mSIP: Extension of SIP for Soft Handover with Bicasting

Seok Joo Koh and Wook Hyun

Abstract—**This Letter proposes an extension of Session Initiation Protocol (SIP) to support soft handover with bicasting, named mobile SIP (mSIP). In the mSIP scheme, the soft handover is achieved by bicasting to the mobile node in the handover region. For this purpose, a new handover' header is defined in the SIP re-INVITE message. The mSIP handover can reduce handover latency and loss, compared to the SIP handover.**

Index Terms—**SIP, handover, mSIP.**

I. Introduction

THE SIP is used for control of real-time multimedia sessions [1]. It can also be used to support a variety of Internet mobility. This Letter will focus on the SIP-based IP handover for terminal mobility. In the existing SIP handover [2], a mobile node (MN) performs IP handover by sending another INVITE (called re-INVITE) method to the correspondent node (CN) after getting a new IP address. This SIP handover tends to give a large handover latency associated with movement detection and IP address configuration [3]. This is mainly because the SIP handover cannot effectively support the 'soft' handover.

This Letter proposes an extension of SIP to support soft handover with 'bicasting'. A recent work on SIP-based bicasting [4] proposed to use a network agent named 'Handover Assisted Server (HOAS)' for bicasting, which is located in the network between MN and CN. However, this Letter will consider an 'end-to-end' bicasting between MN and CN without using any network agent. In the proposed mSIP scheme, MN will communicate with CN using bicasting over two IP addresses in the handover region.

II. mSIP for Soft Handover

A. Existing SIP Handover

We first describe the existing SIP handover [2], as depicted in Figure 1. In the AR_A region, MN is communicating with CN by using IP address A. As it moves into AR_B region, the media channel with IP address A will be disconnected. MN then begins movement detection and address configuration. After getting a new IP address, MN sends an SIP re-INVITE method, which contains the new IP address, and receives the SIP 200 OK from CN. During this handover period, MN cannot receive the media stream from CN, which induces the concerned handover latency. In the figure the SIP ACK method is not shown, since it does not affect the handover latency.

Manuscript received March 4, 2008. The associate editor coordinating the review of this letter and approving it for publication was Prof. Maode Ma.

S. J. Koh is with the School of Electrical Engineering and Computer Science, Kyungpook National University, Daegu, Korea (e-mail: sjkoh@knu.ac.kr).

W. Hyun is with the Protocol Engineering Center, Electronics Telecommunications Research Institute, Daejon, Korea (e-mail: whyun@etri.re.kr).

Digital Object Identifier 10.1109/LCOMM.2008.080335.

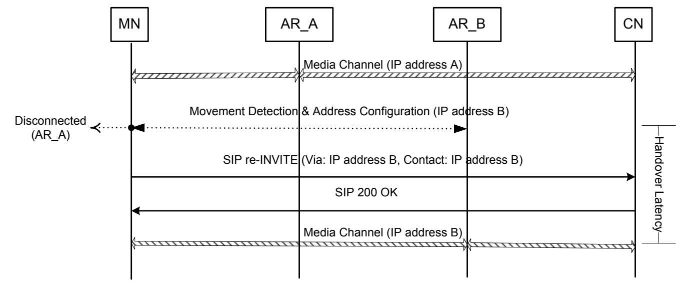

Fig. 1. Information flow of the existing SIP handover.

B. New Headers for mSIP Handover

The mSIP handover is designed to support 'bicasting' of media streams from CN to MN during handover. For this purpose, we define a new SIP 'handover' header, as follows:

```
Handover: add | del;ip=a.b.c.d
```

This new header will be inserted into SIP re-INVITE method. This header instructs CN to add or delete the IP address indicated (`a.b.c.d`) to or from the associated tables used for SIP signaling and media streams. In particular, the '`add`' flag will inform CN to start bicasting to MN (with IP address indicated), in which CN will duplicate and transmits the identical media streams to MN. On the other hand, the '`del`' flag instructs the CN to stop bicasting (i.e., transmission over the old IP address). For backward compatibility with the current SIP protocol, the re-INVITE message may include the 'require' header in the form of "Require: handover." If the CN cannot support the mSIP handover (i.e., handover header), it will respond to MN with "420 (Bad Extension)". In this case, MN may try to perform the exiting SIP handover, as described earlier. It is noted in the mSIP handover that the re-INVITE message also contains the associated Session Description Protocol (SDP) information in the message body so as to describe the characteristics of the media channels associated with handover, as specified in the RFC 3261 [1].

C. Algorithm of mSIP Handover

The mSIP handover is based on 'bicasting' from CN to MN in the handover region. In the proposed scheme, an MN is assumed to exploit the link-layer triggers such as Link-Up and Link-Down. That is, when the MN goes into the handover region, it will initiate the mSIP handover operations with the help of link-layer triggers. On the other hand, it is noted that the existing SIP handover does not use such link-layer information, since the SIP cannot support the soft handover with bicasting. The proposed mSIP handover procedures are illustrated in Figure 2.

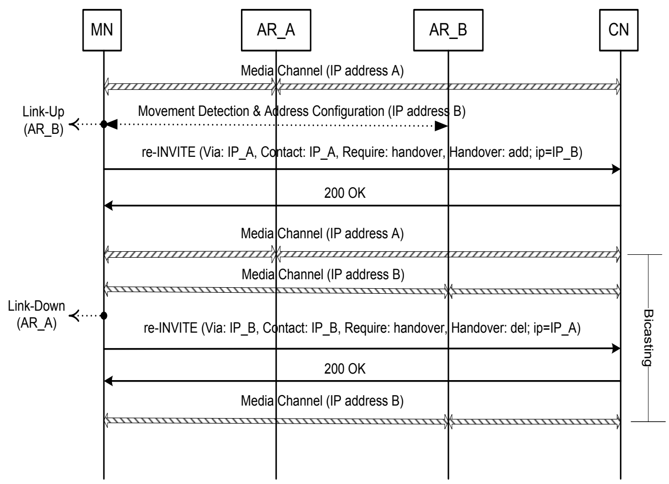

Fig. 2. Information flow of mSIP handover with bicasting.

In the figure, MN initially uses IP address A (IP_A). When it moves into AR_B region, it will detect a Link-Up for AR_B. It then performs movement detection and obtains a new IP address (IP_B) via DHCP or IPv6 address auto-configuration. After getting a new IP address, MN sends an SIP re-INVITE method which contains the information on IP_B and handover header, as specified in the figure. CN will respond with SIP OK message to MN. Since then, CN can transmit an identical media stream to MN over both IP_A and IP_B. That is, CN starts bicasting to MN. In this period, MN transmits its own media stream to CN using either IP_A or IP_B. As MN further moves into the AR_B region, it will detect the Link-Down event for AR_A. MN then sends an SIP re-INVITE message to CN, as specified in the figure, so as to stop bicasting (i.e., transmission over IP address A). After the corresponding OK message is received, MN and CN use only the IP_B address. It is noted that the proposed mSIP handover can be applied to horizontal handover (MN with a single network interface) as well as vertical handover (MN with two or more network interfaces). In the vertical handover, MN can receive duplicated data streams bicast by CN via its two network interfaces (AR_A and AR_B). In case of horizontal handover, MN will 'actually' receive only one data stream via its single network interface (AR_A or AR_B), even though CN bicasts data packets to MN. Even in this case, the mSIP scheme can reduce the probability of handover losses with the help of bicasting.

D. Implementation Considerations

In the application point of view, MN will receive the duplicated media streams from CN in the bicasting period. In this case, MN's application shall select only one of the received two media streams and then discard the other stream. For example, an MN may prefer the media data delivered received from the new IP address to the one received from the old IP address. In the viewpoint of SIP signaling path, the re-INVITE message with `add` flag is transmitted over the old IP address, whereas the re-INVITE message with `del` flag may be delivered over the new IP address. Accordingly, when the CN receives the re-INVITE with add flag, it shall bind its SIP signaling channel to the new IP address as well as the old IP address. After the re-INVITE message with `del` flag is received, the CN can release the old IP address from its SIP signaling channel.

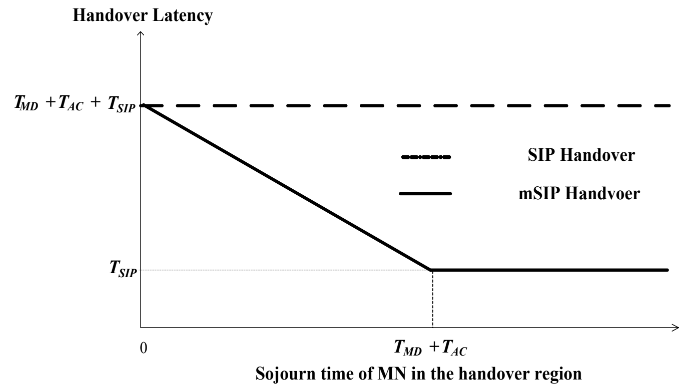

Fig. 3. Handover latency of SIP handover and mSIP handover.

III. Performance Analysis

The handover performance can be measured as handover loss and latency. It is clear that the mSIP handover can reduce the handover loss probability, compared to the SIP handover, with the help of bicasting. In this section, we will analyze the handover latency taken for SIP handover and mSIP handover. The handover latency of SIP handover (T_{SIP_HO}) can be calculated as:

$$T_{SIP\_HO} = T_{MD} + T_{AC} + T_{SIP} \tag{1}$$

In the equation, T_{MD} represents the movement detection (MD) delay taken for MN to detect its movement in the new subnet, T_{AC} is the address configuration (AC) delay required for configuration of a new IP address via DHCP or stateless address auto-configuration, and T_{SIP} is for the delay of exchanging the SIP re-INVITE and OK messages. It is noted that the T_{MD} and T_{AC} will depend on the MD and AC schemes employed in the network, whereas T_{SIP} is equal to the round trip time (RTT) between MN and CN. In case of mSIP handover, the MN performs the handover in the handover (overlapping) region that is located between the two concerned networks. In the handover region, MN can still receive the media streams from CN using bicasting, even when the MD and AC operations are performed. Accordingly, the mSIP handover latency is equal to only the delay taken for MN to exchange the SIP re-INVITE and OK messages with CN. Then, the mSIP handover latency (T_{mSIP_HO}) can be summarized as:

$$T_{mSIP\_HO} = T_{SIP} \tag{2}$$

In fact, the mSIP handover latency in the real networks will depend on the sojourn time of MN in the handover region, as shown in Fig 3.

As shown in the figure, if the sojourn time of MN in the handover region is less than $T_{MD} + T_{AC}$, the mSIP handover latency may increase up to the $T_{MD} + T_{AC} + T_{SIP}$, which is equal to the handover latency of the existing SIP handover.

In the meantime, only if the sojourn time in the handover region is large enough to complete the MD and AC operations, the mSIP handover latency can be reduced to only T_{SIP}. Such a performance gain comes because the mSIP handover can enable MN to receive the media streams from CN using bicasting even during the handover.

IV. Conclusion

This Letter presents an extension of SIP for soft handover with bicasting. The mSIP handover with bicasting can reduce handover loss and latency during handover, compared to the existing SIP handover. In particular, the proposed mSIP scheme seems to be more beneficial to handover of MN with a 'ping-pong' movement pattern in the handover region. Future works will be needed on performance evaluation of the proposed scheme with a simulation tool or prototype implementation.

Acknowledgment

This research was supported by the MKE (Ministry of Knowledge Economy) of Korea, under the ITRC support program supervised by the IITA (IITA-2008-C1090-0801-0026).

References

[1] J. Rosenberg *et al.*, "SIP: Session Initiation Protocol," IETF RFC3261, 2002

[2] H. Schulzrinne and E. Wedlund, "Application layer mobility using SIP," *ACM SIGMOBILE Mobile Computing and Commun. Rev.*, vol. 4, no. 3, pp. 47–57, 2000.

[3] N. Nakajima *et al.*, "Handoff delay analysis and measurement for SIP based mobility in IPv6," in *Proc. IEEE International Conference on Communications (ICC)*, pp. 1085–1089, 2003.

[4] H. Izumikawa and R. Lillie, "SIP-based bicasting for seamless handover between heterogeneous networks," *IETF Internet Draft, draft-izumikawa-sipping-sipbicast-00*, Nov. 2007.