T(a, b)$	G_1	70	60	15	30	5	35
	G_2	10	10	80	30	15	25
	G_3	15	60	20	30	10	20
R_{M_b}		--	--	--	--	0.67	--

Figure 15. System condition at interval *k+2* after the state 4 operation

Chapter 4

Simulation

4.1 Simulation Environment

We conduct simulation results to evaluate the performance of the proposed relocation algorithm. Our simulations were conducted using the ns-2 simulator [15]. The topology of WiMAX End-to-End Network Architecture under test is shown in Figure 16. There are four subnets, and each subnet has one ASN-GW connected to a base station by a 10 Mbps link. We use Constant Bit Rate (CBR) to be our traffic model. Every one second, four new MSs arrives at each BS. There are 1000 MSs in the topology at most. Exponential distribution with mean from 5s to 100s is used to model the residence time of each MS. That is, the mean of residence time for each MS was generate from [5, 100] randomly. Uniform distribution is used to decide which ASN the mobile station is going to next. We observe the load of ASN-GWs and the number of relocation at each ASN-GW. According to the observations, we analyze the performance of different methods of mobility. Table 5 shows the experiment parameters. We compare our ART-based relocation method with the following three methods:

(1) **Pure R3 mobility**: The MS always performs the CSN anchored mobility for inter-ASN handoff.

(2) **Pure R4 mobility**: The MS always performs ASN anchored mobility for inter-ASN handoff and the ASN-GW does not perform relocation.

(3) **Non-predictive ASN-GW relocation method** [12]: The MS performs ASN anchored mobility for inter-ASN handoff. Then the ASN-GW performs non-predictive relocation when its load is heavy.

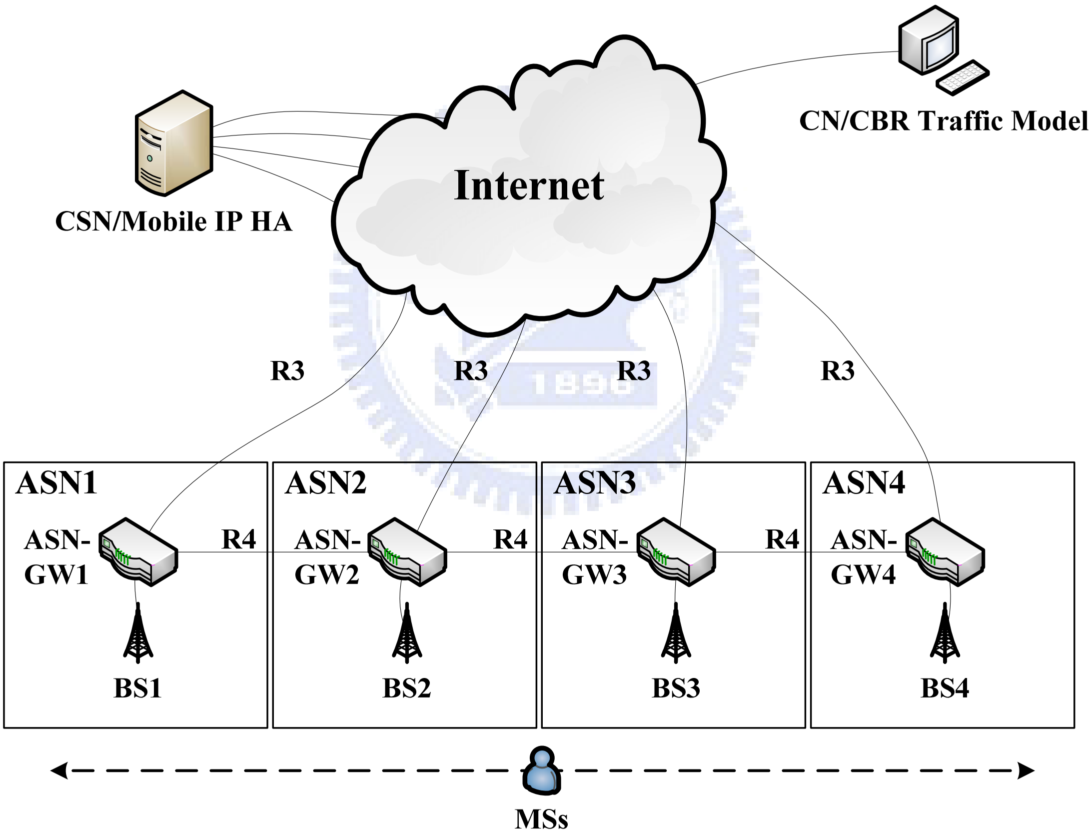


Figure 16. Simulation architecture

Table 5. Experiment parameters

MS's inter-arrival time	1 second
Maximum number of MS	1000
Experiment duration	2000 seconds
β	0.75
R_{th}	5
ℓ_H	0.60
ℓ_M	0.45
ℓ_L	0.30
Detection Interval	5 seconds
MAX Relocation MS numbers	16

4.2 Simulation Results

Figure 17 shows the load of each ASN-GW for 2000 seconds. From Figure 17, we found that the load of the pure R4 mobility is the heaviest among these mobility methods. This is because the pure R4 mobility method never changes MS's the anchor ASN-GW to release gateway's load. On the contrary, the load of the pure R3 mobility is the lowest. Using pure R3 mobility method, MSs always change their anchored ASN-GWs. The anchor ASN-GW is also the serving ASN-GW. While ASN-GWs use non-predictive ASN-GW relocation algorithm [12], MSs will be relocated due to the heavy load. We found that using non-predictive method the load of ASN-GW is going up and down frequently, and the load is not stable. Note that load of ASN-GW using the ART-based relocation method is lower than that of non-predictive method. Besides, the load of ASN-GW using the ART-based relocation method is more stable.

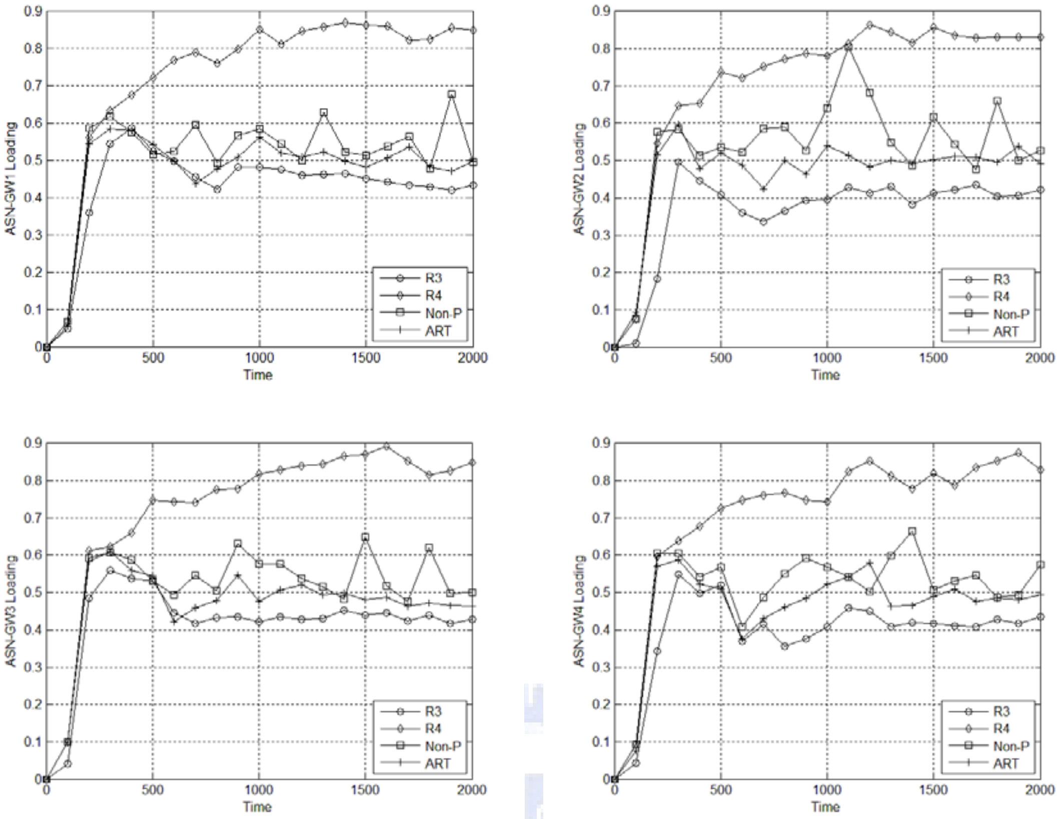


Figure 17. ASN-GW loading vs. Time

Table 6 shows that the mean and the variance of each ASN-GW's load. As previously mentioned, the mean and variance for the load of pure R4 mobility are highest, and they are lowest with pure R3 mobility. Compare to the Non-predictive method, the ART-Based method has lower mean load of ASN-GW. Without a proper strategy to perform relocations, the load of Non-predictive method is unstable. Therefore, the ASN-GW's load variance of Non-predictive is high. The variance of load with the ART-Based method is low and close to the variance of load with pure R3 mobility.

Table 6. Mean and Variance of each ASN-GW's load

(Mean, Variance)	ASN-GW1	ASN-GW2	ASN-GW3	ASN-GW4
Pure R3 Mobility	(0.447 , 0.013)	(0.416 , 0.013)	(0.434 , 0.013)	(0.425 , 0.013)
Pure R4 Mobility	(0.847 , 0.035)	(0.836 , 0.034)	(0.848 , 0.033)	(0.827 , 0.031)
Non-Predictive	(0.546 , 0.019)	(0.585 , 0.025)	(0.538 , 0.018)	(0.545 , 0.020)
ART-Based	(0.502 , 0.015)	(0.504 , 0.013)	(0.485 , 0.013)	(0.498 , 0.014)

Table 7 shows that the average numbers of MSs of each four ASN-GWs. The number includes both of anchored MS and serving MS. Among the four methods, the pure R4 mobility has the highest average number of MSs because the pure R4 mobility does not perform any relocation. On the contrary, the number of the pure R3 mobility is the lowest because it does not have any anchored MS. The ARTBR method has lower number than the Non-predictive method because the ARTBR select suitable mobility method according to the load of the ASN-GW but the Non-predictive method only use ASN anchored mobility.

Table 7. Average number of MSs of each ASN-GW

	ASN-GW1	ASN-GW2	ASN-GW3	ASN-GW4
Pure R3 Mobility	266	238	253	243
Pure R4 Mobility	442	438	443	438
Non-Predictive	341	346	326	331
ART-Based	334	334	325	320

The relationships between the average throughput and the number of serving MSs for the four methods are shown in Figure 18. In general, more serving MSs the ASN-GW serves more throughput the ANS-GW achieves. Among the four methods, the pure R3 mobility has the highest average throughput. The ART-based relocation method achieves better average throughput than the non-predictive method. Note that the average throughput for the pure R4 mobility drops down when the number of serving MSs is greater than 200. This is because the ANS-GW is overloaded.

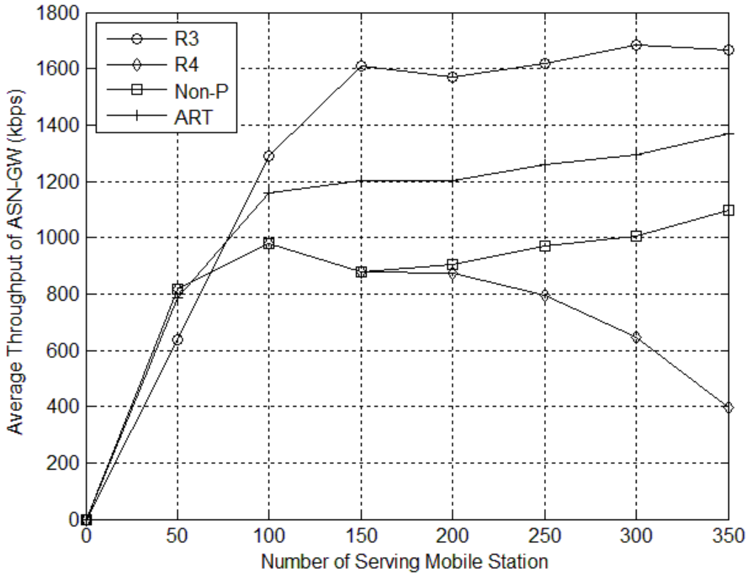


Figure 18. Average throughput vs. number of serving MS

When the ping-pong phenomenon occurs, a R3 tunnel with short lifetime will be constructed. If ping-pong phenomenon happens seriously, the average R3 tunnel lifetime will be short. Thus, we use the lifetime of R3 tunnel to represent the condition of ping-pong phenomenon. Table 8 shows the average R3 tunnel lifetime of the four relocation methods. The average R3 tunnel lifetime of the pure R4 mobility is extremely high because it does not perform any relocation and MSs never change their anchored ASN-GW. Thus, no ping-pong phenomenon will happen. The R3 tunnel lifetime of the pure R3 mobility is shortest because the MS movement triggers relocation and the ping-pong phenomenon occurs seriously. The non-predictive method does not select suitable MS to perform relocation. Therefore, the ping-pong phenomenon is still happened. The R3 tunnel lifetime of the ART-based method is longer than other methods with relocation strategy because our method can choose proper MSs to do relocations. MSs anchor at the ASN-GW where they have longer ART. The result of average R3 tunnel time shows the ART-based method can avoid the ping-pong phenomenon and every R3 tunnel will be used longer.

Table 8. Average R3 tunnel lifetime

	Pure R3 Mobility	Pure R4 Mobility	Non-predictive	ART-based
Avg. R3 Tunnel Lifetime	48.89 (sec)	1,505.45 (sec)	110.75 (sec)	188.88 (sec)

Table 9 shows that the delay time of handoff and relocation during the simulation. The pure R3 mobility has longest delay time because the relocation always trigger by the movement of MSs. The pure R4 mobility never performs relocation, and it does not take time to perform relocation. Thus, the delay of pure R4 mobility is shortest. The ART method has shorter delay time than the non-predictive method because the ART method has proper strategy to perform relocations and it can avoid unnecessary relocations happening.

Table 9. Handoff and relocation delay

	ASN Anchored Mobility	CSN Anchored Mobility	Relocation	Average
Pure R3 Mobility	0	119,604.58 (sec)	0	4.39 (sec)
Pure R4 Mobility	56,055.23 (sec)	0	0	2.06 (sec)
Non-predictive	55,142.87 (sec)	0	35,190.27 (sec)	3.32 (sec)
ART-based	35,877.74 (sec)	42,511.89 (sec)	4989.55 (sec)	3.06 (sec)

Table 10 shows that the numbers of two mobility management methods used by ASN-GWs. The pure R3 mobility has used CSN anchored mobility for 27,235 times without using any ASN anchored mobility. On contrary, the pure R4 mobility has used ASN anchored mobility for 27,235 times without using any CSN anchored mobility. The Non-predictive method is always using ASN anchored mobility while MSs handoff and it may perform relocations because of the heavy load of ASN-GW. For this reason, the Non-predictive method has used ASN anchored mobility for 27,235 times and also has performed relocations for 13,216 times. The ART-based selects suitable mobility method and properly performs relocations. Therefore, the ART-based method has used ASN anchored mobility for 17,844 times, 9,391 times for using CSN anchored mobility, and 1,874 times for relocations.

Table 10. Numbers of mobility used and relocation of each method

	ASN Anchored Mobility	CSN Anchored Mobility	Relocation
Pure R3 Mobility	0	27,235	0
Pure R4 Mobility	27,235	0	0
Non-predictive	27,235	0	13,216
ART-based	17,844	9,391	1,874

We use the number of handoff control message as a criterion for the overhead of each four methods. According to the fully controlled handoff procedure of WiMAX End-to-End Network Systems Architecture [4], the number of handoff control message of ASN anchored mobility, CSN anchored mobility and relocation are showing at Table 11.

Table 11. Numbers of handoff control message

	ASN Anchored Mobility	CSN Anchored Mobility	Relocation
MS ↔ BS	3	3	0
BS ↔ ASN-GW	14	14	0
ASN-GW ↔ ASN-GW	7	13	6
ASN-GW ↔ CSN	0	2	2

From table 11 we can find out that there are 24 messages will be transmitted between network components when MSs using ASN anchored mobility during inter-ASN handoff. CSN anchored mobility has 32 messages should be handled when MSs move between ASNs. Compare to ASN anchored mobility, there are eight additional messages for CSN anchored mobility and these extra messages are used for MSs to change their anchored ASN-GW. The eight extra messages also are the messages should be transmitted when an ASN-GW performs the relocation. Two of the extra messages are transmitted between ASN-GW and CSN to register the new Mobile IP tunnel and the registration may take a moment to complete. We apply the numbers of handoff control messages to the number of mobility method used in the experiment and calculate the total number of handover control message of each four methods. The table 12 shows the overhead of four methods.

Table 12. Numbers of handoff control message between network components

	Pure R3 Mobility	Pure R4 Mobility	Non-predictive	ART-based
MS ↔ BS	81,705	81,705	81,705	81,705
BS ↔ ASN-GW	381,290	381,290	381,290	381,290
ASN-GW ↔ ASN-GW	354,055	190,645	269,941	258,235
ASN-GW ↔ CSN	54,470	0	26,432	22,530

As the table 12 showing, with the same movement data in the experiment, the number of handoff control message of four methods between MS, BS, and ASN-GW are equal. Among the four methods, the pure R3 mobility has to deal with the most handoff control message between ASN-GWs and between ASN-GW and CSN because MSs change their anchored ASN-GW for every movement. On contrary, the pure R4 mobility handles the lowest number between ASN-GWs of handoff control messages with no relocations. However, the Non-predictive method does not perform the relocations properly and there are many unnecessary relocations will occur. Thus, with Non-predictive method, the number of relocation is large and also the number of the handoff message is high. The ART-based method can select suitable mobility method for the MSs when they are leaving to other ASNs and it also can relocate appropriate MSs to their serving ASN-GW. Therefore, the MSs using ART-based method has lower number to change their anchored ASN-GW and the number of handoff control messages is also low.

Chapter 5

Conclusion

In this thesis, we propose an ASN-GW relocation algorithm based on average residence time (ART-based). Using the ART-based method, ASN-GWs can select suitable mobility method when mobile stations perform inter-ASN handoff. Moreover, it also can select suitable mobile stations to relocate and choose proper timing to perform relocations. The simulation results show that the ART-based method makes the ASN-GWs have lower and stable loading than pure R4 mobility and Non-predictive ASN-GW relocation algorithm. Furthermore, the ART-based also has low overhead than pure R3 mobility and Non-predictive method. In conclusion, the ART-based method is one of the best solutions of trade-off between the pure R3 mobility with stable and lower load and pure R4 mobility with lower overhead handling handoff procedure.

Reference

[1] IEEE Std 802.16-2004 TM, IEEE Standard for Local and Metropolitan Area Networks - Part 16: Air Interface for Fixed Broadband Wireless Access Systems, Oct 2004.

[2] IEEE Std 802.16e-2005TM, IEEE Standard for Local and Metropolitan Area Networks - Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems, Feb 2006.

[3] WiMAX Forum. [Online] http://www.wimaxforum.org

[4] WiMAX Forum Proprietary, ”WiMAX End-to-End System Architecture (Stage 2: Architecture Tents, Reference Model and Reference Points, Stage3: Detailed Protocols and Procedures),” January 11, 2008 Release 1.2.1.

[5] C. Perkins, Ed., ”IP Mobility Support for IPv4,” RFC 3344, Internet Engineering Task Force, August 2002.

[6] S Khan, W Yao, P Wolfgang, A Khalid, ”Home agent load balancing in mobile ipv6 with efficient home agent failure detection and recovery,” In *Proceeding of IEEE International Conference on Emerging Technologies (ICET 2006)*, pp. 513-520, Nov. 2006.

[7] A. Vasilache, J. Li, and H. Kameda, “Threshold-Based Load Balancing for Multiple Home Agents in Mobile IP Networks”, *Telecommunications Systems*, vol. 22, issue 1-4, pp. 11-31, January-April, 2003.

[8] A. Vasilache, J. Li and H. Kameda, ”Load Balancing Policies for Multiple Home Agents Mobile IP Networks” In *Proceedings of IEEE International Conference on Web Information System Engineering (WISE 2001)*, pp. 178-185, Dec 2001

[9] F. Heissenhuber, W. Fritsche, and A. Riedl, ”HA Redundancy and Load Balancing in Mobile IPv6,” in *Proceedings of the 5th International Conference Broadband Communications*, pp. 235-244, Hong Kong, Nov 1999.

[10] S. Pack, T. Kwon, and Y. Choi, ”A mobility-based load control scheme at mobility anchor point in hierarchical mobile IPv6 networks,” in *Proceedings of the 47th annual IEEE Global Telecommunications Conference (GLOBECOM 2004)*, vol. 6, pp. 3431-3435, Nov 2004.

[11] A.-C. Pang, Y.-B. Lin, H.-M. Tsai, and P. Agrawal, ”Serving radio network controller relocation for UMTS All-IP networks,” *IEEE J. Select.Areas Commun.*, vol. 22, May 2004, pp. 617-629.

[12] Z.-H. Liu, S.-Y. Pan, and J.-C. Chen, ”Access service network (ASN) gateway relocation algorithms in WiMAX networks,” In *Proceedings of IEEE International Conference on Communications (ICC 2008)*, pp. 2674-2679, May 2008.

[13] S. Floyd and V. Jacobson, ”Random early detection gateways for congestion avoidance,” *IEEE/ACM Transactions on Networking*, vol. 1, issue 4, pp. 397-413, Aug. 1993.

[14] M. R. Spiegel, Schaum’s Outline of Statistics. McGraw-Hill, Dec. 1998.

[15] ”The network simulator - ns-2.” http://www.isi.edu/nsnam/ns/.