Trombone Routing Mitigation Techniques for IMS/MMD Networks
Tsunehiko Chiba, Hidetoshi Yokota, Akira Idoue
KDDI R&D Laboratories, Inc., 2-1-15 Ohara, Fujimino, Saitama 356-8502 Japan
Ashutosh Dutta, Kyriakos Manousakis, Subir Das
Telcordia Technologies, Inc., 1 Telcordia Drive, Piscataway, NJ 08854 USA
Henning Schulzrinne
Columbia University, 1214 Amsterdam Avenue, New York, NY 10027 USA

Abstract – Real-time services such as VoIP and multimedia streaming are affected during a mobile’s rapid handoff due to the associated delay resulting from associated handoff operations of discovery, detection, configuration, registration, media redirection, and processing at different network nodes. Redundant communication path of any signaling and media adds to the delay during session setup and media delivery. This paper highlights the overall problem associated with redundant routing in a 3GPP2-based MMD (Multimedia Domain) environment and proposes several mechanisms to mitigate these problems. We also compare these mechanisms and select one for existing IMS/MMD networks that use MIPv4 as the mobility protocol. We present the implementation details of the selected redundant routing mitigation technique in the experimental MMD environment and verify the effectiveness by analyzing the measurement results.

I. INTRODUCTION

To provide ubiquitous wireless services, wireless service providers need to build an infrastructure that can support a variety of operations, such as quality of service, security and mobility. IMS (IP Multimedia Subsystem) [1] and Multimedia Domain (MMD) [2] architecture as defined by 3GPP (3rd Generation Partnership Project) and 3GPP2 (3rd Generation Partnership Project 2), respectively, are accepted by the industry as the common infrastructure for NGN (Next Generation Networks). These architectures involve several components such as P-CSCF (Proxy Call Session Control Function), I-CSCF (Interrogating CSCF), and S-CSCF (Serving CSCF). Figure 1 shows a general MMD network that supports macro mobility with MIP (Mobile IP) [3] [4].

Currently, in 3GPP2-based MMD networks, media sessions are setup by SIP (Session Initiation Protocol) [5] and mobility is taken care of by MIP. Fast session setup and efficient media delivery are key factors in optimizing services and performance in wireless networks. The number of messages and the communication distance between the MN (Mobile Node) and other communicating entities in the network often contribute to the overall delay for the completion of session setup and handoff. Therefore, it is important to reduce the delay caused by the exchange of signaling messages between the entities by limiting the communication distance. In order to avoid the triangular routing associated with media traffic in the case of MIPv4, route optimization had been proposed [6]. However, it was never standardized. Although MIPv6 takes care of route optimization for media traffic, other signaling messages (e.g., SIP) are still affected due to reverse tunneling between the MN and HA (Home Agent). Reverse tunneling is a bi-directional tunneling between the FA and HA or between MN and HA, which forces the packets to go via HA. Trombone routing is a phenomenon that contributes to the levels of indirection to the communication path. The additional routing caused by reverse tunneling often amplifies the trombone routing effect. Many of the mobility protocols such as MIPv4 and MIPv6 introduce the trombone routing problems to a different degree, where the packets always need to traverse to the home network. In this paper we propose a few mechanisms that can be used to provide solutions for trombone routing problem for SIP signaling in existing MIPv4-based MMD networks.

Figure 1: IMS/MMD Architecture

The remainder of the paper is organized as follows. Section II describes some of the related work in the area of mobility optimization. Section III explains the trombone routing problem in MMD networks for two different mobility protocols such as MIPv4 and MIPv6. Section IV explains several mitigation techniques for MIPv4. Section V explains the cost analysis and provides the implementation details of
one of the mitigation techniques and the associated results. Finally, Section VI concludes the paper.

II. RELATED WORK

Culp [7] introduced the concept of trombone routing in cellular handoffs and illustrated how the inclusion of additional networking components during inter-system handoff is not desirable. Reference [8] discusses the optimal routing in 3GPP networks, which eliminates the trombone routing. The extent of the effect of trombone routing varies from one mobility protocol to another. There are several mobility protocols, such as Cellular IP [9] and HAWAII [10], which work in conjunction with macro mobility protocols (e.g., MIP) and help limit the traversal of signaling messages during MN’s movement within a domain. Thus, it reduces the effect of trombone routing for intra-domain movement. On the other hand, SIP-based mobility [11] avoids the trombone routing altogether because it does not involve any HA or FA (Foreign Agent). Kim et al [12] provide a solution where SIP signaling is piggy-backed over MIPv6 binding update, however in this case, MIPv6 HA and SIP registrar must cooperate. Although a variety of solutions use a combination of network layer and application layer mobility protocols [13], there is no solution that can avoid the trombone routing problem for MIPv4 in IMS/MMD networks. Since many carriers still use MIPv4 as the installed base, it is desirable to examine a mechanism that can avoid trombone routing within MIPv4 and can provide optimized mobility operation.

III. OVERVIEW OF TROMBONE ROUTING IN MMD NETWORKS

Trombone routing occurs in an MMD network because the signaling and media travel via the HA, which usually resides in the home network. In this section, we briefly describe how trombone routing affects both the registration and call setup procedures for two different types of mobility protocols, MIPv4 and MIPv6.

A. Effect of Trombone Routing on MIPv4

Figure 2 shows the inefficiency associated with trombone routing in MIPv4. When P-CSCF is located in the same network as the MN, an access network provider can implement better policy control. In this case, all SIP signaling messages have to traverse the home network via HA, before being routed to the P-CSCF. This inefficiency is partly due to the reverse tunneling associated with MIP. Similarly, any incoming call (e.g., INVITE) from CN (Correspondent Node) or response to SIP registration traverses via P-CSCF and gets redirected to HA in the home network before being delivered to the MN in the visited network. This also results in an increased call setup delay. Creating a security association between MN and P-CSCF and a media session context at P-CSCF is mandatory before the media can pass through the new visited network. Upon movement of the MN to the new network, SIP REGISTER messages that help create a security association using AKA (Authentication and Key Agreement) [14] procedure and re-INVITE messages that help create a session context, are delayed due to the trombone routing problem discussed above. Thus, delayed creation of the

security association and media session context increases the handoff delay as well.

B. Effect of Trombone Routing on MIPv6

Similarly, trombone routing also affects the efficiency of MMD networks when MIPv6 is used as the mobility protocol. Unlike MIPv4, there is no FA used in this case. MN obtains a new care-of-address from the access router by using stateless auto-configuration during each handoff. There is a tunnel between the MN and the HA. However, while using IPv6 in an MMD network, it is often customary to use MN’s home address in the contact address field during SIP registration. Thus, during the re-registration process, only the address of the new P-CSCF is updated at the S-CSCF. The HA keeps a mapping of MN’s home address and its most recent care-of-address by means of the MIP registration. Since there is a reverse tunnel established between MN and HA, both SIP call setup and registration (re-registration) processes are subjected to trombone routing. Thus, just as in MIPv4, this trombone routing also affects the handoff delay because of the long signaling path traversed via the home network.

IV. MITIGATION TECHNIQUES

In this section, we introduce some mitigation techniques applied to SIP signaling, especially when MIPv4 is used as the mobility protocol. Although 3GPP2 provides support for IPv4 as well as IPv6 in MMD networks, we focus our discussion on MIPv4. We describe three relevant mitigation techniques for MIPv4 and describe the prototype details of one of those techniques in Section V. These techniques, when applied to media traffic, can also reduce the one-way delay of the data packets because of the direct path between the communicating hosts. Some of these techniques may need to be modified in order to take care of trombone routing for MIPv6 since the tunnel is between MN and HA instead of between FA and HA as in case of MIPv4.

A. First Approach: Selective Tunneling

Figure 3 illustrates the selective tunneling approach. Reverse tunneling at the FA partly gives rise to the trombone routing problem. RFC 3024 [15] specifies a means to make use of the encapsulated delivery style to perform selective reverse
tunneling. This is intended to support packet delivery to local resources. Packets meant to be reverse tunneled are sent using encapsulated delivery style (via the MN-FA tunnel) by the MN. The FA must reverse tunnel these to the HA. Packets not meant to be reverse tunneled are sent using direct delivery style (not encapsulated), the FA will forward these and will not use reverse tunnel to send these to the HA. The MN can send all packets meant for the P-CSCF using normal IP routing and the FA will forward these as regular packets.

This approach solves one part of the trombone routing problem by optimizing the route from the MN to the P-CSCF. However, packets from the P-CSCF to the MN will be still routed via the HA. In addition, this selective reverse tunneling with encapsulated delivery style approach assumes changes to be made in the MIP protocol behavior. Thus this feature may not be desirable for already installed MIPv4 infrastructure.

We propose a similar technique to optimize packet delivery from MN to P-CSCF and P-CSCF to MN in the visited network. The proposed approach provides an encapsulation technique between FA and P-CSCF. It installs packet interceptors and mangling modules both in FA and P-CSCF to perform selective tunneling operation in both directions and also establishes an IP-IP tunnel between the P-CSCF and the FA for all the packets destined for the MN from the P-CSCF. Packets received at the FA via the P-CSCF-FA tunnel will be decapsulated at the FA and forwarded to the MN. This is identical to the manner in which encapsulated packets received at the FA via the HA-FA tunnel are processed. In addition, it does not require any changes in MIP functional behavior. Figure 3 illustrates the selective tunneling approach and shows how the SIP signaling packets are not affected by trombone routing in either direction.

Figure 3: Selective tunneling in IMS/MMD network

B. Second Approach: Piggybacking SIP registration over MIP

Figure 4 shows an approach that uses piggybacking of SIP registration over the MIP signaling. In this approach, MIP and SIP signaling (i.e., control plane) are integrated so that SIP messages can be delivered as part of the MIP control messages and thus bypass the FA-HA tunnel:

1. MN receives the FA advertisement and obtains a new P-CSCF address using DHCP [16]

Figure 4: Piggybacking SIP registration over MIP

2. The MIP process on the MN (MIP-MN) reports the detection of a new P-CSCF to the SIP User Agent (SIP-UA) on the MN.
3. SIP-UA initiates a SIP registration message by signaling the MIP-MN.
4. MIP-MN invokes MIP and piggy-backs SIP registration message through the MIP-specific messaging operation.
5. FA invokes both the MIP and SIP registration messages in parallel; SIP registration message to P-CSCF and MIP registration message to HA. Note that, here FA is not only an IP layer-forwarding node, instead, it also acts as an application-specific relaying node.
6. SIP registration message is delivered to the S-CSCF.
7. SIP response message is delivered back to the P-CSCF.
8. The P-CSCF sends the SIP response message to the FA.
9. FA passes over the MIP or SIP response message to the MIP-MN.
10. Finally, the MIP-MN reports the notification of the SIP response message to the SIP-UA and the SIP registration is completed.

Consequently, SIP messages from the MN to P-CSCF and vice versa traverse according to the routes set by the regular IP routing and not by MIP, and no FA-HA tunnel is used for forwarding the messages.

C. Third Approach: Dynamic Home Agent Assignment

The problem of additional routing can also be mitigated by introducing HAs closer to MN’s visiting networks. These dynamic HAs are commonly known as Mobility Agents (MA). Dynamic home agent assignments are supported by the MIPv4 base protocol [17]. By placing HAs close to FAs, one can minimize the routing path to a great extent. Figure 5 depicts the scenario where MAs are being deployed in the core network of the visited domain, and thus the MA helps to reduce the path traversal for both signaling and media. It is important to note that it is not necessary to deploy MA in every subnet. Depending upon the topology and size of the visited network, multiple MAs can be deployed. However, one MA can handle multiple FAs since it will usually be placed one level higher than the subnet level. Several published mobility optimization protocols, such as MIP with Regional Registration [18] and IDMP [19] use a similar concept.
We now describe some of the advantages and disadvantages of each of these approaches. Selective reverse tunneling and interceptor at P-CSFC makes use of the standard mechanism, and there is no need to change the MIP protocol. However, it needs a filtering mechanism at FA and P-CSFC and thus adds additional processing delay at FA and P-CSFC.

Piggybacking SIP messages on MIP reduces the number of signaling messages and combines two protocol operations into one. However, both FA and MN need to be modified. The FA also needs to be equipped with SIP user agent functionality so that it can act as a proxy on behalf of the MN.

The dynamic home agent assignment approach reduces the number of signaling during MN’s movement within a domain and thus is suitable for roaming. It also follows the IETF (Internet Engineering Task Force) standard specification. However, this approach introduces the need for installing an MA in the visited networks. Discovery of MA and additional signaling between MA and HA need to be taken care of when deploying this approach.

Consequently, we chose to implement the selective tunneling technique because no changes are required on the MN for this approach and it seems more suitable for an operator’s deployment perspective, as it does not involve installation of any additional networking element.

V. IMPLEMENTATION AND MEASUREMENT RESULTS

Figure 6 shows the effect of both trombone routing and its optimized version in the experimental MMD environment where we implemented the mitigation technique. Specifically, when the MN sends SIP REGISTER to S-CSFC or INVITE to a CN for setting up a session, it has to send the signaling via the P-CSFC. In the non-optimized case, the SIP signaling transmitted by the MN is picked up by the corresponding FA and then is forwarded to the HA, and finally the HA forwards the message to P-CSFC.

A similar effect appears in the reverse path for SIP signaling from P-CSFC to the MN. In the reverse case, the SIP signaling is forwarded to MN’s home address and is picked up by the HA. The HA continues to forward the SIP signaling to MN’s current FA, which eventually gets delivered to the MN.

The forward and reverse SIP signaling follow these paths even if the FA and the targeted P-CSFC entities belong to the same subnet. As we mentioned earlier, trombone routing delays the delivery of SIP signaling significantly affecting MN’s handoff delay. The greater the delay between FA and HA, and between HA and P-CSFC, the more significant the handoff delay is due to the trombone routing effect.

On the other hand, in the optimized case, the forward path (MN to P-CSFC) of SIP signaling is routed directly from FA to the corresponding P-CSFC, without having to traverse the FA-HA-P-CSFC path. Similarly, for optimizing the reverse SIP signaling path, it is routed directly from P-CSFC to the corresponding FA without having to traverse the P-CSFC-HA-FA path. The FA-HA-P-CSFC and reverse paths are much longer compared to the P-CSFC-FA paths resulting in unnecessarily long delays during handoff, affecting the performance of the services provided and customer satisfaction.

Figure 7 shows the SIP messages between different network entities when trombone routing is in effect. For the sake of simplicity, we have not illustrated all the messages used in IMS/MMD, such as PRACK and UPDATE, but have focused only on the SIP REGISTER and INVITE messages. Because of the trombone routing effect, initial registration and call setup are delayed in visited network 1. Similarly, after the MN moves to the visited network 2, it performs the re-registration and sends a re-INVITE to generate the new context. Thus, the creation of context at P-CSFC enables the media gate on PDSN (Packet Data Serving Node), which also acts as FA, to open up right after the re-INVITE completion.

Figure 8 shows the flow when the trombone routing mitigation technique (approach1) is applied. Because of this technique, the messages do not travel via the HA anymore. This helps to reduce the time taken for the registration and call setup operation in visited network 1 and handoff time to the visited network 2.

A. Cost Analysis

We provide a simple calculation of the delay based on the cost due to traversal of signaling messages and processing cost in the networking nodes. We do not include all the processing costs in each networking node in this analysis. We assume that the communication distance between MN and FA is d1,
between FA and HA is $d_2$, between HA and P-CSCF is $d_3$, between P-CSCF and I-CSCF is $d_4$, and between I-CSCF and S-CSCF is $d_5$.

![Figure 7: SIP signaling flow without mitigation technique](image)

The communication distance between P-CSCF and S-CSCF is $d_6$, and between FA and P-CSCF is $d_7$. Without loss of generality we assume that $d_1$, $d_2$, and $d_3$ are smaller than the other distances (i.e., entities are close to each other). The associated cost for traversing these communication distances are $t_1$, $t_2$, $t_3$, $t_4$, $t_5$, $t_6$, and $t_3$, respectively. We now analyze both the trombone routing and mitigation cases. We can assume processing cost at HA and FA can be designated as $P_{HA}$ and $P_{FA}$ respectively when the mitigation technique is not applied. On the other hand, when the mitigation technique is applied, there is an additional processing cost due to the mangling of packets and the additional look up at FA and P-CSCF. We can assume this additional processing cost to be $P_{mitigate}$ for each message. Processing at HA is completely avoided in the mitigation case as the signaling does not pass through HA.

First, we discuss the case where mitigation technique is not applied. Before the move, the MN is in visited network 1 and is subjected to the registration and call setup delays. Referring to Figure 7, these delays can be calculated as follows. The SIP registration cost is $2(t_1+t_2+t_3+t_4+t_5)+2(P_{HA}+P_{FA})$. Similarly, call setup consists of three SIP-based signaling, such as INVITE, OK, and ACK. This cost consists of the costs due to traversal and processing operation that amounts to $3(t_1+t_2+t_3+t_4+t_5)+3(P_{HA}+P_{FA})$. When the MN moves to the visited network 2, it needs to re-register. There are other common set of operations that are part of the handoff, such as PPP (Point-to-Point Protocol) setup and DHCP operation to discover P-CSCF, which are same for both with and without the mitigation technique. Thus, the registration cost in the visited network 2 is the same as that in visited network 1 and amounts to $2(t_1+t_2+t_3+t_4+t_5)+2(P_{HA}+P_{FA})$. Since the re-INVITE and 200 OK signaling help create the new context in the P-CSCF that opens up the gate for media at the corresponding FA, call setup cost in the visited network 2 is calculated as $t_1+t_2+t_3+2t_6+P_{HA}+P_{FA}$. The total handoff delay D1 in case of no mitigation is calculated as $3(t_1+t_2+t_3)+2(t_1+t_2)+3(P_{HA}+P_{FA})$.

Next, we analyze the similar cost when mitigation technique is applied. Registration cost in the visited network 1 is $2(t_1+t_2+t_3+t_4)+2P_{mitigate}$ and call setup delay is $3(t_1+t_2+t_3+t_4+t_5)+3P_{mitigate}$. Basically both of these values are smaller than the values when the mitigation technique is not applied.

![Figure 8: SIP signaling flow with mitigation technique (1)](image)

Referring to Figure 8, when the MN moves to the visited network 2, the re-registration cost is $2(t_1+t_2+t_3+t_4)+2P_{mitigate}$ and the call setup cost is $t_1+t_2+2t_6+P_{mitigate}$. Thus, the delay D2 during handoff from visited network 1 to visited network 2, in case of mitigation, is $3(t_1+t_2)+2(t_1+t_2)+2t_6+3P_{mitigate}$.

The handoff delay gain due to the mitigation technique is calculated to be $(D1-D2)=3(t_1+t_2+t_3)+3(P_{HA}+P_{FA}+P_{mitigate})$.

The effect of the mitigation technique is felt more when the distance between the visited network and home network is greater. When the distance is small, the benefit of the trombone routing mitigation is offset by the additional processing time at FA and P-CSCF during packet capture and encapsulation operation.

### B. Trombone Routing Mitigation Implementation Details

In this section we provide the implementation details for optimizing trombone routing for SIP signaling and measurement results. The implementation of optimization of trombone routing is applied only to SIP signaling and does not include the paths followed by the media initiated or terminated at the MN. The implementation consists of two collaborating modules; interceptor modules and mangling modules for SIP signaling. The SIP signaling packets are captured by utilizing Linux `iptables`, which is used to create rules for packet filtering and NAT modules. While technically, `iptables` serves as a tool that controls the packet filtering and NAT components within the kernel, the name `iptables` is often used to refer to the entire infrastructure, including `netfilter`, connection tracking, and NAT, as well as the tool itself. The `iptables` tool is a standard part of all modern Linux distributions. After the packets have been filtered and captured using `iptables`, they are manipulated by the software module we developed. This software module accesses, manipulates, and releases the captured packets for standard kernel processing. The objective for manipulating the SIP signaling packets is that the routing conforms to the optimized trombone routing specifications represented by Figure 3. The
manipulation process appropriately replaces the IP address headers for routing. P-CS-CSCF keeps track of the associated FA.

C. Analysis of Measurement Results
We have increased the communication distances $d_2$, $d_3$, $d_4$, and $d_5$ by an additional 500 ms delay using the NIST delay simulator [20] and measured the performance in the experimental environment. The additional NIST delay between home network and visited network emulates a real deployment scenario. Figure 9 shows the handoff time for both without and with mitigation techniques, respectively.

![Figure 9: Handoff delay comparison](image)

In particular it shows the breakdown of delay due to several operations such as layer 2 handoff, PPP configuration, Mobile IP registration, DHCP INFORM for P-CS-CSCF discovery and SIP related signaling. It is important to note that the delay due to non-SIP related signaling does not differ for both the cases, with and without mitigation. Thus, we focus the comparison here on the improvement related to SIP signaling only. The amount of delay due to SIP signaling is actually a large fraction of the total handoff delay. For example, when the mitigation technique is not applied, the delay attributed due to SIP signaling is 6.5 sec out of total delay of 9.4 sec, almost 70% of the total delay. When the mitigation technique is applied, the delay due to SIP signaling is reduced to 4.9 sec. The mitigation technique involves some additional amount of processing time at FA and P-CS-CSCF because of additional packet capturing and mangling operations. Thus there is some tradeoff between this additional processing time at FA and P-CS-CSCF and time saved due to mitigation of tombone routing. This measurement analysis has not considered other required SIP signaling messages such as PRACK and UPDATE. When these signaling messages are included in real deployment network, effect of mitigation technique can be felt more as the difference in handoff time will become more apparent.

VI. CONCLUSIONS
The effect of tombone routing is undesirable in an IMS/MMD environment because it delays the signaling and media delivery and hence contributes to degradation in the quality of service. Based on the type of mobility protocol, different mitigation techniques can be devised. We proposed a few mitigation techniques for MIPv4 and provided the experimental details and analysis of the results of one mitigation technique in the experimental MMD environment. Results of this experiment show that effect of mitigation techniques appears to be more effective when the home network is far from the visited network. Carriers who plan to deploy MIPv4 in their networks may like to introduce these techniques to get the optimized operation. We plan to conduct similar experiments for MIPv6. While tombone routing mitigation helps reduce the delay caused by SIP signaling, optimization for other functional handoff components also needs to be studied for reliable operation of real time communication.

REFERENCES
[20] [http://www-x.antd.nist.gov/nistnet/](http://www-x.antd.nist.gov/nistnet/)