Load Balancing for Proxy Mobile IPv6 in SAE-based Mobile Networks

Moneeb Gohar · Sang-Il Choi · Seok-Joo Koh

In the existing Proxy Mobile IPv6 (PMIP) scheme for mobile networks based on the System Architecture Evolution (SAE), the Mobile Access Gateway (MAG) of PMIP is deployed at the Serving Gateway (S-GW) and the Local Mobility Anchor (LMA) of PMIP is employed at the PDN Gateway (P-GW). In this scheme, P-GW shall process data traffic as well as control traffic for binding update. Such a mobility scheme tends to give large traffic overhead at P-GW and increased operational costs. In this paper, we propose the load balancing schemes for PMIP in the SAE-based mobile networks. In the proposed schemes, the data delivery function and the mobility control function are separated, in which the mobility control function for binding update and query will be performed by a newly introduced Mobility Control Agent (MCA), and the data delivery function is done by P-GW. Before data transmission, an optimal data path will be obtained from MCA by using the binding query function. As per the location of MCA, the proposed schemes are divided into the two cases: 1) MCA over P-GW of SAE and 2) MCA over Mobility Management Entity (MME) of SAE. By numerical analysis, the two proposed schemes are compared with the existing scheme. From the numerical results, we see that the proposed load balancing PMIP schemes can give better performance than the existing PMIP scheme in terms of traffic overhead and transmission delay. In particular, it is shown that the PMIP load balancing scheme with MCA over MME provides the best performance among the candidate schemes.

Keywords: LTE/SAE, Mobile networks, Proxy MIPv6, Load balancing, Data/control separation

I.INTRODUCTION

As the Internet becomes the infrastructure of information society, the number of mobile Internet users has been rapidly increasing with wide popularity of smart phones and appearance of various mobile networks. It is reported that the number of mobile Internet users will be 1.6 billion in around 2014 and thus exceed the number of desktop users [1].

The System Architecture Evolution (SAE) with the Long-Term Evolution (LTE) has been used as a key technology for the next generation mobile networks. It was originally designed as a hierarchical architecture to support circuit-based voice traffics. However, an ever-increasing demand of mobile Internet traffic has imposed non-hierarchical or flat structure on mobile networks so as to provide a better cost and performance on data services [2], [3], [4].

Most of the recently proposed mobility schemes are based on a centralized approach, as shown in the Mobile IP (MIP) [5] and Proxy Mobile IP (PMIP) [6] protocols, in which Home Agent (HA) or Local Mobility Anchor (LMA) are used as a mobility anchor which processes all control and data packets. This centralized mobility anchor allows a mobile host to be reachable, when it is away from its home domain, by ensuring the forwarding of data packets destined to or sent from the mobile host. However, such a scheme may be vulnerable to several problems.

논문번호: TR13-102, 논문접수일자:2013.11.11, 논문수정일자:2014.04.02, 논문게재확정일자:2014.05.20
Moneeb Gohar, Sang-Il Choi, Seok-Joo Koh: Kyungpook National University

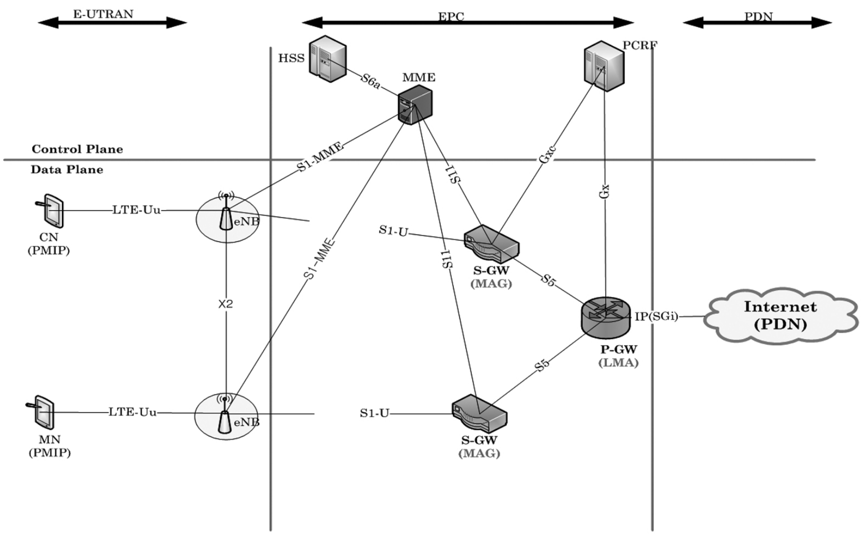

Figure 1. Network model for existing PMIP-SAE scheme

First, the centralized mobility anchor tends to induce unwanted data traffics into core networks, which may give a big burden to network operator due to large operational costs. Next, a single point of failure of central node may affect severe degradation of overall system performance and also increased costs of network engineering. Moreover, the centralized mobility control tends to induce non-optimal routes [7], [8].

In this paper, we present the load balancing schemes for PMIP in SAE-based mobile networks. In the proposed schemes, the data delivery function and the mobility control function for binding update and query will be separated, in which the mobility control function is performed by the Mobility Control Agent (MCA) and the data delivery function is done by PDN gateway. Before data transmission, an optimal data path will be obtained from the MCA by using the binding query function. As per the location of MCA, the proposed schemes are divided into the two cases: 1) MCA over PDN gateway (P-GW) of SAE, and 2) MCA over Mobility Management Entity (MME) of SAE.

The rest of the paper is organized as follows. In Section II, we discuss the existing PMIP scheme in SAE-based mobile networks. Section III describes the two proposed load balancing schemes for PMIP mobility control. Section IV analyzes the performance of the proposed schemes in terms of the traffic overhead and data transmission delay. Section V concludes this paper.

II. EXISTING PMIP IN SAE-BASED MOBILE NETWORKS

We first review the existing PMIP scheme in SAE-based mobile network, which is based on the 3GPP Release 10 [9] and denoted by PMIP-SAE in this paper. To support the PMIP protocol, the SAE system uses the two functional entities: PDN Gateway (P-GW) and Serving Gateway (S-GW). P-GW gives an access to Mobile Nodes (MNs) to different data networks and also allocates an IP address for MN. P-GW is also used as a Local Mobility Anchor (LMA) of PMIP. S-GW is used to detect the movement of MN, when it moves into the 3GPP access network. S-GW also works as a Mobile Access Gateway (MAG) of PMIP. For the PMIP binding update, S-GW will send a Proxy Binding Update (PBU) message to P-GW. Then, P-GW allocates an IP address for MN and responds a Proxy Binding Acknowledgement (PBA) message to S-GW.

The network model for PMIP-SAE is shown in Figure 1, in which both Mobile Node (MN) and Correspondent Node (CN) are located in the same mobile domain. In the figure, S-GW functions as MAG of PMIP, and P-GW works as LMA of PMIP.

The binding update and data delivery operations of PMIP-SAE are illustrated in Figure 2 [9], [10], [11]. When MN establishes a radio link with eNB, it sends an *Attach request* to Mobility Management Entity (MME). Then, the security-related procedures are performed between MN and MME (Step 1, 2, 3). MME sends the *Update location request* to the associated Home

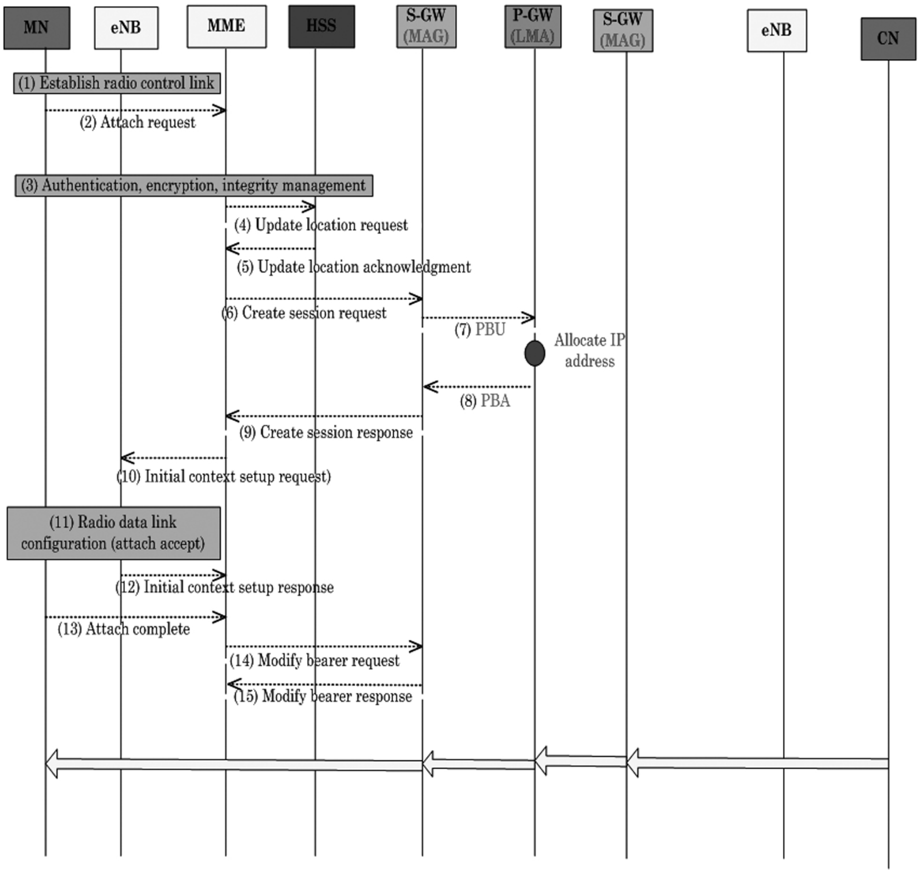

Figure 2. Binding update and data delivery operations in PMIP-SAE

Subscriber Server (HSS). Then, HSS will respond with the *Update location acknowledgement* to MME (Step 4, 5).

To establish a transmission path to PDN, MME sends a *Create session request* to S-GW. When S-GW receives the request from MME, it will send a *PBU* message to P-GW. The P-GW will allocate an IP address for MN and respond with a *PBA* message to S-GW. Then, S-GW will respond with a *Create session response* to MME (Step 6, 7, 8, 9). Now, MME sends the information received from S-GW to eNB with in the *Initial context setup request* message. This signaling message also contains the *Attach accept notification*, which is the response of *Attach request* in Step 2 (Step 10, 11). Then, eNB responds with an *Initial context setup response* to MME. Then, MN sends the *Attach complete message* to MME (Step 12, 13). Then, MME sends the *Modify bearer request* message to S-GW, and S-GW will respond with the *Modify bearer response* to MME (Step 14, 15).

For data delivery, CN will send a data packet to P-GW. Then, P-GW finds the location of MN from its database, and it will forward the data packet to MN.

To provide an optimal data path for two mobile nodes in PMIP, the PMIP with localized routing (PMIP-LR) was recently proposed [12]. In the scheme, the data path between MN and CN will be optimized after the control operation for localized routing, in which the Localized Routing Initiation (LRI) and Localized Routing ACK (LRA) messages are exchanged between MAGs and LMA. However, this scheme was designed for general IP networks, and the use of PMIP-LR over LTE/SAE-based mobile networks has not been studied yet.

III. LOAD BALANCING SCHEMES FOR PMIP IN SAE-BASED MOBILE NETWORKS

In this section, we describe the two load balancing schemes for PMIP in SAE-based mobile networks, named PMIP-LB-SAE-PGW and PMIP-LB-SAE-MME in this paper.

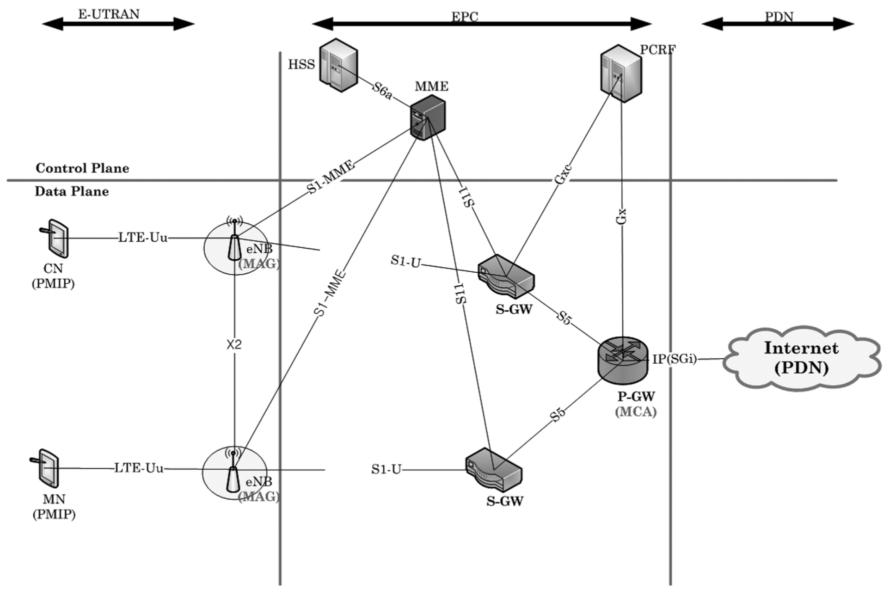

Figure 3. Network model for PMIP-LB-SAE-PGW

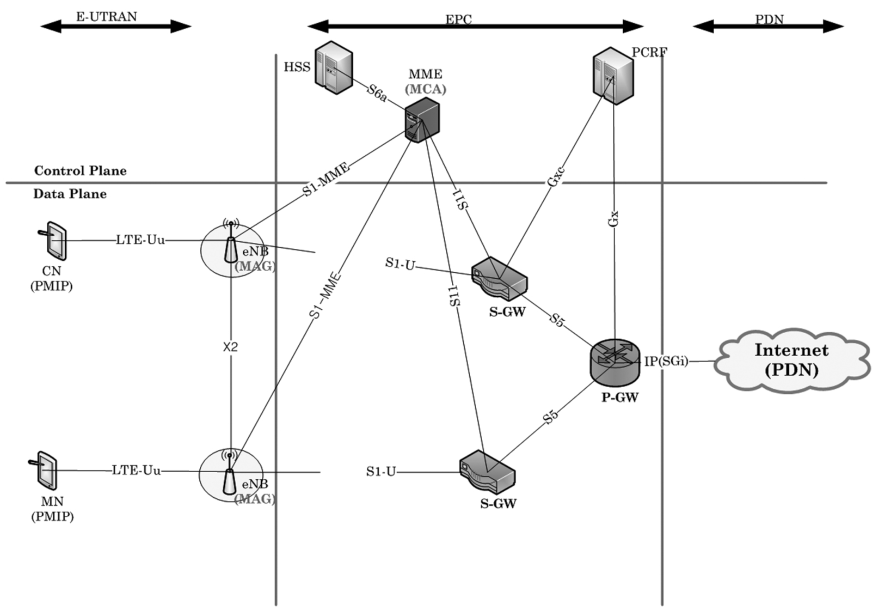

Figure 4. Network model for PMIP-LB-SAE-MME

1. Network Models

The network models for PMIP-LB-SAE-PGW and PMIP-LB-SAE-MME are shown in Figure 3 and Figure 4. In the proposed schemes, each eNB will function as the MAG of PMIP. For separation of data delivery function and mobility control function, we define a new mobility control entity, Mobility Control Agent (MCA). In the proposed PMIP-LB-SAE-PGW scheme, P-GW works as MCA, whereas MME functions as MCA in the proposed PMIP-LB-SAE-MME scheme, as illustrated in Figure 3 and Figure 4.

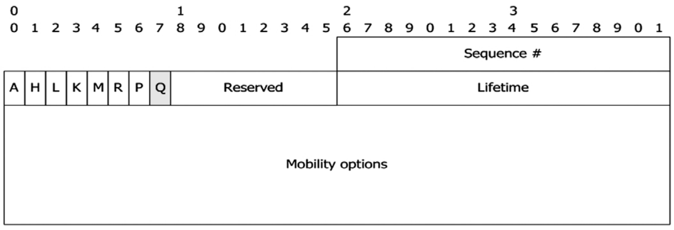

Figure 5. Proxy Binding Query (PBQ)

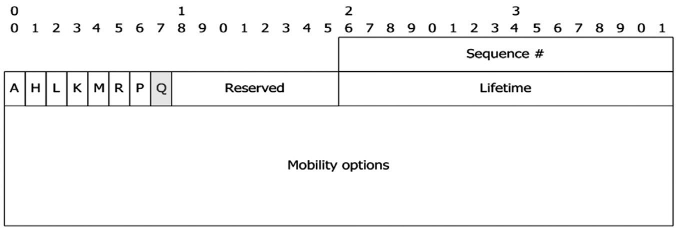

Figure 6. Proxy Query ACK (PQA)

Table 1. Comparison of PMIP schemes in SAE-based mobile networks

Schemes	Binding Update	Binding Query	Location of MAG	Location of Mobility Agent
PMIP-SAE	Used	Not Used	S-GW	P-GW
PMIP-LB-SAE-PGW	Used	Used	eNB	P-GW
PMIP-LB-SAE-MME	Used	Used	eNB	MME

Before going into further description of the proposed schemes, let us compare the proposed and existing schemes in the architectural perspective, as described in Table 1.

In PMIP-SAE, the P-GW performs both the mobility control functions (binding update and query) and the data delivery function. That is, S-GW works as MAG of PMIP, and P-GW works as LMA of PMIP. The data packets will be delivered to P-GW and then delivered to MN.

PMIP-LB-SAE-PGW is a load balancing scheme in which the control plane is separated from the data plane. Each eNB works as MAG, and P-GW works as MCA. MAG (or eNB) will perform the binding update operation with MCA. For data delivery, the MAG of CN performs the binding query operation with MCA so as to find the location of MN. For this purpose, the following control messages are newly defined: Proxy Binding Query (PBQ) and Proxy Query ACK (PQA), which will be described in Figure 5 and Figure 6.

PMIP-LB-SAE-MME is also a load balancing scheme in which the control plane is separated from the data plane. Each eNB works as MAG, and MME works as MCA. MAG (or eNB) performs the binding update operation with MCA. For data delivery, the MAG of CN will perform the binding query operation with MCA so as to find the location of MN.

For binding query operation of the two proposed schemes, PMIP-LB-SAE-PGW and PMIP-LB-SAE-MME, we define the two new messages, PBQ and PQA, by adding the 'Q' flag bit into the existing PMIP Proxy Binding Update (PBU) and Proxy Binding ACK (PBA) packets, as shown in Figure 5 and 6.

2. Protocol Operations

In this paper, we focus on the protocol operations of the proposed schemes in the case of two mobile nodes.

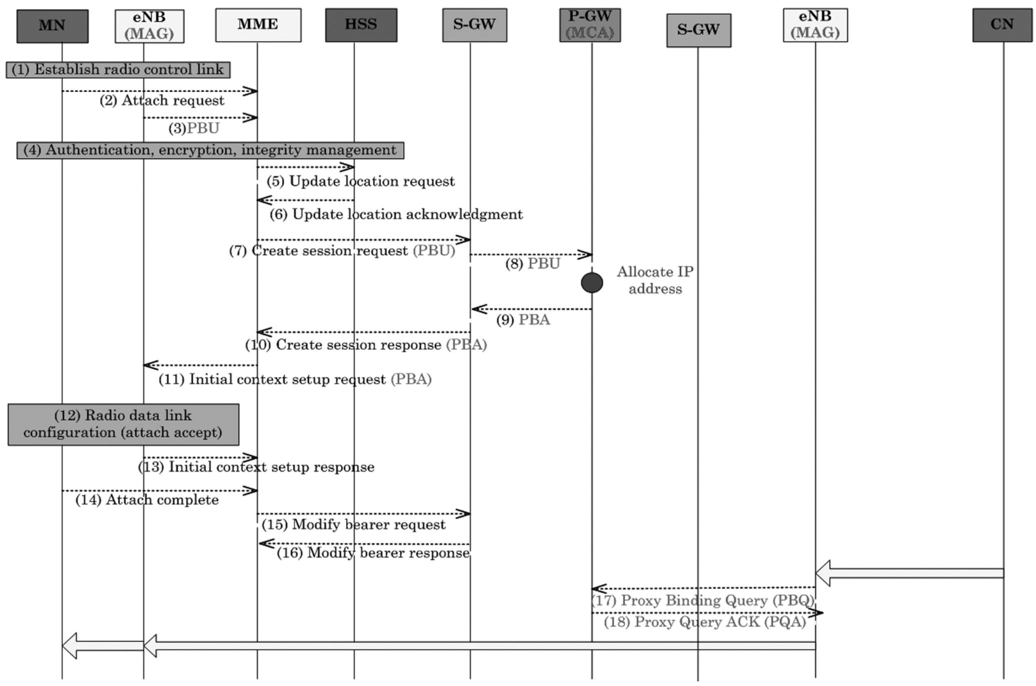

Figure 7. Operationsof PMIP-LB-SAE-PGW

That is, both MN and CN will be in the mobile networks. It is noted that the communication with an external host in the Internet will follow the existing PMIP/LTE procedures.

2.1. PMIP-LB-SAE-PGW

The binding update and data delivery operations of PMIP-LB-SAE-PGW are shown in Figure 7.

When MN establishes a radio link with eNB, it sends an *Attach request* to MME. The eNB will send a *PBU* message to MME. Then, the security procedures are performed between MN and MME (Step 1, 2, 3, 4). Now, MME sends the *Update location request* to HSS. Then, HSS responds with the *Update location \ acknowledgement* to MME (Step 5, 6). To establish the transmission path to P-GW, MME sends a *Create session request* with a PBU message to S-GW. When S-GW receives the request from MME, it sends the *PBU* message to P-GW. The P-GW will allocate an IP address for MN and respond with a *PBA* message to S-GW. Then, S-GW will respond with a *Create session response* containing the PBA message to MME (Step 7, 8, 9, 10).

MME sends the information received from S-GW to eNB in an *Initial context setup request* with the *PBA* message. This signaling message also contains the *Attach accept notification*, which is the response of *Attach request* in Step 2 (Step 11, 12). Then, eNB responds with *Initial context setup response* to MME. Then, MN sends the *Attach complete* message to MME (Step 13, 14). The MME sends the *Modify bearer request* message to S-GW, and then S-GW responds with a *Modify bearer response* to MME (Step 15, 16).

For data delivery, CN sends a data packet to MN. Then, the MAG of CN sends a PBQ message to MCA (P-GW) to find the MAG of MN (Step 17). Then, MCA (P-GW) responds with a PQA message to the MAG of CN (Step 18). Now, the MAG of CN can send the data packet to the MAG of MN. Finally, the data packet is forwarded to MN.

2.2. PMIP-LB-SAE-MME

The binding update and data delivery operations of PMIP-LB-SAE-MME are shown in Figure 8. When MN establishes a radio link with eNB, it sends an *Attach request* to MME. The eNB will also send a PBU to MME. Then, the security procedures are performed between MN and MME (Step 1, 2, 3, 4).

MME sends the *Update location request* to HSS, and the HSS responds with the *Update location acknowledgement* to MME (Step 5, 6). To establish a

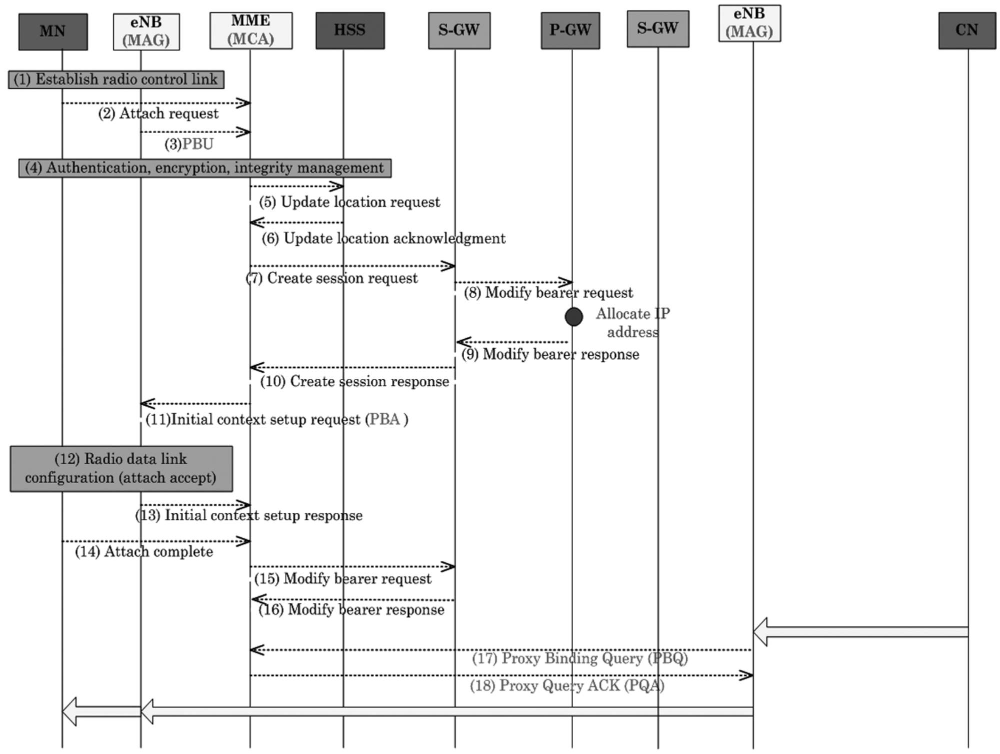

Figure 8. Operations of PMIP-LB-SAE-MME

transmission path to PDN, MME sends a *Create session request* to S-GW. When S-GW receives the request from MME, it will send the *Modify bearer request* message to P-GW. The P-GW will allocate an IP address for MN and then respond with a *Modify bearer response* message to S-GW. Then, S-GW will respond with a *Create session response* to MME (Step 7, 8, 9, 10). MME sends the information received from S-GW to eNB in an *Initial context setup request* with a PBA message. This signaling also contains the *Attach accept notification*, which is the response of *Attach request* in Step 2 (Step 11, 12). Then, eNB responds with *Initial context setup response* to MME. Then, MN sends an *Attach complete* message to MME (Step 13, 14). MME sends the *Modify bearer request* message to S-GW, and then S-GW responds with the *Modify bearer response* to MME (Step 15, 16).

In the data delivery operation, CN sends a data packet to MN. Then, the MAG of CN sends a PBQ message to MCA (MME) to find the MAG of MN (Step 17). Then, MCA (over MME) responds with a PQA message to the MAG of CN (Step 18). Now, the MAG of CN can send the data packet to MAG of MN. Finally, the data packet is forwarded to MN.

It is noted that the proposed schemes provide a different route optimization scheme from the PMIP-LR scheme [12]. In PMIP-LR, an initial data path between MN and CN will be established by way of LMA (P-GW). During data transmission, the data path is changed to an optimal path after exchanging the LRI and LRA messages between MAG and LMA. In the meantime, the proposed schemes will provide an optimal data path from the beginning by using the binding query operation with MCA before the initial data transmission.

IV. PERFORMANCE ANALYSIS

To evaluate the performance, we analyze *Traffic Overhead* (TO) and *Transmission Delay* (TD) for three candidate schemes: PMIP-SAE, PMIP-LB-SAE-PGW, and PMIP-LB-SAE-MME.

1. Analysis Model

For performance analysis, we use the following notations, as shown in Table 2.

We consider a network model for analysis, as illustrated in Figure 9. In the figure, we denote $T_{x-y}(S)$ by

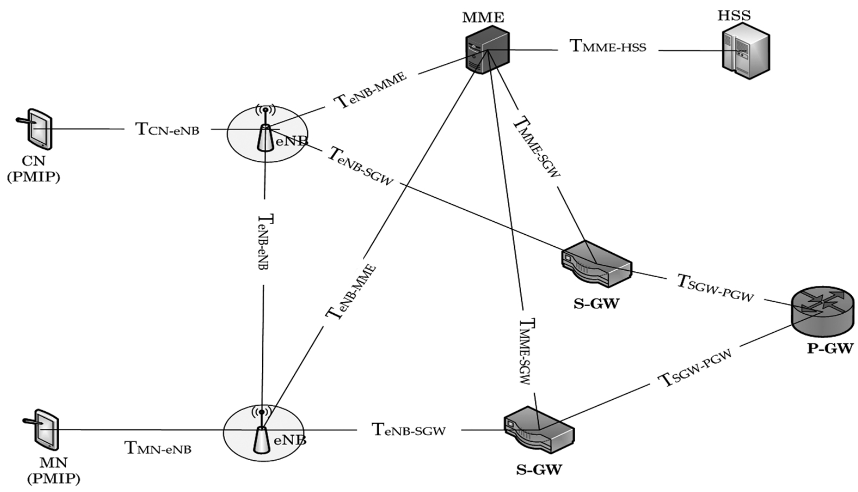

Figure 9. Network model for performance analysis

Table 2. Parameters used for Performance analysis

Parameters	Description
S_c	Size of control packets (bytes)
S_d	Size of data packets (bytes)
B_w	Wired link bandwidth (Mbps) between eNB and S-GW, between S-GW and P-GW, etc
B_{wl}	Wireless bandwidth(Mbps) between host and eNB
L_w	Wired link delay (ms) between eNB and S-GW, between S-GW and P-GW, etc
L_{wl}	Wireless link delay (ms) between host and eNB
H_{a-b}	Hop count between nodes *a* and *b* in the mobile network
N_{host}	Total number of hosts in the mobile network
N_{data}	Average number of data packets transmitted by a host
q	Link failure probability for a wireless link
T_q	Average queuing delay at each node

the transmission delay of a message with size *S* sent from *x* to *y* via the '*wireless*' link. Then, $T_{x-y}(S)$ can be expressed as $T_{x-y}(S)=[(1-q)/(1+q)]\times[(S/B_{wl})+L_{wl}]$. In the meantime, we denote $T_{x-y}(S,H_{x-y})$ by the transmission delay of a message with size *S* sent from *x* to *y* via '*wired*' link, where H_{x-y} represents the number of wired hops between node *x* and node *y*. Then, $T_{x-y}(S,H_{x-y})$ is expressed as $T_{x-y}(S, H_{x-y}) = H_{x-y}\times[(S/Bw) + Lw + Tq]$, which is based on the works in [13].

On the other hand, q represents the probability of wireless link failure, and T_q is the average queuing delay at each node in the network. B_{wl} is the bandwidth of wireless link between host and eNB, whereas B_w is the bandwidth of wired links among eNB, S-GW, P-GW, MME and HSS. L_{wl} is the wireless link delay between MN and eNB, while L_w is the wired link delay among eNB, S-GW, P-GW, MME and HSS.

2. Analysis of Traffic Overhead (TO)

To analyze the performance of the candidate schemes, we evaluate the PMIP Traffic Overhead (TO) to be processed at P-GW or MME. It is noted that we consider only the PMIP-related messages in the analysis.

2.1. PMIP-SAE

In PMIP-SAE, TO is measured by the number of control/data messages to be processed at P-GW. It is assumed that the hosts are equally distributed in the mobile network. For binding update between Care-of Address (CoA) and Home Address (HoA), S-GW sends a binding update message to P-GW for each host. Thus, the binding update messages of $S_c \times N_{host}$ shall be processed by P-GW. For data transmission, all data packets are first delivered to P-GW. Thus, the data packets of $S_d \times N_{host} \times N_{data}$ shall be processed by P-GW. N_{data} represents the total number of data packets transmitted by CN. Accordingly, we get the TO of PMIP-SAE as follows.

$$TO_{PMIP\text{-}SAE} = S_c \times N_{host} + S_d \times N_{host} \times N_{data}$$

2.2. PMIP-LB-SAE-PGW

In PMIP-LB-SAE-PGW, we calculate TO as the number of control/data messages to be processed by P-GW. For binding update, eNB will send a binding update message to P-GW for each host. Thus, the binding update messages of $S_c \times N_{host}$ shall be processed by P-GW. For data transmission, each eNB sends a binding query messages to P-GW. Thus, the binding query messages of $S_c \times N_{host}$ shall be processed by P-GW. Accordingly, we get TO of PMIP-LB-SAE-PGW as follows.

$$TO_{PMIP\text{-}LB\text{-}SAE\text{-}PGW} = S_c \times N_{host} + S_c \times N_{host}$$

2.3. PMIP-LB-SAE-MME

In PMIP-LB-SAE-MME, we calculate TO as the number of control/data messages to be processed by MME. For binding update, eNB will send a binding update message to MME. Thus, the binding update messages of $S_c \times N_{host}$ shall be processed by MME for each host. For data transmission, each eNB sends a binding query messages to MME. Thus, the binding query messages of $S_c \times N_{host}$ shall be processed by MME. It is noted that the traffic overhead of PMIP-LB-SAE-MME is the same as PMIP-LB-SAE-PGW. So, we get TO of PMIP-LB-SAE-MME as follows.

$$TO_{PMIP\text{-}LB\text{-}SAE\text{-}MME} = S_c \times N_{host} + S_c \times N_{host}$$

3. Analysis of Transmission Delay (TD)

The transmission delays are divided into Binding Update Delay (BUD), Binding Query Delay (BQD), and Data Delivery Delay (DDD). Then, the Transmission Delay (TD) can be represented as TD = BUD + BQD + DDD.

3.1. PMIP-SAE

In PMIP-SAE, the binding update operations are performed as follows. When MN enters an eNB region, it establishes a radio link and sends *Attach request* message to MME. This operation takes $T_{MN\text{-}MME}(S_c)$, where $T_{MN\text{-}MME}(S_c) = T_{MN\text{-}eNB}(S_c) + T_{eNB\text{-}MME}(S_c)$. Then, MME performs the *Update location operation* with HSS by exchanging the *Update location request* and *response* messages. This operation takes $2 \times T_{MME\text{-}HSS}(Sc)$. MME also sends a *Create session request* message to S-GW. This operation takes $T_{MME\text{-}SGW}(Sc)$. S-GW performs the *PBU* and *PBA* operations with P-GW. This operation takes $2 \times T_{SGW\text{-}PGW}(S_c)$. Then, S-GW responds with a *Create session response* message to MME, after the PBU and PBA operations. This operation takes $T_{MME\text{-}SGW}(S_c)$. MME will perform the *Initial context setup* operation with eNB by exchanging the *Initial context setup request* and *response* messages. This operation takes $2 \times T_{eNB\text{-}MME}(Sc)$. Then, MN will send the *Attach complete* message to MME, which takes $T_{MN\text{-}MME}(S_c)$, where $T_{MN\text{-}MME}(S_c) = T_{MN\text{-}eNB}(S_c) + T_{eNB\text{-}MME}(S_c)$. MME will perform the *Modify bearer* operation with S-GW by exchanging the *Modify bearer request* and *response* messages. This operation takes $2 \times T_{MME\text{-}SGW}(S_c)$. Accordingly, the Binding Update Delay (BUD) of PMIP-SAE can be represented as follows.

$$\begin{aligned}
&BUD_{PMIP\text{-}SAE} \\
&= T_{MN\text{-}MME}(S_c) + 2 \times T_{MME\text{-}HSS}(S_c) + T_{MME\text{-}SGW}(S_c) \\
&\quad + 2 \times T_{SGW\text{-}PGW}(S_c) + T_{MME\text{-}SGW}(S_c) + 2 \\
&\quad \times T_{eNB\text{-}MME}(S_c) + T_{MN\text{-}MME}(S_c) + 2 \times T_{MME\text{-}SGW}(S_c) \\
&= T_{MN\text{-}eNB}(S_c) + T_{eNB\text{-}MME}(S_c) + 2 \times T_{MME\text{-}HSS}(S_c) \\
&\quad + 2 \times T_{MME\text{-}SGW}(S_c) + 2 \times T_{SGW\text{-}PGW}(S_c) + 2 \\
&\quad \times T_{eNB\text{-}MME}(S_c) + T_{MN\text{-}eNB}(S_c) + T_{eNB\text{-}MME}(S_c) \\
&\quad + 2 \times T_{MME\text{-}SGW}(S_c) \\
&= 2 \times T_{MN\text{-}eNB}(S_c) + 4 \times T_{eNB\text{-}MME}(S_c)
\end{aligned}$$

$+ 2\times T_{MME\text{-}HSS}(S_c) + 4\times T_{MME\text{-}SGW}(S_c)$

$+ 2\times T_{SGW\text{-}PGW}(S_c)$

In PMIP-SAE, the Binding Query Delay (BQD) is 0. Thus, we get

$$BQD_{PMIP\text{-}SAE} = 0$$

For data delivery in PMIP-SAE, CN sends a data packet to P-GW (LMA), and P-GW will forward the data packet to MN. Then, the Data Delivery Delay (DDD) of PMIP-SAE can be represented as

$$\begin{aligned} DDD_{PMIP\text{-}SAE} &= T_{CN\text{-}PGW}(S_d) + T_{MN\text{-}PGW}(S_d) \\ &= T_{CN\text{-}eNB}(S_d) + T_{eNB\text{-}SGW}(S_d) + T_{SGW\text{-}PGW}(S_d) \\ &\quad + T_{SGW\text{-}PGW}(S_d) + T_{eNB\text{-}SGW}(S_d) + T_{MN\text{-}eNB}(S_d) \\ &= T_{CN\text{-}eNB}(S_d) + 2\times T_{eNB\text{-}SGW}(S_d) + 2\times T_{SGW\text{-}PGW}(S_d) \\ &\quad + T_{MN\text{-}eNB}(S_d) \end{aligned}$$

So, we obtain the overall Transmission Delay (TD) of the PMIP-SAE as follows.

$$TD_{PMIP\text{-}SAE} = BUD_{PMIP\text{-}SAE} + BQD_{PMIP\text{-}SAE} + DDD_{PMIP\text{-}SAE}.$$

3.2. PMIP-LB-SAE-PGW

In PMIP-LB-SAE-PGW, the binding update operations are performed as follows. When MN enters a new eNB region, it will establish a radio link and sends *Attach request* message to MME. This operation takes $T_{MN\text{-}MME}(S_c)$, where $T_{MN\text{-}MME}(S_c) = T_{MN\text{-}eNB}(S_c) + T_{eNB\text{-}MME}(S_c)$. eNB will send a PBU message to MME, which takes $T_{eNB\text{-}MME}(S_c)$. Then, MME performs the *update location* operation with HSS by exchanging the *Update location request* and *response* messages. This operation takes $2\times T_{MME\text{-}HSS}(S_c)$. MME also sends a *Create session request* with a PBU message to S-GW. This operation takes $T_{MME\text{-}SGW}(S_c)$. S-GW will perform the PBU and PBA operations with P-GW. This operation takes $2\times T_{SGW\text{-}PGW}(S_c)$. The S-GW will respond with a *Create session response* with a PBA message to MME. This operation takes $T_{MME\text{-}SGW}(S_c)$. MME will perform the initial context setup operation with eNB by exchanging the *Initial context setup request* and *response* messages. This operation takes $2\times T_{eNB\text{-}MME}(S_c)$. Then, MN sends the *Attach complete* message to MME, which takes $T_{MN\text{-}MME}(S_c)$, where $T_{MN\text{-}MME}(S_c) = T_{MN\text{-}eNB}(S_c) + T_{eNB\text{-}MME}(S_c)$. MME will perform the modify bearer operation with S-GW by exchanging the *Modify bearer request* and *response* messages. This operation takes $2\times T_{MME\text{-}SGW}(S_c)$. Accordingly, the binding update delay of PMIP-LB-SAE-PGW can be represented as follows.

$$\begin{aligned} BUD_{PMIP\text{-}LB\text{-}SAE\text{-}PGW} &= T_{MN\text{-}MME}(S_c) + T_{eNB\text{-}MME}(S_c) + 2\times T_{MME\text{-}HSS}(S_c) \\ &\quad + T_{MME\text{-}SGW}(S_c) + 2\times T_{SGW\text{-}PGW}(S_c) + T_{MME\text{-}SGW}(S_c) \\ &\quad + 2\times T_{eNB\text{-}MME}(S_c) + T_{MN\text{-}MME}(S_c) + 2\times T_{MME\text{-}SGW}(S_c) \\ &= T_{MN\text{-}eNB}(S_c) + T_{eNB\text{-}MME}(S_c) + T_{eNB\text{-}MME}(S_c) \\ &\quad + 2\times T_{MME\text{-}HSS}(S_c) + 2\times T_{MME\text{-}SGW}(S_c) + 2 \\ &\quad \times T_{SGW\text{-}PGW}(S_c) + 2\times T_{eNB\text{-}MME}(S_c) + T_{MN\text{-}eNB}(S_c) \\ &\quad + T_{eNB\text{-}MME}(S_c) + 2\times T_{MME\text{-}SGW}(S_c) \\ &= 2\times T_{MN\text{-}eNB}(S_c) + 5\times T_{eNB\text{-}MME}(S_c) + 2 \\ &\quad \times T_{MME\text{-}HSS}(S_c) + 4\times T_{MME\text{-}SGW}(S_c) + 2 \\ &\quad \times T_{SGW\text{-}PGW}(S_c) \end{aligned}$$

The binding query delay for PMIP-LB-SAE-PGW can be calculated as follows. First, CN sends a PBQ message to P-GW so as to find the location of MN. Then, P-GW responds to CN with a PQA message. This takes $2\times T_{eNB\text{-}SGW}(S_c) + 2\times T_{SGW\text{-}PGW}(S_c)$. Thus, the binding query delay of PMIP-LB-SAE-PGW can be represented as follows.

$$BQD_{PMIP\text{-}LB\text{-}SAE\text{-}PGW} = 2 \times T_{eNB\text{-}SGW}(S_c) + 2 \times T_{SGW\text{-}PGW}(S_c)$$

For data delivery in PMIP-LB-SAE-PGW, CN sends the data packet directly to MN over an optimal data path. Then, the data delivery delay of PMIP-LB-SAE-PGW can be represented as follows.

$$DDD_{PMIP\text{-}LB\text{-}SAE\text{-}PGW} = T_{CN\text{-}eNB}(S_d) + T_{eNB\text{-}eNB}(S_d) + T_{MN\text{-}eNB}(S_d)$$

So, we obtain the overall transmission delay of the PMIP-LB-SAE-PGW as

$$TD_{PMIP\text{-}LB\text{-}SAE\text{-}PGW} = BUD_{PMIP\text{-}LB\text{-}SAE\text{-}PGW} + BQD_{PMIP\text{-}LB\text{-}SAE\text{-}PGW} + DDD_{PMIP\text{-}LB\text{-}SAE\text{-}PGW}.$$

3.3. PMIP-LB-SAE-MME

It is noted that the binding update delay of PMIP-LB-SAE-MME is the same with that of PMIP-LB-SAE-PGW. Thus we get

$$BUD_{PMIP\text{-}LB\text{-}SAE\text{-}MME} = 2 \times T_{MN\text{-}eNB}(S_c) + 5 \times T_{eNB\text{-}MME}(S_c) + 2 \times T_{MME\text{-}HSS}(S_c) + 4 \times T_{MME\text{-}SGW}(S_c) + 2 \times T_{SGW\text{-}PGW}(S_c)$$

The binding query delay for PMIP-LB-SAE-MME can be calculated as follows. First, CN sends a PBQ message to MME so as to find the location of MN. Then, MME responds to CN with a PQA message. This takes $2 \times T_{eNB\text{-}MME}(S_c)$. Thus, the binding query delay of PMIP-LB-SAE-MME can be represented as follows.

$$BQD_{PMIP\text{-}LB\text{-}SAE\text{-}MME} = 2 \times T_{eNB\text{-}MME}(S_c)$$

The data delivery delay of PMIP-LB-SAE-MME is the same with that of PMIP-LB-SAE-PGW. Thus, we get

$$DDD_{PMIP\text{-}LB\text{-}SAE\text{-}MME} = T_{CN\text{-}eNB}(S_d) + T_{eNB\text{-}eNB}(S_d) + T_{MN\text{-}eNB}(S_d)$$

So, we obtain the overall transmission delay of the PMIP-LB-SAE-MME as

$$TD_{PMIP\text{-}LB\text{-}SAE\text{-}MME} = BUD_{PMIP\text{-}LB\text{-}SAE\text{-}MME} + BQD_{PMIP\text{-}LB\text{-}SAE\text{-}MME} + DDD_{PMIP\text{-}LB\text{-}SAE\text{-}MME}.$$

4. Numerical Results

Based on the analysis given so far, we compare the performance of the existing and proposed schemes. For numerical analysis, we configure the parameter values, as described in Table 3, in which some of the values are taken from [14].

The following notations are used for analysis. S_c is the size of control packets, and S_d is the size of the data packets. L_w is the wired link delay among eNB, S-GW, P-GW, MME and HSS, whereas L_{wl} is the wireless link delay between host and eNB. $H_{a\text{-}b}$ is the hop count between the two nodes, *a* and *b*, in the mobile network. B_{wl} is the bandwidth of wireless link between host and eNB, whereas B_w is the bandwidth of wired link among eNB, S-GW, P-GW, MME and HSS. N_{host} is the total number of hosts in the mobile network, whereas N_{data} is the average number of data packets transmitted by each host. In the meantime, q is the probability of wireless link failure and T_q is the average queuing delay at each node.

For numerical analysis, the default values of the associated parameters are configured, as indicated in Table 3. Among these parameters, we note that $H_{eNB\text{-}SGW}$, $H_{SGW\text{-}PGW}$, N_{host}, N_{data}, and T_q may depend on the mobile network conditions. Thus, we will compare the performance of candidate schemes by varying these parameter from the minimum value to the maximum value.

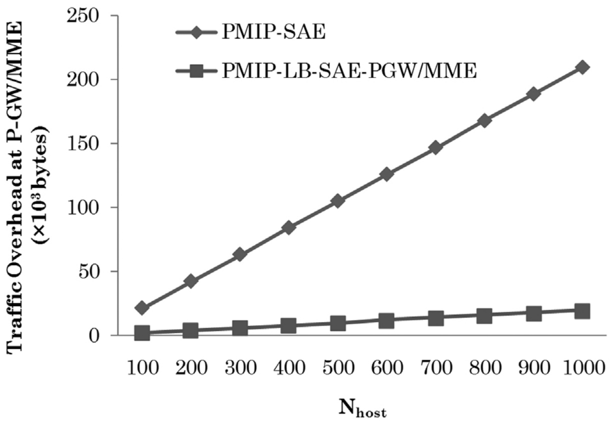

Figure 10. Impact of N_{host} on the PMIP traffic overhead

Table 3. parameter values used for analysis

Parameter	Default	Minimum	Maximum
T_q(ms)	5	1	55
$H_{eNB\text{-}SGW}$	2	1	55
N_{host}	500	100	1000
N_{data}	10	1	100
$H_{SGW\text{-}PGW}$	3	1	55
q	0.5		
$H_{eNB\text{-}MME}$	2		
$H_{MME\text{-}HSS}$	3		
$H_{MME\text{-}SGW}$	2		
$H_{eNB\text{-}eNB}$	2		
L_{wl}(ms)	10ms		
Lw	2ms		
S_c	96bytes		
S_d	200bytes		
B_{wl}	11Mbps		
B_w	100Mbps		

4.1. Traffic Overhead (TO)

Figure 10 and Figure 11 compare the traffic overhead (the total number of control and data packets) to be processed by P-GW/MME. It is noted that the two proposed schemes give the same traffic overhead, since the binding update/query operations are the same for those two schemes, and the only difference is the location of MCA (over PGW or MME).

Figure 10 shows the impacts of the number of hosts in the networks (N_{host}) on the traffic overhead. From the figure, we can see that the proposed schemes provide lower traffic overhead than the existing scheme. This is because all of the control and data messages shall be processed by P-GW in PMIP-SAE, whereas in the proposed PMIP schemes only the control traffics for binding update/query are processed by P-GW or MME, and the data traffics are processed by eNB. The gaps of performance between the existing and proposed schemes get larger, as the number of hosts in the network increases.

Figure 11 shows the impacts of the number of data packets (N_{data}) on the traffic overhead. From the figure,

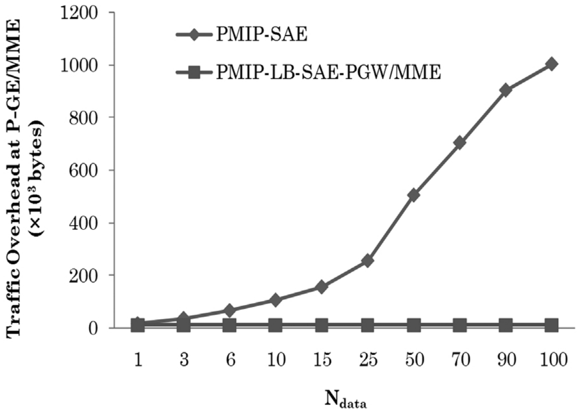

Figure 11. Impact of N_{data} on the PMIP traffic overhead

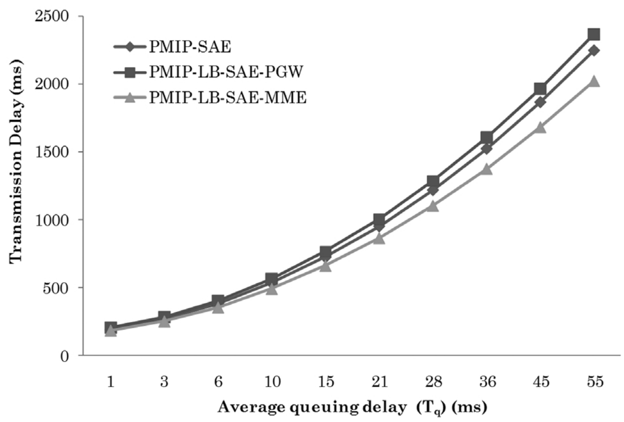

Figure 12. Impact of T_q on transmission delay

we can see that N_{data} gives significant impacts on traffic overhead for the existing scheme. This is because the traffic overhead of the existing PMIP-SAE scheme depends on a centralized P-GW in the mobility control and data delivery operations, and all data traffics should be delivered to P-GW. In the meantime, the traffic overhead of the proposed PMIP-LB-SAE-PGW/MME schemes are not affected by the number of data packets, since all data traffics are delivered directly between eNBs over an optimal data path.

4.2. Transmission Delay (TD)

Figure 12 compare the transmission delay for different average queuing delay (T_q) at each node. It is shown in the figure that the transmission delay linearly increases, as T_q gets larger at each node, for all candidate schemes. However, PMIP-LB-SAE-PGW gives worse performance than PMIP-LB-SAE-MME. This is because the PMIP-LB-SAE-PGW performs the binding query operation with P-GW that is usually more distant from eNB, compared to MME. However, it is noted that the proposed PMIP-LB-SAE-MME scheme gives the best performance among the candidate schemes.

Figure 13 shows the impact of the hop counts between S-GW and P-GW ($H_{SGW\text{-}PGW}$). In the figure, we can see that the transmission delay linearly increases for all of the candidate schemes, and that the proposed PMIP-LB-SAE-PGW and PMIP-LB-SAE-MME schemes give better performance than the existing PMIP-SAE schemes. This is because all of the control and data traffics go through S-GW and P-GW in the existing scheme, whereas in the

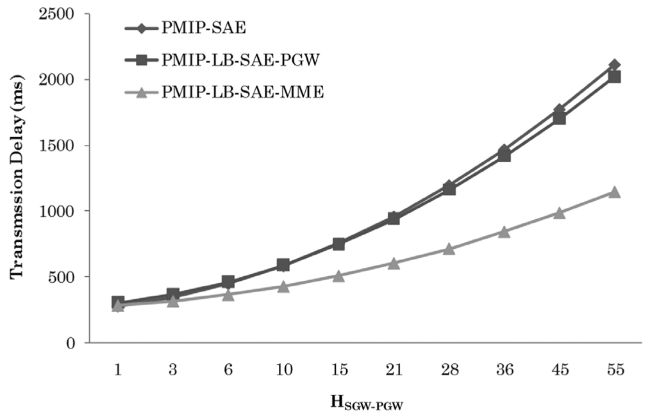

Figure 13. Impact of $H_{SGW-PGW}$ on transmission delay

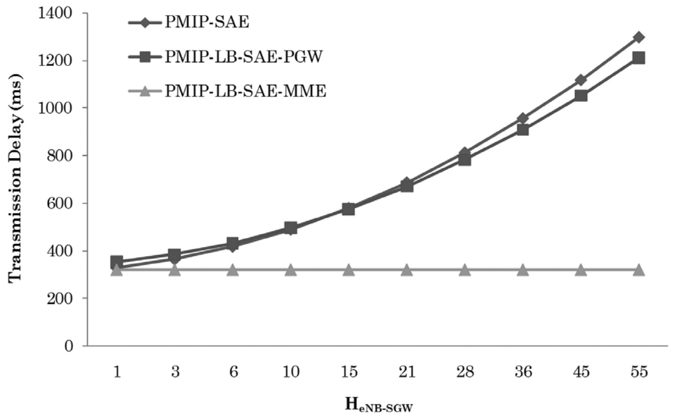

Figure 14. Impact of $H_{eNB-SGW}$ on transmission delay

proposed schemes only the binding control traffics are processed at P-GW or MME. The proposed PMIP-LB-SAE-MME scheme gives the best performance among the candidate schemes. The gaps of performance get larger, as $H_{SGW-PGW}$ increases.

Figure 14 compares the transmission delay of the candidate schemes for different hop counts between eNB and S-GW ($H_{eNB-SGW}$). In the figure, we can see that $H_{eNB-SGW}$ gives significant impacts on transmission delay for the existing PMIP-SAE scheme and the proposed PMIP-LB-SAE-PGW scheme. This is because all of the binding control and data traffic in these schemes go through eNB and P-GW. In the meantime, the PMIP-LB-SAE-MME scheme is not affected by $H_{eNB-SGW}$, and it provides the best performance among all the candidate schemes. The gaps of performance get larger, as $H_{eNB-SGW}$ increases.

V. CONCLUSIONS

In this paper we proposed the load balancing schemes for PMIP mobility control in SAE-based mobile networks. The proposed schemes are featured by the separation of data delivery function and mobility control function. Each eNB, instead of S-GW, will function as the MAG of PMIP. For the binding update and query functions, we define a new Mobility Control Agent (MCA). As per the location of MCA, the proposed schemes are divided into the PMIP-LB-SAE-PGW and PMIP-LB-SAE-MME. Before data transmission, eNB will obtains an optimal data path by using the binding query function with MCA.

For performance analysis, we compared the two proposed load balancing schemes with the existing PMIP scheme. From the numerical results, we see that the proposed PMIP load balancing schemes can give better performance than the existing PMIP scheme in the SAE-based mobile networks in terms of traffic overhead and transmission delay. In a certain network condition, the PMIP-LB-SAE-PGW scheme may give worse performance than the existing PMIP-SAE scheme. However, it is noted that the proposed PMIP-LB-SAE-MME scheme can give the best performance among the candidate scheme.

Acknowledgment

This research was partly supported by the Basic Science Research Program of NRF(2010-0020926).

[References]

[1] Morgan Stanley Report, *Internet trends*, Apr. 2010.

[2] P. Bosch, et al., "Flat cellular (UMTS) networks," *Conference of WCNC, Hong Kong*, Mar. 2007.

[3] K. Daoud, *et al.*, "UFA: Ultra Flat Architecture for high bit rate services in mobile networks," *Conference of PIMRC*, Sep. 2008.

[4] ZoltánFaigl, *et al.*, "Evaluation and Comparison of Signalling Protocol Alternatives for the Ultra Flat Architecture," *Conference of ICSNC*, Aug. 2010.

[5] D. Johnson, *et al.*, Mobility Support in IPv6, IETF RFC 3775, Jun. 2004.

[6] S. Gundavelli, *et al.*, Proxy Mobile IPv6, IETF RFC 5213, Aug. 2008.

[7] H. Chan, *et al.*, Requirements for Distributed Mobility Management, IETF Internet Draft, draft-ietf-dmm-requirements-15, Mar. 2014.

[8] H. Yokota, *et al.*, Distributed Mobility Management: Current Practices and Gap Analysis, IETF Internet-Draft, draft-ietf-dmm-best-practices-gap-analysis-03.txt, Feb. 2014.

[9] 3GPP TR 23.402, Technical Specification Group Services and System Aspects: Architecture enhancements for non-3GPP accesses, V10.7.0, Mar. 2012.

[10] Julien Laganier, *et al.*, "Mobility Management for All-IP Network," *NTT DOCOMO Technical Journal*, Vol. 11, No. 3, Dec. 2009, pp. 34-39.

[11] ETSI TR 123 401, LTE; General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access, V10.5.0, Nov. 2011.

[12] S. Krishnan, *et al.*, Localized Routing for Proxy Mobile IPv6, IETF RFC 6705, Sep. 2012.

[13] Makaya, C. and Pierre, S., "An analytical framework for performance evaluation of IPv6-based mobility management protocols," *IEEE Transaction on Wireless Communication*, Vol. 7., No. 3, 2008, pp. 972-983.

[14] M. Gohar, *et al.*, "A Distributed Mobility Control Scheme in LISP Network," *Wireless Networks*, Vol. 20. No. 2, Feb. 2014, pp. 245-259.

Moneeb Gohar

He received B. S. degree in Computer Science from University of Peshawar, Pakistan, and M. S. degree in Technology Management from Institute of Management Sciences, Pakistan, in 2006 and 2009, respectively. He received Ph. D. degree from the school of Computer Science and Engineering in the Kyungpook National University, Korea, in 2012. He is now with the Kyungpook National University, as a postdoctoral researcher. His current research interests include Network Layer Protocols, Wireless Communication, Mobile Multicasting and Internet Mobility.

E-mail: moneebgohar@gmail.com

Sang-Il Choi

He received B.S and M.S. degrees in School of Computer Science and Engineering from Kyungpook National University in 2010 and 2012, respectively. Since March 2012, he is with the School of Computer Science and Engineering from Kyungpook National University as a Ph. D. candidate. His current research interests include mobile communication and lighting control network.
E-mail: overcycos@gmail.com

Seok-Joo Koh

He received the B.S. and M.S. degrees in Management Science from KAIST in 1992 and 1994, respectively. He also received Ph.D. degree in Industrial Engineering from KAIST in 1998. From August 1998 to February 2004, he worked for Protocol Engineering Center in ETRI. He has been as a professor with the school of Computer Science and Engineering in the Kyungpook National University since March 2004. His current research interests include mobility management in the future Internet, IP mobility, multicasting, LED-based visible lights communication, IoT and SCTP. He has so far participated in the international standardization as an editor in ITU-T SG13 and ISO/IEC JTC1/SC6.
E-mail: sjkoh@knu.ac.kr