J Korea Industr Inf Syst Res Volume 18 Number 1
http://dx.doi.org/10.9723/jksiis.2013.18.1.001

A Caching Scheme to Support Session Locality in Hierarchical SIP Networks†

KwangHee Choi*, Hyunwoo Kim**

Abstract Most calls of a called user are invoked by the group of calling users. This call pattern is defined as call locality. Similarly Internet sessions including IP telephony calls have this pattern. We define it session locality. In this paper, we propose a caching scheme to support session locality in hierarchical SIP networks. The proposed scheme can be applied easily by adding only one filed to cache to a data structure of the SIP mobility agent. And this scheme can reduce signaling cost, database access cost and session setup delay to locate a called user. Moreover, it distributes the load on the home registrar to the SIP mobility agents. Our performance evaluation shows the proposed caching scheme outperforms the hierarchical SIP scheme when session to mobility ratio is high.

Key Words : mobility, caching, SIP, SMR

1. Introduction

As the Internet service is popular, the IP networks are extended in not only wired network but also wireless network [1]. One of the most important challenges IP networks face is mobility support [2]. There are two approaches for mobility support. One is Mobile IP (MIP), the other is Session Initiation Protocol (SIP). Mobile IP [3] and Mobile IPv6 (MIPv6) [4] enhance the network layer, so that IP hosts can change location and retain their communication session. Mobility in IP networks can be alternatively supported by application layer mobility protocols that rely on higher layer signaling to achieve the sought results. Some of these efforts include the use of SIP [5]. Motivation and description of SIP functionality to support mobility can be found in [6] and [7]. According to these proposals, SIP can be used to provide terminal mobility to Internet multimedia applications, with the appropriate SIP extensions to the base SIP specification. Specifically, SIP signaling is used after the handover for the end-to-end session re-establishment of SIP ongoing sessions between the communicating users.

The major argument for using SIP to achieve terminal mobility in SIP environments is the reuse of existing SIP infrastructure like SIP proxies, SIP registrars and SIP back-to-back user agents for the functionality required by mobile nodes [8, 9, 10, 11]. The deployment of Mobile IP leads to some extent in a duplicated network functionality and stored user data. Both SIP and Mobile IP have their own mechanism for registration/location update and their own database for storing location information. Additionally, the use of SIP could also compensate for the current lack of wide deployment of Mobile IP [12, 13].

† 이 논문은 2012년도 경일대학교 일반연구비 지원에 의하여 수행된 것임.
* LG유플러스, 제1저자
** 경일대학교 경영학부, 교신저자(e-mail: kimhw@kiu.ac.kr)

The need of micro-mobility support in SIP networks increases recently [14]. Micro-mobility protocol can handle local movement (e.g., within a domain) of mobile nodes without interaction with the SIP registrar. This scheme is defined as a hierarchical SIP (HSIP). HSIP can be achieved by using border router enabled SIP proxy, B2BUA and registrar functionality. This has the benefit of reducing delay and packet loss during handover and eliminating registration between mobile nodes and their registrar. As a result, HSIP can reduce the signaling overhead and support seamless handover [15, 16].

Caching schemes in PCS networks are proposed to reduce signaling cost and database access cost to locate called users [17, 18]. These schemes use the call pattern which most calls of a called user are invoked by the group of calling users. This call pattern is defined as call locality. Similarly Internet sessions including IP telephony calls have this pattern. We define it session locality.

Therefore we propose a caching scheme to support session locality in hierarchical SIP networks. The key idea of proposed scheme is to reduce signaling cost, database access cost and session setup delay using session locality feature. To apply this scheme, we add a filed to cache to a data structure of the SIP mobility agent. In this scheme, we can reduce the load of forwarding messages between the home registrar and SIP mobile agents because SIP mobile agent queries a cached domain address prior to querying the home registrar.

The remainder of this paper is organized as follows: Section 2 describes the overview of hierarchical SIP. Section 3 describes the proposed caching scheme in more detail. We explain procedures of two cases (cache hit case and cache miss case). Section 4 describes performance analysis. From performance analysis, cache hit ratio according to session to mobility ratio and cost functions for HSIP and the proposed caching scheme are generated. And also numerical results in detail are generated. Finally, section 5 assesses the proposed scheme, presents conclusions.

2. HSIP Overview

Hierarchical SIP scheme is proposed by Dimitr Vali [15, 16]. This scheme is a micro mobility management scheme in SIP networks which is similar to MIP-RR [19], HMIPv6 [20].

Fig.1 shows the network architecture of HSIP which is composed of a home registrar (HR), SIP mobility agents (SIP MAs) and access routers (ARs). A HR is responsible for globally handling inter-domain mobility. SIP MAs can be deployed in a domain and they are responsible for locally handling intra-domain mobility and managing routes in its domain. For simplicity reason, we assume that each domain is controlled by a single SIP MA situated at the domain boundary. An AR is deployed in every cell and is responsible for wireless transmission.

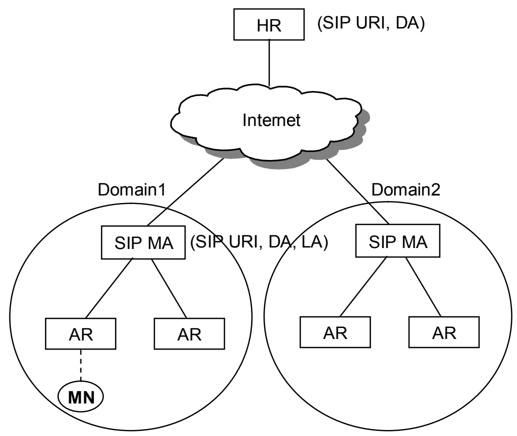


Fig. 1. HSIP network reference architecture.

Fig. 2 shows the SIP MA which is a border router enhanced with the functionality of a SIP proxy/B2BUA and a registrar. HSIP can support intra-domain mobility using SIP MA. HSIP allocates two addresses to each MN. One is local address (LA) and the other is global domain address (DA). The LA is an IP address reflecting the current point of attachment of the MN and it is allocated to the MN by the serving AR. A new LA is allocated to the MN each time it performs an intra-domain movement. The DA is a globally routable IP address that uniquely identifies the MN for the whole duration of moving inside the same access domain. Each MN is allocated a different DA by the SIP MA, which has a pool of globally routable IP addresses associated with it. The SIP MA is responsible for maintaining and managing mappings among the SIP URI, the DA and the LA for each mobile that moves inside the domain. The MN registers locally with its SIP MA in the case of intra-domain movement. However it registers not only locally with its SIP MA but also globally with its HR in the case of inter-domain movement [15, 16].

Fig. 3 shows the session delivery procedures in HSIP when a calling MN and a called MN are located in the same domain. The local registrar in the SIP MA can locate the called MN without querying the home registrar.

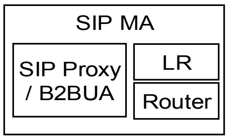


Fig. 2. SIP MA architecture.

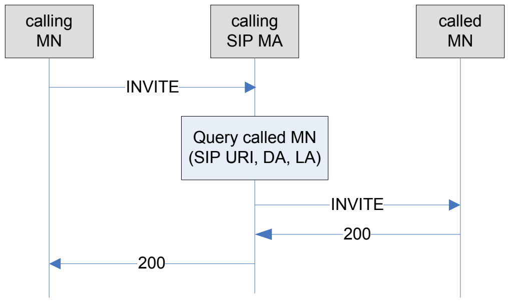


Fig. 3. session delivery procedures in HSIP (a calling MN and a called MN are in the same domain)

Fig. 4 shows the session delivery procedures in HSIP when a calling MN and a called MN are located in different domains. The request message to locate the called MN is forwarded to the home registrar via the SIP MA.

3. A Caching Scheme in HSIP

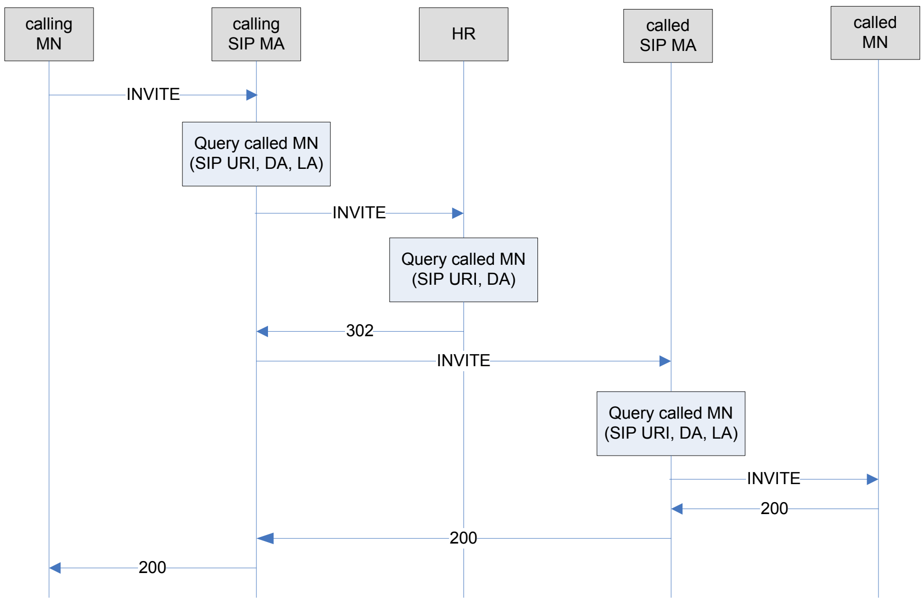


Fig. 4. session delivery procedures in HSIP (a calling MN and a called MN are in different domains)

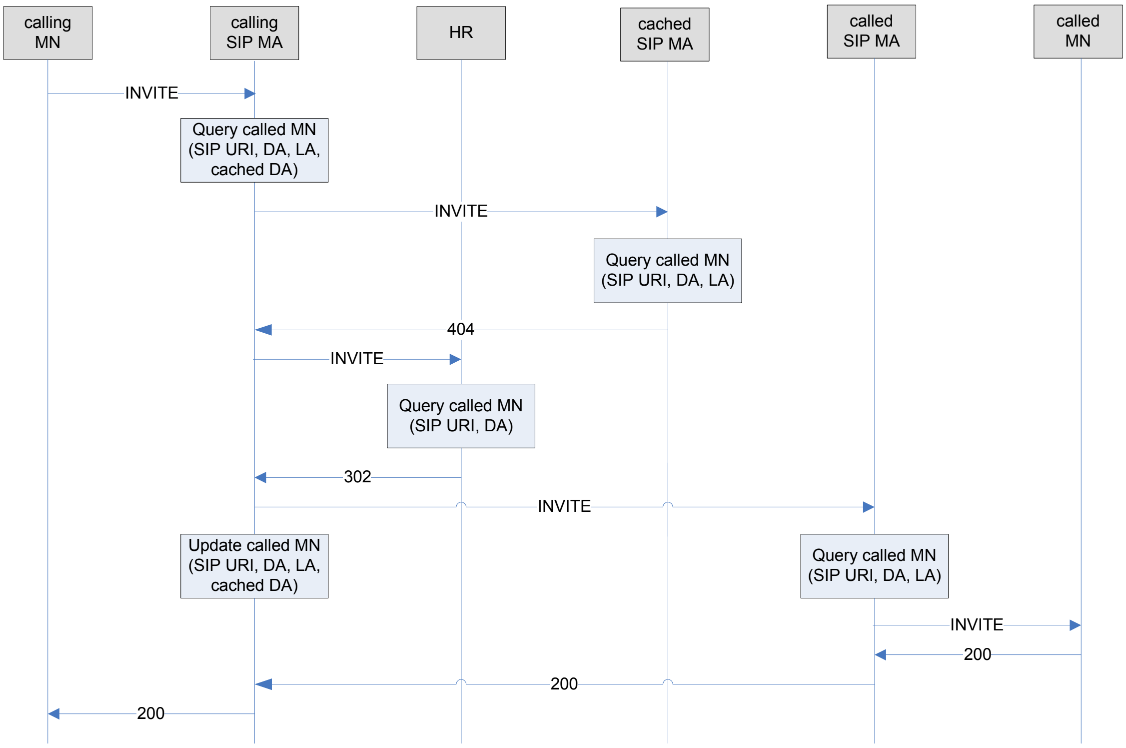


Fig. 7. session delivery procedures in CHSIP (cache miss case)

In this section, we describe the proposed caching scheme in HSIP (CHSIP) in detail. Most procedures for location registration in CHSIP are exactly the same as those of HSIP scheme.

As mentioned before, caching schemes in PCS networks are proposed to reduce signaling cost and

MN's DA	MN's LA	MN's operation Mode (active/idle)	LifeTime	MN's Cached DA

Fig. 5. SIP MA's data structure for MN

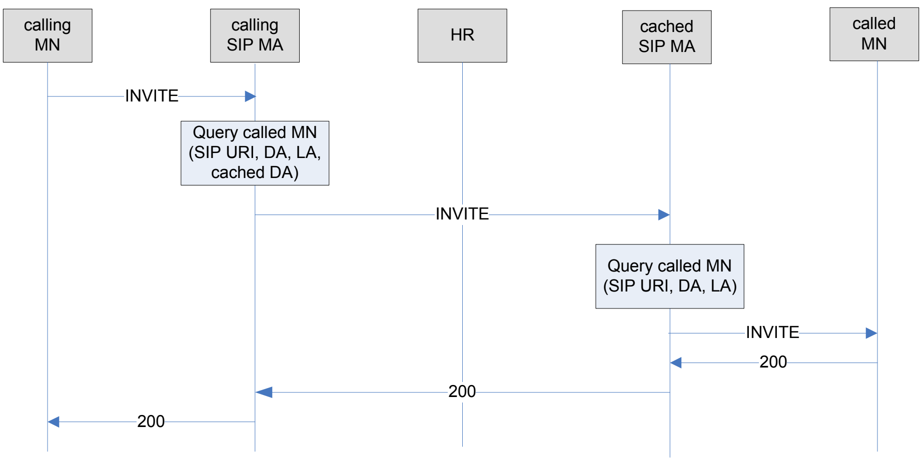


Fig. 6. session delivery procedures in CHSIP (cache hit case)

database access cost to locate called users [17, 18]. These schemes use the call pattern which most calls of a called user are invoked by the group of calling users. This call pattern is defined as call locality. Similarly Internet sessions including IP telephony calls have this pattern. We define it session locality.

CHSIP scheme can reduce cost to locate called user for session delivery by exploiting session locality. In order to apply caching scheme to HSIP infrastructure, we add a filed to cache to a data structure of the SIP mobility agent.

Fig. 5 shows the data structure which the SIP MA manages MNs. It has a global domain address (DA), a local address (LA), an operation mode, a life time and a cached DA for each MN.

Fig. 6 and Fig. 7 show session delivery procedure to locate called user. Fig 6 is the cache hit scenario and Fig. 7 is the cache miss scenario.

In Fig.6, a calling MN sends INVITE to its accessible SIP MA. The SIP MA queries a cached DA of the called MN. If a valid cached DA exists, the SIP MA forwards INVITE to the cached SIP MA without querying HR. If cache is hit, INVITE is forward to the called MN. The SIP MA can check the validation of a cached DA for each MN by using timeout mechanism. It can be helpful to increase the cache hit ratio.

In Fig. 7, the cached SIP MA to receive INVITE tries to query the called MN. If cache is missed, the cached SIP MA responses 404 NOT FOUND. Then the calling SIP MA forwards INVITE to the HR. The HR responses 302 TEMPORARILY MOVED. This response includes the called SIP MA to contact. The calling SIP MA updates the cached DA field of the called MN and sends INVITE to the called SIP MA. Then the called SIP MA forwards INVITE to the called MN. And if cache is frequently missed, signaling message transmission load in CHSIP can become higher than that in HSIP.

4. Performance Analysis

In this section, we analyze the performance in an aspect of signaling message transmission cost and database processing cost mathematically for session delivery procedure. Because location update procedure of two schemes is the same.

The basic assumptions for the performance analysis are as follows:

- The session arrivals to an MN follow a Poisson distribution with mean λ_s
- The domain residence time of an MN follows a Gamma distribution with mean $1/\lambda_m$, where λ_m is the movement rate.

Table 1. Performance analysis parameters

Parameter	Description
l_{MN}	Signaling message transmission cost between the SIP MA and the MN
l_{MM}	Signaling message transmission cost between two SIP MAs
l_{HM}	Signaling message transmission cost between the HR and the SIP MA
α_{MA}	Database access cost for a query or an update at the SIP MA
α_{HR}	Database access cost for a query or an update at the HR
p	Cache hit ratio at the SIP MA
q	MN's session to mobility ratio

Table 1 shows parameters and their description used for the performance analysis.

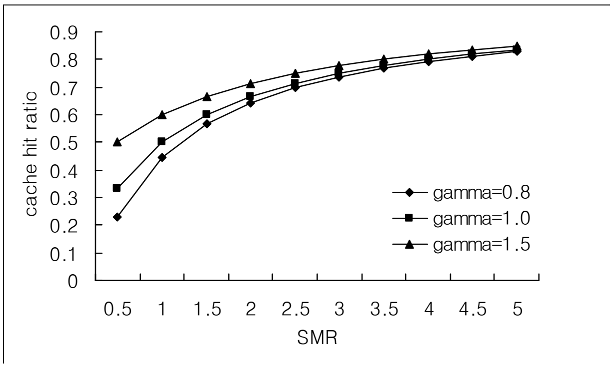


Fig. 8. Cache hit ratios

4.1 Cache Hit Ratio

The cache hit ratio is determined by a called MN's SMR. The SMR is defined as the expected number of sessions to a called MN from a given originating SIP MA during the period that the called MN visits a domain. Thus, the SMR can be denoted by λ_s/λ_m. We assume that the domain residence time follows a Gamma distribution. The Gamma distribution is selected for its flexibility and generality. By setting appropriate parameters, a Gamma distribution can be also used to represent the distribution for a set of measured data.

The Laplace transform, $f_m^*(s)$, of a Gamma distribution with mean $1/\lambda_m$ and variance V is as follows:

$$f_m^*(s) = \left(\frac{\lambda_m \gamma}{s + \lambda_m \gamma}\right), \qquad \gamma = \frac{1}{V{\lambda_m}^2} \tag{1}$$

In order to simplify the analysis, we can set $\gamma = 1$ such that the domain residence time of an MN follows an exponential distribution. Thus, it can be expressed as $f_m^*(\lambda_c) = \left(\frac{\lambda_m}{\lambda_c + \lambda_m}\right)$. Then, we will investigate the effect of the variance of the domain residence time on the performance of the cache scheme by setting different γ values (see Fig. (8)). The cache hit ratio can be obtained by probability that called MN resides in a domain between two consecutive sessions.

$$p = 1 - \frac{1}{q}\left(1 - f_m^*(\lambda_c)\right) \tag{2}$$

4.2 Cost in HSIP

We derive the signaling transmission cost, the database access cost and the total cost for session delivery procedure in HSIP scheme.

From fig, the signaling transmission cost in HSIP scheme is:

$$SC_H = 4l_{MN} + 2l_{HM} + 2l_{MM} \tag{3}$$

And the database access cost in HSIP scheme is:

$$DC_H = 2\alpha_{MA} + \alpha_{HR} \tag{4}$$

From (3) and (4), we can get the total cost in HSIP as follows:

$$TC_H = SC_H + DC_H \tag{5}$$

4.3 Cost in CHSIP

We derive the signaling transmission cost, the database access cost and the total cost for session delivery procedure in the proposed caching scheme. The signaling transmission cost is expressed as:

$$SC_C = pSC_{CH} + (1-p)SC_{CM} \quad (6)$$

where SC_{CH} and SC_{CM} are the signaling costs when a cache hit and a cache miss occur, respectively.

$$SC_{CH} = 4l_{MN} + 2l_{MM} \quad (7)$$

$$SC_{CM} = 4l_{MN} + 2l_{HM} + 4l_{MM} \quad (8)$$

The database access cost is expressed as:

$$DC_C = pDC_{CH} + (1-p)DC_{CM} \quad (9)$$

where DC_{CH} and DC_{CM} are the database access costs when a cache hit and a cache miss occur, respectively.

$$DC_{CH} = 2\alpha_{MA} \quad (10)$$

$$DC_{CM} = 4\alpha_{MA} + \alpha_{HR} \quad (11)$$

From (10) and (11), we can derive the total cost in the caching scheme as follows:

$$TC_C = SC_C + DC_C \quad (12)$$

4.4 Numerical Results

In this section, we conduct numerical results based on performance analysis. Table 2 shows the used parameter values. Signaling message transmission costs are dependent of the distance between origination node and destination node. And the database access costs are dependent of the number of users. Therefore table 2 is reasonable. We evaluate signaling communication cost, database access cost and total cost for different gamma values.

Table 2. parameters and their values

Parameter	Value
l_{MN}	4
l_{MM}	6
l_{HM}	8
α_{MA}	10
α_{HR}	20
q	0.5 ~ 5

Fig. 9 shows the relative signaling communication cost for session delivery. It means the cost of CHSIP assuming the cost of HSIP to be 1. As SMR is increased, the relative signaling communication cost is decreased. And as gamma is increased, the relative signaling communication cost is decreased. Because high SMR and high gamma value derive high cache hit ratio. However low SMR and low gamma value can derive low cache hit ratio and the cost of CHSIP can be more than that of HSIP because of cache fail overhead.

Fig. 10 shows the relative database access cost for session delivery. Fig. 11 shows the relative total cost. They are similar with Fig. 9. And if cache is frequently missed, all the cost of CHSIP can become higher than those of HSIP.

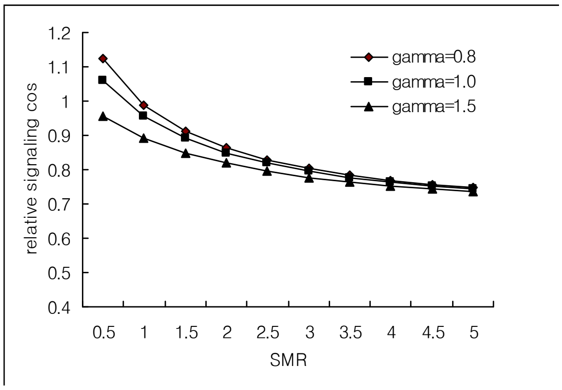


Fig. 9. Signaling communication cost for session delivery

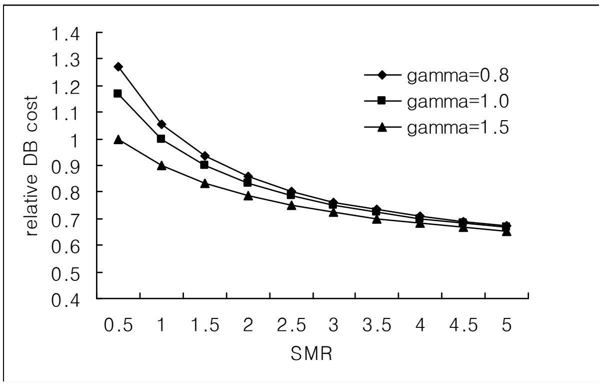


Fig. 10. Database access cost for session delivery

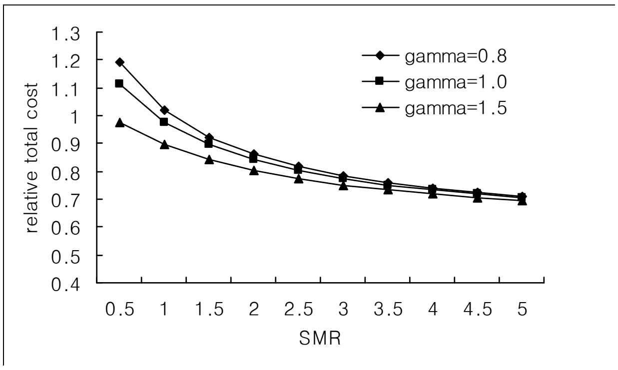


Fig. 11. Total cost for session delivery

5. Conclusion

In this paper, we propose a caching scheme to support session locality in hierarchical SIP networks. The main idea is to exploit a user's session locality at session delivery. In order to apply the caching scheme to HSIP infrastructure, we propose the way to modify the SIP mobility agent simply. This scheme can reduce signaling cost, database access cost and session setup delay to locate a called user. Moreover, it distributes the load on the home registrar to the SIP mobility agents. The analytical results have indicated that the caching scheme significantly outperforms HSIP scheme when the SMR is high or the signaling traffic to the home registrar is heavy.

References

[1] B. Ahn and Y. Joo, "Design of Location Based Social Network Service Model Centering around Smart Phone", Journal of the Korea Industrial Information System Society, vol.16, no.5, pp.55-62, 2011.

[2] I. Al-Surmi, M. Othman, and B. M. Ali, "Mobility management for IP-based next generation mobile networks: Review, challenge and perspective", Journal of Network and Computer Applications, vol. 35, Issue 1, pp. 295-315, Jan. 2012.

[3] C. Perkins ed., "IP Mobility Support for IPv4", RFC 3344, August 2002.

[4] D. Johnson, C. Perkins, and J. Arkko, "Mobility Support in IPv6", RFC 3775, June 2004.

[5] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler, "SIP: Session Initiation Protocol", RFC 3261, June 2002.

[6] H. Schulzrinne and E. Wedlund, "Application layer mobility using SIP", Mobile Computing and Communications Review, vol. 4, no. 3, pp.47-57, July 2000.

[7] E. Wedlund and H. Schulzrinne, "Mobility support using SIP", 2nd ACM/IEEE International Conference on Wireless and Mobile Multimedia, Seattle, Washington, Aug. 1999.

[8] I. Chang and Y. Mai, "Mobile multimedia supports with hierarchical multicast SIP framework on heterogeneous wireless networks", Computers & Electrical Engineering, vol. 32, pp.78-101, 2006.

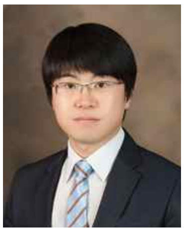
[9] Y.C. Yee, K.N. Choong, Andy L.Y. Low, and S.W. Tan, "SIP-based proactive and adaptive mobility management framework for heterogeneous networks", Journal of Network and Computer Applications, vol. 31, Issue 4, pp. 771-792, Nov. 2008.

[10] F. Leu, "A novel network mobility handoff scheme using SIP and SCTP for multimedia applications", Journal of Network and Computer

Applications, vol.32, Issue 5, pp. 1073-1091, Sep. 2009.
[11] R. Libnik, A. Svigelj, and G. Kandus, “A novel SIP based procedure for congestion aware handover in heterogeneous networks”, Computer Communications, vol. 33, pp.2176-2184, Dec. 2010.
[12] J. Lin and M. Yang, “Fault-tolerant design for wide-area Mobile IPv6 networks”, Journal of Systems and Software, vol. 82, Issue 9, pp. 1434-1446, Sep. 2009.
[13] Y. Ma and M. Song, “A novel enhanced authentication scheme supporting forwarding mode in heterogeneous hierarchical mobile IPv6 networks”, The Journal of China Universities of Posts and Telecommunications, vol.16, pp.86-91, Sep. 2009.
[14] K. Cho and D. Kim, “Performance Analysis of Improved Distance-based Location Registration Scheme in Mobility Model”, Journal of the Korea Industrial Information System Society, vol.11, no.2, pp.1-8, 2006.
[15] D. Vali, S. Paskalis, and A. Kaloxylos, "A SIP-based Method for Intra-Domain Handovers", Proceedings of IEEE VTC-Fall 2003, Orlando, FL, USA, October 2003.
[16] D Vali, S. Paskalis, A. Kaloxylos, and L. Merakos, "An Efficient Micromobility Solution for SIP Networks", Proceedings 8of IEEE Globecom 2003, San Francisco, CA, USA, November 2003.
[17] K. Kong, J. Gil, Y. Han, U. Song, and C. Hwang, "A Forwarding-Pointer Based Cache Scheme for Reducing Location Management Cost in PCS Networks," Journal of Information Science and Engineering, vol.18, no.6, pp.1011-1025, Nov. 2002.
[18] D. Lee, J. Kim, and J. Song, “Performance of Location Caching Scheme to Support Call Locality in Wireless Networks,” Journal of High Speed Networks, LNCS 1961, pp.43-57, Dec. 2003.
[19] E. Gustafsson, A. Jonsson, and C. Perkins, “Mobile IP Regional Registration”, Internet Draft, March 2001.
[20] H. Soliman, C. Castelluccia, K. E. Malki, and L. Bellier, “Hierarchical Mobile IPv6 mobility management (HMIPv6)”, RFC4140, August 2005.

최 광 희 (KwangHee Choi)

• 고려대학교 컴퓨터학과 이학사
• 고려대학교 컴퓨터학과 이학석사
• 고려대학교 컴퓨터학과 이학박사
• LG유플러스 기업솔루션개발팀 차장
• 관심분야 : 모바일컴퓨팅, VoIP, 유헬스



김 현 우 (Hyunwoo Kim)

• 정회원
• 한국과학기술원 산업경영학과 공학사
• 한국과학기술원 산업공학과 공학석사
• 한국과학기술원 산업공학과 공학박사
• LG유플러스 기술연구원 책임연구원
• 경일대학교 경영학부 조교수
• 관심분야 : 무선통신시스템, 정보보호, 서비스품질

논문접수일 : 2012년 11월 19일
1차수정완료일 : 2013년 02월 08일
2차수정완료일 : 2013년 02월 22일
게재확정일 : 2013년 02월 22일