Wireless Personal Communications (2005) 35: 269–287
DOI: 10.1007/s11277-005-6175-1

Integrated Mobility Management Methods for Mobile IP and SIP in IP based Wireless Data Networks

HYEJEONG LEE[1], JEE-YOUNG SONG[1], SUN-HO LEE[2], SUNGWON LEE[2] and DONG-HO CHO[1]

[1]*Korea Advanced Institute of Science and Technology (KAIST)*
[2]*Samsung Electronics Co., Ltd.*

Abstract. With developments in voice over IP (VoIP), IP-based wireless data networks and their application services have received increased attention. While multimedia applications of mobile nodes are served by Session Initiation Protocol (SIP) as a signaling protocol, the mobility of mobile nodes may be supported via Mobile IP protocol. For a mobile node that uses both Mobile IP and SIP, there is a severe redundant registration overhead because the mobile node has to make location registration separately to a home agent for Mobile IP and to a home registrar for SIP, respectively. Therefore, we propose two new schemes that integrate mobility management functionality in Mobile IP and SIP. We show performance comparisons among the previous method, which makes separate registration for Mobile IP and SIP without integration, and our two integrated methods. Numerical results show that the proposed methods efficiently reduce the amount of signaling messages and delay time related to the idle handoff and the active handoff.

Keywords: mobility management, Mobile IP, SIP, IP based networks, mobile networks

1. Introduction

The development of next generation wireless systems, which are characterized by seamless worldwide communication and support of various multimedia services, has been accelerated along with the development of wireless internet. Standardization for fourth generation (4G) systems has already been discussed, and the future 4G systems are expected to be based on an all-IP solution. Moreover, strategies are now being devised to deliver wireless data applications and services to the mobile user over a packet-switched IP network and, ultimately, to reduce conventional circuit switching and cellular infrastructure. In such IP-based wireless communication environments, it is critical to support seamless mobility of mobile terminals.

Currently, most solutions for wireless internet mobility are based on the network layer solution using the Mobile IP standard of IETF [1]. When a mobile node (MN) is in motion between IP subnets, Mobile IP is effective for providing transparent mobility, which hides the changes of IP addresses for the application layer of the MN and the correspondent node (CN). Therefore, Mobile IP makes it possible to keep an ongoing TCP connection alive when the IP subnet of an MN is changed. Moreover, Mobile IP doesn't require any correspondent nodes to adopt Mobile IP protocol. However, Mobile IP has a number of drawbacks, such as data encapsulation overhead and the usage of a unique IP address for each MN. In addition, Mobile IPv4 cannot avoid triangular routing, which increases routing delay for traffic destined for the MN and causes traffic tunnelling overhead. Further, Mobile IP is not efficient for supporting delay-sensitive multimedia traffic, because it takes much time to exchange registration/acknowledgement messages between mobility agents that are located far from each other. To overcome this problem, much research on micro-mobility protocol, such as Cellular IP [2], HAWAII [3], Regional Registration [4] and IDMP [5] has been conducted.

In Mobile IPv6, the packets from the correspondent node can be routed directly to the care-of address of the mobile node by using Route Optimization. Once the mobile node moves to a different subnet, it informs the correspondent node of its new care-of address by sending a "Binding Update" message. When sending a packet to a mobile node, the correspondent node sets the Destination Address in the IPv6 header to the care-of address of the mobile node and adds the original home address by using IPv6 "Home Address" option. Once the packet arrives at the care-of address, the mobile node can retrieve its home address from the routing header. Therefore, the inclusion of home addresses in these packets makes the use of the care-of address transparent above the network layer. Although Route Optimization can solve a triangular problem and packet transmission overhead, it is required for the correspondent node to be modified for processing of binding update messages and maintenance of binding update table.

The application layer mobility solution, which uses the SIP [6], has been studied by several researchers [7, 8]. Although SIP was developed initially for the signaling of multimedia sessions, it can be used for mobility support without much modification because it fundamentally supports personal mobility and service mobility for mobile nodes that will change their locations within networks but will not move frequently during their sessions. When using SIP for mobility support, packet tunneling and modification of the protocol stack in the MN are no longer required. Moreover, SIP mobility is efficient to support mid-call mobility by using *SIP re-INVITE* message [7]. However, it is still problematic for the SIP-based mobility support method to completely replace conventional Mobile IP, because SIP-based mobility support requires a certain amount of modification at the kernel level to trigger location registration to the application layer protocol on detection of Layer 2 handoff [8]. In addition, the SIP-based mobility support method lacks the ability to maintain transport layer connections in the event of handoff, because it cannot support transparent mobility at the transport layer such as TCP and UDP. Especially, the performance of TCP flow control will be severely degraded because transport layer connections should be re-established between CN and MN using a newly allocated IP address or can be maintained by an additional agent that monitors ongoing TCP connections [9].

In this paper, we propose efficient mobility management methods that integrate the mobility support functions of Mobile IP and SIP. The rest of this paper is organized as follows. We review the previous work supporting mobility in Section 2. In Section 3, we describe the proposed efficient IP mobility management methods that integrate Mobile IP and SIP. We analyze the performance of proposed schemes in Section 4 and provide numerical results in Section 5. Section 6 presents conclusions.

2. Related Work

Until recently, most research related to SIP mobility has focused mainly on how to support user mobility using SIP or how to solve handoff disruption in the SIP-based approach [7–9]. Only a few papers consider the comparison between Mobile IP and SIP [10] and the coexistence problem of SIP based telephony and conventional PSTN networks [11].

In [10], Kwon compares mobility management methods for VoIP services, which are based on the Mobile IP and SIP. He introduces a new concept called Shadow Registration to reduce disruption time in the interdomain handoff, which is based on the prior establishment of security association between the MN and the AAA (Authentication, Authorization and Accounting) server before the actual handoff occurs. However, this paper just provides a performance comparison between mobility solutions based on Mobile IP and SIP, and does not consider the coexistence problem of Mobile IP and SIP.

Although IP is considered to be the ultimate end-to-end approach to support future wireless multimedia services, it will take considerable time for all communication services to become IP-based. To solve the interworking problem between conventional UMTS-based wireless telephone networks and SIP-based internet telephony networks, Unified Mobility Manager (UMM) has been proposed in [11]. The UMM approach, which combines UMTS home location register (HLR) and SIP proxy functionality in one logical entity, can eliminate unnecessary signaling exchange related to interworking and reduce average call delivery delay. The UMM-based mobility support method provides an important motivation to unify the mobility protocol when different mobility methods exist concurrently.

Although several wireless technical forums, such as 3GPP, 3GPP2 and MWIF, have chosen SIP as the signaling protocol of the mobile internet, SIP still has several difficulties in supporting user mobility, as discussed in the previous section. Therefore, it is likely that most mobile nodes in future wireless networks will adopt both Mobile IP for mobility support and SIP for multimedia signaling protocol, respectively. Moreover, recent paper [12] considers a hybrid Mobile IP/SIP environment that uses both Mobile IP and SIP in order to efficiently support different traffic characteristics such as real-time and non-real-time services. Thus, in this paper, we basically assume the wireless data networks in which mobile nodes adopt both Mobile IP and SIP.

Figure 1 describes the architecture of an IP-based mobile data network and Figure 2 illustrates the signaling procedures when the MN adopts both Mobile IP and SIP protocols without integration of mobility protocols (henceforth, NO-INT method). The signaling procedures are divided into three major events such as the location registration (location update), SIP session setup and active handoff. We call it *active handoff* when the mobile node has an ongoing SIP session while it crosses the subnet boundary. Compared with the active handoff, we hereafter

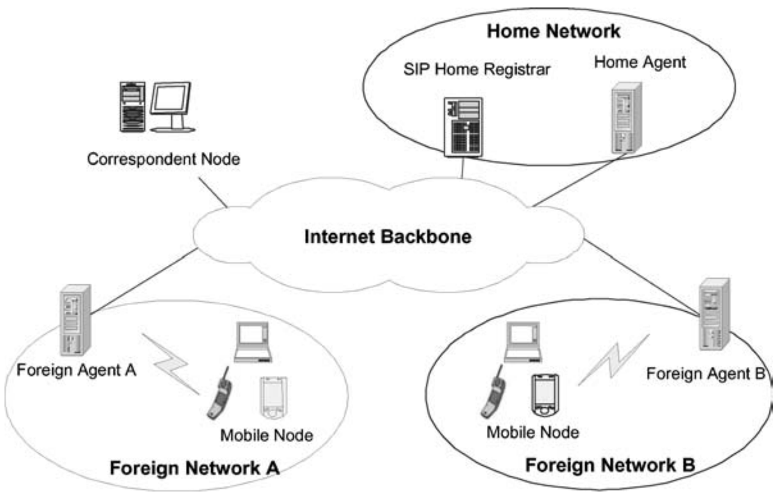

Figure 1. Architecture of IP based mobile data network.

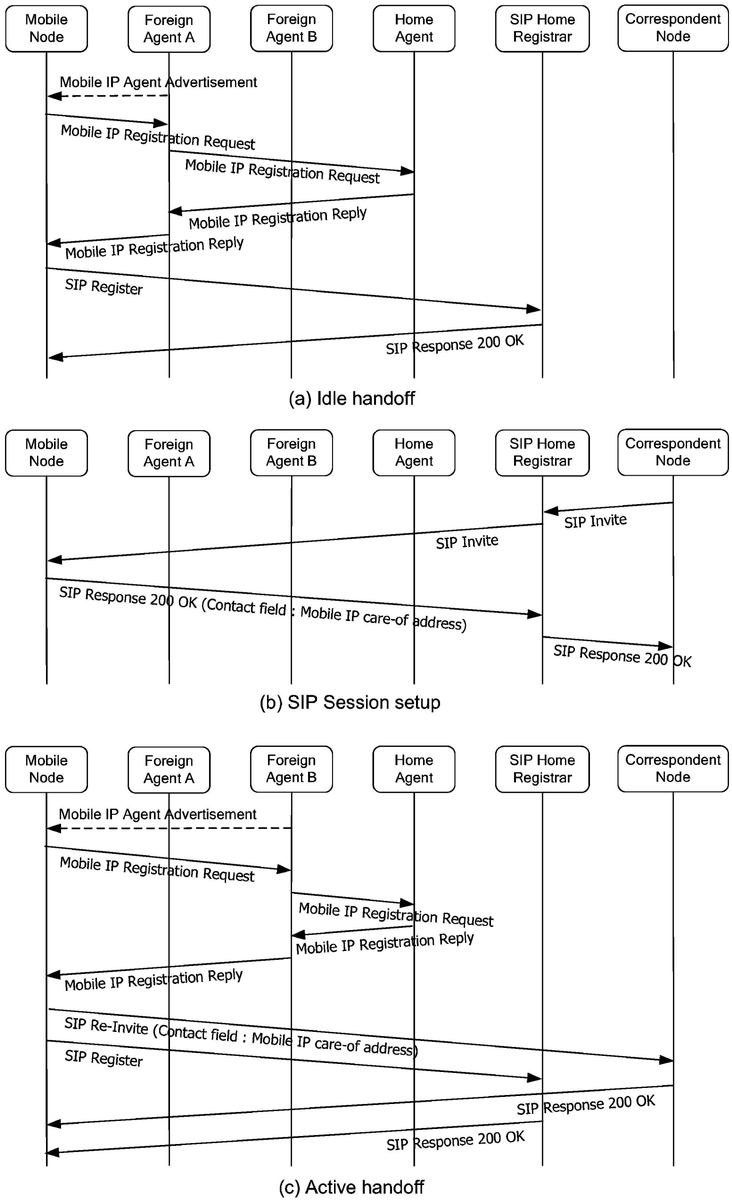

Figure 2. Signaling procedure based on NO-INT method (using both Mobile IP and SIP).

use the term *idle handoff* for the event that the MN performs the location registration/update without any ongoing SIP session. As shown in Figure 2(a), if an MN using both Mobile IP and SIP moves from one subnet coverage area to another subnet coverage area without any active SIP session, it performs two separate idle handoff procedures, one to the Mobile IP home agent (HA) and the other to the SIP home registrar (HR), respectively. However, the information included in the registration message is similar for both Mobile IP registration and SIP registration, thus separate registrations for both Mobile IP and SIP are redundant and inefficient with respect to utilizing radio resources and the battery power of the mobile node. In addition, this complex registration procedure raises severe packet loss and delay in the case of handoff. Therefore, it is critical to devise an approach that integrates the functionalities related to mobility support in Mobile IP and SIP. Such an approach will ensure that the mobility of a mobile node that adopts both Mobile IP and SIP can be supported efficiently in IP-based mobile data networks.

3. Proposed Mobility Management Method Integrating Mobile IP and SIP

In order to optimize redundant signaling traffic related to the location update for idle handoff and active handoff, we propose two mobility management methods that integrate the mobility support functionality of both Mobile IP and SIP. Mobile nodes that have both Mobile IP and SIP are the concern of our integrated mobility management methods. The proposed methods should make as few modifications as possible to conventional Mobile IP and SIP, to maintain compatibility with other mobile nodes that do not use our methods. That being so, our proposed methods make only a minor modification to the operating algorithm of mobile node and do not make any changes to the protocol stacks of the agent or server. The proposed methods, which are differentiated according to the MN's IP address used in the *Contact* field of the SIP message header, are described in the following subsections.

3.1. Integrated Mobility Management Method using the Mobile IP Home Address (INT-HOA Method)

The INT-HOA method integrates the mobility support functions based on the Mobile IP. The mobile node using the INT-HOA performs only Mobile IP registration when the terminal moves and registers on the SIP home registrar only when the personality or service characteristic of the mobile node is changed. Therefore, the mobile node does not make any registration on the SIP home registrar even in the event of subnet change. In the INT-HOA method, the SIP home registrar maintains the home address of the mobile node.

Figure 3 illustrates the signaling procedures when the INT-HOA method is applied in the network represented in Figure 1. Figure 3(a) shows the signaling procedure of idle hand-off for the INT-HOA method. When a mobile node enters foreign network *A*, the mobile node performs registration to the home agent with a new care-of address (COA) allocated in the foreign network *A*. Thus, the home agent can keep track of the location of the mobile node.

Then, we consider the SIP session setup procedures of the INT-HOA method, as shown in Figure 3(b). Because the INT-HOA method does not modify Mobile IP, it is not necessary to consider normal IP packet transfer using Mobile IP. When a correspondent node initiates SIP session setup toward the mobile node, it transmits an *SIP INVITE* message to the home address of the mobile node stored in the SIP home registrar. Therefore, the *SIP INVITE* message is transferred to the current location of the mobile node through the Mobile IP protocol. When the mobile node receives the *SIP INVITE message*, the mobile node sends an *SIP Response 200 OK* message to the SIP home registrar directly. In the INT-HOA method, we intentionally set the *Contact* field of the *SIP Response* message header to the home address of the mobile node, even if the mobile node is not located in its home network. By doing so, the INT-HOA method can operate efficiently in the event of handoff.

Figure 3(c) shows the signaling flows of active handoff procedure. If the mobile node moves into another foreign network *B* from foreign network *A* with an ongoing SIP session, the mobile

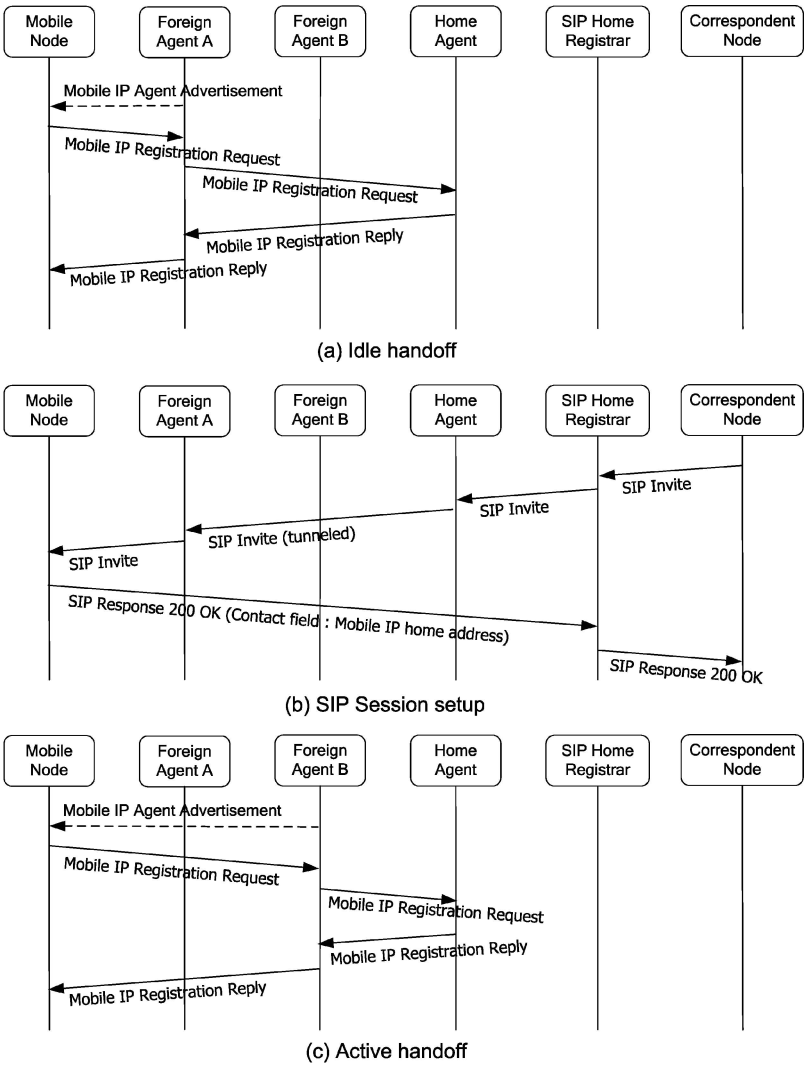

Figure 3. Signaling procedure based on INT-HOA method.

node performs the Mobile IP registration procedure. Shortly after completion of the Mobile IP registration, the mobile node can receive subsequent packets, because the correspondent node may transmit all packets to the home address of the mobile node without regard to the location change of the mobile node. Therefore, the INT-HOA method can effectively reduce active handoff signaling and delay.

The INT-HOA method has the drawback that all packets from the correspondent node should be delivered via the Mobile IP home agent using IP tunneling. Therefore, if handoff events don't occur more frequently compared to the session arrival, tunneling overhead in the INT-HOA method may not be acceptable. However, if our INT-HOA method is implemented in the IPv6 based network, the problem for media transmission overhead can be easily resolved by using Route Optimization of Mobile IPv6. The reason is that all packets destined to the mobile node can be directly routed through the care-of address that is saved in the correspondent node by binding update procedure. Although we have explained our proposed INT-HOA method based on Mobile IPv4, the proposed method can be simply adopted in Mobile IPv6 based networks. Moreover, the INT-HOA method doesn't cause tunneling problem in delivery of most signaling messages because other signaling messages except *SIP INVITE* don't involve media transmission.

3.2. Integrated Mobility Management Method using the Mobile IP Care-of Address (INT-COA Method)

The INT-COA method integrates the mobility support functions based on the Mobile IP, which is similar to the INT-HOA method. A mobile node that uses the INT-COA performs only Mobile IP registration in the event of mobility, and SIP registration is executed only in the event of personal or service mobility. Therefore, the SIP home registrar maintains the home address of the mobile node and the SIP session is established using the home address. However, the INT-COA method informs the correspondent node of the current location of the mobile node by setting the *Contact* field of the *SIP Response* message header to the care-of address. Thus, subsequent packets of the SIP session can be delivered to the mobile node directly by using the new care-of address of the mobile node.

Figure 4 illustrates the signaling procedures when the INT-COA method is applied in the network represented in Figure 1. When a mobile node enters foreign network A during idle state, the mobile node performs only the Mobile IP registration procedure to the home agent, as shown in Figure 4(a).

Figure 4(b) shows the signaling procedure of SIP session setup in the INT-COA method. In the event that a correspondent node initiates an SIP session toward the mobile node, the *SIP INVITE* message can be routed to the mobile node through the home address stored in the SIP home registrar according to the Mobile IP protocol. When the mobile node receives the *SIP INVITE* message, the mobile node sends an *SIP Response 200 OK* message to the SIP home registrar directly. In this case, the current care-of address of the mobile node is set in the *Contact* field of the *SIP Response* message header. As a result, it is possible for the correspondent node to transmit packets without IP tunneling.

Next, we consider the signaling flows of active handoff as shown in Figure 4(c). When the mobile node roams into another foreign network *B* with an ongoing SIP session, the mobile node performs the Mobile IP registration procedure. Once the Mobile IP registration is completed, the mobile node sends an *SIP Re-INVITE* message to the correspondent node in order to inform the correspondent node of the changed care-of address. Then, subsequent packets can be transmitted between the mobile node and the correspondent node by using the new care-of address.

The INT-COA method does not raise the IP tunneling overhead in delivering data traffic, except for delivering of the first session initiation message (*SIP INVITE*). However, compared to the INT-HOA method, the INT-COA method causes additional delay and signaling message exchanges to inform the correspondent node of IP address change for the mobile node when perfoming active handoff.

4. Performance Analysis

In this section, we make analytic comparisons among the INT-HOA method, the INT-COA method, and the conventional NO-INT method. We investigate the overall performance in

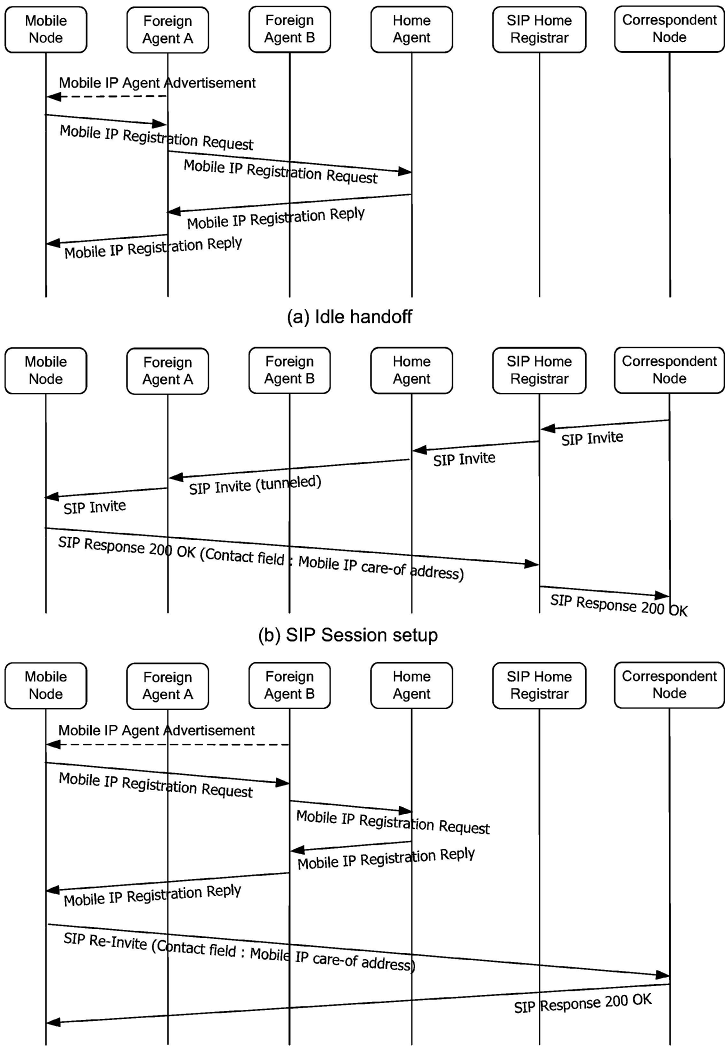

Figure 4. Signaling procedure based on INT-COA method.

terms of signaling cost and delay for three events of idle handoff, SIP session setup and active handoff. Signaling cost is defined as the message traffic exchanged between network nodes during mobility management procedures. Delay is defined as the time difference from the start time to the end time of a procedure.

4.1. Total Signaling Cost for the Mobility Management

To analyze the performance of three comparative mobility support methods, we define the signaling cost parameter between node A and node B as c_{AB}. For simplicity, we use subscripts as follows: m for the mobile node, c for the correspondent node, f for the foreign agent, h for the home agent and r for the home registrar, respectively. According to the signaling procedures in Figures 2, 3 and 4, the signaling costs of the INT-HOA, INT-COA and NO-INT methods can be evaluated. Then, the signaling cost for the idle handoff in each method, C_i, is given as follows.

$$C_i(\text{NO-INT}) = 2c_{mf} + 2c_{fh} + 2c_{mr} \tag{1}$$
$$C_i(\text{INT-HOA}) = 2c_{mf} + 2c_{fh} \tag{2}$$
$$C_i(\text{INT-COA}) = 2c_{mf} + 2c_{fh} \tag{3}$$

In addition, the signaling cost involved in the SIP session setup, C_s, is given as follows.

$$C_s(\text{NO-INT}) = 2c_{rc} + 2c_{mr} \tag{4}$$
$$C_s(\text{INT-HOA}) = 2c_{rc} + c_{hr} + c_{fh} + c_{mf} + c_{mr} \tag{5}$$
$$C_s(\text{INT-COA}) = 2c_{rc} + c_{hr} + c_{fh} + c_{mf} + c_{mr} \tag{6}$$

Let λ_s be the SIP session arrival rate for a mobile node and λ_u be the average rate for subnet boundary crossings, which is proportional to user mobility. Based on the signaling costs in Equations (1)–(6), the total signaling cost including signaling costs of the idle handoff and SIP session setup, C_{total}, can be computed by

$$C_{\text{total}} = \lambda_u \cdot C_i + \lambda_s \cdot C_s. \tag{7}$$

4.2. Average Signaling Cost for Location Update

Now, we investigate the effect on the average signaling cost for location update of the active handoff. In comparison to the signaling cost for the idle handoff C_i in Equations (1)–(3), the signaling cost for the active handoff in each method, C_a, can be obtained as follows.

$$C_a(\text{NO-INT}) = 2c_{mf} + 2c_{fh} + 2c_{mc} + 2c_{mr} \tag{8}$$
$$C_a(\text{INT-HOA}) = 2c_{mf} + 2c_{fh} \tag{9}$$
$$C_a(\text{INT-COA}) = 2c_{mf} + 2c_{fh} + 2c_{mc} \tag{10}$$

Then, taking into account whether or not the mobile node has an ongoing session while it performs location update, we analyze the location update which is composed of the idle handoff and the active handoff. Assuming that a mobile node can have arbitrary number of active sessions during handoff, we first derive active handoff probability $\alpha(k)$ that the mobile node will have k ongoing sessions while it moves into another subnet. $\alpha(k)$ is given by the characteristics of a mobile node, such as session arrival rate and mean residence time in a subnet, and can be derived as follows.

Suppose that the incoming session arrivals at a mobile node have a Poisson process with average arrival rate λ_s. Then, the probability density function of the new session inter-arrival time t_s is given by

$$f_s(t_s) = \lambda_s e^{-\lambda_s t_s}, \quad \text{for} \quad t_s \geq 0. \tag{11}$$

We assume that the session duration time is exponentially distributed with mean value $1/\tau_d$ and the mobile node has k active sessions during handoff. Let $t_{d,i}$ be the duration time of the ith session among active k sessions during active handoff ($1 \leq i \leq k$). Assuming $t_{d,i}$ to be independent identically distributed, $t_{d,i}$ has an exponential distribution given by following.

$$f_d(t_{d,i}) = \tau_d e^{-\tau_d t_{d,i}}, \quad \text{for} \quad t_{d,i} \geq 0 \quad \text{and} \quad 1 \leq i \leq k \tag{12}$$

Since the residence time in a subnet, t_u, can be assumed to be Gamma distributed with a mean of $1/\lambda_u$ (i.e., λ_u: idle handoff rate) and a variance of V_u [13], the probability density function of the subnet residual time $f_u(t_u)$ is described by

$$f_u(t_u) = \frac{\eta\lambda_u e^{-\eta\lambda_u t_u}(\eta\lambda_u t_u)^{\eta-1}}{\Gamma(\eta)}, \quad \text{for} \quad t_u \geq 0, \tag{13}$$

where η is the shape parameter with value $1/(V_u {\lambda_u}^2)$ and $\Gamma(\eta) = \int_{z=0}^{\infty} z^{\eta-1}e^{-z}dz$ is the Gamma distribution function. We can derive the Laplace transform of $f_u(t_u)$, which is represented as follows.

$$f_u^*(u) = \left(\frac{\lambda_u\eta}{s + \lambda_u\eta}\right)^{\eta} \tag{14}$$

From Equations (12) to (14), we can derive $\alpha(k)$ as

$$\begin{aligned} \alpha(k) &= Pr[t_s < t_u] \cdot \prod_{i=1}^{k} Pr[t_{d,i} > t_u] \\ &= Pr[t_s < t_u] \cdot (1 - Pr[t_d \leq t_u])^k. \end{aligned} \tag{15}$$

In Eqn. (15), the probability $Pr[t_s < t_u]$ and $Pr[t_d \leq t_u]$ are derived as follows.

$$\begin{aligned} Pr[t_s < t_u] &= \int_{t_u=0}^{\infty}\int_{t_s=0}^{t_u} \lambda_s e^{-\lambda_s t_s} \cdot f_u(t_u) \ dt_s dt_u \\ &= \left(1 - f_u^*(s)\right)\big|_{s=\lambda_s} \\ &= 1 - f_u^*(\lambda_s) \end{aligned} \tag{16}$$

$$\begin{aligned} Pr[t_d \leq t_u] &= \int_{t_u=0}^{\infty}\int_{t_d=0}^{t_u} \tau_d e^{-\tau_d t_d} \cdot f_u(t_u) \ dt_d dt_u \\ &= \left(1 - f_u^*(s)\right)\big|_{s=\tau_d} \\ &= 1 - f_u^*(\tau_d) \end{aligned} \tag{17}$$

From Eqn. (15), the average signaling cost for the location update, C_{update}, can be obtained from the weighted sum of C_i and C_a and given by

$$\begin{aligned} C_{\text{update}} &= \alpha(0) \cdot C_i + \left(\sum_k k \cdot \alpha(k)\right) \cdot C_a, \\ &= \left(1 - \sum_k \alpha(k)\right) \cdot C_i + \left(\sum_k k \cdot \alpha(k)\right) \cdot C_a, \end{aligned} \tag{18}$$

where $\alpha(0)$ is the idle handoff probability of a mobile node and $\alpha(k) = (f_u^*(\tau_d))^k \cdot (1 - f_u^*(\lambda_s))$.

4.3. Delay Analysis

Now we make an analytic comparison for the mobility management methods in terms of delay in the event of idle handoff or SIP session setup, and disruption time in active handoff, respectively. Delay time consists of the transmission time between network nodes and the processing time in a network node. Defining the total processing time as T_{proc} and the total transmission time as T_{trans}, respectively, we can derive the total delay time taken in each scenario, T_{total}, by summing T_{trans} and T_{proc}.

First of all, we derive the total transmission time (T_{trans}). Let l_{AB} be the average distance between node A and node B in terms of the number of hops that packets travel. It can be generally assumed that the transmission time is proportional to the distance between two nodes. Thus, the transmission time between node A and node B, τ_{AB}, can be expressed as

$$\tau_{AB} = l_{AB} \cdot \delta, \tag{19}$$

where δ is the proportionality constant that transforms the distance parameter into the time taken in that distance interval. We define the proportionality constant as follows: δ_d for a wired link, δ_r for a wireless link and δ_b for a mixed link of wired/wireless links, respectively. The transmission time between two network components is summarized in Table 1. Since the transmission time of the wireless link is usually longer than that of the wired link, it can be assumed that

$$\delta_d \quad \leq \quad \delta_r \quad \leq \quad \delta_b. \tag{20}$$

We then determine the total transmission delay taken in each scenario, such as idle handoff, session setup and active handoff as shown in Figures 2, 3 and 4 by summing the transmission time of each message required to perform each scenario. The total transmission time in each scenario is summarized in Table 2.

Next, we analyze the processing time (T_{proc}) taken in completing each scenario such as idle handoff, SIP setup and active handoff. We can assume that each network component is an M/M/1 queuing system [11, 14]. To determine total processing time, we need to know the average sojourn time of signaling messages in each network element. When each signaling

Table 1. Transmission time between two network components

Notation	Transmission time
τ_{mf}	$l_{mf} \cdot \delta_r$
τ_{fh}	$l_{fh} \cdot \delta_d$
τ_{hr}	$l_{hr} \cdot \delta_d$
τ_{mr}	$l_{mr} \cdot \delta_b$
τ_{mc}	$l_{mc} \cdot \delta_b$
τ_{rc}	$l_{rc} \cdot \delta_b$

Table 2. Transmission time for idle handoff, SIP session setup and active handoff

	Idle handoff	SIP session setup	Active handoff
T_{trans}(NO-INT)	$2\tau_{mf}+2\tau_{fh}+2\tau_{mr}$	$2\tau_{rc}+2\tau_{mr}$	$2\tau_{mf}+2\tau_{fh}+2\tau_{mc}$
T_{trans}(INT-HOA)	$2\tau_{mf}+2\tau_{fh}$	$2\tau_{rc}+\tau_{hr}+\tau_{fh}+\tau_{mf}+\tau_{mr}$	$2\tau_{mf}+2\tau_{fh}$
T_{trans}(INT-COA)	$2\tau_{mf}+2\tau_{fh}$	$2\tau_{rc}+\tau_{hr}+\tau_{fh}+\tau_{mf}+\tau_{mr}$	$2\tau_{mf}+2\tau_{fh}+2\tau_{mc}$

message arrives at a network element, such as foreign agent, home agent or home registrar, the average sojourn time is determined from the waiting time in the queue and the service time in each network component.

Let λ be the average occurring rate of events and $\mu_{(\cdot)}^{\text{msg}}$ be the service rate of message msg in a network element $(\cdot)$. Then, the processing load of a network element $(\cdot)$, $l_{(\cdot)}$, is given by

$$l_{(\cdot)} = \sum \frac{1}{\mu_{(\cdot)}^{\text{msg}}}. \tag{21}$$

Using Eqn. (21), total processing load of each scenario, L, can be obtained by summing processing load of network elements that are used to complete each scenario, and is given as follows.

$$L = \sum l_{(\cdot)} \tag{22}$$

For fair comparison of network scenarios under different mobility management schemes, we obtain the average message arrival rate in each network element by normalizing the processing load of each network element with total processing load in each scenario. Thus, the average message arrival rate in each network element $(\cdot)$, $\lambda_{(\cdot)}$, can be obtained by

$$\lambda_{(\cdot)} = \lambda \frac{L}{l_{(\cdot)}}. \tag{23}$$

For example, if we denote the idle handoff rate as λ^i and total processing load for idle handoff as L^i, the average message arrival rate in home agent can be expressed by $\lambda_h = \lambda^i \cdot L^i / l_h$.

From M/M/1 system modeling, the utilization of the network element, $\rho_{(\cdot)}$, is given by

$$\rho_{(\cdot)} = \lambda_{(\cdot)} \sum \frac{1}{\mu_{(\cdot)}^{\text{msg}}}. \tag{24}$$

Thus, the average sojourn time of message msg in a network element $(\cdot)$, $\sigma_{(\cdot)}^{\text{msg}}$, is given by

$$\sigma_{(\cdot)}^{\text{msg}} = \frac{1}{\mu_{(\cdot)}^{\text{msg}}(1-\rho_{(\cdot)})}. \tag{25}$$

We then determine the total processing time (T_{proc}) by summing the sojourn time of each signaling message at the network components through which each message is passed. T_{proc} is given as follows.

$$T_{\text{proc}} = \sum \sigma_{(\cdot)}^{\text{msg}} \tag{26}$$

As an example, from Fig. 3(a), the total processing time of idle handoff in INT-HOA method, $T_{\text{proc}}^{\text{idle handoff}}$, is expressed as

$$T_{\text{proc}}^{\text{idle handoff}}(\text{INT-HOA}) = \sigma_m^{\text{req}} + \sigma_f^{\text{req}} + \sigma_h^{\text{req}} + \sigma_h^{\text{rep}} + \sigma_f^{\text{rep}} + \sigma_m^{\text{rep}} = \left(\frac{1}{\mu_{\text{req}}} + \frac{1}{\mu_{\text{rep}}}\right)\left(\frac{1}{1-\rho_m} + \frac{1}{1-\rho_f} + \frac{1}{1-\rho_h}\right) \quad (27)$$

In a similar way, we can easily obtain the total processing time for NO-INT, INT-HOA and INT-COA method during each mobility scenario.

Finally, by using T_{trans} and T_{proc}, we can obtain total delay time, T_{total} as follows.

$$T_{\text{total}} = T_{\text{trans}} + T_{\text{proc}} \quad (28)$$

5. Numerical Results

In this section, numerical results are provided to demonstrate the performance of the proposed mobility management methods that integrate Mobile IP and SIP. For the reliable results, we consider two sets of signaling cost parameters given in Table 3. In the case of Set 1, the signaling cost is the same as the total sum of the number of exchanged signaling messages during the signaling procedure, because all signaling cost parameters are assumed to be 1. Set 2 denotes the case in which the signaling costs between the foreign network and the home network are assumed to be high, considering the distance between two networks [15].

First, we consider the effect of the Call-to-Mobility Ratio (CMR) on the total signaling cost (C_{total}) derived in Eqn. (7). Similar to the work in [15], we define the CMR as the ratio of the SIP session arrival rate λ_s to the idle handoff rate λ_u. When the CMR (λ_s/λ_u) is higher than 1, the SIP session arrival rate is high compared to the mobility rate, thus the signaling cost for SIP session setup dominates. By contrast, when the CMR is lower than 1, the mobility rate is higher than the arrival rate of SIP sessions and the signaling cost for idle handoff is more critical in reducing the total signaling cost.

Figure 5 shows the total signaling cost (C_{total}) for the parameter Set 1 in Table 3 as a function of CMR. It can be seen that the total signaling costs of the INT-HOA method and INT-COA method are lower than that of the NO-INT method, when the CMR is smaller than 1. Our proposed schemes can reduce signaling cost for the idle handoff, thus they are very effective when the idle handoff rate of a mobile node dominates the session arrival rate. In contrast, the NO-INT method has a lower signaling cost for high CMR. Since the NO-INT method

Table 3. Sets of signaling cost parameters

Signaling cost parameters	Set 1	Set 2
c_{mf}	1	1
c_{fh}	1	2
c_{hr}	1	1
c_{mr}	1	3
c_{rc}	1	1
c_{mc}	1	3

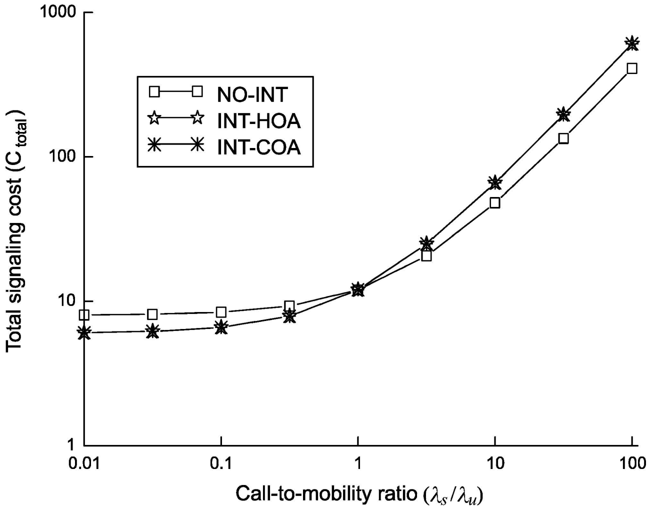

Figure 5. Total signaling cost vs. CMR for parameter Set 1.

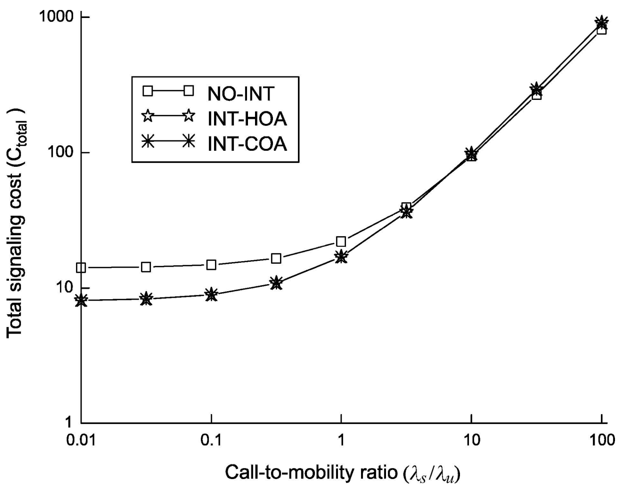

Figure 6. Total signaling cost vs. CMR for parameter Set 2.

performs registration separately to both the Mobile IP home agent and SIP home registrar for the idle handoff, a signaling message related to SIP session setup can be routed to the mobile node without accessing the home agent of the MN. When the signaling cost between foreign network and home network is assumed to be high compared to that of intra-network by using the parameter Set 2 in Table 3, the result in Figure 6 indicates that both the INT-HOA and INT-COA methods outperform the NO-INT method, regardless of the value of CMR.

Figure 7 represents the average signaling cost for the location update (C_{update}) derived in Eqn. (18) as a function of an active handoff probability α of a mobile node. Since the INT-HOA method performs only Mobile IP registration, whether the mobile node has ongoing session or not, the average signaling cost of the INT-HOA method is constant and the lowest among three methods, without regard to the value of α. The signaling cost in the event of idle handoff for the INT-COA method is the same as that of the INT-HOA method, because both

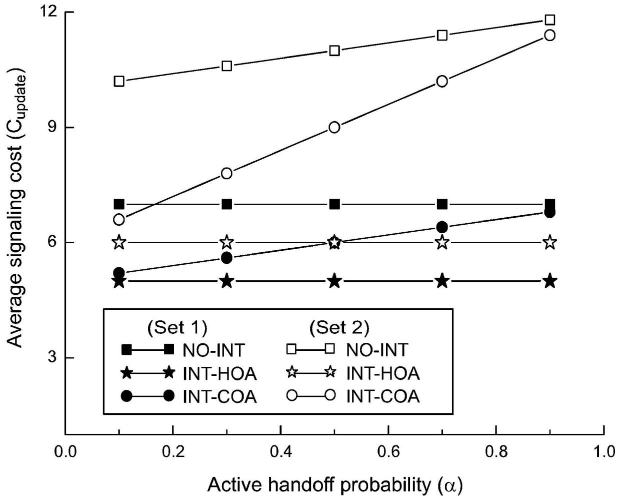

Figure 7. Average signaling cost for location update vs. active handoff probability.

methods operate in the same manner when a mobile node moves from one network to another without continuing its session. However, the signaling cost of the INT-COA method increases according to the increase of α because a mobile node that adopts the INT-COA method should execute an additional *SIP Re-INVITE* procedure to maintain an ongoing session. The NO-INT method has the highest signaling cost for all ranges of active handoff probability because it performs complex registration procedures for both Mobile IP and SIP protocols. Moreover, the NO-INT method should register to the SIP home registrar upon the completion of the handoff signaling between correspondent node and mobile node.

Next, we demonstrate the delay performance improvement of the proposed INT-HOA and INT-COA schemes compared to the conventional NO-INT scheme. The service time of each signaling message in network nodes depends on the configuration of the network element and has an implementation-specific value. However, we can generally assume that the service time of request/query messages ($\frac{1}{\mu^{\text{req}}}$) is higher than that of acknowledgement/reply messages ($\frac{1}{\mu^{\text{rep}}}$) because acknowledgement/reply messages do not require processing in network nodes and are just transferred to the next network node. Thus, it can be assumed that

$$\frac{1}{\mu^{\text{req}}} = \omega \cdot \frac{1}{\mu^{\text{rep}}}, \quad \omega < 1 \tag{29}$$

where ω is the ratio of the service time for request message to the service time for reply message. We assume that $\frac{1}{\mu^{\text{req}}} = 10$ ms and $\omega = 0.25$.

Actual distances between network nodes can be determined by network implementation and the distance between mobile nodes will vary when they are in motion. In addition, since data and signaling packets may take different paths each time according to network environments such as traffic load, it is difficult to know the exact distance between network nodes. For simplicity, we consider two distance parameter sets and assume that they have fixed values listed in Table 4, which are similar to [16]. The proportionality constant δ_d, δ_r and δ_b are set to 0.001, 0.003 and 0.005, respectively. In fact, the transmission time in real communication environments will be different from our assumption. However, since our goal is to investigate the changes in processing time caused by different mobility management methods, we set the transmission time to have similar values to the processing time.

Table 4. Sets of distance parameter

Distance parameter (# of hops)	Set 1	Set 2
l_{mf}	1	1
l_{fh}	10	20
l_{hr}	2	3
l_{mr}	12	20
l_{mc}	12	20
l_{rc}	5	10

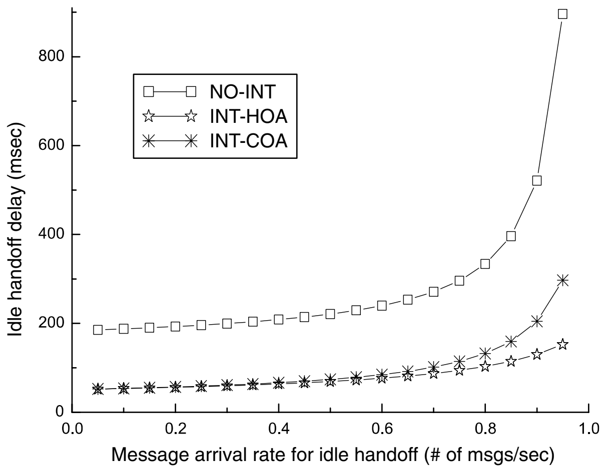

Figure 8. Idle handoff delay for distance Set 1.

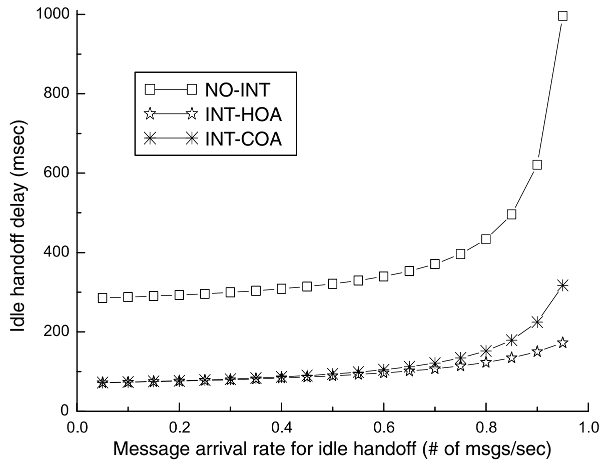

Figure 9. Idle handoff delay for distance Set 2.

Figures 8 and 9 show the idle handoff delay ($T_{\text{total}}^{\text{idle handoff}}$) of a mobile node when the message arrival rate for idle handoff varies. Here, the message arrival rate represents the number of messages arriving at a network element, thus increasing arrival rate means an increase in the traffic load of the network elements. When the traffic load increases, the idle handoff delay also increases because the processing time in network elements becomes higher. Note that our

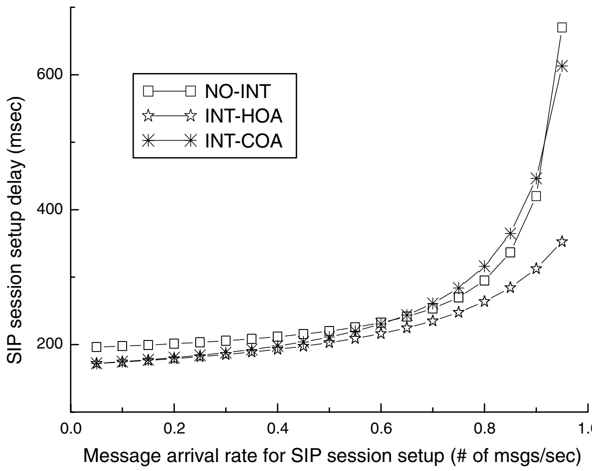

Figure 10. SIP session setup delay for distance Set 1.

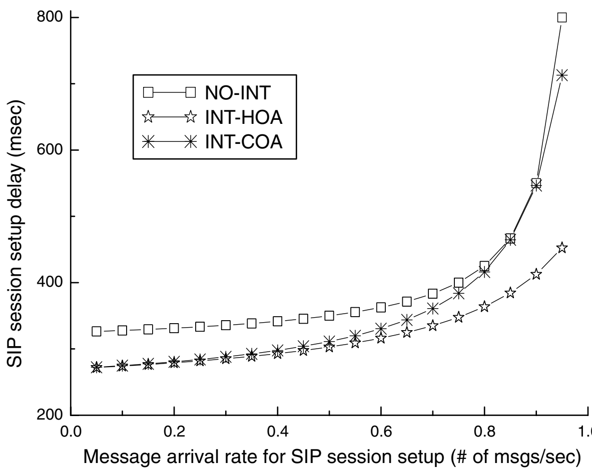

Figure 11. SIP session setup delay for distance Set 2.

proposed schemes can significantly reduce idle handoff delay for all ranges of message arrival rate compared with the NO-INT scheme.

Figures 10 and 11 show the SIP session setup delay ($T_{\text{total}}^{\text{SIP setup}}$) for two different distance sets. When the message arrival rate is low, the differences between SIP session setup delay times are lower than 1% for three comparative schemes. Since the overall number of signaling messages is the lowest in the INT-HOA scheme, the total delay time of SIP session setup in the INT-HOA scheme increases most slowly according to increasing traffic load. By contrast, the delay time for SIP session setup in the NO-INT and INT-COA schemes increases significantly when the traffic load becomes higher.

In Figures 12 and 13, we show the active handoff disruption time ($T_{\text{total}}^{\text{active handoff}}$) versus the message arrival rate for active handoff. We can observe that the handoff disruption time is the lowest in the INT-HOA method. The reason is that the signaling procedure for active handoff of the INT-HOA method is the simplest. Thus, active handoff disruption time can be effectively

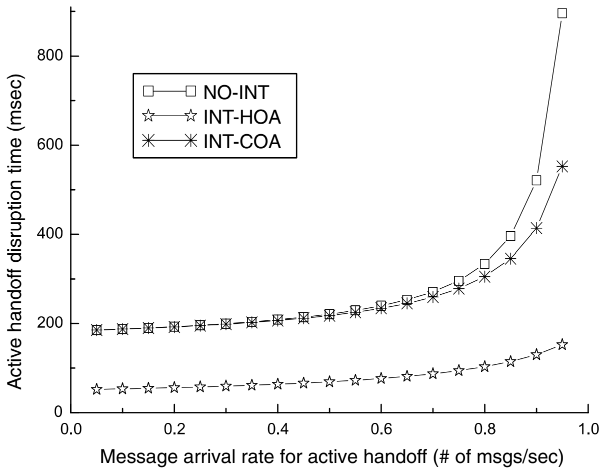

Figure 12. Active handoff disruption time for distance Set 1.

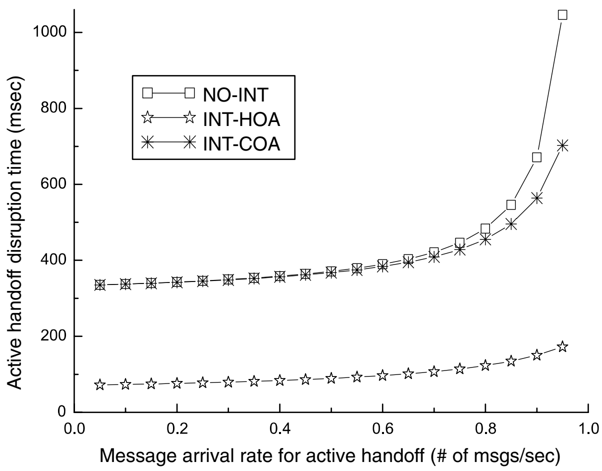

Figure 13. Active handoff disruption time for distance Set 2.

reduced when the INT-HOA scheme is applied to the packet data service for the delay-critical applications.

6. Conclusions

In this paper, we considered the problem of mobility support for mobile nodes with multimedia sessions. In order to support mobility efficiently, it is desirable that mobile nodes using multimedia sessions should adopt both SIP and Mobile IP protocols. To reduce the redundancy of mobility support between Mobile IP and SIP, we introduced two approaches that integrates mobility support functions in Mobile IP and SIP.

Compared to the conventional NO-INT method, both the INT-HOA and INT-COA methods can efficiently reduce signaling messages exchanged during idle handoff and active handoff. Performance evaluation also shows that our INT-HOA and INT-COA methods can significantly reduce idle handoff delay time and active handoff disruption time. In addition, we can see that the INT-HOA method, that uses only Mobile IP without SIP mobility, is a better way to support terminal mobility of the nodes adopting both Mobile IP and SIP, in view of overall protocol layers.

Our INT-HOA and INT-COA methods can be easily implemented in the mobile node with a simple algorithm that disables the registration function of SIP in the event of location change. Moreover, the INT-HOA method can significantly reduce the signaling traffic for idle handoff and active handoff as well as maintain compatibility. Even if the INT-HOA method causes media transmission overhead in the IPv4 based networks, the INT-HOA method is expected to be a promising candidate for the mobility support method in the next generation IPv6 based networks that adopt both Mobile IPv6 and SIP. Although we have considered the IPv4 based network, the proposed INT-HOA and INT-COA methods can be simply adopted to the IPv6 networks.

References

1. C. Perkins, "IP Mobility Support for IPv4," *IETF Request for Comments 3344*, Aug. 2002.
2. A. Campbell et al., "Cellular IP," IETF Internet Draft, draft-ietf-mobileip-cellularip-00.txt, Jan. 2000.
3. R. Ramjee and T.F. La Porta et al., "IP-Based Access Network Infrastructure for Next-Generation Wireless Data Networks," *IEEE Pers. Commun.*, Vol. 7, no. 4, pp. 34–41, Aug. 2000.
4. E. Gustafsson, A. Jonsson, and C. Perkins, "Mobile IP Regisonal Registration," IETF Internet Draft, draft-ietf-mobileip-reg-tunnel-05.txt, Sep. 2001.
5. S. Das, A. McAuley, and A. Dutta et al., "IDMP: An Intra-Domain Mobility Management Protocol for Next-Generation Wireless Networks," *IEEE Wireless Commun.*, Vol. 9, no. 3, pp. 38–45, Jun. 2002.
6. J. Rosenberg and H. Schulzrinne et al., "SIP: Session Initiation Protocol," *IETF Request for Comments* 3261, Jun. 2002.
7. H. Schulzrinne and E. Wedlund, "Application-Layer Mobility using SIP," *ACM Mobile Computing and Communications Review*, Vol. 4, no. 3, pp. 47–57, Jul. 2000.
8. N. Nakajima and A. Dutta et al., "Handoff Delay Analysis and Measurement for SIP Based Mobility in IPv6," in *Proc. of IEEE International Conference on Communications (ICC) 2003*, Vol. 2, pp. 1085–1089, 2003.
9. F. Vakil and A. Dutta et al., "Supporting Mobility for TCP with SIP," IETF Internet Draft, draft-itsumo-sipping-mobility-tcp-00.txt, Jun. 2001.
10. T.T. Kwon and et al., "Mobility Management for VoIP Service: Mobile IP vs. SIP," *IEEE Wireless Commun.*, Vol. 9, no. 5, pp. 66–75, Oct. 2000.
11. O. Haase, K. Murakami, and T.F. Laporta, "Unified Mobility Manager: Enabling Efficient SIP/UMTS mobile network control," *IEEE Wireless Commun.*, Vol. 10, no. 4, pp. 66–75, Aug. 2003.
12. C. Politis, K.A. Chew, and N. Akhtar et al., "Hybrid Mobility Management with AAA Context Transfer Capabilities for All-IP Networks," *IEEE Wireless Commun.*, Vol. 11, no. 4, pp. 76–88, Aug. 2004.
13. M.M. Zonoozi and P. Dassanayake, "User Mobility Modeling and Characterization of Mobility Patterns," *IEEE J. Select. Areas Commun.*, Vol. 15, no. 7, pp. 1239–1252, Sep. 1997.
14. G. Willman and P. Kuhn, "Performance Modeling of Signaling System no. 7," *IEEE Commun. Mag.*, Vol. 28, no. 8, pp. 44–56, Jul. 1990.
15. J.S.M. Ho and I.F. Akyildiz, "Dynamic Hierarchical Database Architecture for Location Management in PCS Networks," *IEEE/ACM Trans. on Networking*, Vol. 5, no. 5, pp. 646–660, Oct. 1997.
16. J. Xie and I. F. Akyildiz, "A Novel Distributed Dynamic Location Management Scheme for Minimizing Signaling Costs in Mobile IP," *IEEE Trans. Mobile Comp.*, Vol. 1, no. 3, pp. 163–175, Jul. – Sep. 2003.