Research on Home NodeB Gateway Load Balancing Mechanism

Xiaojie Wang, Lanlan Rui, Peng Yu, Xingyu Chen
State Key Laboratory of Networking and Switching Technology
Beijing University of Posts and Telecommunications
Beijing, China
wangxiaojie.work@gmail.com

***Abstract*—Home NodeB Gateway (HNB-GW) is a mobile operator's equipment through which Home NodeB (HNB) gets access to the core network. Without a Load Balancing (LB) mechanism, HNB-GWs' load distribution is determined by the random access of HNBs. HNB-GWs with too many HNBs may become overloaded easily. Once a HNB-GW is overloaded, other entering HNBs cannot be accepted and serviced due to the one-to-many relationship between HNB-GW and HNB in the traditional network architecture. To solve the problems above, this paper firstly brings a HNB-GW Pool concept to optimize the traditional architecture. Then in order to balance the load of HNB-GWs in the pool, this paper proposes two HNB-GW LB mechanisms, one is based on HNB-GW selection control and the other is based on periodic HNB migration. Through analysis in specific scenario, both mechanisms can effectively balance the load between HNB-GWs. LB mechanism based on HNB-GW selection control can select the best HNB-GW for a HNB when it needs to register, but its load adjust rate is relatively slow. LB mechanism based on periodic HNB migration is characterized by fast load adjust rate, but HNB migration may bring negative influence on the related HNBs, such as upper business interruption. Finally, a comprehensive solution which colligates the advantages of both mechanisms is brought forward.**

Keywords- Home NodeB Gateway; Home NodeB Gateway Pool; Load Balancing; Home NodeB Gateway selection control; periodic Home NodeB migration

I. INTRODUCTION

HNBs are low-power, low-cost cellular base stations designed to serve a very small area, such as a home or office environment [1]. HNB connects a UE over UTRAN wireless air interface to a mobile operator's network using a broadband IP backhaul [2]. With HNBs, the operators can increase the capacity of their network and provide better service for in-building areas without deploying as many macro base stations which have both real estate costs and high equipment costs [3]. HNB-GW is a mobile operator's equipment through which HNB gets access to mobile operator's core network.

A. Problems Need to be Solved

HNBs are deployed in indoor environments, in order to improve radio coverage and system capacity, by the end-customers [4]. Since end-customers may frequently switch on and off HNBs, HNBs may access to HNB-GWs at random. Without a HNB-GW Load Balancing (LB) mechanism, the load of HNB access network is totally determined by the random access and free handover of HNBs, and the HNB distribution of different HNB-GWs may have significant difference. A HNB-GW with too many HNBs may become overloaded easily, and therefore other entering HNBs cannot be accepted and serviced due to the one-to-many relationship between HNB-GW and HNB in traditional architecture. So the HNB-GW with heavy load may be the bottleneck to network performance, the service may be in a free and uncontrolled state, the overall utilization state of network resource may be low. Overall, the network may have a poor stability and a low tolerance to the network undulation.

To avoid problems above and make the performance of HNB access network better, the one-to-many relationship between HNB-GW and HNB should be improved, and effective HNB-GW LB strategies must be adopted.

B. Related Works

3GPP has defined a series of corresponding technical specifications and reports. As for network architecture aspect, [5] describes the architectural aspects of specific HNB, [6] specifies the UTRAN architecture for HNB, [7] specifies the RANAP User Adaption between HNB and HNB-GW, [8] specifies the HNB Application Part between HNB and HNB-GW, and [9] defines SON OAM solution architecture for HNB. As for LB aspect, [10] provides a MME LB functionality based on MME selection for MME Pool. If HNB-GWs can be organized as a HNB-GW Pool, we can design a similar HNB-GW LB mechanism based on HNB-GW selection.

There are also some papers related. As for network architecture aspect, [11] puts forward Pool of Gateway, which groups several gateways into a pool and shares the resources of the connected base stations, and [12] designs a policy based HNB-GW self-configure structure in which each HNB can connect to a HNB-GW group for selecting route path instead of a single HNB-GW. To take advantage of Pool of Gateway and the HNB-GW group mentioned above, we can extend the traditional structure of the HNB access network by introducing a HNB-GW Pool concept. As for LB aspect, paper [11] introduces an inter-GW load balancing approach for next generation mobile network with Flat Architecture. However, it is not applicable for HNB access network very well, since it

This work was supported by the Funds for Creative Research Groups of China (60821001), NSFC (60973108, 60802035), National S&T Major Project (2011ZX03003-002-01, 2011ZX03005-004-02) and Chinese Universities Scientific Fund (BUPT2009RC0504).

does not consider the fact that, different from base stations, HNBs can be switch on and off frequently by end user.

C. Work in the Paper

The main works in the paper are:

(1) Proposing the HNB-GW Pool concept to optimize the one-to-many relationship between HNB-GW and HNB in the traditional architecture.

(2) Proposing two HNB-GW LB mechanisms: one based on HNB-GW selection control, which will select an appropriate HNB-GW for a HNB during its registration procedure, and another based on periodic HNB migration, which will move part of HNBs from higher load HNB-GWs to lower ones periodically.

(3) Analyzing these two LB mechanisms in specific scenario, making a contrast of them, and giving a comprehensive solution which combines them together.

II. HNB Access Network Architecture

In the traditional architecture of HNB access network, the one-to-many relationship between HNB-GW and HNB may lead to some problems: (1) Once the HNB-GW is at fault, all the HNBs connected will lose connection with the network and cannot get service from HNB-GW anymore. (2) Once the HNB-GW is overloaded, other HNBs entering the network cannot be accepted and serviced by the network.

To improve the traditional architecture, the paper extends it by introducing HNB-GW Pool. Fig. 1 illustrates the HNB-GW Pool concept. A plurality of HNB-GWs forms one HNB-GW Pool. HNBs within the area of HNB-GW Pool can connect to any HNB-GW in the Pool.

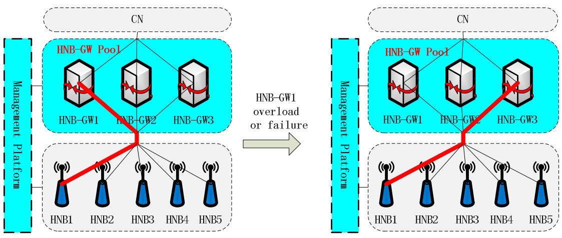

Figure 1. HNB-GW Pool

This extension has lots of advantages: (1) HNB-GW Pool can provide redundancy, in case of a HNB-GW overload or failure, where HNB can be redirected over to any other HNB-GW in the pool. (2) The capacity of the pool can be optimized easily by adding or removing HNB-GWs.

III. HNB-GW LB Mechanisms

This section formulizes the HNB-GW LB problem and then proposes two different HNB-GW LB mechanisms.

A. Problem Formulation

To describe the model of HNB access network, we assume a communication graph $S=(G,H)$, where $G=\{g_1,g_2,\dots,g_m\}$ is the set of HNB-GWs in the pool, $H=\{h_1,h_2,\dots,h_n\}$ is the set of HNBs within the area of the pool. Assume that $E(t)$ is the expectation for summation of absolute values of the difference between every HNB-GW instantaneous capacity utilization rate and the mean rate. It can reflect the load difference of HNB-GWs. The HNB-GW LB problem can be formalized as below:

$$
\begin{aligned}
\min\ & E(t)=\mathbb{E}\left\{\sum_{i=1}^{m}\operatorname{abs}\left(R_i(t)-R_{ave}(t)\right)\right\} \\
s.t.\ & R_i(t)=L_i(t)/C_i, \quad \forall i=1,\cdots,m, \\
& R_{ave}(t)=\sum_{i=1}^{m}L_i(t)\Big/\sum_{i=1}^{m}C_i, \\
& L_i(t)=\sum_{j=1}^{n}r_{ij}(t), \quad \forall i=1,\cdots,m, \\
& L_i(t)\le C_i, \quad \forall i=1,\cdots,m, \\
& r_{ij}(t)\in\{0,\ 1\}, \quad \forall i=1,\cdots,m, \forall j=1,\cdots,n, \\
& \sum_{i=1}^{m}r_{ij}(t)\le 1, \quad \forall j=1,\cdots,n,
\end{aligned}
\tag{1}
$$

where $abs(x)$ function returns the absolute value of x, $R_i(t)$ represents the instantaneous capacity utilization rate of g_i at time t, and $R_{ave}(t)$ is the average capacity utilization rate of all HNB-GWs in G at time t. The first constraint defines the calculation formula of $R_i(t)$, where L_i is the instantaneous load of g_i, which means the quality of HNBs connected, and C_i is the capacity of g_i, which means the maximum number of HNBs allowed to be connected. The second constraint defines the calculation formula of $R_{ave}(t)$. The third constraint defines the calculation formula of L_i, where $r_{ij}(t)$ represents the association between h_j and g_i. The fourth constraint corresponds to the fact that the instantaneous load of the HNB-GW cannot exceed its capacity. The fifth constraint limits $r_{ij}(t)$ to be a binary variable, which equals 1 when h_j is registered on g_i, otherwise equals 0. The sixth constraint limits each HNB to select one serving HNB-GW.

B. LB Mechanism based on HNB-GW Selection Control

This part proposes a HNB-GW selection control based LB mechanism, which will direct the entering HNB to an appropriate HNB-GW in a manner that achieves LB between HNB-GWs in the pool.

1) LB based on HNB-GW Selection Control

[10] provides a MME LB functionality based on MME selection to permit UEs entering into an MME Pool Area to be directed to an appropriate MME in a manner that achieves LB between MMEs. To achieve this, a weight factor is set for each MME, and the probability of the eNodeB selecting an MME is proportional to its weight factor.

Similarly, we can design a HNB-GW LB mechanism based on HNB-GW selection control, which can select an appropriate HNB-GW for a HNB during its registration procedure, to balance the load between HNB-GWs in the pool. This is achieved by setting a weight factor for each HNB-GW, such that the probability of the HNB selecting a HNB-GW is proportional to its weight factor.

The weight factor is typically set according to the instantaneous load and capacity of a HNB-GW relative to other HNB-GWs. And the weight factor shall be changed after the HNB registration since the load has been changed. A newly installed HNB-GW should be given a higher weight factor for an initial period of time to increase its load faster.

Assume that $W_i(t)$ is the weight factor of g_i at time t. $W_i(t)$ can be obtained by dividing the capacity C_i to the instantaneous load $L_i(t)$. For a newly installed HNB-GW with no HNB connected, $W_i(t)$ can be obtained by dividing the capacity C_i to a number $\alpha \in (0,1]$, which can be determined by prediction based on recorded information. In a word, $W_i(t)$ can be obtained by the equation below:

$$W_i(t) = \begin{cases} C_i / L_i(t), & L_i(t) \neq 0 \\ C_i / \alpha, & L_i(t) = 0 \end{cases}, \quad \forall i = 1, \cdots, m. \qquad (2)$$

a) HNB-GW Selection Control Algorithm

When a HNB h_j needs to register, the Management Platform (MP) selects an appropriate HNB-GW according to HNB-GW Selection Control Algorithm as follows:

(1) Every $g_i \in G$, calculates the value of $W_i(t)$ and sends $W_i(t)$ and C_i to MP.

(2) MP makes a comparative study of the weight factors of all these HNB-GWs. If there is a sole HNB-GW g^* in G with largest $W_i(t)$, g^* should be selected.

(3) Else if there are several HNB-GWs with the same highest weight factor $W_i(t)$, then compare their capacity. If there is a sole HNB-GW g^* in G with largest C_i, g^* should be selected.

(4) Else if there are several HNB-GWs with the highest capacity C_i, then select any one g^*.

(5) MP sends the information of g^* to h_j.

Fig. 2 presents the simplified scenario in which the HNB-GW with the highest Weight Factor is selected.

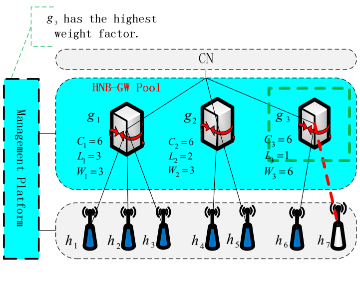

Figure 2. HNB-GW Selection Control

C. LB Mechanism based on Periodic HNB Migration

This part proposes an alternative LB mechanism based on periodic HNB migration, which will move part of HNBs that are already registered on the higher load HNB-GWs to the lower ones periodically.

1) Periodic Load Migration

To describe the periodic load migration based LB mechanism, we can divide the load state of the HNB-GW into three kinds, lightly loaded, modestly loaded, and heavily loaded according to its instantaneous capacity utilization rate.

Heavily loaded HNB-GWs should send heavy load information to MP to request HNB migration to other HNB-GWs. Lightly loaded HNB-GWs should send light load information to MP to indicate that it can still accept HNBs from other HNB-GWs. MP should maintain a global lightly loaded HNB-GW table to store the information of all lightly loaded HNB-GWs.

Assume $S_i(t)$ is the load state of g_i, so $S_i(t)$ can be obtained by the equation below:

$$S_i(t) = \begin{cases} L, & R_i(t) \in [0,\ R_{ave}(t) - \delta) \\ M, & R_i(t) \in [R_{ave}(t) - \delta,\ R_{ave}(t) + \delta], \\ H, & R_i(t) \in (R_{ave}(t) + \delta,\ 1] \end{cases} \quad \forall i = 1, \cdots, m, \qquad (3)$$

where $\delta \in [0,1)$, which can be determined by prediction based on recorded information.

Assume T_{gllhg} is the global lightly loaded HNB-GW table. As for the organization of T_{gllhg}, the last HNB-GW sending light load information should be put into the front. So T_{gllhg} should a stack, which can enable that each time the lightly loaded HNB-GW got from T_{gllhg} is the most effective one.

As is shown in Fig.3, the LB mechanism should include two parts: load state sending and HNB migration scheduling.

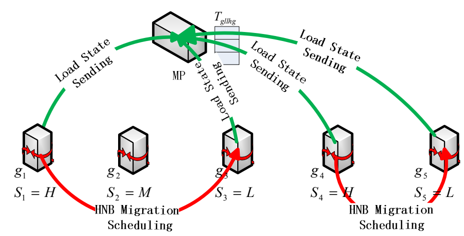

Figure 3. LB mechanism based on Periodic HNB Migration

a) Load State Sending

The load state sending procedure can be described as follows:

(1) Every $g_i \in G$, reads the value of $S_i(t)$ periodically.

(2) If $S_i(t)=H$ *or* $S_i(t)=L$, g_i sends the $S_i(t)$ to MP.

b) HNB Migration Scheduling

The HNB migration scheduling procedure can be described as follows:

(1) Once receiving $S_i(t)$ from g_i, MP reads the value.

(2) If $S_i(t)=L$, record it to T_{gllhg}.

(3) Else if $S_i(t)=H$, check if T_{gllhg} is null.

(4) If T_{gllhg} is not null, get the first HNB-GW g_{ll} out from T_{gllhg} and send its information to g_i. Then do the HNB migration from g_i to g_{ll} until g_i is not heavily loaded anymore or g_{ll} changes to be modestly loaded.

(5) Else if T_{gllhg} is null, send HNB migration rejecting information to g_i.

Fig. 4 presents the simplified HNB migration scheduling scenario in which part of HNBs of the heavily loaded HNB-GW g_1 are moved to the lightly loaded HNB-GW g_3.

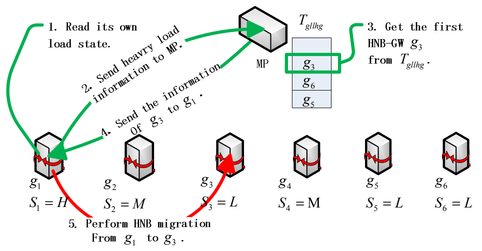

Figure 4. HNB Migration Scheduling

IV. Analysis of the Two LB Mechanisms

This section analyzes the two LB mechanisms in specific scenario, makes a contrast of them, and then presents a comprehensive solution which combines them together.

Assume that $G=\{g_1,g_2,g_3,g_4,g_5\}$, $H=\{h_1,h_2,\ldots,h_{8000}\}$, $C_1=C_2=C_3=C_4=C_5=2000$, $L_1=700$, $L_2=800$, $L_3=1000$, $L_4=1200$, $L_5=1300$, and 1000 HNBs will arrive at random in the next 100 minutes. Assume that LB Mechanism based on periodic HNB migration is run every 10 minutes and $\delta=0.05$. We will collect data every 10 minutes. Fig.5 shows $E(t)$'s changing curve *of* the two algorithms, and it is shown that both algorithms can decline the value of $E(t)$ and therefore balance the load between HNB-GWs in the pool. But the two algorithms express different features.

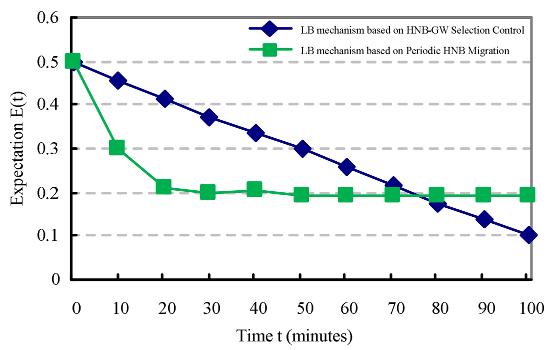

Figure 5. $E(t)$'s Changing Curve

LB mechanism based on HNB-GW selection control works on HNBs entering into the network. Every time a HNB needs to register, the mechanism will make the most appropriate HNB-GW selection for it according to the load state of all HNB-GWs at that time. So a HNB can always get the HNB-GW with the best performance at that moment. But the mechanism can control the direction of only one HNB at a time, it cannot reduce the load immediately. So the adjust rate is relatively slow. And at extreme conditions when no HNB enters, the load will keep being unbalanced all the time.

In contrast with the first mechanism, LB mechanism based on periodic HNB migration works on HNBs already registered. Since the load of different HNB-GWs can be quickly adjusted by HNB migration, this mechanism is characterized by fast adjust rate. But HNB migration may also bring some negative influence on related HNBs, such as the risk of migration failures and upper business interruption.

Both mechanisms have their own relative merits. The operator can choose any one or combine them together to be a comprehensive solution according to specific requirements.

To describe the combination procedure, we assume $K(t)$ to be the slope of $E(t)$. $K(t)$ is negative when $E(t)$ is declining. So the smaller $K(t)$ is, the faster $E(t)$ is declining. The comprehensive solution can be described as follows:

(1) Periodically calculate $K_1(t)$ of LB mechanism based on HNB-GW selection control and $K_2(t)$ of LB mechanism based on periodic HNB migration.

(2) If $K_1(t)> K_2(t)$, use LB mechanism based on HNB-GW selection control. Otherwise, use mechanism based on periodic HNB migration.

Fig.6 shows $E(t)$'s changing curve of the comprehensive solution. It is shown that this solution colligates advantages of both mechanisms and performs better.

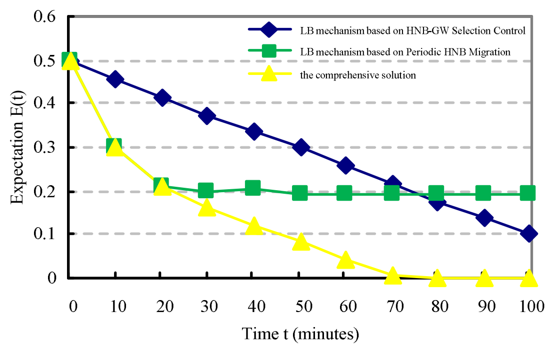

Figure 6. $E(t)$'s Changing Curve of the comprehensive solution

V. Conclusion and Future Work

The paper firstly proposes an extending architecture of HNB access network by using the HNB-GW Pool concept. Then it brings forward two alternative HNB-GW LB mechanisms, LB mechanism based on HNB-GW selection control and LB mechanism based on periodic HNB migration. Through analysis in specific scenarios, both alternatives are proved to be effective in LB between HNB-GWs in the pool. Since both mechanisms have their own relative merits, a comprehensive solution, which combines them together, is finally put forward. The future work includes: Predict the value of α and δ; Define corresponding LB protocol; Evaluate the communication overhead of implementation.

REFERENCES

[1] Shaunak Joshi, Ray C.C. Cheung, Pooya Monajemi, and John D. Villasenor, “Traffic-Based Study of Femtocell Access Policy Impacts on HSPA Service Quality,” in Proc. of GLOBECOM 2009. IEEE

[2] 3GPP TS 22.220 V10.4.0, “Service requirements for Home NodeB (HNB) and Home eNodeB (HeNB)”

[3] Hyung-Deug Bae, and Nam-Hoon Park, “Impact of reading system information in inbound handover to LTE femtocell,” in Proc. of Communications (APCC), 2010 16th Asia-Pacific Conference on

[4] Guillaume de la Roche└y, A´ kos Lada´nyiy, David Lo´pez-Pe´rezz, Chia-Chin Chong└, and Jie Zhangy, “Self-organization for LTE enterprise femtocells” in Proc. of GLOBECOM Workshops (GC Wkshps), 2010 IEEE

[5] 3GPP TR 23.830 V9.0.0, “Architecture aspects of Home NodeB and Home eNodeB”

[6] 3GPP TS 25.467 V9.3.0, “UTRAN architecture for 3G Home NodeB (HNB);Stage 2”

[7] 3GPP TS 25.468 V9.3.0, “UTRAN Iuh Interface RANAP User Adaption (RUA) signalling”

[8] 3GPP TS 25.469 V9.3.0, “UTRAN Iuh interface Home NodeB (HNB) Application Part (HNBAP) signalling”

[9] 3GPP TR 32.821 V9.0.0, “Study of Self-Organizing Networks (SON) related Operations, Administration and Maintenance (OAM) for Home NodeB (HNB)”

[10] 3GPP TS 23.401 V10.1.0, “3rd Generation Partnership Project;Technical Specification Group Services and System Aspects;General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access(Release 10)”

[11] Cheng Xue, Jijun Luo, Ruediger Halfmann, Egon Schulz, and Christian Hartmann, “Inter GW Load Balancing for Next Generation Mobile Networks with Flat Architecture,” in Proc. of 69th IEEE Vehicular Technology Conference (VCT 09), Barcelona, SPAIN, IEEE Press, 2009, pp. 2645-2649.

[12] Yunge Liang, and Zhipeng Gao, “Policy based Home NodeB gateway self-configure structure,” Research Challenges in Computer Science, 2009, pp. 116-119