Journal of Information Processing Vol. 17 26–38 (Jan. 2009)

Regular Paper

Network-Controlled Route Optimization for Heterogeneous Mobile IP Networks

Shinta Sugimoto,[†1,†2] Ryoji Kato[†1] and Toshikane Oda[†1]

In this paper, we propose a new solution for route optimization that can be applied to heterogeneous mobile IP networks that support client-based and network-based IP mobility management. The proposed solution aims to satisfy important requirements such as an efficient network resources utilization, an improvement of the end-to-end quality of service, and the capability to monitor user traffic for policy control and charging. In the proposed solution, the management of the route optimization context is performed by the mobility anchor, while the IP packet processing for route optimization is performed by the nodes located around the edge of the networks. We provide a comparative evaluation of the proposed solution and existing solutions in terms of: 1) the ability to maintain route optimization when the mobile node moves across a boundary between network domains where different types of IP mobility management are employed, 2) the range of application, 3) the latency required to continue route optimization upon handover, and 4) the amount of signaling. The evaluation results show that our proposed solution best satisfies the route optimization requirements for the heterogeneous mobile IP networks that we consider in this paper.

1. Introduction

Mobile networks are evolving towards a fully converged IP network that can support various access technologies such as 3G wireless access, Long Term Evolution (LTE) and Wireless LANs [1]. The mobile networks need to be able to provide IP-based mobility management that enables a mobile node (MN) to roam across different types of access networks and change its attachment point to the IP network.

Network-based IP mobility protocols such as the one based on Proxy Mobile IPv6 (PMIPv6)[2] have gained a wide attention. These are designed to reduce signaling over the air, accomplish location privacy, and put IP mobility management totally under the control of the mobile network provider (MNP). In the Evolved Packet Core (EPC) architecture discussed in 3GPP [1], PMIPv6 has been selected as the protocol for supporting IP mobility across the different types of access networks. However, the MNs still need to be able to work with client-based mobility management such as Dual Stack Mobile IPv6 (DSMIPv6)[3] which needs to be activated when a network does not support network-based IP mobility management. For example, the 3GPP EPC is designed to support both client-based and network-based IP mobility management systems.

Because of an increasing demand for real-time IP based applications and a need for handling vast volumes of user traffic, an efficient packet routing will become more and more important. The end-to-end latency of user traffic should be minimized, for instance, to satisfy the requirements of interactive applications such as on-line gaming. In 3GPP EPC, there is a serious requirement for supporting local breakout (LBO) by which the user traffic can be routed without being transferred through a home mobile network to avoid long and redundant packet transfer routes. Hence route optimization in the IP mobility framework is becoming more important, and solutions to achieve the required route optimization need to be developed.

Mobile IPv6 (MIPv6)[4] specifies a mechanism for route optimization that enables the MN to establish a context for route optimization with a peer node with which it does not have any trust relationship a priori [5]. However, in mobile networks such as 3GPP EPC, the plain MIPv6 route optimization mechanism cannot be applied because the MN would not be aware of its topological location when attached to a PMIPv6 network. Therefore, various solutions for route optimization in PMIPv6 have been proposed. A common approach taken in some of the existing solutions [6,7] is to relocate the route optimization functionality from the MN to the Mobile Access Gateway (MAG). However, this makes it necessary to transfer the route optimization context between the MAGs, and more importantly, it appears infeasible for an MN to continue route optimization when it moves across a boundary between client mobile IP (CMIP) and proxy mobile IP (PMIP) networks.

The objective of this paper is to propose a solution for route optimization that is applicable to mobile networks where domains differ by the type of IP mobil-

†1 Nippon Ericsson K.K.
†2 Keio University

ity management being used (client-based or network-based) and in which MNs move between these different domains while route optimization is in use. We first describe the mobile network model that we consider in this paper and the requirements for route optimization while taking into account the characteristics of the mobile networks. Then we present our proposed solution for route optimization which is designed to meet these requirements. We discuss the features of the proposed solution and provide a comparative evaluation of the advantages of this solution compared to other existing solutions. Based on the analyses and the evaluation, we conclude that our proposed solution for route optimization is the most suitable for mobile networks.

2. Problems of Route Optimization in Heterogeneous Mobile IP Networks

In this section, we describe the model of mobile networks that we consider in this paper and explain its characteristics. Then we describe the requirements and issues of route optimization in these mobile networks.

2.1 Architectural Overview

Figure 1 shows a basic model of a mobile network in which the IP mobility is provided for MNs using the Mobility Management (MM) anchor. In the model, the MM anchor assigns a home address (HoA) to each MN and forwards user traffic sent to or from this home address. An MN moves around the access networks and gains the IP connectivity from the Access Router (AR). IP mobility management is performed either in a client-based or a network-based manner. In the former case, the MN registers its topological location with the mobility anchor in the same way as in the case of a Mobile IP. In the latter case, network entities such as ARs perform IP mobility management functions in such a way that the MN can behave as a normal IP host without being aware of its topological location. In our network model, a protocol such as DSMIPv6 [3] is assumed as the CMIP protocol and a protocol such as PMIPv6 [2] is assumed as the PMIP protocol. In the example shown in Fig. 1, the CMIP protocol is used on MN1 because there is no support for network-based IP mobility in the associated AR. In contrast, MN2 is attached to an AR called a Mobile Access Gateway (MAG) that provides network-based IP mobility for MNs by terminating IP tunnels and

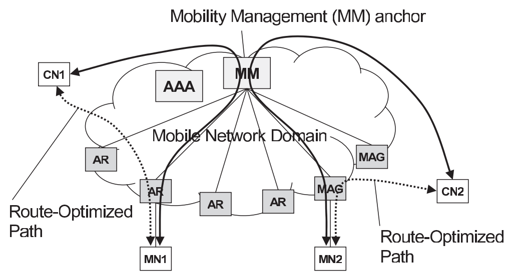

Fig. 1 An architectural overview of the basic network model.

carrying out mobility signaling. In both cases, as shown in Fig. 1, the user traffic can be route-optimized when a proper mechanism is applied.

As commonly seen in current mobile networks, we assume that the MNPs in our model supports roaming. That is, MNPs have agreements with each other and a roaming MN can be authorized to connect to a visited network. **Figure 2** shows a scenario where roaming MNs (MN1 and MN2) are each authorized to connect to visited networks. IP mobility is provided by the Global Mobility Management (GMM) anchor and the Local Mobility Management (LMM) anchor for the roaming MN. The LMM anchor serves local IP mobility for the MN and assigns a local home address (LHoA). The HoA assigned by the GMM anchor is hereafter referred to as a “global HoA” (GHoA) to differentiate it from an LHoA. The user traffic sent from or to the LHoA is forwarded by the LMM anchor, while the user traffic sent from or to the GHoA is forwarded by the GMM anchor. However, this multi-homed environment is not preferable from the viewpoint of the upper layer protocols which usually have no idea what characteristics a chosen IP address would have (global mobility or local mobility) and what routing path a chosen IP address would result in. In this respect, it is more advantageous to simply present the GHoA to the upper layer protocols to avoid confusion.

An MN may get several IP addresses assigned at different topological locations, particularly when it is roaming. Route optimization can be executed at a network entity where the user traffic sent to or from any of the MN’s IP addresses is anchored. At the point where route optimization is executed, IP packet processing such as IP encapsulation is performed to redirect IP flow. For instance, potential

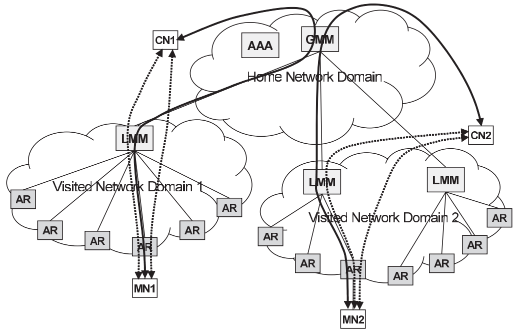

Fig. 2 An architectural overview of a roaming scenario.

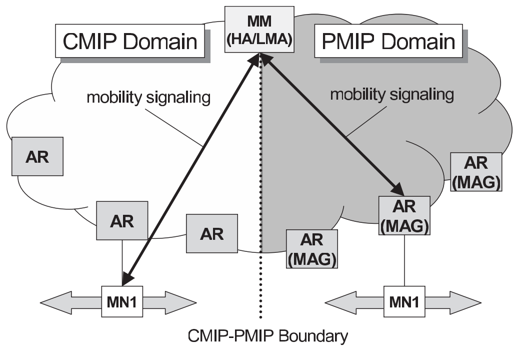

Fig. 3 Heterogeneity of IP mobility management.

points for executing route optimization for the user traffic between MN1 and CN1 in Fig. 2 are the LMM anchor and the AR in Visited Network Domain 1.

Figure 3 shows how the mobile networks can be logically divided into CMIP and PMIP domains depending on the type of IP mobility management. The GMM anchor in our model should be IP reachable from both CMIP and PMIP domains so as to provide constant global IP mobility for MNs. Hence, as shown in Fig. 3, the GMM anchor can function as both a Home Agent (HA) in the CMIP domain and a Local Mobility Anchor (LMA) in the PMIP domain at the same time. When the MN moves from the CMIP domain to the PMIP domain, the CMIP protocol needs to be inactivated. On the other hand, when the MN detects that the access network does not support PMIP, the CMIP protocol needs to be activated.

One should note that the two major characteristics of the mobile networks mentioned above, a) the scope of IP mobility management and b) the heterogeneity of IP mobility management, are orthogonal. In our model, we assume that the mobile networks bear both characteristics as seen also in the 3GPP EPC.

2.2 Requirements for Route Optimization

Regarding the route optimization solution required in mobile networks that have the above-mentioned features, we envision that the following requirements should be satisfied.

- Req-1: Route optimization should be possible no matter which type of network domain (CMIP or PMIP) the MN is attached to.
- Req-2: Route optimization should be continued even when the MN moves across the boundary between CMIP and PMIP domains. A soft state called "route optimization context" is created and maintained during the route optimization. A route optimization context comprises 1) the mapping of home address (HoA) and care-of address (CoA) of the MN, and/or 2) the mapping of HoA and CoA of the peer node, and 3) the authentication information for the correspondent binding registration. Re-establishing the route optimization context from scratch should be avoided so as to make the handover seamless.
- Req-3: The applicability of route optimization should not be limited to inside the mobile networks. It should be possible to perform route optimization with a standard MIPv6 correspondent node (CN) and/or MN.
- Req-4: MNPs should be able to have full control over route optimization. This means that MNPs should be able to activate or deactivate route optimization and to allocate the route optimization execution point freely.
- Req-5: MNPs should be able to monitor all the user traffic for the purpose of policy and charging control even if the traffic is route-optimized.
- Req-6: The credentials of the MN used to create the route optimization context should not be disclosed to any nodes which have no trust relationship with the MN.

Note that the above requirements are distinct from the ones stated in Jeong, et al.[8] because the scope of applicability is different. The assumption in our case is that the MN moves across CMIP and PMIP domains and performs route optimization with different types of peer nodes but the scope of route optimization in Jeong, et al.[8] is limited to inside a local network domain.

2.3 Issues of Route Optimization

Given the requirements mentioned above, specific issues arise when designing a route optimization solution for the mobile networks in our model.

Issue-1 A straightforward way of enabling route optimization in PMIP domain is to have the MAG perform signaling and management of route optimization contexts. Such a design leads to the requirement that a context transfer mechanism such as Context Transfer Protocol (CXTP)[9] transfers the route optimization context from the old MAG to the new MAG. However, such an approach cannot satisfy Req-2 because the MN has no information about the old MAG when it moves from the PMIP domain to the CMIP domain and thus the context transfer cannot be performed. The old MAG would not be aware of the new location of the MN, either. For the above reasons, any route optimization solutions that require any kind of context transfer between the MN and MAGs are considered infeasible to satisfy Req-2.

Issue-2 In order to satisfy Req-3, the route optimization between the concerned MN and its peer node such as a MIPv6 CN should be possible. This seems to be fulfilled by applying the correspondent binding registration mechanism defined in standard MIPv6[4] to the target mobile system. However, there is an issue that the home MNP cannot have full control of the route optimization if the MN performs the correspondent registration by itself when attached to the CMIP domain, which means that Req-4 and Req-5 are not satisfied. Therefore, a solution that can satisfy not only Req-3 but also Req-4 and Req-5 is needed.

Issue-3 Taking Req-6 into account, there is an issue about where to store an MN's credential for authenticating the correspondent Binding Update (BU) message. In PMIP domain, where mobility signaling required for MN is proxied by any network entity, the ownership of the HoA must be proven by the sender of the correspondent BU message to prevent redirection attacks[5]. Therefore, the sender of the BU message should have access to any credential (e.g., binding management key in standard MIPv6) that can be used for proving address ownership of the HoA. If the HoA is a Cryptographically Generated Address (CGA), the address ownership can be proven by the receiver of the BU message by verifying the CGA signature[10]. In this case, the private key of the MN is considered as a credential. Considering Req-6, there is a concern if any credential of the MN is stored at the MAG because there is no trust relationship between the MN and the MAG in the PMIPv6 network model. In addition, storing an MN's credential at a MAG increases the danger of the credentials being exposed to malicious third parties because MAGs are distributed in the PMIP domain.

3. Proposed Solution for Route Optimization

In this section, we propose a new solution for route optimization in mobile networks that use our model. Our proposed solution aims to meet the requirements and solve the issues concerning route optimization addressed in Section 2.

3.1 Solution Overview

The key design feature of the proposed solution is to have the mobility anchor take the primary role in managing the route optimization context. The signaling required to establish the route optimization context is performed solely by the mobility anchor on behalf of the MN, and the mobility anchor remains responsible for maintaining the contexts. The Return Routability procedure of MIPv6 is used to authenticate correspondent BU messages to keep the system interoperable with conventional MIPv6 entities (MNs and CNs). Note that route optimization can be initiated either from the MN side or the peer node side, or both sides simultaneously. The proposed mechanism can work with both client-based and network-based IP mobility protocols.

Figure 4 shows an architectural overview of a mobile network where our proposed solution is used for route optimization. As shown in the figure, two functional modules called the Primary Route Optimization Module (P-ROM) and the Secondary Route Optimization Module (S-ROM) are introduced in the proposed solution. The P-ROM performs signaling for route optimization and maintenance of the route optimization context on behalf of the MN, and it is deployed at the GMM anchor. The S-ROM performs IP packet processing for route optimization

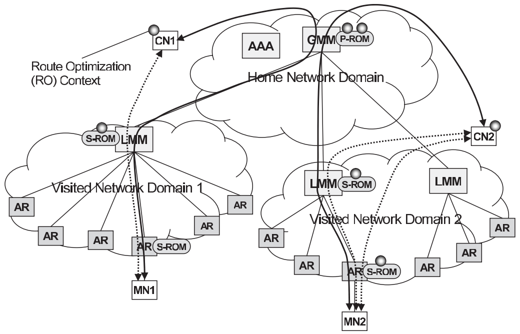

Fig. 4 An architectural overview of the proposed solution.

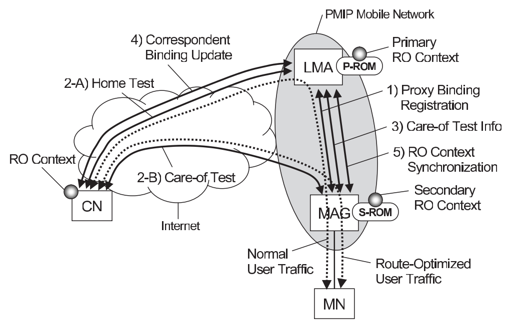

Fig. 5 Procedure for initiating route optimization.

(the procedure is hereafter called "address switching") based on the route optimization context provided by the P-ROM. The P-ROM and the S-ROM work coordinately and exchange route optimization contexts between each other. In Fig. 4, the IP flow between MN1 and CN1 is routed via the LMM. The route optimization can be executed at the LMM but not at the AR because the route optimization context is present only at the S-ROM on the LMM. In contrast, the IP flow between MN2 and CN2 can be route-optimized either at the LMM anchor or the AR since the route optimization context is present at the S-ROMs on both of the two network entities. In this way, the S-ROM is deployed at the network entity where the user traffic of the MN is anchored. Also, the distribution of S-ROMs is handled dynamically by the MNP.

Address switching at the S-ROM requires the manipulation of IP header information such as the encapsulation and the re-writing of the IP header. For instance, if the peer node is a MIPv6 Correspondent Node (CN), address switching is performed by processing IPv6 extension headers, namely the type 2 routing header and the destination option [4].

3.2 Initiating Route Optimization

Figure 5 shows how the route optimization can be initiated by the mobility anchor in the proposed solution. Note that, in this example, the assumption is that the concerned MN is visiting a PMIP domain and the peer node (CN) is a MIPv6 CN. The network hierarchy and the CMIP-PMIP boundary are omitted in this figure for simplicity. The procedure for establishing route optimization context is as follows.

(1) The proxy binding registration is performed by the MAG in order to register the MN's Proxy CoA (PCoA) with the LMA. The LMA informs the MAG that it needs to initiate route optimization.

(2) The LMA performs a home test, while the MAG performs a care-of test in response to the request from the LMA. (steps 2-A and 2-B in Fig. 5, respectively)

(3) The MAG sends the result of the care-of test to the LMA.

(4) Based on the home keygen token and the care-of keygen token, the LMA performs the correspondent binding registration with the CN to create a route optimization context. The P-ROM holds the primary route optimization context.

(5) The P-ROM transfers the route optimization context to the S-ROM which is placed at the MAG where address switching is performed. The S-ROM holds a secondary route optimization context.

(6) Any subsequent user traffic sent to or from the MN's HoA is route-optimized. That is, the user traffic is sent directly to or from the MN's PCoA.

A secondary route optimization context contains the HoA and CoA mappings for a given pair of nodes. On the other hand, a primary route optimization context contains the address mapping and the credential by which the address ownership of the HoA can be proven. That is, a secondary route optimization context is a subset of a primary route optimization context.

3.3 Responding to Route Optimization Request

Figure 6 shows how a route optimization request sent from the peer node side can be processed by the mobility anchor on behalf of the MN. Note that, in this example, the assumption is that the concerned MN (MN1) is visiting a PMIP domain and the peer node (MN2) is a MIPv6 MN. The proposed solution is deployed on the side of MN1. The intention of using this example is to demonstrate that the proposed solution can work with conventional mobile systems that use MIPv6 [4] for IP mobility management. The procedure for responding to a request to establish route optimization context is as follows.

(1) Return Routability procedure is initiated by the peer node. That is, the MN2 initiates a home test and care-of test.

(2) The LMA detects that route optimization has been initiated by the peer node of the MN and responds to the Return Routability procedure by sending separate Home Test and Care-of Test messages back to the MN2. (steps 1-A and 1-B in Fig. 6)

(3) After the Return Routability procedure is completed, the MN2 sends a correspondent BU message to MN1. Note that the BU message is sent to the HoA of MN1 and thus it is routed to LMA.

(4) The LMA intercepts the BU message and validates the message as per the MIPv6 specification [4]. If the validity of the message is confirmed, a primary route optimization context is created at the P-ROM on the LMA. (step 2 in Fig. 6)

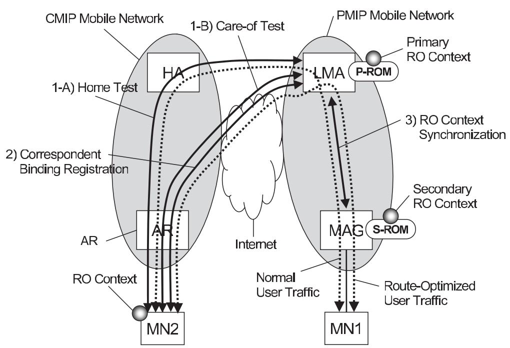

Fig. 6 Procedure of responding to route optimization request.

(5) Optionally, the P-ROM on the LMA signals to the S-ROM on the MAG to notify it to synchronize the route optimization context (step 3 in Fig. 6) so that the S-ROM can also perform the address switching.

(6) It becomes possible to route user traffic using the route-optimized path.

3.4 Route Optimization Context Synchronization

In the proposed solution, a procedure called Route Optimization Context Synchronization is defined so that the P-ROM and S-ROM can synchronize route optimization contexts. The route optimization context is initially created at the P-ROM and is then distributed to the S-ROMs in accordance with the hierarchy of IP mobility management. The context is to be configured at the S-ROM where route optimization is actually executed. Where to execute route optimization is determined by the home and/or visited MNPs based on their roaming agreement and policy settings. The data transfer for synchronization is performed either by the out-of-band signaling mechanism or by an extension to the IP mobility protocol. For instance, one can extend Mobility Header Signaling Messages [11] in order to support Route Optimization Context Synchronization for MIPv6/PMIPv6.

3.5 Summary of the Proposed Solution

In summary, the characteristics of the proposed solution fulfill all the requirements described in Section 2.2. Req-1 is satisfied because a common framework of route optimization can support route optimization for a given MN no matter which network domain (CMIP or PMIP) the concerned MN is attached to. Req-2 is satisfied thanks to the design choice whereby the proposed solution stores the route optimization context at the GMM anchor (HA/LMA) which continuously serves IP mobility management for MNs. Note that the proposed solution does not require any context transfer between MAGs and MNs and thus Issue-1 is solved. Thanks to the design choice of using the Return Routability procedure of MIPv6 for authenticating correspondent BU messages, the system is interoperable with MNs and CNs that use standard MIPv6. By allocating the S-ROM at the MN, the home MNP can continuously have full control of route optimization even when the MN is attached to the CMIP domain. Therefore, Req-3 is satisfied and Issue-2 is solved. Req-4 is satisfied thanks to the network-centric design of the proposed solution whereby the home MNP of the MN can activate or deactivate route optimization. This is possible because the proposed solution manages the primary route optimization contexts at the GMM anchor. Req-5 is satisfied thanks to the functional split between P-ROM and S-ROM which enables MNPs to freely select the best execution point for route optimization and to monitor user traffic even when it is route-optimized. Req-6 is satisfied because MN credential for proving address ownership are solely managed by the GMM anchor (the P-ROM) which has a trust relationship with the MN. Even if the system is further extended to use CGA for authenticating BU messages, the private key of the MN which is considered as a credential, can be managed securely by the GMM anchor in the proposed solution. Therefore, Issue-3 is solved.

4. Evaluation

In this section, we evaluate the proposed solution in comparison with existing solutions in terms of a set of important criteria.

4.1 Categories of Route Optimization Solutions

There are various existing solutions for route optimization. Those solutions can be categorized based on the allocation of the functional components required for route optimization. The functional components that are useful to characterize each category of solutions are: 1) route optimization signaling with a peer node and context management function, and 2) address switching function, as was described in Section 3.1. **Table 1** shows four categories of route optimization solutions, each of which takes a different approach to allocating the functional components.

The first category is called MN-based route optimization. In this category, the two functions are both performed at the MN. The basic route optimization mechanism defined in standard Mobile IPv6 [4] is an example solution in this category. The second category is called MAG-based route optimization. In this category, the two functions are both performed at MAG. Example solutions in this category are Sarikaya, et al.[6], and Jeong, et al.[7]. The third category is called LMA-based route optimization. In this category, the functions of signaling and management of the route optimization context are performed by the mobility anchor (HA in CMIP, or LMA in PMIP) while the address switching function is performed at different network entities. Our proposed solution described in Section 3 falls into this category. The fourth category is called Hybrid route optimization. In this category, the functions of signaling and management of route optimization context are performed by LMA or MAG while the address switching function is performed by MAG. Examples solutions in this category are Liebsch, et al.[12] and Dutta, et al.[13].

Table 1 Categorization of route optimization solutions.

	Signaling and Context Management	Address switching
1.MN-based [4]	MN	MN
2.MAG-based [6,7]	MAG	MAG
3.LMA-based	LMA (HA)	LMA (HA), MAG (AR)
4.Hybrid [12,13]	LMA, MAG	MAG

4.2 Evaluation Criteria

We consider the following three criteria for evaluating route optimization solutions.

CMIP-PMIP Route Optimization Continuity The ability to seamlessly continue route optimization by maintaining route optimization context when the MN moves across the CMIP-PMIP boundary is referred to as "CMIP-PMIP route optimization continuity" in this evaluation. If the route optimization context is completely lost at a handover, the procedure required to re-start or continue route optimization cannot be performed right after moving across the boundary because there is no information to identify the peer node with which the route optimization has been performed.

Applicability The ability to perform route optimization with a standard MIPv6 MN and/or CN is referred to as "applicability" in this evaluation. Having a wide range of applicability is important because the peer node of the concerned MN may not always be located inside the same mobile network.

Handover Latency The latency required to setup the route optimization context when the concerned MN performs the handover is referred to as the "handover latency" in this evaluation. In order to continue route optimization, the route optimization context needs to be properly configured at the new execution point of the concerned route optimization. Handover latency is an important criterion because it directly relates to the disruption time for user traffic. Note that until the route optimization context is set up, user traffic would be mis-routed to the old point of the execution of route optimization.

4.3 Applicability

The solution proposed by Liebsch, et al.[12] is designed to function inside a network domain operated by a single MNP. There are two operational modes defined; 1) RO Proxy mode and 2) RO Direct mode. In RO Proxy mode, the LMA performs the function of signaling on behalf of the MAG. It should be noted that, in both of the two operational modes, a pre-established security association is required between the LMAs. In RO Direct mode, a pre-established security association is also required between MAGs. The signaling messages are protected by IPsec ESP based on the pre-established security association between the network entities.

One should note that the assumption of pre-established security association between network entities in Liebsch, et al.[12] limits the range of applicability only to the inside of a single network domain. Hence the solution of Liebsch, et al.[12] cannot be applied to the communication between an MN inside a mobile network and a peer node which is located outside the mobile network. In contrast, our proposed solution is designed so that route optimization can be performed between the MN and various kinds of peer nodes including the standard MIPv6 MNs and CNs, without requiring any pre-established security association. This is possible due to the design choice to use Return Routability procedure defined in MIPv6 for authenticating correspondent BU messages. Modifications to the Return Routability procedure were carefully made in the proposed solution so as to keep the changes transparent to the peer side. Hence there are no changes required at the peer.

4.4 Handover Latency

Next, we compare the handover latencies of the solution proposed by Liebsch, et al.[12] and our proposed solution.

Figure 7 shows the handover procedure sequence in Liebsch, et al.[12]. Note that it is assumed herein that the system operates in RO Proxy mode in order to make a fair comparison of the handover latency in the respective proposals. First, route optimization is initialized by the RO-initiating LMA (LMA1 in Fig. 7) which signals the correspondent LMA (LMA2) and the RO-initiating MAG (nMAG1). Note that each RO Init message can be sent independently. Next, the RO Report and RO Setup messages are handled by the correspondent LMA (LMA2) and MAG (MAG2) respectively. After the successful creation of the route optimization state at the correspondent MAG (MAG2), the route optimization state is finally activated at the RO-initiating MAG (nMAG1).

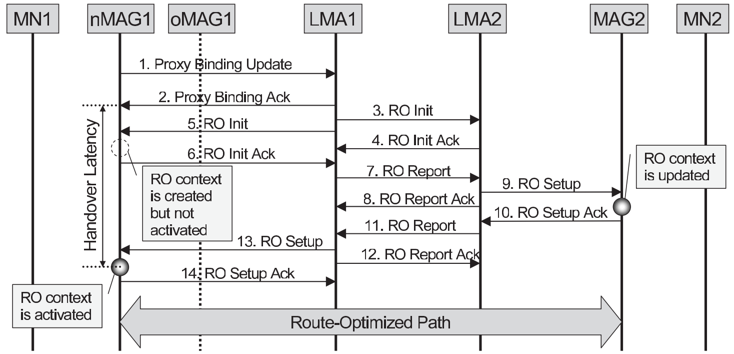

Fig. 7 Handover sequence in Liebsch, et al.[12].

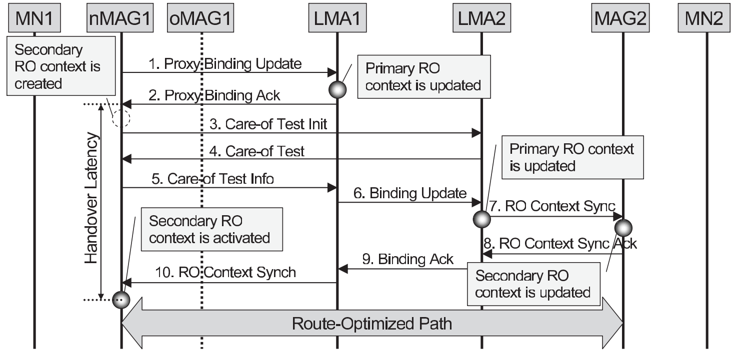

Fig. 8 Handover sequence in the proposed solution.

Figure 8 shows the handover procedure sequence in our proposed solution. Route optimization is invoked by the RO-initiating LMA (LMA1 in Fig. 8) by conveying the information about route optimization, such as the IP address of the peer node on the Proxy Binding Acknowledgment (PBA) message. After the RO-initiating MAG (nMAG1) receives the PBA message, it initiates a care-of test against the MN's peer node (MN2). Later, the result of the care-of test is transferred to the RO-initiating LMA (LMA1) by the Care-of Test Info message. Then, LMA1 for MN1 becomes ready to send a correspondent BU message to the peer node on behalf of MN1. Once the binding is created at the correspondent LMA (LMA2), the route optimization context is distributed to the edge of the networks. In this example, the route optimization context is transferred from LMA2 to MAG2 and from LMA1 to nMAG1, as described in Section 3.4. Finally, a route-optimized path is established between nMAG1 and MAG2.

With regard to the latency required for the RO-initiating MAG (nMAG1) to signal the correspondent MAG (MAG2), there is no significant difference in the two proposals. The minimum latencies required to complete the sequence of steps in the respective proposals (steps 6, 7, 9, 10, 11 and 13 in Fig. 7, and steps 5, 6, 7, 8, 9 and 10 in Fig. 8) are considered identical in terms of signaling transfer delay. However, a difference can be seen in the latency incurred by the steps before beginning the procedure mentioned above. In Liebsch, et al.[12)], nMAG1 waits until the RO Init from LMA1 is received (step 5 in Fig. 7). In our proposed solution, nMAG1 waits until the care-of test is completed (steps 3 and 4 in Fig. 8). This difference is likely to be very small and will not overwhelm the advantages of our proposed solution.

4.5 CMIP-PMIP Route Optimization Continuity

In our proposed solution, route optimization is maintained when the MN moves across a CMIP-PMIP boundary. This is because the route optimization context is stored by the GMM anchor (the P-ROM) and therefore it is transfered to the S-ROM located at the new execution point of route optimization. The transfer of context is performed by the Route Optimization Context Synchronization procedure described in Section 3.4. On the other hand, route optimization cannot be maintained by MAG-based solutions such as in Sarikaya, et al.[6)] and in Jeong, et al.[7)], as described in Section 2.3. One should note that hybrid solutions such as Liebsch, et al.[12)] and Dutta, et al.[13)] are also able to accomplish CMIP-PMIP route optimization continuity provided that there is a trust relationship between the LMAs.

The characteristics of the solutions with respect to route optimization continuity vary. Examples of this include: 1) latency required to transfer or re-establish the route optimization context for an MN that moves across a CMIP-PMIP boundary and 2) signaling overhead, as represented by the number of signaling messages and signaling volume. Regarding the latency, the analysis shown in Section 4.4 is applicable because the handover procedure for an MN moving across a CMIP-PMIP boundary is identical to that performed in the PMIP domain. Therefore, the same conclusion can be drawn. That is, there is no significant difference between Liebsch, et al.[12)] and the proposed solution in terms of latency. On the other hand, a comparative evaluation in terms of the signaling overhead is left for further study. This will be possible when sufficient details of the signaling messages required for Liebsch, et al.[12)] and Dutta, et al.[13)] become available.

4.6 Analysis of Protocol Overhead

In this section, we present a quantitative analysis of the protocol overheads of the proposed solution by investigating: 1) the additional memory required, and 2) the additional volume of signaling messages.

Table 2 summarizes details of the route optimization context that needs to be stored by the GMM anchor (HA/LMA) in the proposed solution. As shown, the GMM anchor requires at least 123 bytes of memory for a route optimization context to establish and maintain a bi-directional route-optimized session. For example, a mobility anchor deployed in the current cellular network is equipped with several gigabytes of main memory to serve for a large number of MNs. Therefore, it is judged that the overhead of our proposed solution in terms of memory usage is sufficiently small in practice.

To evaluate the signaling overhead, we analyze the signaling volume for standard MIPv6[4)], Jeong, et al.[7)] and our proposed solution to compare the different solutions and identify the differences. In this evaluation, we consider connectivity and handover scenarios where the MN communicates with its peer node which is either an MN or a stationary node. It is assumed that the route optimization has been established either from the MN side or from the peer side, and the concerned MN moves around CMIP or PMIP domains. Four scenarios of MN movements are considered to cover both intra-domain handovers (CMIP-to-CMIP or PMIP-to-PMIP) and inter-domain handovers (CMIP-to-PMIP or

Table 2 Additional memory required in GMM anchor when using the proposed solution.

Data Item	BU sender	BU receiver	Size (bytes)
MN's home address	required	required	16
Peer node's address*	required	required	16
RO CoA of the MN	required	-	16
CoA of the peer MN	-	required	16
Seqno of sending BU	required	-	2
Seqno of receiving BU	-	required	2
Flag of sending BU	required	-	1
Flag of receiving BU	-	required	1
RO policy flag	required	required	1
Home init cookie	required	-	8
Home keygen token	required	-	8
Care-of init cookie	required	-	8
Care-of keygen token	required	-	8
Kbm	required	-	20
Total (bytes)	104	52	123

Note:
- The peer node can be either an MN or stationary node.
- "BU sender" means that the correspondent BU message is sent from the MN side.
- "BU receiver" means that the correspondent BU message is sent from the peer node side.

PMIP-to-CMIP). The comparison of intra-domain handover scenarios within a PMIP domain and within a CMIP domain use the Jeong, et al.[7] and standard MIPv6[4] solutions, respectively. The reason for selecting Jeong, et al.[7] as an existing solution in our comparison of PMIP inter-domain handover scenario is that it is designed based on the existing standard IP mobility management protocols (i.e., MIPv6 and PMIPv6) and is considered as a basic approach to enabling route optimization in PMIPv6 networks. There is no existing solution which is comparable with our proposed solution in intra-domain handover scenarios because none of the existing solutions have intra-domain handover scenarios inside their scope. **Table 3** and **Table 4** show comparisons of the volume of signaling messages for the above handover scenarios.

Table 3 summarizes types and lengths of the messages exchanged between the MN side and the peer node side for re-starting or continuing route optimization when the direction is from the MN side (that is, when the correspondent BU message is sent from the MN side). As shown in the table, our proposed solution consumes a constant volume (548 bytes) of signaling messages in all four handover scenarios. It can be seen from the table that, in intra-domain handover scenarios, the Jeong, et al.[7] and standard MIPv6[4] solutions incur smaller signaling volumes than our proposed solution. However, with regard to the amount of signaling over the air, there is no significant difference between the different solutions in the respective handover scenarios. Table 4, on the other hand, summarizes the types and lengths of the messages exchanged when re-starting or continuing route optimization when the direction is from the peer node (that is, the peer node is also an MN and the correspondent BU message is sent from the peer node). Table 4 clearly shows that our proposed solution requires significantly smaller amount of signaling compared to the Jeong, et al.[7] solution in the PMIP intra-domain handover scenario. The tables show that the amounts of signaling overhead incurred by the respective solutions vary and that overall signaling overhead depends on mobility and route optimization scenarios inside the mobile networks.

4.7 Summary of Evaluation

Table 5 summarizes the qualitative evaluation of the solutions. As shown in the table, the Dutta, et al.[13] and Liebsch, et al.[12] solutions are equivalent to our proposed solution in terms of providing continuity of route optimization. However, our proposed solution is the only one that provides a wide range of applicability, in that it can operate with standard MIPv6 MNs and CNs.

In our proposed solution, the GMM anchor consumes 123 bytes of memory to store the route optimization context of a given route-optimized session. However, this overhead is considered practically small from the viewpoint of the capacity of mobility anchors in current cellular networks. Regarding the handover latency, the evaluation result shows that there is no significant difference between the existing solution Liebsch, et al.[12] and our proposed solution. Regarding the signaling overhead, it is shown that our proposed solution has both advantages and disadvantages compared to existing solutions depending on mobility and route optimization scenarios. In the case where route optimization is requested from the MN side (the correspondent BU message is sent from the MN side), our proposed solution consumes a larger volume of signaling messages than existing solutions[4),7] under intra-domain handover scenarios. On the other hand, in the case where route optimization is requested from the peer node side (the

Table 3 Comparison of signaling overhead at handover (the correspondent BU message is sent from the MN side).

	Inter-domain Handover		Intra-domain Handover			
	CMIP to PMIP	PMIP to CMIP	PMIP to PMIP		CMIP to CMIP	
Message type	Proposed solution (bytes)	Proposed solution (bytes)	Jeong, et al. (bytes)	Proposed solution (bytes)	Standard MIPv6 (bytes)	Proposed solution (bytes)
Care-of Test Init	56	56*	56	56	56*	56*
Care-of Test	64	64*	64	64	64*	64*
Home Test Init	-	-	56	-	-	-
Home Test	-	-	64	-	-	-
Binding Update	116	116	96	116	96*	116
Binding Ack	96	96	96	96	96*	96
Care-of Test Info	120	120*	-	120	-	120*
RO Context Synch	96	96*	-	96	-	96*
Signaling over the air	0	336	0	0	312	336
Total	548	548	432	548	312	548

Note: An asterisk indicates that the signaling message is sent over the air.

Table 4 Comparison of signaling overhead at handover (the correspondent BU message is sent from the peer node side).

	Inter-domain Handover		Intra-domain Handover			
	CMIP to PMIP	PMIP to CMIP	PMIP to PMIP		CMIP to CMIP	
Message type	Proposed solution (bytes)	Proposed solution (bytes)	Jeong, et al. (bytes)	Proposed solution (bytes)	Standard MIPv6 (bytes)	Proposed solution (bytes)
Care-of Test Init	-	-	56	-	-	-
Care-of Test	-	-	64	-	-	-
Home Test Init	-	-	56	-	-	-
Home Test	-	-	64	-	-	-
Binding Update	-	-	96	-	-	-
Binding Ack	-	-	96	-	-	-
Care-of Test Info	-	-	-	-	-	-
RO Context Synch	96	96*	-	96	-	-
Signaling over the air	0	96	0	0	0	0
Total	96	96	432	96	0	0

Note: An asterisk indicates that the signaling message is sent over the air.

correspondent BU message is sent from the peer node side), our proposed solution requires a smaller volume of signaling messages than the existing solution [7]. In all, it can be concluded that our proposal gives a novel solution that overcomes all the formulated problems, yet its protocol overhead is reasonably small in practice.

5. Conclusions

In this paper, we proposed a novel solution for route optimization for future mobile networks where IP mobility is managed either by a client-based or a network-based IP mobility protocol. Our proposal was intended to provide a new solution that satisfies the important requirements for the future mobile networks and solve the specific issues relevant to the requirements. The key requirements are the sup-

Table 5 Qualitative evaluation of route optimization solutions.

RO Solution	Applicability	CMIP-PMIP RO Continuity
RFC 3775[4)]	N/A	N/A
Sarikaya, et al.[6)]	X	X
Jeong, et al.[7)]	X	X
Dutta, et al.[13)]	X	O
Liebsch, et al.[12)]	X	O
Proposed solution	O	O

port of backward-compatible route optimization with legacy MNs and CNs that use standard MIPv6, and the seamless continuation of route optimization for the MNs that move around CMIP and PMIP domains. The first key requirement is satisfied by a key design feature of applying a new authentication mechanism for correspondent binding registration based on the Return Routability procedure defined in MIPv6. Another key design feature is the functional split between the management of the route optimization context and the execution of address switching. The former is performed solely by the GMM anchor and the latter is performed by other network entities located around the edge of networks. This key design feature yields important characteristics of route optimization that are essential to overcome the unsolved problems in an MNP-friendly manner. Firstly, continuation of route optimization is achieved in a more secure and realistic way compared to existing solutions. Secondly, the proposed solution enables MNPs to have full control of route optimization. Thirdly, it also enables the MNPs to freely allocate points of executing route optimization inside the mobile network. The paper presented a comparative evaluation of the proposed solution and existing solutions. It was shown from the evaluation results that the proposed solution is the only one that meets the formulated requirements and solves the issues while its protocol overhead is reasonably small in practice. Therefore, it is concluded that our proposed solution has significant advantages over the existing solutions and can best fulfill the route optimization requirements for the heterogeneous mobile IP networks.

Acknowledgments We express our thanks to Dr. Christian Vogt from Ericsson Research, Dr. Ryuji Wakikawa, Professor Rodney Van Meter and Professor Fumio Teraoka from Keio University for their valuable technical comments. We sincerely thank Jan Höller and Yuri Ismailov from Ericsson Research, Professor Jun Murai, Professor Keisuke Uehara and Professor Osamu Nakamura from Keio University for their continuous support and encouragement to our research activity.

References

1) 3rd Generation Partnership Project: Architecture Enhancements for non-3GPP Accesses, TS 23.402 Release 8 (2007).
2) Gundavelli, S., Leung, K., Devarapalli, V., Chowdhury, K. and Patil, B.: Proxy Mobile IPv6, draft-ietf-netlmm-proxymip6-18.txt (2008).
3) Soliman, H.: Mobile IPv6 support for dual stack Hosts and Routers (DSMIPv6), draft-ietf-mext-nemo-v4traversal-04.txt (2008).
4) Johnson, D., Perkins, C. and Arkko, J.: Mobility Support in IPv6, RFC 3775 (2004).
5) Nikander, P., Arkko, J., Aura, T., Montenegro, G. and Nordmark, E.: Mobile IP Version 6 Route Optimization Security Design Background, RFC 4225 (2005).
6) Sarikaya, B., Qin, A., Huang, A. and Wu, W.: PMIPv6 Route Optimization Protocol, draft-qin-netlmm-pmipro-00.txt (2008).
7) Jeong, S. and Wakikawa, R.: Route Optimization Support for Proxy Mobile IPv6 (PMIPv6), draft-jeong-netlmm-ro-support-for-pmip-00.txt (2007).
8) Jeong, S., Vogt, C., Wakikawa, R., Liebsch, M., Sugimoto, S. and Sarikaya, B.: Problem Statement and Requirements for Route Optimization in PMIPv6, draft-jeong-netlmm-pmipv6-roreq-01.txt (2007).
9) Loughney, J., Nakhjiri, M., Perkins, C. and Koodli, R.: Context Transfer Protocol (CXTP), RFC 4067 (2005).
10) Arkko, J., Vogt, C. and Haddad, W.: Enhanced Route Optimization for Mobile IPv6, RFC 4866 (2007).
11) Haley, B. and Gundavelli, S.: Mobility Header Signaling Message, draft-haley-mip6-mh-signaling-02.txt (2007).
12) Liebsch, M., Le, L. and Abeille, J.: Route Optimization for Proxy Mobile IPv6, draft-abeille-netlmm-proxymip6ro-01.txt (2007).
13) Dutta, A., Das, S., Yokota, H., Chiba, T. and Schulzrinne, H.: ProxyMIP Extension for Inter-MAG Route Optimization, draft-dutta-netlmm-pmipro-00 (2008).

(Received March 31, 2008)
(Accepted October 07, 2008)
(Released January 7, 2009)

Shinta Sugimoto received a master degree in Media and Governance from Keio University in 2001 and joined Nippon Ericsson K.K. He has been engaged in IP Mobility research and made contributions to the IETF, especially in the mobility-related working groups. He is also a Ph.D. student at Keio University.

Ryoji Kato earned a Master's Degree in Information Processing Technology at Osaka University in 1991. He joined Nippon Ericsson K.K. in 1998. He has worked in the area of IP mobility and security system.

Toshikane Oda has been with Nippon Ericsson K.K. in Tokyo since 2000, where he was the Manager of Market Support and New Technology Department, and is currently the Manager of Ericsson Research Japan and engaged in IP network and service layer technology research. He had worked in R&D on network design and services in KDD, and also involved in the standardizations especially in ITU-T SG2. He received the Ph.D. in Communication Engineering from Osaka University.