Performance Analysis of A Novel Inter-Networking Architecture for Cost-Effective Mobility Management Support

Myungseok Song[1], Jongpil Jeong[2]

[1]Network R&D Team, BIZtelecom Corp.
#1203 1st LG TWINTEL, 157-8 Samsung-dong Gangnam-gu, Seoul, 135-880, Korea
[e-mail : bestsong21@hotmail.com]
[2]College of Information and Communications, Sungkyunkwan University
2066 Seobu-ro Jangan-gu, Suwon, Kyunggi-do, 440-726, Korea
[e-mail: jpjeong@skku.edu]
[*]Corresponding Author: Jongpil Jeong

Received September 9, 2013; revised December 26, 2013; revised February 28, 2014; accepted March 25, 2014; published April 29, 2014

Abstract

Mobile traffic is increasing a masse because of the propagation of the Internet and the development of wireless mobile technology. Accordingly, the Network Local Mobility Management (NETLMM) working group [1] of the Internet Engineering Task Force (IETF) has standardized Proxy Mobile IPv6 (PMIPv6) [2] as a protocol for accomplishing the transmissibility of mobile terminals. PMIPv6 is a network-led IP-based mobility management protocol, which can control terminal mobility without depending on the type of access system or the capability of the terminal. By combining PMIPv6 and the mobility of Session Initiation Protocol (SIP), we can establish terminal mobility and session mobility through a more effective route. The mobility function can be improved and the overlap of function reduced as compared to that in the case of independent operation. PMIPv6 is appropriate for a non-real-time service using TCP, and SIP is appropriate for a real-time service using RTP/UDP. Thus, in the case of a terminal using both services, an effective mobility management is possible only by using PMIPv6 together with SIP. In order to manage mobility in this manner, researches on PMIPv6-SIP are in progress. In line with this trend, this paper suggests a new PMIPv6-SIP architecture where when a mobile terminal conducts a handover, a network-led handover while maintaining the session without the addition of a special function or middleware is possible along with effective performance evaluation through mathematical modeling by comparing the delay and the packet loss that occur during the handover to the Pure-SIP.

Keywords: PMIPv6, MIPv6, PMIPv6-SIP, MIPv6-SIP, Pure-SIP, Mobility Management

http://dx.doi.org/10.3837/tiis.2014.04.011

1. Introduction

As Mobile Nodes (MNs) such as tablet PCs and smart phones have advanced, wireless LAN services have become widely used. Technologies to handover various types of wireless networks with session maintenance are also being introduced. In the current public wireless LAN environment, however, the IP address may change as one wireless LAN router after another is used while the user moves. Since the IP address of the MN is changed when the MN perform the handover, it is difficult to maintain the TCP connection resulting in frequent disconnections. It is vital, therefore, to prevent communication from being disconnected during the handover of the MN. To implement an advanced handover technology, Mobile IPv6 (MIPv6) [3] and SIP [4] have been developed. However, middle ware for the MN is necessary to be able to use MIPv6 and SIP. MIPv6 detours the data to the Home Agent (HA), the server that controls the MN location and hides the change-over of the IP address from the mobile device. This approach maintains the session during handover execution. Middle ware is necessary for the MN and HA detouring is essential for communication, which causes the communication route to be expanded. In the case of SIP, which is a protocol to establish, modify, and cut off the session, it can set up the optimal communication route between terminals and also enable terminals to communicate directly. For mobile SIP, however, the SIP session needs to be reset when the MN's IP address is changed, which disconnects the communication line.

Evolved Packet Core (EPC) specified by 3rd Generation Partnership Project (3GPP) Release 8 System Architecture Evolution (SAE) standardization adopts PMIPv6 [2], an IP-based mobile management protocol. PMIPv6 is a common mobile management method that supports handovers in various environments including standardized wireless protocols such as WLAN, World interoperability for Microwave Access (WiMAX) [6], 3rd Generation Partnership Project2 (3GPP2) as well as relocation of terminals in Long Term Evolution (LTE) telecommunication [7]. PMIPv6 executes a network-led handover through a Mobile Access Gateway (MAG) in a network infrastructure. An MN does not need other special functions for the handover between the Local Mobility Anchor (LMA) and MAG (network-led handover). However, the communication route is longer since it always needs to detour LMA.

This study suggests a new PMIPv6-SIP architecture that enables a network-led handover without disconnection and without any special function or middle ware for the MN handover. The suggested architecture adopts PMIPv6 to manage terminal mobility, and SIP to manage the centralized session mobility. New information is added to SIP methods for this architecture. The model to which the suggested new architecture and Pure-SIP are added is analyzed, and the mathematical analysis model is presented to evaluate performance.

The components of this study are as follows: Chapter 2 describes related studies. Chapter 3 suggests a new method to manage mobility. This method includes SIP-applied PMIPv6-SIP. Chapter 4 presents the result of performance evaluation and numerical analysis. Chapter 5 presents the conclusion.

2. Related Work

2.1 IMS-based Session Management

Universal Mobile Telecommunications System (UMTS) release 5 introduced IP Multimedia Subsystem (IMS) in the Core Network (CN) [8] to control multimedia sessions for the first time. The core elements of IMS include Call State Control Function (CSCFs) and Home Subscriber Server (HSS). HSS, which is an advanced version of the Home Location Register (HLR), functions as the central database to manage member information. It handles such tasks as user registration, modification, authentication, authorization, session routing, charging, etc. CSCF functions to handle calls and sessions. The major tasks include the incoming call gateway, call controlling, Serving Profile Database (SPD), address handling, etc. CSCFs are divided into Proxy-CSCF (P-CSCF), Interrogating-CSCF (I-CSCF), and S-CSCF Serving-CSCF (S-CSCF). P-CSCF is the first point that the User (UE) meets when accessing IMS through General Packet Radio Service (GPRS) access. This is also the first point where UE accesses IMS, and exists in such domains as Gateway GPRS Support Node (GGSN). I-CSCF plays a role as the connecting point for all calls coming in within the network and as the connecting point with roaming members in another network. In the latter case, it functions as a firewall. I-CSCF determines S-CSCF by inquiring HSS, and allots S-CSCF to UE in the registration process. S-CSCF not only manages all sessions of IMS, but also receives member profiles in association with HSS, and executes major functions to handle calls. S-CSCF is an SIP server that registers users at IMS and manages the sessions. IMS is the basic protocol that manages signaling, session, and mobility by means of SIP. In general, 3GPP2 or IMS have been adopted. For multimedia session handling, IMS has been introduced in 3GPP2 or Packet Data Subsystem (PDS) as in PDS [9]. Although the initial version of 3GPP2-IMS was affected by the specific IMS list of 3GPPS, there are several differences as summarized below:

3GPP-IMS
☐ uses SIP for session mobility.
☐ is based on IPv6 CN.
☐ manages the user data by means of HSS.
☐ adopts PDP context activation for P-CSCF Discovery.
☐ has P-CSCF and Associated Home of Visiting GGSN in the same network.

3GPP2-IMS
☐ uses MIP for IP mobility.
☐ uses SIP for session mobility management.
☐ secures flexibility in the use of IPv4 or IPV6.
☐ handles subscriber information by means of Home AAA server and database.
☐ does not use PDP Context activation for P-CSCF Discovery.
☐ has no need to keep P-CSCF and PDSN within the same network.

2.2 Proxy Mobile IPv6 Overview

MIPv6 is a mobility management technology that informs HA of the relocation of a MN. It is possible to exchange packets with the communication partner even after relocation by informing him of the combination of Care-of Address (CoA) on the network as well as Home Address (HoA). While MIPv6 is a terminal-led mobility technology, PMIPv6 is a network-led mobility controlling technology, as in **Fig. 1** [10].

Since MAG, the network node, detects relocation of a MN, there is no need to control mobility by means of the mobile terminal. The PMIPv6-Domain, a network to which PMIPv6 is applied, detects access by a mobile device by means of MAG, which is the network node, and informs LMA of the PMIPv6-Domain of the MN's Home Network Prefix (HNP) allotment and of the Proxy-CoA, which is the IP address of the MAG.

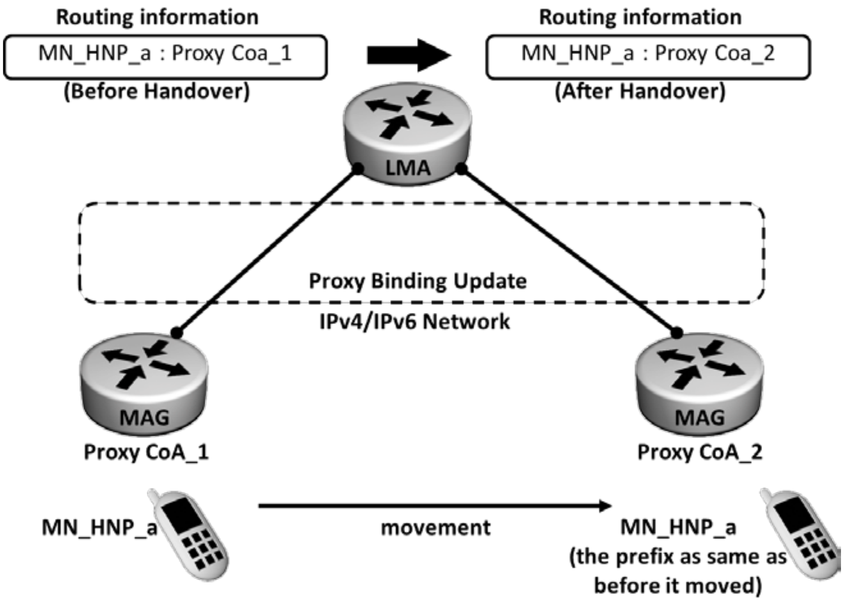

Fig. 1. Proxy Mobile IPv6 Brief.

The packets transmitted to LMA are capsulized through the MAG that the mobile device accesses so that they can be transmitted to the MN. Since PMIPv6 does not involve a MN for controlling mobility, the three types of MNs - IPv4, IPv6, and Dual-Stack are supported.

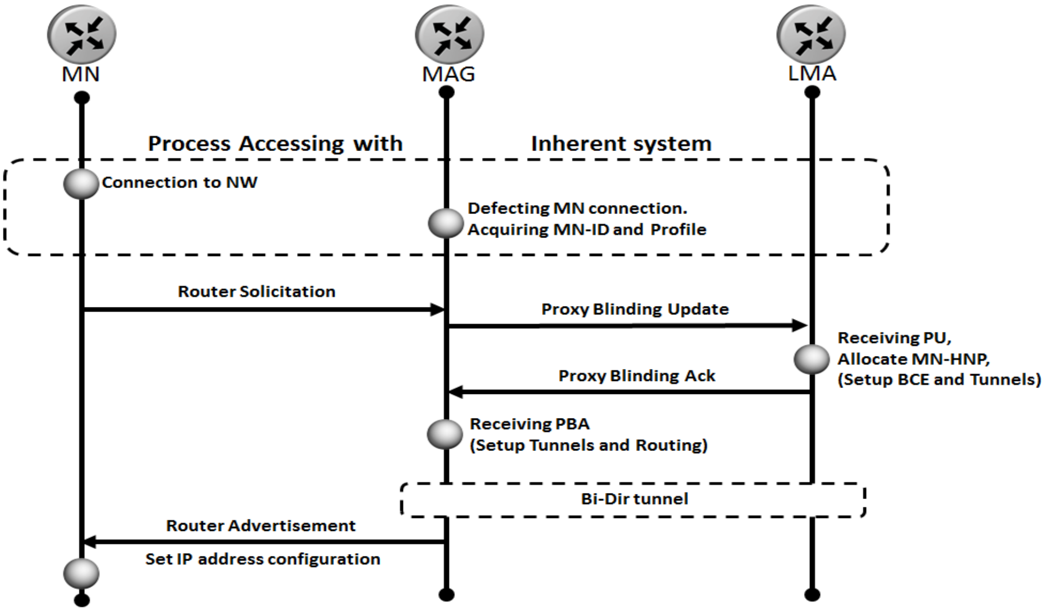

Fig. 2. Information Flow when the initial connection occurs.

Fig. 2 shows the flow of information when a MN accesses the PMIPv6-Domain in the initial stage. MAG detects a MN access and transmits Proxy Binding Update (PBU) to the LMA. Router Solicitation (RS) messages arrive randomly after the MN's access, and are controlled separately from PBU transmission. LMA confirms the validity of PBU, allots HNP to the MN, and transmits to the MAG Proxy Binding Acknowledgement (PBA). The interactive tunnel between the MAG and LMA is set up by writing a Binding Cache Entry (BCE). The MAG sets up the interactive tunnel with the LMA upon receiving the PBA, and then establishes the channel to the data traffic of the MN. The MAG transmits the Router Advertisement (RA) to the mobile device and informs the HNP received from the LMA. Based on the HNP acquired by the RA, the MN sets up the stated or stateless address and completes the address acquisition. In the process above, the HNP-based address is set in the MN while the HNP routing information is set in MAG and LMA. In the PMIPv6-Domain, the MAG functions as the default gateway for the MN. The packets transmitted from the MN to the MAG are sent to the LMA through the tunnel. The LMA removes the capsulizing head of the received packets, and transmits it to an external network or into the PMIPv6-Domain depending on the routing information of the LMA. Fig. 3 shows the information flow during a MN handover between MAGs. When a MN goes into the control range of a new MAG that is moving, the latter detects the MN's access and handles the process in the order illustrated in Fig. 3. This way, an interactive tunnel between the new MAG and LMA is established. The new MAG sends RA to RS received from the MN, and informs the MN of HNP from LMA.

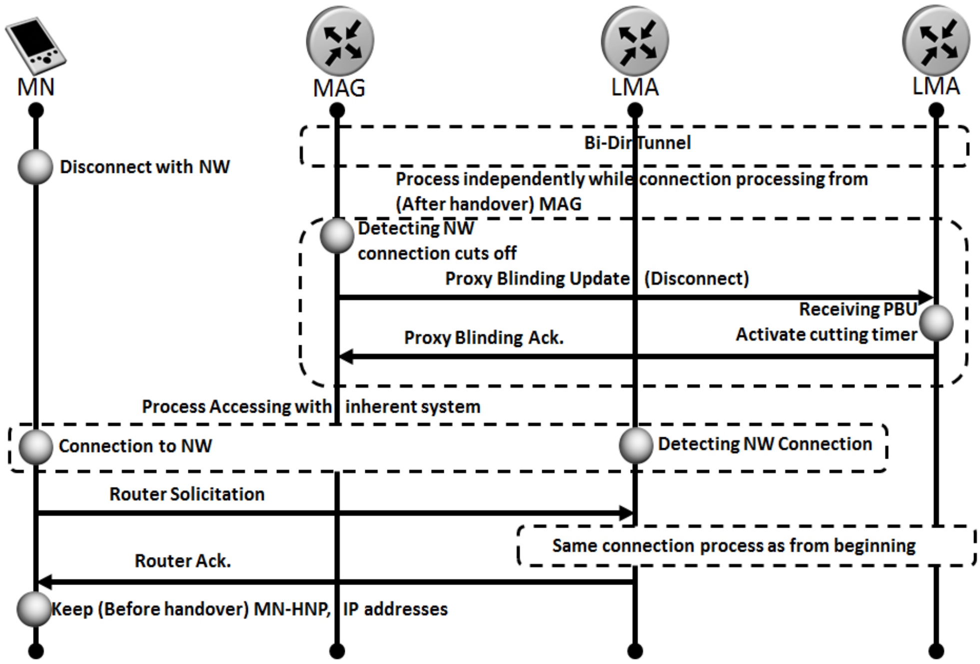

Fig. 3. The Flow of Information during Handover.

As the MN shares the same information with the HNP, the communication maintains the same IP address. The MAG connected before the relocation detects the disconnection with the MN, transmits the PBU to the LMA, and requests the deletion of the BCE. After a while, the LMA deletes the BCE and transmits the PBA to the MN. Each step of the process above may be executed independently.

Mobility management [8] has to solve the problem of delays before and after a handover that do not satisfy the requirements of an application in an IP-based voice data network where real-time applications such as voice call are necessary. Mobility management handles PMIPv6 domains. LMA and MAG are defined in PMIPv6, and the mobility in the network is supported in cooperation between these two elements. In this case, mobility management includes setting the communication channel in advance prior to resuming communication in the case of a handover in the PMIPv6 domain. This minimizes the delay when resuming communication. The specific steps are as follows.

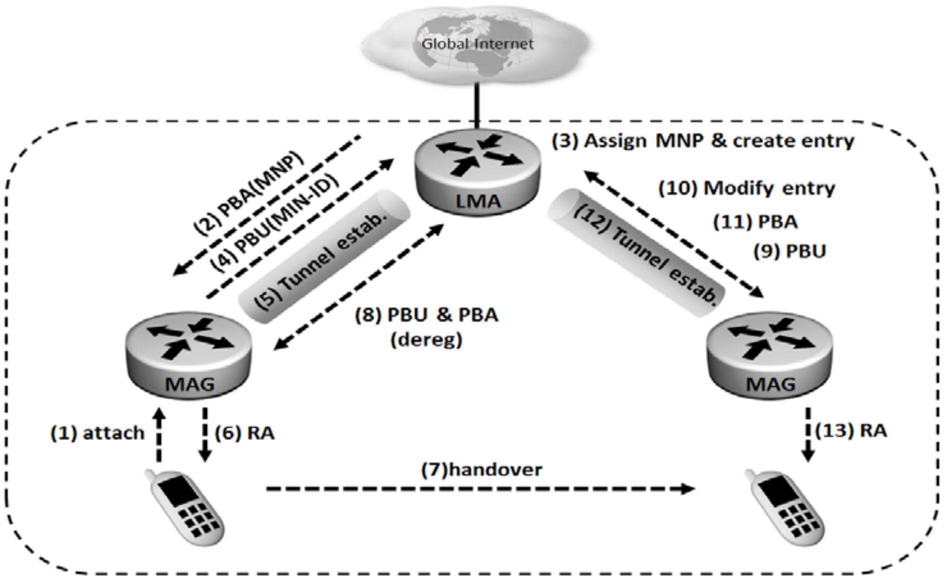

Fig. 4. Handover Brief.

For mobility management, a server that manages the entire network that is called the Proxy Information Server (PIS) is arranged within the domain. The PIS acquires all MAG information within the domain among all of the data sets in the network, and then grasp MAG that maintains the connectivity with each MAG. When the MN tries to cut off the communication from the MAG upon handover, the LMA is informed by the MAG of the timing, and then informs every adjacent MAG of the handover information. Upon handover of the CN, the LMA accesses a new MAG, which sends a PBA message to the LMA to resume communication with the MN.

(1) A MN accesses MAG-1. The MAG in PMIPv6 recognizes the access of a MN using link layer technology. (2) MAG-1 transmits PBU messages to the LMA that include the MN-ID, which is an indicator of the MN. (3) The LMA allots HNP(64bit_prefix) which corresponds to the MN-ID, and writes the BCE. (4) The LMA transmits to the MAG-1 PBA and includes the HNP. (5)The process above establishes a tunnel between the LMA and MAG-1. (6) MAG-1 transmits to the MN RA that includes the HNP. Once the MN receives the RA, it generates the IPv6 address, and becomes available for communication. Packets arriving at the MN arrive at the LMA. The LMA transmits the packets to MAG-1 through the tunnel. MAG-1 removes the tunnel header from the received packets, and then transmits them to the MN. Packets from the MN are sent in reverse order. (7) The MN is handed over to the MAG-2 Area. (8) MAG-1 senses the disconnection of the MN communication, transmits the PBU to the LMA for De-Registration, and then the LMA returns the PBA. This process removes the tunnel between the LMA and MAG-1. (9) MAG-2 senses the access of the MN and transmits the PBU to the LMA. (10-11) The LMA renews the BCE, and returns the PBA that includes the HNP. (12) The process above establishes a tunnel between the LMA and MAG-2. (13) MAG-2 sends to MN RA which includes the HNP. Since the HNP involves no change upon handover, the MN can maintain communication regardless of handover. The handover procedure from (7) to (13) is the procedure from PMIPv6.

2.3 Comparison of Mobility Management Protocols

MIP and SIP are mobility management protocols that are appropriate for a TCP-based service and a UDP-based service, respectively. Thus, when a user uses both real-time and non-real-time services, a method that simultaneously uses the MIP and the SIP is required for effective mobility management. Related studies are being conducted with a focus on the interworking and the optimization of signaling between the MIP and the SIP; some examples of these studies are those on the EVOLUTE structure [11], the Multilayered Mobility Management (MMM) structure [12], and the Integrated MIP-SIP structure [13]. EVOLUTE is the most basic MIP-SIP interlocking method where MH selects between the MIP and the SIP according to the type of traffic of the session currently being serviced.

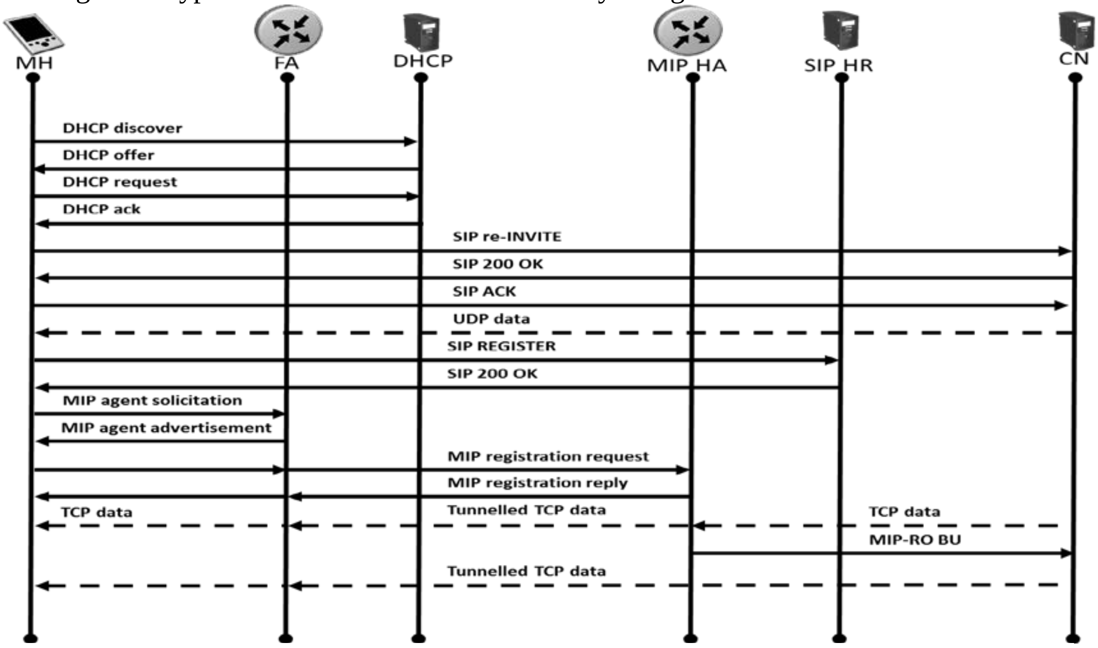

Fig.5. Mobility Management Procedure of EVOLUTE [11].

Fig.5 illustrates the operational process of EVOLUTE. That is, EVOLUTE uses the operation of the existing MIP and SIP, and is the simplest form of a multi-layered mobility management technique that merely combines the operational processes of the two protocols.

As shown in **Fig. 6**, MIP-LR achieves improved performance by performing routing optimization for CH in a similar manner as MIPv6. Further, MMM manages IP addresses through Dynamic and Rapid Configuration Protocol (DRCP), which is an enhanced DHCP. In other words, although the signaling process of MMM is similar to that of EVOLUTE, it seeks the optimization of the overall mobility management through various improved techniques.

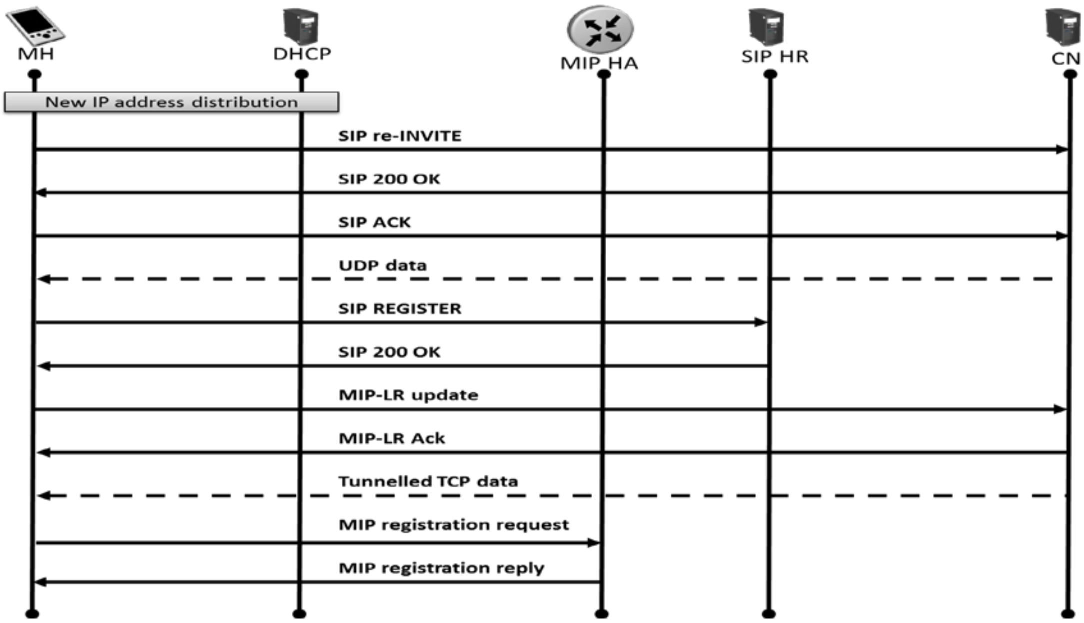

Fig.6. Mobility Management Procedure of MMM [12].

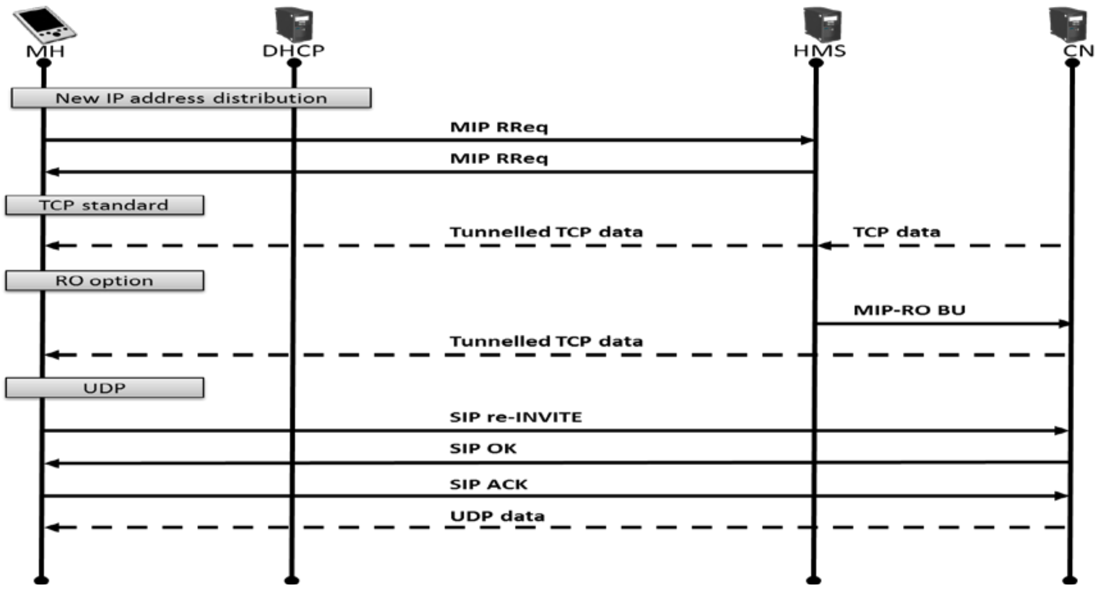

Fig. 7. Mobility management procedure of Integrated MIP-SIP [13].

The integrated MIP-SIP structure goes beyond simply combining the MIP and the SIP to perform effective mobility management by integrating the portions performing overlapping roles in the MIP and the SIP into a single portion. As evident from **Fig. 7**, the integrated MIP-SIP structure uses a home mobility server (HMS), which is a combination of MIP HA and SIP HR. While both the MIP and the SIP need to perform new registration in the home domain during the handover process, the integrated MIP-SIP structure can integrate the MIP and the SIP home servers through HMS and conduct it as a single process. That is, MH simultaneously conducts the home registrations of the MIP and the SIP in HMS through MIP RReq. Then, MH selectively performs the appropriate handover process depending on the type of service section that MH possesses. The above-described SIP-MIP interworking techniques are still at the stage of simply combining the two protocols. That is, after performing a network layer handover based on MIPv4, SIP mobility management with respect to the UDP session is performed once again. However, such a method poses several problems. First, it is difficult to guarantee the QoS of the UDP session. In the case where the MIP-RO option is not used, UDP sessions are delivered via tunneling by HA until the route optimization of the UDP session is completed through Re-INVITE, which leads to a prolonged transmission route and overhead generated through encapsulation. This makes it difficult to guarantee the QoS of the UDP session that transmits the real-time service. Further, SIP mobility management is a difficult handover technology that stops and resets the session and causes the user to experience service disconnections. In the SIP mobility management technology researched thus far, the period of service disconnection could be prolonged to several seconds, causing a significant inconvenience for the user using the service. This is well described in the results of [11], which uses SIP mobility management. In the next section, we suggest a technique that can perform macro-mobility management more effectively through information sharing between the MIP and the SIP, and signaling optimization

3. A New Interworking Architecture for Mobility Management

For a new interworking architecture to manage mobility, WiMAX accesses the ALL-IP core network through a Connectivity Services Network (CSN) gateway, UMTS through GGSN, CDMA2000 through PDSN, WLAN through GGSN emulator, LTE through S-GW (MAG) and P-GW (LMA) respectively. Each network has a MAG at one local P-CSCF. Other elements of SIP and PMIPv6-MAG are located in the home network of the MN. Thus, SIP is used to manage the central session mobility, while PMIPv6 is used to manage the terminal mobility.

The suggested internetworking architecture is described in **Fig. 8**. Each network accesses the ALL-IP core network through the gateway. The data flows detour the home network, moving from the departure point to the destination. SIP based on session controlling signals such as Call Setup, Call Termination, and Session Management are transmitted through the home network. The session controlling signals are transmitted to S-CSCF through I-SCSF of the home network by means of P-CSCF of the target network. The scenario of the session handover from UMTS to LTE is as follows:

☐ A data pipeline is set at the UMTS system.
☐ IP addresses are allotted to each MN through GGSN.
☐ PMIPv6 registration is requested through HA.
☐ SIP registration messages are transmitted to S-CSCF through P-CSCF.
☐ Upon approval, a proper S-CSCF is allotted.
☐ The member information is transmitted to the designated S-CSCF.

Thereafter, Packet Data Protocol (PDP) Context and service registration are activated, and then the MN is ready for a session.

As in **Fig. 9**, SIP INVITE messages are transmitted from the UMTS interface to S-CSCF through P-CSCF, and then finally to the destination. This is because it involves the request for a call flow model so that QoS, which is a prerequisite before a session is established, can be satisfied. Then the response to Session Description Protocol (SDP) and 183 Session Progress

Messages are sent to the destination. The confirmation of the temporary response upon request for this Precondition Acknowledgement (PRACK) is as follows:

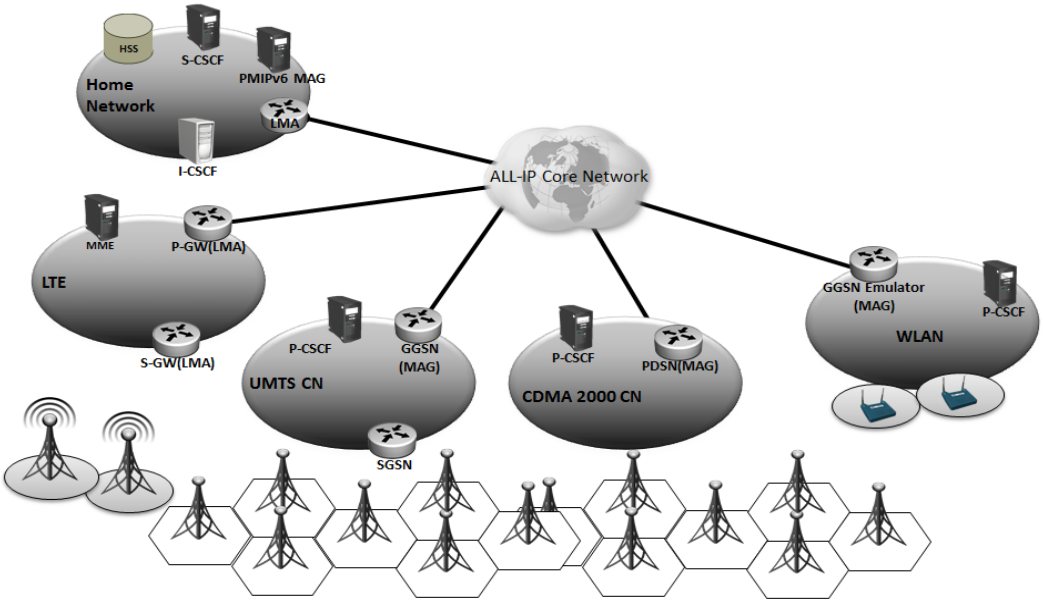

Fig. 8. The Proposed Interworking Architecture.

Once the request for PRACK arrives at the destination, 200 OK responses arrive with SDP responses. Then a request for updates that confirm resource reservation is transmitted at the next departure point. Once the destination receives the request for update, 200 OK Responses are generated, and then the session is executed through UMTS interface. When the MN moves to LTE, a session link to LTE is requested. The standard LTE registration procedure and PMIPv6 registration procedure with PMIPv6-MAG are implemented. This request is transmitted to PMIPv6-LMA, and the home IP address allotted to the LTE interface is acquired. Then to prevent the packets transmitted by the other node from detouring HA rather than being transmitted directly to the next node (Triangular Routing), the Binding Update Message of PMIPv6 is exchanged between the MN and the core network.

The next step is the SIP session handover procedure. This sends the Call-ID of the session in progress and the SIP Re-INVITE that includes other indicators to the destination. Then prerequisites for the LTE interface are reserved. Once this step is successfully implemented, a new session can be connected. It is noteworthy that while a new data transmission is initiated through the LTE interface, data transmission through the UMTS interface remains active for soft handover. In this study, therefore, the communication channel is connected, and then the existing channel is cut off according to the handover mechanism (make before break). Once the session handover technology of the core network is centralized, a handover is readily applied between UMTS-LTE-WiMAX-CDMA 2000-WLAN [15].

It is necessary to involve an SIP server with SIP messages extended to handle handovers between wireless LAN routers [16]. A standardized SIP server has mapping of IP addresses and SIPURI of an SIP terminal. The information on the terminal and session establishment is added to each SIP terminal's SIPURI and IP addresses. The terminal information refers to the link local address of the MN that accesses a wireless LAN router or a SIP terminal.

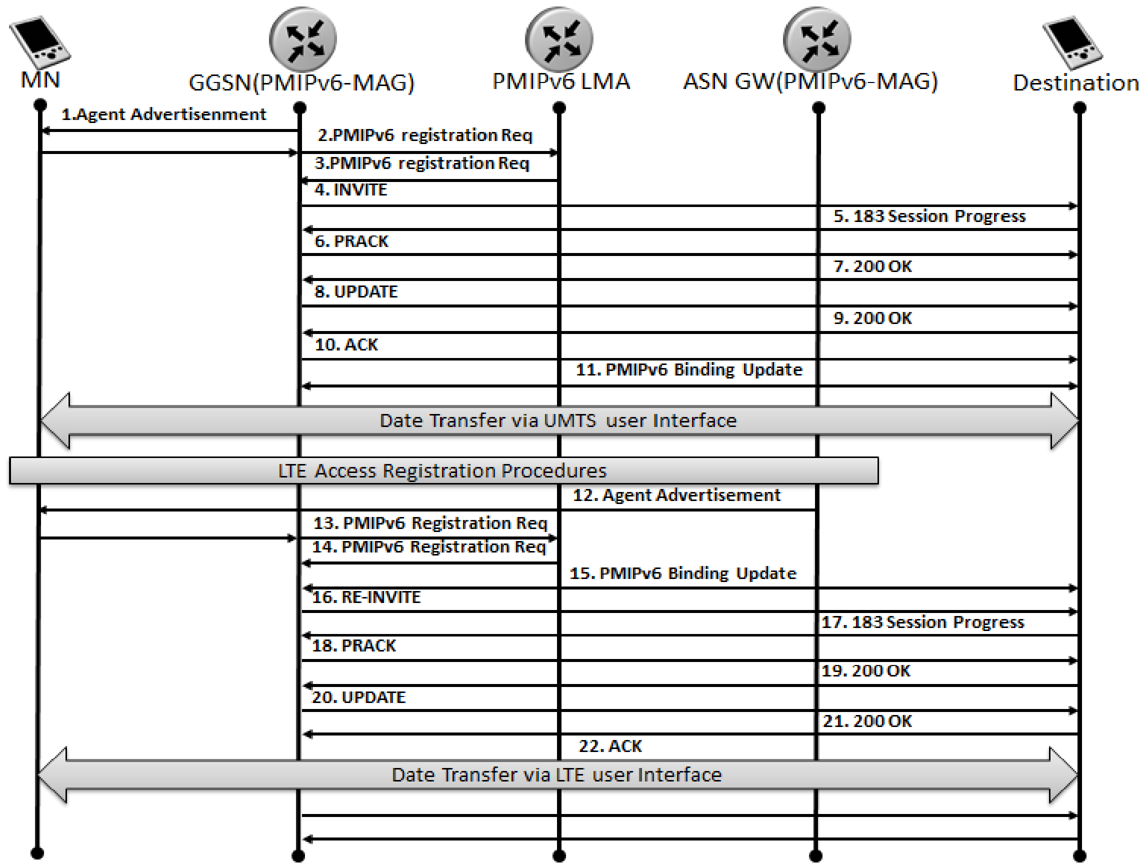

Fig. 9. PMIPv6-SIP Signaling.

Information on session establishment indicates the SIPURI of the wireless LAN router that the other terminal accesses. In the condition above, the location of a MN can be managed by means of the SIP server. SIP messages need to be extended to register this information to the SIP server. SIP messages may involve various methods, and the information is added to REGISTER, a message for the registration request, and INVITE, a message for the establishment request. REGISTER adds to the message the link local address of the MN connected to the wireless LAN router while INVITE adds to the message the IP address of the other party. This way, the SIP server maintains the session establishment information and MN information depending on each method, as in **Fig.10** [17].

The suggested communication environment utilizes PMIPv6 and SIP in an inter-terminal communication as in **Fig. 11**. With SIP introduced to a wireless LAN router environment, the session between wireless LAN routers is established, modified, and cut off for handover. The same IP address is allotted to the MN in order to hide the change of the IP address and to avoid adding more functions for handover to the MN. This way, the inter-terminal communication channel is optimized without adding special functions for handover to MNs, and to maintain the TCP connection.

```
REGISTER sip:[2001:db8::10] SIP/2.0
To: sip:user@example.com
From: sip:user@example.com;tag=81x2
Via: SIP/2.0/UDP [2001:db8::9:1];branch=z9hG4bKas3-111
Call-ID: SSG9559905523997077@hlau_4100
Max-Forwards: 70
Contact: "Caller" <sip:caller@[2001:db8::1]>
CSeq: 98176 REGISTER
Content-Length: 0
```

```
INVITE sip:user@[2001:db8::10] SIP/2.0
To: sip:user@[2001:db8::10]
From: sip:user@example.com;tag=81x2
Via: SIP/2.0/UDP [2001:db8::20];branch=z9hG4bKas3-111
Call-ID: SSG9559905523997077@hlau_4100
Contact: "Caller" <sip:caller@[2001:db8::20]>
CSeq: 8612 INVITE
Max-Forwards: 70
Content-Type: application/sdp
Content-Length: 268

v=0
o=assistant 971731711378798081 0 IN IP6 2001:db8::20
s=Live video feed for today's meeting
c=IN IP6 2001:db8::20
t=3338481189 3370017201
m=audio 6000 RTP/AVP 2
a=rtpmap:2 G726-32/8000
m=video 6024 RTP/AVP 107
a=rtpmap:107 H263-1998/90000
```

Fig. 10. IPv6 SIP Format (RFC5118)

MIPv6 detours the data at the location management server of the MN that is called the HA in order to hide the change of the IP address from the MN during the handover with the TCP connection. However, this requires middleware for the MN, and the communication channel is extended. SIP, a protocol to set up, modify, and cut off a session, establishes the optimal communication channel between terminals, and makes it possible for two terminals to communicate directly with each other. As for SIP mobility during session modification, the change of the MN's IP address makes it difficult to maintain the TCP connection to reestablish the session that uses the IP. There have been studies on SIP mobility, in which the change of the IP address is hidden from the Transport Layer of the MN, and the handover is realized while maintaining the TCP connection [18] but this too involves middleware for the MN. PMIPv6 realizes a network-led handover at the network by means of a MAG allotted between the LMA and the wireless LAN router. Handling handover between the LMA and the MAG removes the need to add some special functions to the MN for handover. Since the communication always involves the LMA, however, the communication channel is extended even though the change in the IP address is hidden from the MN and the TCP connection is not necessarily cut off. In a communication with the destination (CN) by means of the suggested method, the process of handover to the router after the MN moves from one to the next router is described in **Fig. 11**. In the initial status, the MN is connected to the former router, while the destination is connected to the other router.

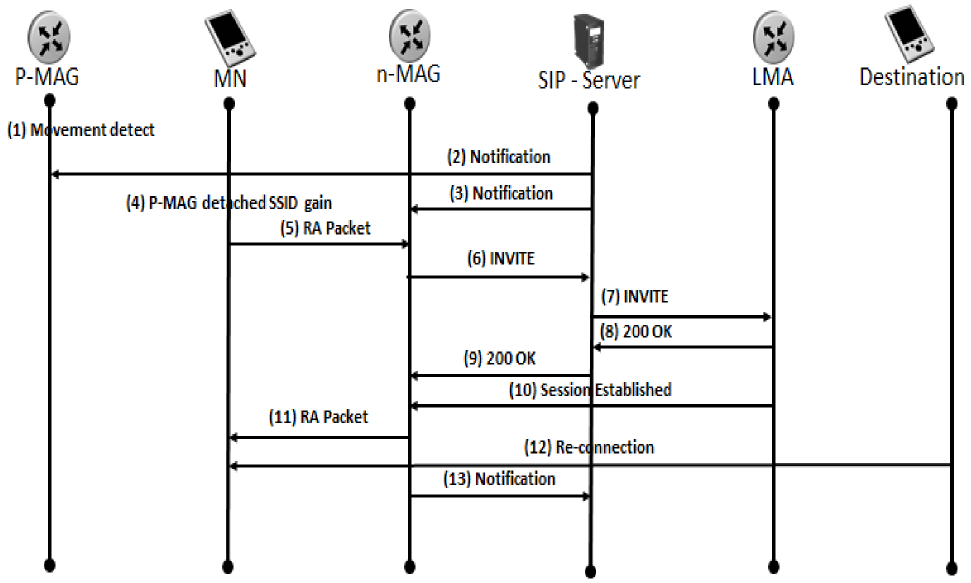

Fig. 11. Handover Procedure.

The suggested method is introduced to every router for the same function. Steps (4), (5) and (11) indicate functions of common MNs. The former router detects MNs that are likely to move based on the RSSI values, and informs the next router of the terminal information and session establishment information [steps (1) ~ (3)]. Since the information on the MN is sent to the next router, the router can establish a session without inquiring of the SIP server about the information on the MN. Once the MN is relocated and accesses the router [step (4) ~ (5)], the next router establishes a session with the other router by means of the information sent through step (3) [step (6) ~ (10)]. This process makes it possible for the next router to control the communication channel of the MN. Once the session is established, the link local address of the MN is allotted so that the MN at step (11) can keep the same address before and after the relocation. After the communication is resumed [step (12)], the next router informs the SIP server to renew the location information of the MN [step (13)].

4. Performance Evaluation

4.1 System Modeling

This section suggests an analysis model for mobility management assessment. First of all, it is assumed that each of the communication partners in a user session is independent, and that a handover method in which the communication channel is connected before the existing channel is cut off is used (make before break). The traffic is generated independently by each individual as part of all the traffic sets following the Poisson arrival process [19]. In addition, since background traffic may affect the suggested model, the worst scenario, in which subsets of traffics are transmitted irregularly on a small scale, was taken into consideration [20]. In **Table 1**, the standard vertical handover on mobility while a session is established is as follows:

Table 1.. Parameters for Performance Evaluation.

Parameter	Description
D_1	Handover delay in a link layer
D_2	Delay in detecting a relocation
D_3	Delay due to allotting a new address
D_4	Delay in resetting a session
D_5	Delay in packing retransmitting [21]

A vertical handover delay in an upper layer did not consider D_1 in the calculation, D_3 and includes and D_4. Allotting an address by means of DHCP was not considered. Thus, the major cause of a vertical handover delay of a network layer is D_4, and this can be represented by the following expression:

$$D_4 = D_{wl} + D_w + L_{wl} + L_w \tag{1}$$

In expression (1), D_{wl} indicates the total delay in a wireless link while D_w represents the total delay in a wire link, L_{wl} represents the waiting time in a wireless link, and L_w represents the waiting time in a wire link. To induce D_{wl} and D_w, M/M/1 queuing model in the data routing route and signaling was applied to a wireless Base Station (BS) and other networking elements.

In analysis of M/M/1, some assumptions were required. The most important of them is that the service time that packets experience in another node should be independent. Since the service time is in proportion to the packet length, and the length of the packet while going through a network is the same, this assumption is false. This assumption of independence may be valid in a large scale network environment [22]. In the queuing theory used, D_{wl} and D_w may be represented by the following expression:

$$D_w = \frac{1}{\mu_{wl} - \lambda_{wl}} \tag{2}$$

μ_{wl} indicates the processing ratio in a wireless interface, and λ_{wl} the arrival ratio in a wireless interface. To make it clear, the unit of μ_{wl} was changed from packet/sec to bit/sec. When the Probability Density Function (PDF) of χ (packet size) is represented with bits, and the average packet length of $1/\mu$ bit/packet is $\mu e^{-\mu\chi}$, the capacity of i (communication channel) is C_i bits/sec, the arrival ratio of the communication channel i becomes λ_{wl} packet/sec, and μC_i becomes the service ratio. Thus, the communication channel i may be represented as follows:

$$D_{wl} = \frac{1}{\mu C_i - \lambda_{wl}} \tag{3}$$

D_{wl} includes queuing delay and transmission delay. The average packet size is not affected by the capacity and input ratio of a communication channel. D_w may be represented as a set of delays of various M/M/1 queues. When various M/M/1 server outputs are supplied to another server's input queue, the input process result would correspond to Poisson theory, and the sum of the supply process average value would be the same as well [22]. Thus, the total delay of packets in a wire network can be described by the following expression:

$$D_w = \frac{1}{\lambda_w} \sum_j \lambda_j \left(\frac{1}{\mu C_j - \lambda_j} \right) \tag{4}$$

λ_w indicates the total arrival ratio of packets, λ_j the arrival ratio of j^{th}, and μC_i the service ratio of the j^{th} node respectively. Thus, the combination of expressions (1), (3) and (4) may lead to the following result:

$$D_4 = \left(\frac{1}{\mu C_i - \lambda_{wl}} \right) \frac{1}{\lambda_w} \sum_j \lambda_j \left(\frac{1}{\mu C_j - \lambda_j} \right) + L_{wl} + L_w \tag{5}$$

In reference to expression (5), $D_{handover}$ on the message flow as in **Fig. 9** may be described by expression (6).

$$D_{Handover} = \sum_{i=1}^{j+1} (D_{PMIPv6_i} + D_{SIP_i}) + \Delta \tag{6}$$

Δ indicates the waiting time for HSS searching. Expression (5), however, does not accommodate errors that may cause the loss of various types of messages. To successfully establish a session, the entire message flow must occur. If any of the messages were damaged or lost, the handover process would fail. To reduce a delay in handover that involves large delay values, the handover must connect the communication channel before the existing channel is cut off. The total loss of packets while the session is available may be defined as the sum of packet loss when the MN is relocated with data packets being downloaded. The loss of packets may begin when Layer 2 handover is detected, and all packets that occur during the relocation are regarded as lost during the vertical handover. This can be described as follows: The signaling expense of mobility management during the vertical handover can be analyzed based on the accumulated traffic load of signaling messages exchanged during the MN session.

$$Pkt_{loss} = \frac{1}{[(2T_{ad} + D_{Handover})]} \times \lambda_d \times N_m \tag{7}$$

$$Cost = P \times S_{message} \times H_{a-b} \tag{8}$$

P indicates the possibility of each handover, S refers to the average size of signaling messages, and H_{a-b} refers to the average number of hops between a and b respectively.

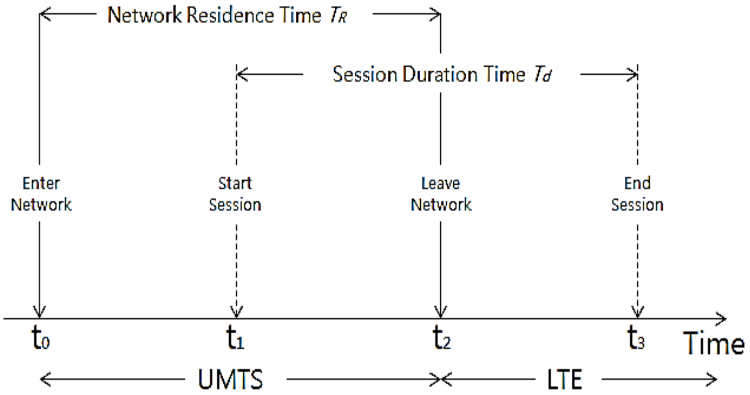

Fig. 12. Timing Diagram.

For a successful session handover when a MN is relocated from one network to another, it is important to keep it activated as the MN moves out of a network as shown in **Fig. 12** [23]. A probability function is created based on this assumption. It is assumed that the Poisson process is implemented in the average arrival ratio of sessions, λ and the probability of a session arriving during the Time Period $t = \lambda te^{-\lambda}$ follows the Poisson process with an average ratio of λ. P_1 is the session arrival probability for a MN that goes through a handover between networks t_0 and t_2, and this can be described as follows:

$$P_1(t_0 \le t_1 \le t_0 + T_R) = \int_0^\infty \lambda te^{-\lambda t} f_{T_R}(t)dt \tag{9}$$

$f_{T_R}(t)$ indicates the PDF of T_R, the network waiting time. In the given network, T_R, the waiting time of the MN, has the exponential distribution of $1/\eta$ on average. Here, η indicates the network flow rate. Hence, $f_{T_R}(t)$ can be described as follows:

$$f_{T_R}(t) = \mathrm{ne}^{-nt} \tag{10}$$

In expression (9), which involves the permutation of $f_{T_R}(t)$, the numerical expression P_1 may be described as follows:

$$P_1(t_0 \le t_1 \le t_0 + T_R) = \int_0^\infty \lambda te^{-\lambda t}\eta e^{-nt}dt$$

$$P_1 = \lambda \frac{\eta}{(\lambda+\eta)^2} \tag{11}$$

The following condition is that the Session Duration Time, or T_D, needs to be greater than the network waiting time, or T_R. T_D indicates the exponential distribution with an average value of $1/\mu$, and μ_{wl} the average message processing service ratio of a wireless link. Hence, the PDF of T_D, or $f_{T_D}(t)$, can be described as follows:

$$f_{T_D}(t) = \mu_{wl} e^{-\mu wlt} \tag{12}$$

The vertical session handover probability, or P_2, can be described as follows:

$$P_2(T_D < T_R) = \int_0^\infty \int_t^\infty \mu_{wl} e^{-\mu wly} fT_R(t) dydt$$

$$P_2(T_D < T_R) = \int_0^\infty \int_t^\infty \mu_{wl} e^{-\mu wly} \eta e^{-nt} dydt$$

$$P_2 = \frac{\eta}{(\mu_{wl}+\eta)} \tag{13}$$

As in numerical expression (3), when μC_i is the service ratio in packet/sec, μC_i is replaced with μ_{wl}, and thus P_2 can be described with the following expression:

$$P_2 = \frac{\eta}{(\mu C_i+\eta)} \tag{14}$$

Now the values from P_1 and P_2 can be replaced with the initially defined signaling value given in expression (8). The average value of η and that of λ are included in the final numerical expression. S_i (SIP INVITE message sequence) is in relation to the session arrival probability and the session arrival ratio, or P_1. S_R (SIP ReINVITE message sequence) is in relation to the vertical handover probability and Inter-Network mobility ratio, or P_2. Hence, in the given data session, the total signaling overhead from the vertical handover can be described as follows:

$$Cost = P_1 \lambda \sum_{i=1}^{n_1} (\mathrm{S}_{Ii} \times H_{(a-b)\mathrm{i}}) + P_2 \eta \sum_{i=1}^{n_2} (\mathrm{S}_{Ri} \times H_{(a-b)\mathrm{i}}) \tag{15}$$

n_1 and n_2 indicate the number of messages that include the order of handover, respectively. When η indicates the average network flow rate of the MN and λ the average session arrival rate respectively, λ / η can be defined as the Call-to-Mobility Ratio (CMR) [22]. Thus, the numerical expression (15) can be rearranged as follows:

$$Cost = \{P_1 \lambda \sum_{i=1}^{n_1} (S_{Ii} \times H_{(a-b)i})\} \frac{\lambda}{\eta} + P_2 \eta \sum_{i=1}^{n_2} (S_{Ri} \times H_{(a-b)i}) \tag{16}$$

4.2 Result of the Numerical Analysis

The following result is based on expressions (5),(7),(15),(16) which utilize the standard 3GPP-SIP message. **Table 2** shows the size of typical SIP messages and other related parameters from references [21][24].

Table 2. PMIPv6-SIP Message Sizes and Parameter Values.

Message	Length (Byte)	Parameter	Value
INVITE	736	L_{wl}	2 ms
ReINVITE	731	L_w	0.5 ms
183 Ses. Pro	847	η	10-100 sec
PRACK	571	μ	1-1000 sec
200OK	558	CMR	1-10
UPDATE	546	Δ	100 ms
ACK	314	T_{ad}	1 sec
Reg. Req.	60	λ_d	GSM Codec
Reg. Rep.	56	$\lambda_w, \lambda_{wl}, \lambda_j$	10-100 pkts/sec
Bind. Update	66	C_i	2-70 Mbps
Bind. ACK	66		

The distance values according to the number of hops are shown in **Fig. 13**, **Fig. 14** shows a delay in vertical handover as it applies to a session handover that increases for UMTS-to-LTE and LTE-to-UMTS. The graph shows a relatively extended handover delay in the case of LTE-to-UMTS. When the session is transmitted to a network with a relatively low link bandwidth, a relatively extended delay in vertical handover is expected in this case.

As for LTE-to-UMTS, the delay value drastically increases as the number of handovers increases from 5 to 6 **[Fig. 14]**. This is because various handovers occur in a relatively narrow bandwidth whose system use rate is ρ and λ / μ is 1. If the handover reaches a point where the packet flow exceeds the maximum capacity of the link, the packet transmission will fail after all. In addition, the average delays in session establishment for UMTS and LTE interfaces are 179 ms and 166 ms, respectively, on a mechanism that includes PMIPv6-SIP according to the presented analysis model. As for LTE-to-UMTS and UMTS-to-LTE, the average delays in vertical handover based on the PMIPv6-SIP mechanism are 94 ms and 180 ms, respectively.

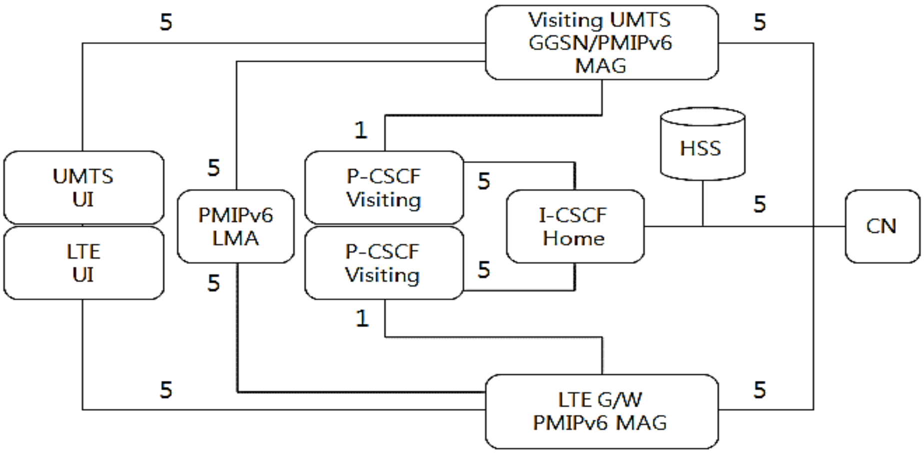

Fig. 13. RelativeDistances in Hops.

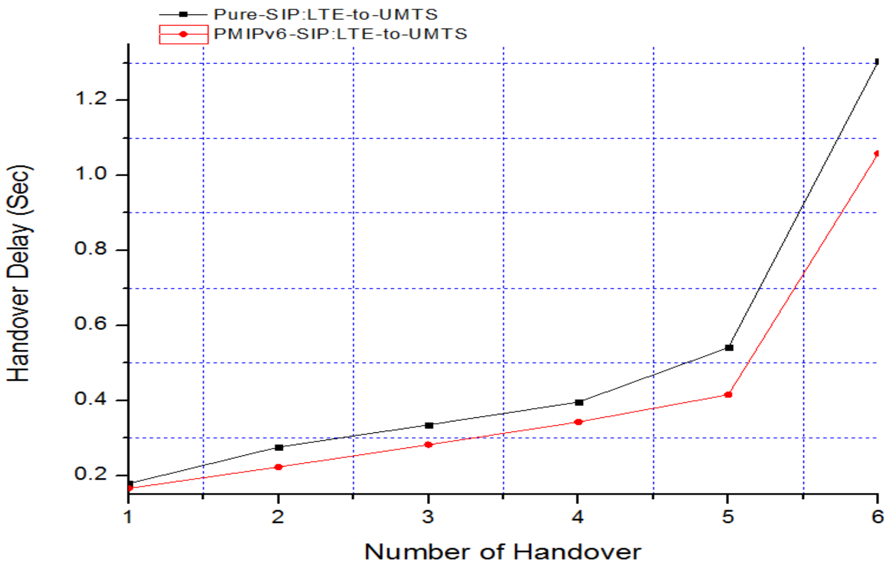

Fig. 14. Vertical Handover Delay vs. Number of Session Handovers.

Lastly, the PMIPv6-SIP based handover mechanism showed a relatively short vertical handover delay in general compared to the Pure-SIP based approach. PMIPv6-SIP based handover involved less vertical handover delay compared to the Pure-SIP based one. This is because the Pure-SIP based approach involves message flows of a relatively larger number of IMS related application layers and corresponding waiting times [24].

Fig. 15 shows the general packet loss during the vertical handover as the number of handovers increases. According to expression (7), an instant packet loss is in direct proportion to the vertical handover delay. Hence, the packet loss graph in **Fig. 15** is similar to that in **Fig. 14**. Specifically, **Fig. 14** shows that the packet loss of Pure-SIP is larger than that of PMIPv6-SIP. One interesting point regarding **Fig. 14** and **Fig. 15** is that the graphs diverge as the packet transmission ratio and number of handover increase. This is because of the IMS-related waiting time added to Pure-SIP, which results in an instant but outstanding loss of packets. **Fig. 16** shows the regularized signaling expenses on CMR when η and μC_i (case 1) and λ and μC_i (case 2).

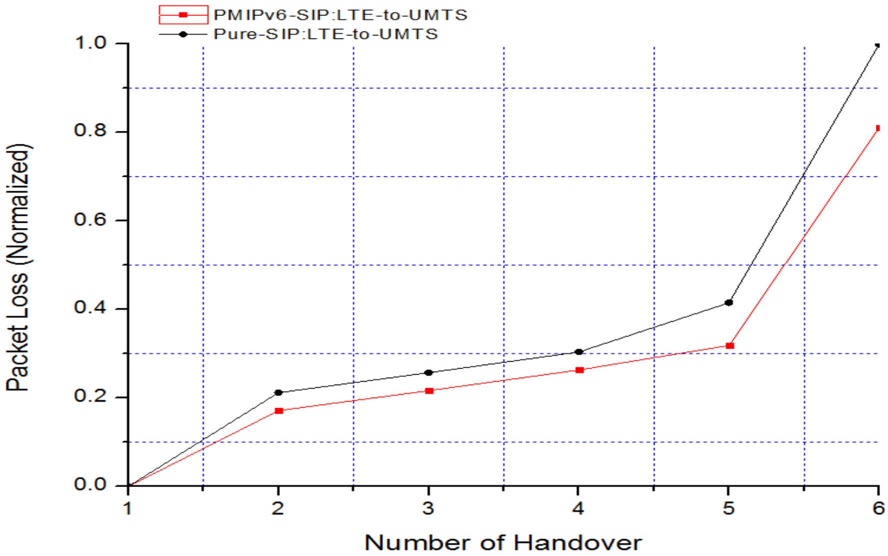

Fig. 15. Transient Packet Loss vs. Number of Handovers.

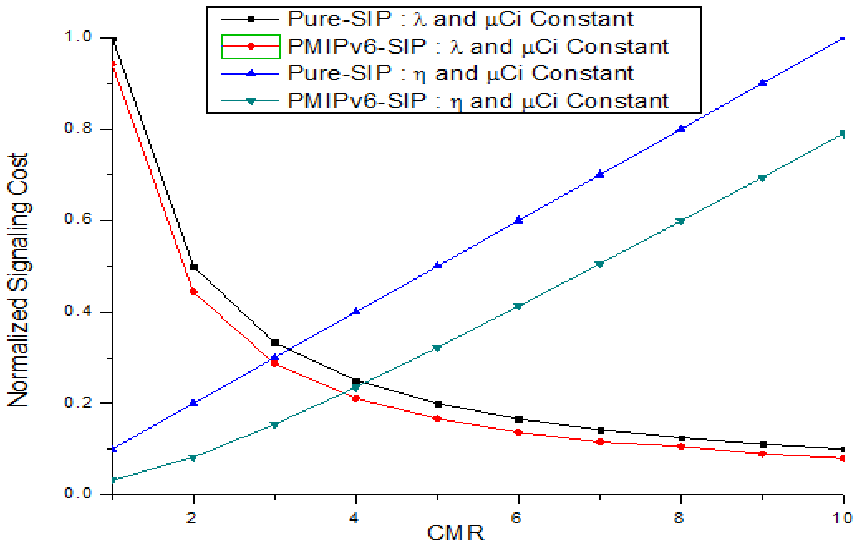

Fig. 16. Cost vs Call-to-mobilityRate (CMR).

In the graph for Case 2, CMR increases while the signal expense decreases when λ and μC_i are stable. **Fig. 17** shows the effect of η on the signaling expense when μC_i and λ are stable. When η increases, P_1 and P_2 increase linearly accordingly. As a result, the signaling expense linearly increases. Lastly, **Fig. 18** shows the effect of λ on the signaling expense when η and μC_i are stable.

As λ increases, the system reaches the saturation point in the end, which results from the stabilization effect of the signaling expense curved line. It is noteworthy that the analysis result above assumes a user that is randomly relocated. There could be some correlation, however, that involves relocations of communication partners or group mobility. As for group mobility, each group has a logical center that defines the orbit of the group movement. In general, a user moves randomly regardless of the logical center but within the geographical range. The movement of a logical center may be defined by means of a pre-determined motion channel or individual mobility model [25]. If there is a pre-determined motion channel, the movement of the logical center would not be random.

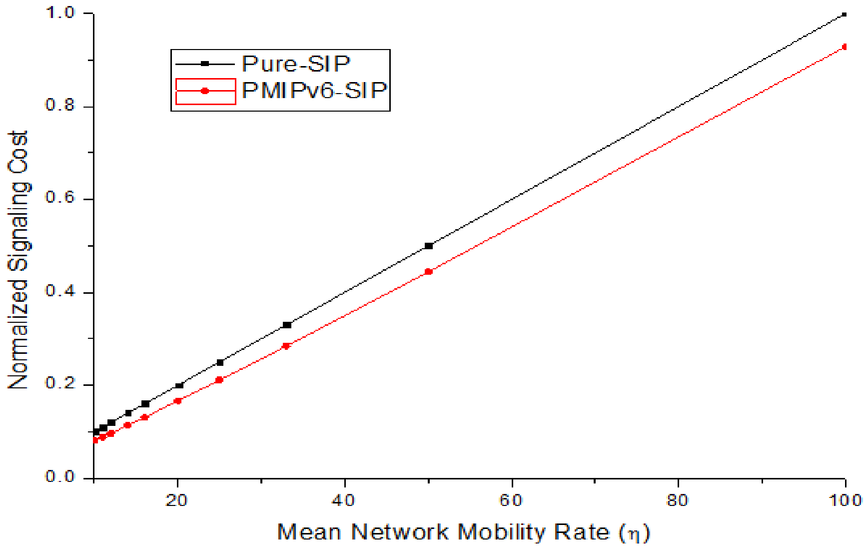

Fig. 17. Cost vs. η when μC_i and λ are constant.

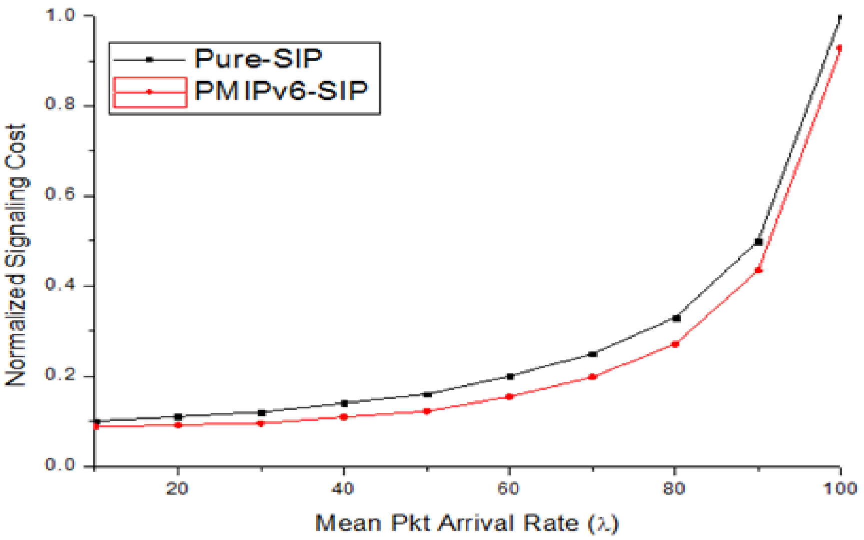

Fig. 18. Cost vs. λ when μC_i and η are constant.

In an exceptional case scenario of group mobility, for example, if the movement of a logical center responds to a mobility model on a MN that is relocated randomly by the walking user, this would be quite close to the mobility model [25] analyzed here.

5. Conclusion

In this paper, we propose a new PMIPv6-SIP architecture based on network-led handover while maintaining the session without a special function addition and middleware during MN handover. In proposed architecture, PMIPv6 for terminal mobility management and SIP for centralized session mobility management are employed. The new information is the SIP method for the proposed architecture as follows: The first is the local address of link layer for MN of SIPURI and IP address for SIP terminal and the second is the session establishment for SIPURI of wireless LAN router that the CN is connection. To register this information with SIP server, we propose the expansion of SIP message. Also, we describe the new inter-working architectures and their detailed performance evaluation. The analysis model is designed to analyze such factors as delay, temporary packet loss, signaling cost, and other vertical handover performance metrics. The numerical results show that additional waiting time in the application layer may significantly affect general performance. So, to establish terminal mobility and session mobility, we combine PMIPv6 with SIP and our proposed method shows higher mobility performance than the existing methods. In the considered scenario, the performance metrics such as handover delay and packet loss are all within the allowed limit in the real-time VoIP communications [26].

Future work will involve expanding the study scope, enhancing the security level of PMIPv6-SIP, and simplifying the procedures to enhance network performance based on the mathematical approach in [27]. In addition, an additional study is in progress on IKEv2 and AAA[28] with regard to PMIPv6 security issues.

Acknowledgement

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2010-0024695). Corresponding author: Jongpil Jeong.

References

[1] J. Kempf and Ed. “Goals for Network-Based Localized Mobility Management (NETLMM),” *RFC 4831*, April, 2007.
[2] S. Gundavelli, Ed., K. Leung, V. Devarapalli, K. Chowdhury, and B. Patil. “Proxy Mobile IPv6,” *RFC 5213*, August, 2008.
[3] D. Johnson, C. Perkins, J. Arkko, “Mobility Support in IPv6,” *RFC3775*, 2004.
[4] J. Rosenberg, H. Shulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, E.Schooler, “SIP :Session Initiation Protocol, ” *RFC3261*, 2002.
[5] S. Gundavelli, K. Leung, V. Devarapalli, K.Chowhury, B. Patil, “Proxy Mobile IPv6,” *RFC4861*, 2007.
[6] Seok-hoon Hong, et al., “A Cross-Layer Approach to Fair Resource Allocation for Multimedia Service in WiMAX” *KSII Transactions on Internet and Information Systems, vol.4, No.5,* pp.1006-1022, December, 2010. Article (CrossRef Link)
[7] Gun-Yeob Kim, et al., “Analysis of Macro-Diversity in LTE-Advanced” *KSII Transactions on Internet and Information Systems, vol.5, No.9,* pp.1596-1612, September, 2011. Article (CrossRef Link)
[8] 3GPP, “IP multimedia subsystem (IMS),” *3GPP TS 23.228 Version 6.10.0 Release 6*, 2005.
[9] 3GPP2, “All-IP core network multimedia domain,” *3GPP2 X.S0013-002-A v1.0*, November, 2005.
[10] Sung-kuen Lee, et al,. “An Enhanced Network-based Mobility Management Protocol for Fast

Mobility Support" *KSII Transactions on Internet and Information Systems, vol.5, No.11,* pp.1997-2015, November, 2011. Article (CrossRef Link)

[11] C. Politis, K. A. Chew, N. Akhtar, M.Georgiades, R. Tafazolli, and T. Dagiuklas, "Hybrid Multilayer Mobility Management with AAA Context Transfer Capabilities for All-IP Networks," *IEEE Wireless Communications,Vol.11, No.4, pp*.76-88, August, 2004. Article (CrossRef Link)

[12] K. D. Wong, A. Dutta, J. Burns, R. Jain, K.Young, and H. Schulzrinne, "A Multilayered Mobility Management Scheme for Autoconfigured Wireless IP Networks," *IEEE Wireless Communications, Vol.10, No.59*, pp.62-69, October, 2003. Article (CrossRef Link)

[13] Q. Wang, and M. A. Abu-Rgheff, "Mobility Management Architectures Based on Joint Mobile IP and SIP Protocols," *IEEE Wireless Communications, Vol.13, No.6*, pp.68-76, December, 2006. Article (CrossRef Link)

[14] K. Kim, H. Jung, and J. Lee, "A Seamless Voice Call Handover Scheme for Next Generation Cellular Network," in *Proc. of APCC 2009*, pp.782-785, October, 2009. Article (CrossRef Link)

[15] Sahana Bhosale, et al., "Multi-criteria Vertical Handoff Decision Algorithm Using Hierarchy Modeling and Additive Weighting in an Integrated WLAN/WiMAX/UMTS Environment-A Case Study" *KSII Transactions on Internet and Information Systems, vol.8, No.1*, pp.35-57, January, 2014.

[16] Dong-Youl Jeong, et al., "Design Call Control Mechanism for MultiParty Conference in SIP and Case Study" *KSII journal of Korean Society for Internet Information, vol.4, No.5*, pp77-86, October, 2003.

[17] V.Gurbani et al., "Session Initiation Protocol (SIP) Torture Test Messages for Internet Protocol Version 6 (Ipv6)," *RFC 5118*, February, 2008.

[18] Brownson Obaridoa Obele and Minho Kang, "Mobility Management: A Proactive QoS-Aware Proxy MIP with improved Handover Latency for End-to-End QoS provisioning in a Proxy MIP domain," In *Proc. Of ICACT 2009*, February, 2009.

[19] M. Karam and F. Tobagi, "Analysis of delay and delay jitter of voice traffic in the Internet," *INFOCOM 2001, Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings, IEEE., vol.2*, pp.824-833, April, 2001. Article (CrossRef Link)

[20] V. Paxon and S. Floyd, "Wide area traffic: the failure of Poisson modeling," *IEEE/ACM Trans. Networking, vol. 3*, pp. 226-244, 1995. Article (CrossRef Link)

[21]S. C. Lo, et al., "Architecture for mobility and QoS support in all-IP wireless Networks," *IEEE J. Select, Areas Commun., vol. 22*, pp. 691-705, 2004. Article (CrossRef Link)

[22] L. Kleinrock, *Queuing Systems: Theory*, John Wiley and Sons, 1975.

[23] U. Narayan and X. Jiang, "Signaling cost analysis of handovers in mixed IPv4/IPv6 mobile environment," *in Proc. of IEEE Globecom, Washington, D.C.*, November, 2007. Article (CrossRef Link)

[24] Q. Wang and M. A. Abu-Rgheff, "Interacting mobile IP and SIP for efficient mobility support in all IP wireless Networks," *in Proc. of IEE Int. Conf. on 3G Mobile Commun., London*, October, 2004. Article (CrossRef Link)

[25] L. Badia and N. Bui, "A group mobility model based on nodes attraction for next generation wireless Networks," *in Proc. of Int. Conf. on Mobile Technol., Applications and Systems, Bangkok*, 2006. Article (CrossRef Link)

[26] Bo-Wei Li, Kuochen Wang, and Yi-Ling Hsieh "A Hierarchical Social Network-based P2P SIP System for Mobile Environments1," 2010.

[27] Jui-Hung Yeh et al, "Fast Intra-Network and Cross-Layer Handover (FINCH) for WiMAX and Mobile Internet," *IEEE Trans. on Mobile Computing, Vol. 8, No. 4*, April, 2009. Article (CrossRef Link)

[28] Seung-Hyun Lee, Dong-Ryeol Shin, and Jongpil Jeong, "Performance Analysis of Proxy-AAA Authentication Scheme in PMIPv6 Networks with Forwarding Mode Supporting," *KSII journal of Korean Society for Internet Information, vol.13, No.1*, pp15-25, February, 2012.

Article (CrossRef Link)

Myougseok Song received his M.S. degree in College of Information and Communication Engineering from sungkyunkwan university, korea, in 2013. Currently he is working in BIZtelecom Corporation as network R&D team manager. His research interests includes mobile, LTE and wireless security.

Jongpil Jeong received his B.S. degree in engineering from Sungkyunkwan University and the M.S. and Ph.D. degrees in computer engineering from Sungkyunkwan University, Suwon, Korea, in 2003 and 2008, respectively.
He was a Research Professor with Sungkyunkwan University in 2008-2009 and 2011, and a visiting professor with the Department of Interaction Science (WCU Program) in Sungkyunkwan University in 2009-2010. He started his academic profession at the Research & Business Foundation of Sungkyunkwan Univeristy, Korea in 2012 as an assistant professor. He received twice Excellent Research Awards from Department of Electrical and Computer Engineering, Sungkyunkwan University, Korea (2007), from KSII (Korea Society for Internet Information), Korea (2011) and from IIBC (2013). His research interests include mobile computing, mobility management for vehicular networks, interaction science, sensor networking, protocol operation based performance analysis, Internet security, MIPv6 and ubiquitous computing. He is a member of the IEEE, KIISC, KSII, KIPS, and IEEK.