

Efficient handover with optimized localized routing for Proxy Mobile IPv6

Ahmad Rasem¹ · Marc St-Hilaire¹ · Christian Makaya²

Published online: 28 October 2015
© Springer Science+Business Media New York 2015

Abstract Proxy Mobile IPv6 (PMIPv6) has been designed by the IETF as a network-based mobility management protocol to support the localized mobility of IP devices. Although several proposals have been made for localized routing optimization, they don't take into account handover management and localized routing simultaneously. In fact, the localized routing state is restored after the handover management procedure, leading to packet loss and signaling overhead. On the other hand, Fast Handovers for PMIPv6 (F-PMIPv6) protocol has been designed to mainly solve the issues of the high handover delay and packet loss during handover that occur with PMIPv6. As a result, this paper looks at enhancing F-PMIPv6 by controlling the handover management with optimized localized routing and proposes an extension of PMIPv6 called optimized PMIPv6. The proposed protocol enhances the performance of PMIPv6 and F-PMIPv6 in terms of route optimization, handover delay, signaling cost and network utilization.

Keywords Route optimization · Localized routing · Handover · Service continuity · PMIPv6 · F-PMIPv6

✉ Marc St-Hilaire
marc_st_hilaire@carleton.ca

Ahmad Rasem
arasem@connect.carleton.ca

Christian Makaya
chrismak@ieee.org

¹ Department of Systems and Computer Engineering, Carleton University, Ottawa, ON, Canada

² IBM Research, Yorktown Heights, NY, USA

1 Introduction

Wireless and cellular networks are evolving towards an all-IP infrastructure such as 4G LTE (Long Term Evolution). As a result, there is a need to manage the mobility of IP-based devices across different networks (independently of the access technology) while maintaining mobile connectivity [9, 15]. The first step towards mobility management was the introduction of Mobile IP (MIP) which can operate in IPv4 and IPv6 networks [22, 23]. In MIP, a Mobile Node (MN) that obtains its IP Home Address (HoA) from its home agent (HA) in its home network, needs to roam with this address within other networks. When this MN moves to another network, referred to as foreign network, it obtains a new temporary IP care-of address (CoA). Then, the mobile node sends a binding update message to its HA. If the route optimization functionality is enabled, the binding update may be sent to all the correspondent nodes (CNs) that are communicating with the MN. The binding update contains the newly acquired care-of address and the home address of the MN. This process is called registration. When the home agent receives the binding update message, it updates its cache table with the combination of care-of address and home address. Therefore, the CNs can either send the data directly to the mobile CoA or to the HoA after which the HA forwards it to the mobile CoA. Due to this binding/registration, the information can continue to be exchanged between the mobile node and its correspondent nodes even after the mobile node hands over to a new access network.

In the context of IPv6, Mobile IPv6 (MIPv6) [23] has been standardized by the IETF. However, MIPv6 has a lot of performance drawbacks such as high handover delay and signaling overhead [18]. As a result, several protocols have been proposed to solve some of the MIPv6 limitations [3, 27]. For example, Proxy Mobile IPv6 (PMIPv6) [3] has been

introduced by the IETF to initially avoid the high amount of signaling and the software modification required on the MN side. Furthermore, scaling benefits were also expected as the signaling work is being done on the network side giving operators more control and flexibility [14]. In fact, the main difference between PMIPv6 and MIPv6 is that MIPv6 is a host-based approach while PMIPv6 is a network-based approach.

Although PMIPv6 exhibits a lot of benefits over MIPv6, it still possesses high handover delay and packet loss. Therefore, Fast Handovers for PMIPv6 (F-PMIPv6) [29] has been introduced. This protocol allows for the network to prepare the mobile node (from the network side) for the handover prior to the actual handover. Due to the nature of PMIPv6 and F-PMIPv6, MN and CN communicate with each other through non-optimal paths [16]. Therefore, a lot of research has been done to establish route optimization (RO)¹ between MN and CN and reduce handover delay [8, 11, 29]. These studies have been done separately and to the best knowledge of the authors, there is no previous work combining localized routing and handover management simultaneously, without reducing the performance of the network and degrading end-users perceived quality of service. It is important to note that depending on the use case, keeping handover (HO) and route optimization separate might be beneficial. For example, one may want to optimize only data routing path regardless of the side effect of HO. However, in our solution, we are addressing the scenario where HO leads to the need of route optimization. Rather than doing these two procedures separately, as it has been demonstrated in the literature to lead to performance and QoS degradation, our solution takes advantage of both mechanisms and provide an optimized procedure for HO and localized routing. More precisely, this paper looks at combining F-PMIPv6 with localized routing for PMIPv6 [8] into one solution called Optimized PMIPv6 (O-PMIPv6). The proposed solution minimizes the signaling messages needed to perform both operations separately. Hence, we can reduce the signaling overhead and the latency of HO and route optimization procedure. The choice of F-PMIPv6 is motivated by the fact that F-PMIPv6 is one of the major enhancements of PMIPv6, that is proved to reduce service disruption during handover through analysis and implementation.

The remainder of this paper is organized as follows. In Sect. 2, an overview of PMIPv6, F-PMIPv6, and localized routing is presented. Section 3 describes the problem statement and the proposed solution denoted O-PMIPv6. Then, Sect. 4 describes the analytical model followed by simulation results in Sect. 5. Finally, Sect. 6 concludes the paper.

¹ Route optimization and localized routing have the same meaning in this paper.

2 Background and related work

There is an interesting paper which is analyzing and comparing existing IPv6 mobility management protocols including host-based (MIPv6, FMIPv6, HMIPv6) and network-based (PMIPv6, FMIPv6) protocols [12]. However, since our solution is combining PMIPv6 and FPMIPv6, we will focus on these two protocols next.

2.1 Proxy Mobile IPv6

In PMIPv6 [3], there are two main entities: local mobility anchor (LMA) and mobile access gateway (MAG). The LMA is the topological anchor point for the MN prefix assignments. It is responsible for maintaining the state of the MN. Usually, it sits on the device that a HA would typically sit on. On the other hand, the MAG sits on the network access and is typically the connecting point between the MN and the network. It is responsible for detecting MN movements and change of attachment and subsequently registering the MN with the network (i.e., LMA).

When MN enters the PMIPv6 domain, it has the option to send a router solicitation (RS) message. The MAG detects its attachment and attempts to contact the LMA by sending a proxy binding update (PBU) message containing relevant information about MN attachment such as MN identification (MN-ID). The LMA processes the PBU message and assigns the MN with home network prefix(es). It sets up a binding cache entry (BCE) in an internal cache table with this information and sends a proxy binding acknowledgement (PBA) to the MAG in which it includes the assigned home prefix(es). In addition, the LMA sets-up a bidirectional tunnel between its LMA address (LMAA) and the MAG proxy care-of address (PCoA) that can be used for forwarding traffic. As soon as the MAG receives the PBA, it updates its binding update list (BUL) entries with the PBA information and sends a router advertisement (RA) message to the MN with the available home network prefix(es) (HNPs) for which MN uses to configure its IP address. The above procedure is depicted in Fig. 1a for clarification.

Once the address is configured on the MN, it becomes ready to send and receive packets. Any packets destined to the MN are received by LMA. The LMA encapsulates the packets with a new header containing its LMAA as a source and the destination PCoA, and forwards them using the bidirectional tunnel to the MAG that the MN is connected to. The MAG removes the outer header of the packets and forwards them to the MN that is attached to it. For packets that are sent from MN, the MAG receives the packets and encapsulates them with a new header containing its PCoA as a source and the destination LMAA. Then the packets are sent through the same tunnel to the LMA. The LMA in turn removes the outer header, encapsulates it again as before and forwards

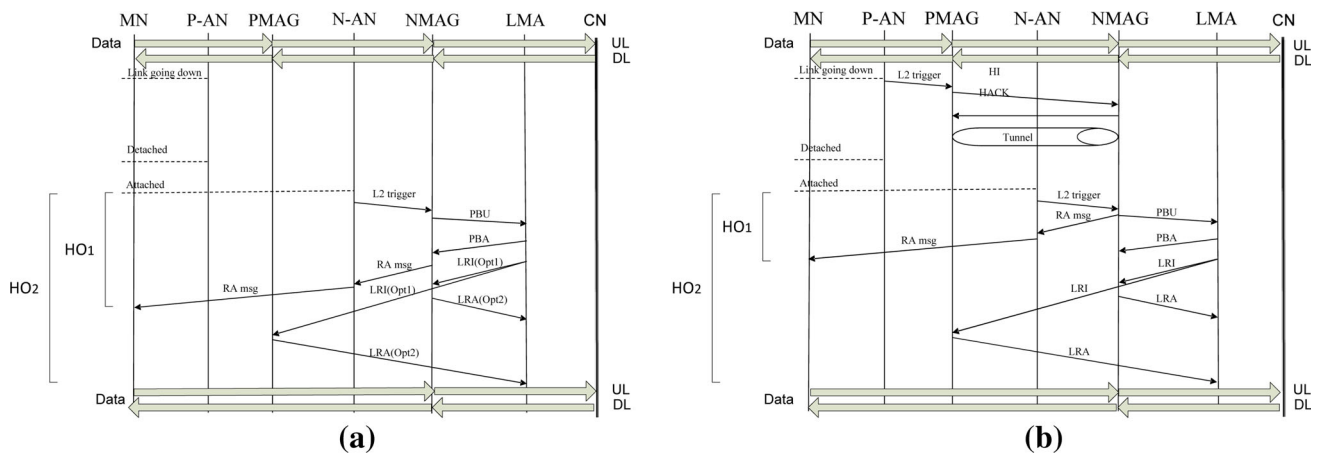


Fig. 1 Localized routing for PMIPv6. The *thick arrows* show the data path of ongoing session between the MN and CN

it to destination node such as CN. When the MN leaves the current access network for the purpose of disconnecting or performing a handover, the MAG will detect its disconnection and signals the LMA of the disconnection by sending a de-registration message. The LMA starts a timer for the entry corresponding to that MN in the binding cache table. If a PBU message for that MN is received before the timer times out, a result of a new attachment during a handover, then the LMA updates the BCE for that MN and the timer is canceled. However, if the time out happens, then the LMA removes the BCE corresponding to that entry from the internal table. When MN changes its point of attachment by performing a handover to another MAG within the same PMIPv6 domain, then the LMA and the new MAG update their internal table and assign the same prefix to MN so that the whole PMIPv6 network appears as a single link.

Unfortunately, PMIPv6 suffers from high handover delay and packet loss during the handover process. The authors in [19] provide an extensive survey on the various approaches to perform the handover process.

2.2 Fast handovers for Proxy Mobile IPv6

F-PMIPv6 [29] was introduced to enhance the performance of PMIPv6 when it comes to handover delay and packet loss during handover. The idea behind F-PMIPv6 is to anticipate the movement of MN and act on it to minimize its downtime during handover. This is done by establishing a bidirectional tunnel between the Previous MAG (PMAG), that the MN is moving from, and the New MAG (NMAG), that the MN is moving to. This tunnel is used to exchange session information between both MAGs. It should be noted that each MAG acts as an access router for MN. If the tunnel is established prior to MN attachment to the NMAG then it is called predictive mode, otherwise it is called reactive mode. In a severe case when the MN is detached from both the old link and the new link, the MAGs have to have the capability

of buffering the packets for future forwarding to reduce the packet loss during handover. In the predictive mode, the MN detects that it is about to perform a handover. Therefore, it reports its MN-ID and the new Access Point Identifier (AP-ID) [7] to which the MN will move to. The PMAG derives the NMAG information from the AP-ID and sends a Handover Initiate (HI) message to the NMAG with the Proxy Flag (*P*) set which includes the MN-ID, its HNP and the LMAA used. The NMAG sends a Handover Initiate Acknowledgment (HACK) to the PMAG with the *P* flag set. As a result, a bidirectional tunnel is established between the PMAG and the NMAG. Any downlink packet that is destined to the MN and received by the PMAG can be forwarded over the tunnel to the NMAG and buffered till the MN is fully attached and handover (HO) is completed. Any packets that are sent from the MN to the NMAG are forwarded to the PMAG through the established tunnel and then sent from there to the MAG-LMA tunnel. Once the MN is attached to NMAG, the NMAG exchanges PBU and PBA with LMA. Finally, all data packets will only go through NMAG and the tunnel between PMAG and NMAG is no longer needed and the resource can be released. The above procedure is depicted in Fig. 1b.

In the case of the reactive mode, the tunnel establishment has to come from the NMAG as the AP-ID is acquired only when the MN moves to the new link. The information can be provided to the NMAG either from the MN on the old link or by means of communication between the serving and target access networks. Once this information is acquired, a similar procedure to the predictive mode is followed. For simplicity and the purpose of this paper, we will be focusing only on the predictive mode.

2.3 Localized routing

Route optimization, sometimes referred to as localized routing (LR), is an important enhancement to PMIPv6. It is used to allow data traffic to take a shorter path when the source and

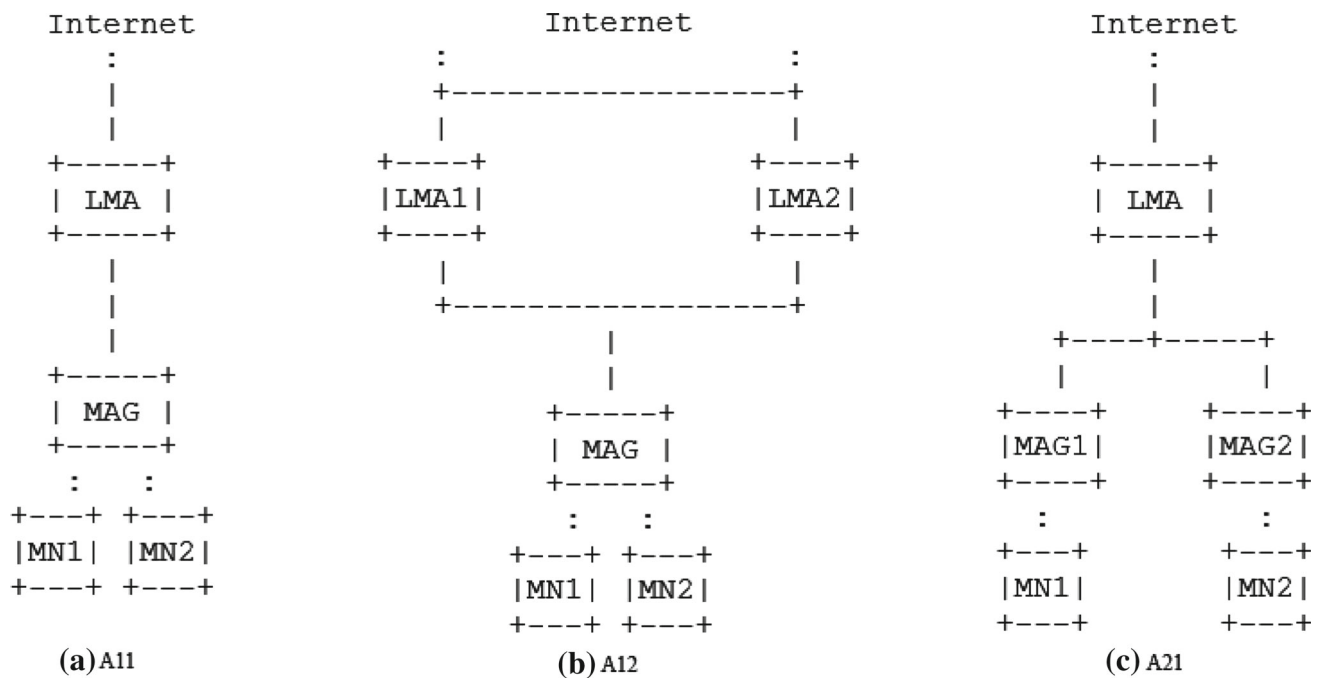


Fig. 2 Graphical representation of the different localized routing scenarios [8]

the destination are located in the same PMIPv6 domain. This results in performance improvement such as minimal packet delivery delay and optimal resources usage. In PMIPv6, data packets always have to go through the LMA even if the LMA sits on a non-optimal path. With LR, this inefficiency can be lessened. According to the PMIPv6 localized routing problem statement [16], there are multiple cases for localized routing and are referred by the following notation: $A[\textit{number of MAGs}][\textit{number of LMAs}]$ as described below and shown in Fig. 2 [8].

- **A11** In this case, there is one MAG and one LMA. The MN and the CN (referred to as MN2 in Fig. 2) are both connected to the same MAG and the latter forwards all packets between MN and CN directly without going through the LMA.
- **A12** In this scenario, there is one MAG and two LMAs that have MN and CN registered with. The common MAG has to forward packets directly between CN and MN and both LMAs have to accommodate this localized routing by considering their policy.
- **A21** This is a very common case where there are two MAGs that are registered with the same LMA. In this case, the MAGs forward packets to each other and respectively forward the packets to their destination without having to go through the common LMA.
- **A22** This is the most complicated case in which there are two MAGs and two LMAs. The MN and the CN are registered with a different MAG and a different LMA. The MN and CN have their data delivered using MAGs

only without involving the LMA in forwarding the data packets. However, the LMA will need to be involved in the setup of the localized routing path. Maintenance of the localized routing states and avoiding race conditions is one of the issues facing such feature.

Multiple solutions have been proposed in an attempt to solve the LR problem. Authors in [2] introduce the idea of route optimization controller which sits on all LMAs. The LMA is the device that initiates the RO process by sending an Init message to one of the MAGs involved in the RO and counts on MAGs communication with each others. This MAG sets up the RO by messaging the other MAG involved. The second MAG acknowledges this to the sender MAG which in turns confirms the establishment of the RO path by informing the LMA of its success. In [10], authors analyze (quantitatively) different routing optimization strategies. Then, the two most promising ones are further investigated in terms of signaling cost, packet delivery cost, total cost and blocking probability. In [17], localized routing optimization schemes are proposed for scenarios where mobile nodes have access to multiple access networks. Even with all these protocols, the IETF decided to go on its own way by using a protocol called Localized Routing for Proxy Mobile IPv6 [8] which has been used in our research. Below, we provide the description of the IETF protocol.

In a single LMA domain, when LMA detects that there is traffic between MN and CN, LMA initiates LR by sending a localized routing initiation (LRI) message to the MAG(s) that the MN and CN are connected to. Option 1 in the LRI

message contains information about the MN and CN that are connected to the same MAG, or the MN and the remote MAG PCoA (primary care-of address) in case of two MAGs. If the MAG(s) is capable of performing route optimization, the MAG(s) sends a localized routing acknowledgement (LRA) message to the LMA. Option 2 in the LRA message contains the same information that was sent in LRI in addition to a status field to indicate the status of LR establishment. As a result, a MAG starts forwarding packets directly to the destination node connected to it in case of single MAG, or to the correspondent node's MAG (CMAG) through tunnel if the CN is connected to the remote MAG. This means that the traffic does not have to go through the LMA when being exchanged between the MN and CN participating in the LR. When MN performs a handover, the LR is turned down and the LR establishment is restarted from the beginning as described above. Figure 1 shows the localized routing procedure for both PMIPv6 and F-PMIPv6.

In [5], authors propose a route optimization mechanism for PMIPv6 by using the routing table at MAG and a policy database in order to find the optimal path. However, the proposed scheme requires routing table to be propagated among MAGs in PMIPv6-domain, otherwise the solution falls back to the legacy PMIPv6 specification. Furthermore, the signaling cost of the mechanism proposed in [5] is higher than in PMIPv6. In [30], fast handover concept and predictive re-establishment of localized routing is proposed. This paper is based on F-PMIPv6; however, it requires duplicate LRI/LRA messages exchange. In fact, the previous MAG notifies the new MAG about existing LR state, but the LR should not be established with the CN's MAG. The localized routing will be established only upon attachment of MN to the new MAG. This proposal induces resource wastage (e.g., bandwidth consumption) and signaling overhead due to the processing of duplicate LRI/LRA messages. Furthermore, uplink packets are delivered through the previous MAG, which increase the delivery delay. Moreover, packets might be delivered through the non-optimized path, i.e., through the LMA.

Fast handover based solutions, such as FPMIPv6, have the potential issue of out-of-order packets and synchronization problem due to the existence of two packets delivery paths. This issue is addressed in a testbed environment in [24], by using the last packet marker to notify the end of data delivery in the old path and the buffering for holding the data delivered in the new path. In [30], the authors proposed a mechanism which combines the fast handover procedure with the re-establishment procedure of localized routing. During the fast handover procedure, the previous MAG will report existing localized routing information to the new MAG so as to quicken the re-establishment procedure of localized routing.

A state-aware pointer forwarding scheme with fast handover support for PMIPv6, called FC-PMIPv6, has been proposed in [28] to enhance performance of PMIPv6. In

FC-PMIPv6, a pointer forwarding chain between MAGs is established to reduce location update traffic to an LMA during handover and guarantees the efficiency of packet transmission. One of the issue with pointer forwarding approach is the length of the forwarding chain. In fact, a longer forwarding chain will degrade significantly the performance of the scheme. Furthermore, management of the forwarding chain requires additional processing cost, thus induces overhead.

3 Problem statement and proposed solution

Let us assume that a MN already has a LR established with the correspond node (i.e., CN) and wishes to perform a handover to another MAG. According to PMIPv6 or F-PMIPv6, during handover, the established LR status is lost. Once the MN completes its handover, and some data packets go through the non-optimal path, the LMA initiates the LR procedure in order to restore the previous LR state. When the LR establishment is done, then the MN and the CN can resume sending packets through the optimal path.

As can be seen, there is big redundancy of re-establishing the LR session which reduces the overall performance of PMIPv6-based handover management and localized routing approaches. Major performance factors are as follows: increased LMA utilization in large scale networks, higher signaling rate, higher packet loss, and higher packet delivery delay until packets take the optimal path.

In our initial work, we looked at combining localized routing and fast handover features in one solution. In [25], the predictive mode was considered. However, when the mobile node moves very fast, the attachment with the target network may happen before the completion of the tunnel setup procedure between the previous and the new access network while using predictive mode. Moreover, the prediction of the attachment could fail for several reasons such as the lost of the detachment notification messages. All of these scenarios require support of reactive mode. In [26], we addressed the reactive mode.

This paper proposes a solution to overcome the above issues by efficiently minimizing handover delay and guaranteeing localized routing state. The proposed solution integrates simultaneously handover management and localized routing concepts. The benefit of such integration would be the ability of sending all the data packets over the optimal path even when the handover is performed, reduced signaling and packet loss, minimized packet delivery delay, and finally reduced utilization of core network elements that are not on the optimal path such as LMA. It should be noted that this performance evaluation with the new solution proposed is applicable independently of the access network technology similarly to IP-based mobility management protocols.

3.1 Proposed solution: O-PMIPv6

To solve the above limitations of PMIPv6 and F-PMIPv6, this paper introduces the Optimized PMIPv6 (O-PMIPv6) protocol. The idea behind this solution is to make the F-PMIPv6 signaling participates in the LR establishment at the moment that the MN handover is triggered by the network. Using this solution, the MN will be able to handover while maintaining its LR state. In other words, there is no need to establish (or restore) a new LR session after the handover. Below we will discuss the operational procedure of O-PMIPv6 in different topologies and introduce the message format.

3.1.1 Protocol operation

The basic idea of O-PMIPv6 is to transmit the LRI/LRA information using the HI/HACK messages introduced in F-PMIPv6 to anticipate the handover. Hence, O-PMIPv6 improves the performance of F-PMIPv6 and PMIPv6 as well as allowing efficient restoration of LR state during handover. O-PMIPv6 operates successfully in any single-LMA PMIPv6 domain no matter how the topology looks like. There are three possible network topologies in which MN and CN can coexist as shown in Fig. 3 and described below.

- *Topology 1* MN and CN1 are connected to MAG1 but MN is moving to MAG2.
- *Topology 2* MN is connected to MAG1 and moves to MAG2 where CN2 is connected to.
- *Topology 3* MN is connected to MAG1 and CN3 is connected to MAG3 but MN performs an handover to MAG2.

Figure 4 shows the message flow for O-PMIPv6 in the three different topologies. When PMAG detects that MN is about

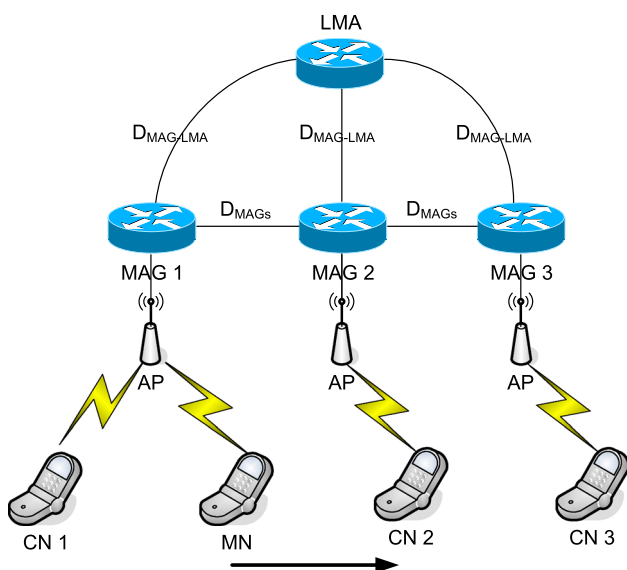


Fig. 3 PMIPv6 possible topologies

to perform a handover (due to a degraded signal strength for example or other metric provisioned in MN’s profile), it sends a HI message to the NMAG. The HI message contains, in addition to the typical HI fields, LR information that is specific to that MN such as CN ID and CN prefix. CMAG address will be included too in case of the third topology. When NMAG receives the HI message, it firstly processes this message according to the F-PMIPv6 specification (such as tunnel establishment with PMAG for all MN data packets). Then, NMAG checks for the O-PMIPv6 flag and if it is set, it processes the HI packet by following O-PMIPv6 specification and creates an entry in its internal LR table for the MN. The NMAG sends back HACK message to PMAG to indicate the success of the operation. In addition, in case of topology 3, it will exchange the LRI/LRA information with CMAG to update the CN LR information in CMAG with NMAG being its new peer LR node.

When MN is attached to NMAG, the NMAG sends the prefix information right away to it. NMAG and LMA exchange the PBU/PBA messages as in PMIPv6. When MN sends a data packet to CN on the new link, even before PBU/PBA is exchanged, then the packet will go on the optimal path from NMAG to CMAG. Therefore, it can be seen that the data packet is not lost as expected by F-PMIPv6 [6] however, it is carried through the the optimal path as expected by O-PMIPv6.

To guarantee interoperability of O-PMIPv6 with PMIPv6 as well as with F-PMIPv6, we assume that the LMA keeps track of LR state as in PMIPv6. In other words, upon completion of LR procedure with O-PMIPv6, when the PBU is sent by NMAG to LMA, for example as mobility options, the LMA is notified for any LR status change for a given MN. The LMA then updates the LR cache accordingly. Hence, the LMA is aware of ongoing LR state without being the main entity managing the LR and is not required to send separate LRI to a MAG in order to setup the LR.

Let’s consider the scenario where the MN fails handover after NMAG’s LRI/LRA exchanges due to various reasons (e.g., user’s abrupt movement change, etc.). As specified in [8], localized routing has a defined lifetime specified by the initiator. In other words, there is a delay before releasing the localized routing state in PMAG. In fact, the in-flight packets will be delivered through PMAG which tunnels the packets to NMAG. CMAG will deliver packets towards NMAG using the updated state only after a successful acknowledgement of MN being attached to NMAG. In case of the scenario above, the initiator of the localized routing can terminate the procedure before the expiry of the lifetime specified in the LRI message, by sending a new LRI message with the lifetime set to zero.

On the other hand, if the time between MN’s detachment from PMAG and attachment to NMAG is too long, packets send by CMAG toward NMAG are buffered at NMAG and

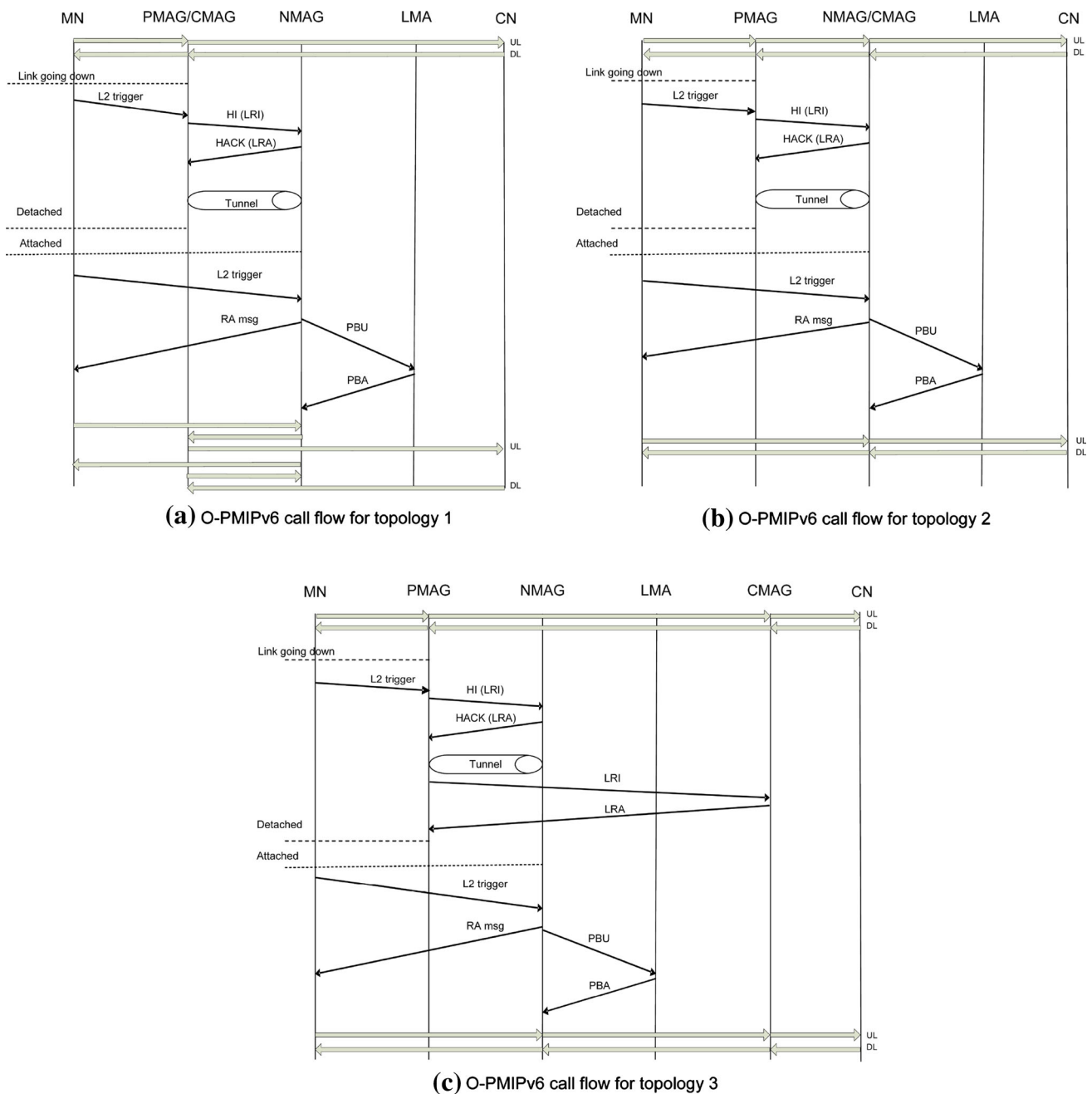


Fig. 4 O-PMIPv6 call flow in different topologies

will be delivered to MN once its attachment with NMAG is complete. In this situation, the initiator can decide to terminate the localized routing by sending a new LRI message with the lifetime set to zero [8]. In order to avoid waste of the resources, if this delay is excessively high, the transient packets in the buffer will be dropped and in this case. This might happen for reasons such as the user switches off its device leading to failure of its attachment to NMAG. This scenario is out of the scope of this paper. The third topology, is picked for the detailed analysis in the subsequent sections.

This choice is made since this topology is more common and also can cover the two other topologies.

3.1.2 Message format

From the network operation and standardization requirements, it is much more complex to change the protocol as they are designed nowadays inside the IETF. As a result, adding a new IPv6 header makes it more difficult to convince the IETF community and makes the network operation more

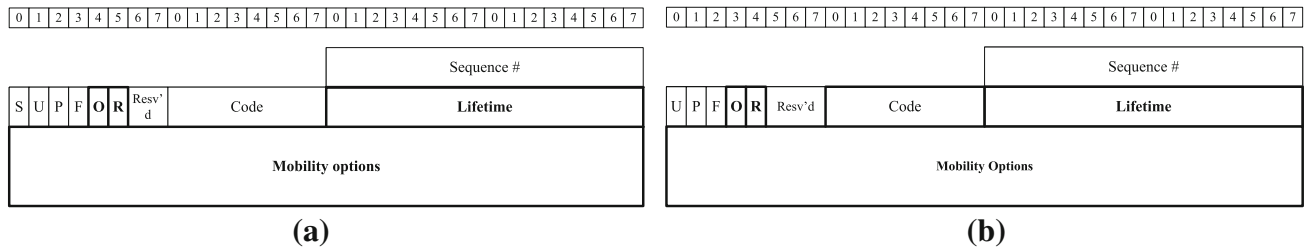


Fig. 5 O-PMIPv6 Message Format. **a** Handover initiate (HI) **b** Handover acknowledgement (HACK)

complex. However, adding a new field in an existing header is much more efficient. For this reason, our choice was to propose the addition of new flags in standardized messages such as HI/HACK. A similar approach has been used for other MIPv6-based protocols.

The only two messages that are modified from F-PMIPv6 are the handover initiate (HI) and the handover acknowledgement (HACK).

Handover initiate (HI) message format:

O-PMIPv6 uses the HI message to prepare the NMAG for the MN that will be attaching to it by providing the MN-ID, MN prefix, and its LR session information. Figure 5 shows the modified version of the original HI message. The fields that are newly introduced or have their possible values modified are indicated in bold and are explained below.

- 'O' flag (O-PMIPv6 flag): This new field is used to distinguish the HI message from F-PMIPv6 defined in [29]. It should be set to 1 to indicate that this packet is O-PMIPv6 and carries LR information. Otherwise, it becomes a F-PMIPv6 message.
- 'R' flag (LRI flag): This new field is set to 1 to indicate it contains relevant LRI information. Otherwise, no LRI information is appended in this message.
- Lifetime (LR supported lifetime): This new field represents the requested time in seconds for which the sender wishes to have local forwarding. It is usually set to the remaining lifetime for the LR session from the sending MAG perspective. A value of 0xffff (all ones) indicates an infinite lifetime.
- Mobility options: This field contains one or more mobility options, whose encoding and formats are defined in [23]. A new combination of options are added in O-PMIPv6 as discussed below:
 - Mobile node identifier (MN1-ID) whose format is defined in [23]. This is already required by F-PMIPv6 [29] to identify the target node and should be used by O-PMIPv6.
 - Mobile Node Home Network Prefix (MN1-HNP) whose format is defined in [23].
 - Mobile node identifier (MN2-ID) whose format is defined in [23]. This identifies the other mobile node involved in the LR, i.e MN2 or CN.

- Mobile node home network prefix (MN2-HNP) whose format is defined in [23]. This is the prefix for MN2 or CN.
- Remote MAG IPv6 (Remote PCoA) as defined by the [8]. This is to identify the CMAG that MN2/CN is attached to. This is required in case of the third topology.

Handover acknowledgement (HACK) message format:

O-PMIPv6 uses the HACK message to acknowledge the receipt of the HI message by the NMAG and provide the LR state information. Figure 5 shows the modified version of the original HACK message. The fields that are newly introduced or have their possible values modified are indicated in bold and are explained below.

- 'O' flag (O-PMIPv6 flag): This new field is used to distinguish the HI message from F-PMIPv6 defined in [29]. It should be set to 1 to indicate that this packet is O-PMIPv6 and carries LR information. Otherwise, it becomes a F-PMIPv6 message.
- 'R' flag (LRI flag): This new field is set to 1 to indicate it contains relevant LRI information. Otherwise, no LRI information is appended to the message.
- Code (status code): In addition to what is defined in [29], the following codes are used for LR purposes: 1 (success) and 2 (localized routing not allowed)
- Lifetime: This new field represents the LR supported lifetime which is lifetime of the LR session that is supported and it is usually copied from the HI message.
- Mobility options: This field contains one or more mobility options, whose encoding and formats are defined in [23]. The same options that were presented in the HI message should exist in this message including any LR information.

4 Analytical model

In this section, a comparative performance analysis is done between PMIPv6, F-PMIPv6 and O-PMIPv6. The factors chosen for this analysis when MN performs a handover are LR establishment latency, signaling cost, packet loss, and network utilization.

4.1 Network model

In order to analyze the performance of the mobility protocols, we propose a single-LMA PMIPv6 domain with an hexagonal model in which the network consists of N cells. Each cell is covered by a MAG and one or more point of attachment (PoA). We assume that the signal strength is strong enough such that there is some overlap coverage at the edges of each cell, meaning that any MN at any location will always have a signal allowing attachment to the network. MAGs and LMA are connected through a number of hops in which a hop can be a normal router. For the purpose of this analysis, we assume that MN is sending/receiving data packets to CN3. Then, MN moves toward MAG2 subnet. This model is depicted in Fig. 3. $D_{MAG,LMA}$ denotes the distance between LMA and MAG in terms of number of hops. The average distance between two MAGs is referred by D_{MAG} . It is typically assumed that the average distance between two MAGs is less than the distance between LMA and MAG [2]. Due to the nature of the hexagonal network, the distance between two MAGs, is equal to \sqrt{N} [4, 13]. An MN is connected to MAG through a PoA. It is assumed that all the links between the major entities (MAG, LMA, and PoA) are wired links while the link between MN and PoA is wireless.

For the purpose of this model, the notations and their descriptions are shown in Table 1. Moreover, The following assumptions are made for the performance evaluation.

- Address configuration is only performed using stateless address auto-configuration, and the time required to combine the network prefix from the RA message to its interface address is negligible.

Table 1 System notation

Notation	Description
$T_{MAG,LMA}$	Transmission delay between MAG and LMA
T_{MAG}	Transmission delay between any two MAGs
$T_{MN,MAG}$	Transmission delay between the MN and MAG
L_w/L_{wl}	Wired/wireless link delay
B_w/B_{wl}	Bandwidth of wired/wireless link
PC_X	Processing cost per packet at node X
W_X	Average waiting time at node X
$D_{X,Y}$	Hop distance between nodes X and Y
PS_X	Packet size of packet type X
L	Session length
λ_s	Arrival rate at any server (# of packets/second)
λ_h	Send/receive rate at the host (# of packets/second)
q	Probability of wireless link failure
D_{wl}	Number of wireless links in a connection path

- There is constant rate traffic between MN and CN. Prior to performing a handover, we assume that an optimal path has been established between MN and CN using LR procedure and all the packets are going through this optimized path.
- The MN starts attaching its interface to the new AP as soon as its previous communication signal strength is weakening. Therefore, we assume that the latency until the AP knows about MN attachment is referred to as L2 handover latency.
- For simplicity, the router solicitation message is not considered here. Thus, only RA message can affect the movement detection of the MN.
- LR is considered completed and packets are optimized when LRA is received (acknowledgement).
- AP is considered a normal router hop with processing cost of PC_R .
- The processing latency of L2 event at NMAG is ignored for simplicity.
- Each hop/router/entity is modeled as an M/M/1 server with mean service time (μ) equal to Processing Cost (PC) of the server and equal arrival rate of λ_s . The system follows a Poisson Process.
- Waiting time of a packet in a system (i.e., server) is defined as queuing delay + service time.
- The link delay in both direction, for any link, is symmetric, i.e. $T(x, y) = T(y, x)$.

The delay experienced by a packet transmitted from source X to the queue of destination Y is equal to the transmission delay and propagation delay at every link between X and Y in addition to the waiting time of the packet at queue. The links can be wired and/or wireless with a failure probability of q . From [18], the total transmission delay between the source and destination, as expressed in (1), is generalized to include the delay at the end points.

$$T_{X,Y} = \frac{1-q}{1+q} \left(\frac{S}{B_{wl}} + L_{wl} \right) * D_{wl} + |D_{X,Y} + 1 - D_{wl}| * \left(\frac{S}{B_w} + L_w \right) + (D_{X,Y} * W) \tag{1}$$

It should be noted that for entities with wired links only (such as MAGs, LMA, etc.), the wireless part of the equation disappears and vice versa.

4.2 Mathematical modeling

After the network model is proposed, each of the performance metrics mentioned above is analyzed for PMIPv6, F-PMIPv6, and O-PMIPv6.

4.2.1 Total localized routing and handover latency

The MN faces downtime when it performs a handover to a new MAG because it is unable to send or receive packets during that operation. We define total localized routing handover latency or delay as the time between the last optimized data packet received prior to MN handover until the first optimized data packet is received after MN hands over to the new network (NMAG). An optimized packet is defined as any packet sent/received by MN after the LR has been established between MN and CN. In other words, the handover latency is given as follows:

$$HO_2 = HO_1 + D_{LR} \quad (2)$$

Where HO_2 is the LR handover latency, HO_1 is the basic handover delay for MN till it is capable of receiving packets and D_{LR} the delay for LR establishment.

We can mark the timestamp of the last data packet received as the moment that the MN detaches completely from PMAG. The MN starts attaching to the NMAG prior to MN full detachment from the PMAG. This is due to the assumption that there is some coverage overlap between the MAGs which prevents MN from being totally disconnected. As a result, the link layer (L2) delay (D_{L2}) is the delay between the moment that MN begins informing the access network of its intention to attach to the new AP and the moment that the NMAG completes this L2 attachment. Since the L2 delay is a common additive delay for all the above mentioned protocols, we can safely ignore this value in our analysis since it does not affect the analytical comparison. However, it is shown in the equations for general consideration. Below, the total LR handover latency is computed as the sum of all the encountered delay during the HO_2 process for each protocol.

(a) PMIPv6

HO_2 , as expressed in (3), is the total latency of MN informing NMAG through the PoA of its attachment (i.e., L2 handover), NMAG sending PBU to LMA and processing it, the maximum of NMAG receiving PBA from LMA, processing it and NMAG sending RA with the prefix to MN through the PoA or LMA sending LRI to NMAG/CMAG, processing it at the MAG, and receiving LRA from both NMAG/CMAG. Finally, NMAG forwards the first optimized data packet to MN through the PoA.

$$HO_2 = D_{L2} + T_{MAG,LMA} + T_{LMA} + \max(T_{MAG,LMA} + T_{MAG} + T_{MN,MAG}, 2T_{MAG,LMA} + T_{MAG}) + T_{Data} \quad (3)$$

(b) F-PMIPv6

HO_2 , as expressed in (4), is the total latency of MN informing NMAG through the PoA of its attachment (i.e.,

L2 handover) and the maximum of NMAG sending RA with prefix to MN through the PoA or NMAG sending PBU to LMA and processing it, and the maximum of NMAG receiving PBA from LMA and processing it, and NMAG sending RA with the prefix to MN through the PoA or LMA sending LRI to NMAG/CMAG and processing it at the MAG, and receiving LRA from NMAG/CMAG. Finally, NMAG forwards the first optimized data packet to MN through the PoA.

$$HO_2 = D_{L2} + \max(T_{MN,MAG}, T_{MAG,LMA} + T_{LMA} + \max(T_{MAG,LMA} + T_{MAG} + T_{MN,MAG}, 2T_{MAG,LMA} + T_{MAG})) + T_{Data} \quad (4)$$

(c) O-PMIPv6

In the case of O-PMIPv6, HO_2 , as expressed in (5), is the total latency of MN informing NMAG through the PoA of its attachment (i.e., L2 handover), NMAG sending RA with the prefix to MN through the PoA and NMAG forwarding the first optimized data packet to MN through PoA.

$$HO_2 = D_{L2} + T_{MN,MAG} + T_{Data} \quad (5)$$

4.2.2 Signaling cost

Another performance parameter used when comparing mobility protocols is the signaling cost as it is considered as an important design consideration for any mobility protocol. Signaling is usually referred by the number of signaling messages used in a protocol and the total message sizes. The higher the cost is, the more processing is required at the different mobility entities, and the more bandwidth is required from the links.

For this analysis, only IP signaling cost is analyzed as L2 signaling is access technology specific and is outside the scope of this analysis. In addition, it is worth noting that the RA message from MAG to MN is considered as one message and the fact that the PoA sits in between and relays the message is ignored.

For the purpose of this analysis, we define two signaling parameters that will be added together to give us the total signaling cost to compare the three protocols.

- HO_s is the total number of signaling messages exchanged that are related to handover,
- LR_s is the total number of signaling messages exchanged that are related to localized routing,
- $TS_s = HO_s + LR_s$ is the total number of signaling messages exchanged for handover and localized routing management

4.2.3 Network utilization

In this network utilization analysis, we focus on the common network entity which is usually the LMA as it is eventually the bottleneck in a PMIPv6-domain. As LMA utilization increases, data packets will experience higher waiting time and thus more delay in addition to delay to signaling packets by other nodes.

When a localized routing is established, data packets go through optimal path which is generally not through LMA. Even though packets might end up going through the LMA if it happens to be on the optimal path, however the LMA will act as a normal router in that case and it has to only do a table lookup on the destination IP address and forwards a packet. On the other hand, when a packet goes through LMA without LMA being a normal router (i.e., non-optimal path) then LMA has to do extra processing such as encapsulation/decapsulation and performs some overheard operation to forward it to its MAG as discussed in PMIPv6 [3].

Since the node utilization for all the mobility protocols after the localized routing has been established is the same, then we only need to study the node utilization over a period that starts with the MN attachment to the new network and the maximum HO_2 we computed previously (i.e., PMIPv6's HO_2). After HO_2 seconds have passed, the node utilization will go to normal operation where it is the same for all the protocols as the LR would have been established by then. We call this duration a session length, L .

We consider the uplink traffic going from MN to CN for this analysis and we assume the traffic is going at constant rate as mentioned above. For the purpose of LR, we assume that LMA does not sit on the optimal path between MN and CN as this is generally the case.

The utilization (ρ) of any node is generally defined as $\rho = \lambda/\mu$ where λ and μ are respectively the arrival rate and service rate of the node.

In order to compute the average arrival rate of packets at the LMA, we need to find out the fraction of data packets, sent by MN, that are actually processed by LMA. In other words, the percentage of packets going through the non-optimal path needs to be computed. This is computed using (6) where $L_{non-opt}$ is the duration in which packets go through non-optimal path.

$$P_{non-opt} = \frac{L_{non-opt}}{L} \tag{6}$$

Any packets traversing the network from the start of the session before the localized routing is established goes through the non-optimal path, i.e., LMA. L is the session length we are monitoring the node utilization over and it is equal to HO_2 for PMIPv6 in this analysis. If all the data packets over duration L are going through LMA, then the LMA will have

an average utilization that can be given by ρ_{node} . Therefore, If we assume that over the same period L only partial number of the total number of packets goes on the non-optimal path through the LMA and the rest of the packets go on the optimal path. Then, we can calculate the average LMA utilization as shown in (7):

$$\rho_{LMA} = P_{non-opt} * \rho_{node} \tag{7}$$

Below is the LMA utilization computed for each of the protocols over L :

(a) PMIPv6

All the data packets in PMIPv6 in any scenario go through non-optimal path. This is due to the fact that LR establishment is completed at the end of handover procedure. This is illustrated in (8).

$$P_{non-opt} = \frac{HO_{2,PMIPv6}}{L} \tag{8}$$

(b) F-PMIPv6

A fraction of the data packets in F-PMIPv6 in any scenario go through the non-optimal path until the LR is established. Therefore, any packets prior to $HO_{2,F-PMIPv6}$ are taking non-optimal path. Then, packets start going through optimal path till the end of the session. This is illustrated in (9).

$$P_{non-opt} = \frac{HO_{2,F-PMIPv6}}{L} = \frac{HO_{2,F-PMIPv6}}{HO_{2,PMIPv6}} \tag{9}$$

By replacing Eqs. (3) and (4) in the above equation, we can obtain the full expression.

(c) O-PMIPv6

All of the data packets in O-PMIPv6 in any scenario go through the optimal path from the start of the session until the end of it. This is due to the fact the LR is established prior to MN performing handover, i.e., attaching to the new network, and therefore from the start of the session. Therefore, $P_{non-opt} = 0$.

4.3 Results and analysis

Each one of the above discussed performance metrics has been plotted to illustrate the improvement gained by implementing O-PMIPv6 compared to PMIPv6 and F-PMIPv6. Table 2 shows the system parameters that have been used where some of the values are taken from the literature [4, 13].

In addition, the control packets sizes for the different protocols have been estimated as shown in Table 3. The size of the packets varies depending on the mobility options used. However, we used the following estimation equation:

$$\begin{aligned} Control\ pkt\ size &= IPv6\ header + Mobility\ header \\ &+ Mobility\ option(s) \end{aligned} \tag{10}$$

Table 2 System parameters

Parameter	Value
L_w/L_{wl}	2 ms / 10 ms
B_w/B_{wl}	100 Mbps / 11 Mbps
PC_R	8
PC_{MAG}	12
PC_{LMA}	24
$D_{MN,MAG}/D_{CMAG,CN}$	1 hop
$D_{MAG,LMA}/D_{CMAG,LMA}$	$a = 10$ hops
$D_{MAG,CMAG}$	\sqrt{N} where N is 9–100 cells When fixed, $N = 30$ cells
λ_s	10–100 pkts/s, when fixed it is 50
λ_h	2–30 pkts/s, when fixed it is 5
q	0.01

Table 3 Control packet sizes

Packet type	Packet size
PS_{PBU}	75 bytes
PS_{PBA}	75 bytes
$PS_{HI-FPMIP}$	57 bytes
$PS_{HACK-FPMIP}$	57 bytes
PS_{LRI}	71 bytes
PS_{LRA}	71 bytes
$PS_{HI-OPMIP}$	71 bytes
$PS_{HACK-OPMIP}$	71 bytes
PS_{RA}	80 bytes (IPv6 header + RA packet size with 8 bytes prefix option)

Where IPv6 header = 40 bytes, mobility header = 6 bytes and the mobility option varies according to the packet (here we used the minimum required).

4.3.1 Total localized routing handover latency

Two different scenarios are used to evaluate the handover performance. In the first scenario, the handover performance is measured against wired link congestion (i.e., the core network). The general congestion level of the network can be varied by changing the value of the packet waiting delay at different network nodes. By increasing the packets arrival rate at any network node, the network congestion increases from the point of view of queuing system. The results are shown in Fig. 6a.

In the second scenario, the handover performance is measured against different number of hops between the major core network entities affecting handover. This is done by modifying the number of hops between MAG and LMA while fixing the congestion level of the network. This is shown in Fig. 6b.

We can see that O-PMIPv6 outperforms both F-PMIPv6 and PMIPv6 in both cases when it comes to the total localized routing handover latency. This is a huge advantage for O-PMIPv6 due to the fact that the LR information was sent prior to performing the handover. It can be seen that the network congestion and the distance between MAG and LMA does not have a major impact on O-PMIPv6, however the impact on PMIPv6 and F-PMIPv6 is much larger than O-PMIPv6. This is due to the fact that all of the O-PMIPv6 packets follow the optimal path between the two MAGs, therefore the distance between LMA and MAG is irrelevant. In the case of other protocols, packets that go through non-optimal path have to go through the path between LMA and MAG. Since packet loss rate is proportional to the handover latency [16], similar results or behavior than those illustrated in Fig. 5 have been obtained. However, the graph for packet loss rate is not displayed.

4.3.2 Signaling cost

Table 4 shows the signaling cost for each protocol when a single handover is performed. HO, LR, and total signaling (TS) are in terms of message exchanged while the total signaling overhead is in bytes. It can be seen that O-PMIPv6 outperforms F-PMIPv6 and PMIPv6 in terms of signaling cost. This benefit is multiplied when multiple MN perform a handover simultaneously since the number of signaling messages adds up.

4.3.3 Network utilization

Two different cases are used to evaluate the LMA utilization. It should be noted that a longer LMA processing time, as shown before, is used for the analysis, therefore low number of nodes and small packet rate is sufficient. This is done for illustration purposes as a proof of concept since it will be easier to show the graphs in the previous sections with same values. Scaling all the above values accordingly to deployment numbers will yield the same results below.

In the first scenario, there are multiple MNs performing handover simultaneously. We vary the number of MNs performing handover simultaneously from 1 to 10 MNs. We assume that each MN is sending packets at the same constant rate to its CN. In addition, we assume that LMA has already 20% steady utilization due to other factors such as packets from other sources, processing, etc.

In the second scenario, there is one MN performing a handover. We vary the MN sending packets rate from 2 to 30 packets/s to its CN. As before, we assume that LMA has already some steady utilization due to other factors.

It can be seen from Fig. 7a, b that LMA utilization in the case of O-PMIPv6 is independent of the number of MN performing a handover and the MN packet sending rate. In

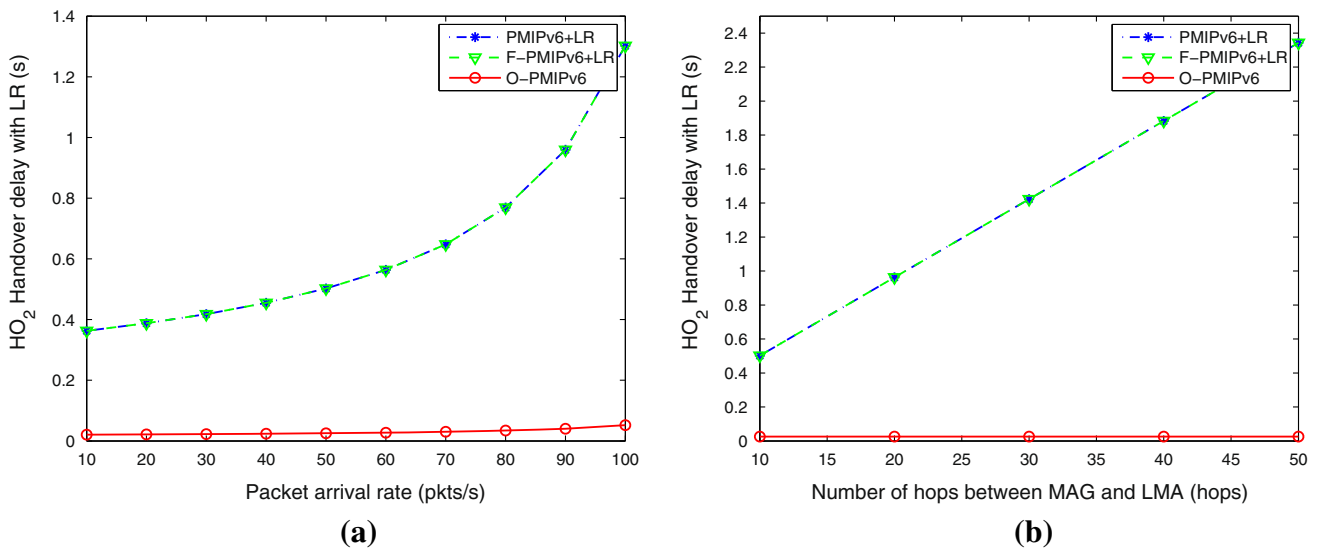


Fig. 6 Total localized routing handover latency. **a** Network congestion versus HO_2 delay **b** distance between MAG and LMA versus HO_2 delay

Table 4 Signaling overhead comparison

	PMIPv6	F-PMIPv6	O-PMIPv6
HO_s	3	5	5
LR_s	4	4	0
TS_s	7	9	5
Signaling overhead (bytes)	514	628	372

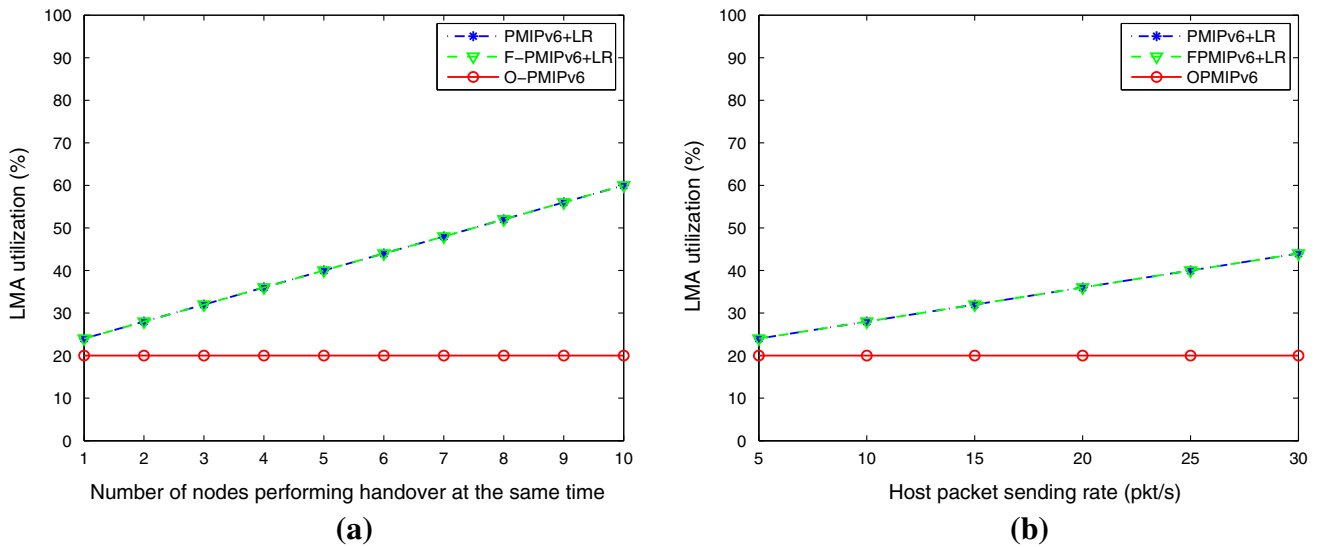


Fig. 7 Network utilization. **a** Number of nodes performing HO versus LMA utilization **b** MN sending packet rate versus LMA utilization

fact, the LMA utilization is not affected by MN performing handover. This is again due to the fact that all the packets go through optimal path which does not have LMA sitting on.

However, in the case of F-PMIPv6 and PMIPv6, the LMA utilization increases as the number of MN performing a handover or the MN packet sending rate increases as partial number of the total number of packets go through non-

optimal path in which LMA sits on (i.e., prior to the LR setup after the handover is completed).

5 Simulation results

In order to further analyze O-PMIPv6 and compare its performance to PMIPv6 and F-PMIPv6, we simulated the three

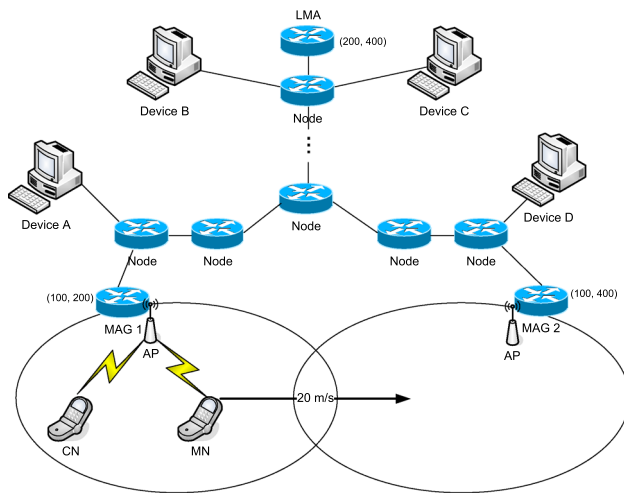


Fig. 8 Default network topology used in the simulation

protocols and acquire the results in a similar fashion as explained in the previous section.

5.1 Simulation setup

NS2.29 with the National Institute of Standards and Technology (NIST) mobility extension [20,21] and initial PMIPv6 implementation [1] have been used as a starting point. PMIPv6 has been modified to include LR and possibility of handling multiple MN handover. The two remaining protocols, namely F-PMIPv6 with LR and O-PMIPv6, have also been implemented. All of the above protocols implementations have been successfully tested by tracing the data packets and their timing to make sure that they are following the correct path from source to destination as expected. The topology shown in Fig. 8 has been the default topology used for simulating the above mobility protocols. As can be seen in the figure, for the sake of simplicity, MAG1 and MAG2 are located 200m apart from each other. Each MAG has 110 m coverage therefore there is a bit of coverage overlap to allow for all-time coverage. The simulation starts with MN and CN connected to MAG1. At random number of seconds, traffic starts going from MN to CN. MN moves at a steady speed of 20 m/s east towards the coverage of MAG2. This is enough time to allow MN to establish LR with its CN prior to its handover as assumed in the mathematical section. When MN reaches the full coverage of MAG2 and is totally disconnected from MAG1, it stays there till the end of simulation. The simulation ends after 19s from the start. It should be noted that the results were taken from running the simulation several times, in order to increase the confidence interval, for each test case where each simulation has packets generated at random type. The other values were kept fixed as much as possible as stated in each respective section in order to have a fair comparison with the mathematical model. The

average value for each set of simulations is taken. Also, it should be kept in mind that a hop refers to a node sitting in between two network entities. The wireless technology used is 802.11 and as a result, the L2 handover technology is specific to NS2/NIST implementation of 802.11 which is outside the scope of this research.

5.2 LR handover latency simulation

As mentioned in the previous sections, the handover delay of MN with LR established is the time elapsed from the timestamp that the last data packet received by CN on the optimal path prior to MN handover till the first data packet received by CN after MN handover. The fact that CN is able to receive the packets successfully on the optimal path means that MN was able to send the packets successfully and therefore, MN handover was completed. The following two cases are tested for LR handover latency.

5.2.1 Case 1: wired link congestion

The handover performance is measured against wired link congestion level (i.e., the core network). The congestion level is varied by inserting extra traffic for external nodes (denoted as device A, B, C, and D in the figure). The sum of the number of packets involved in the communications among all the devices at each core node is varied from 10 pkt/s to 100 pkt/s. Therefore by increasing the packet arrival rate at the core nodes, the network congestion increases accordingly. It can be noted that given the above rate, the wired link bandwidth utilization due to external packets only is varied from 0.08 to 8%. This may not seem a high utilization to the link, however the processing of these packets at the core nodes may be the bottleneck depending on the rate and that what gives the network a sense of congestion desired. The distance between the MAGs is kept at 5 hops while the distance between any MAG and LMA is kept at 10 hops. Also, the packet sending rate of MN is 5 pkt/s where each packet has the size of 1000 bytes. The result of this test case is illustrated in Fig. 9a. As can be seen, the handover delay with LR for both PMIPv6 and F-PMIPv6 increases as the congestion of the network increases. This is due to the fact that packets will start taking the optimized path after the LR is established which happens following the establishment of localized routing after the exchange of PBU/PBA. Therefore, the exchange of these signaling packets take longer time as the congestion level increases. In the case of O-PMIPv6, the handover is relatively lower as the localized routing is established between the MAGs prior to performing the handover. Therefore, the delay faced here is the sum of the L2 attachment delay of MN to MAG2 and the network congestion. However, this is still significantly lower than PMIPv6 and F-PMIPv6. It can be seen that the PMIPv6 and F-PMIPv6 curves don't exactly

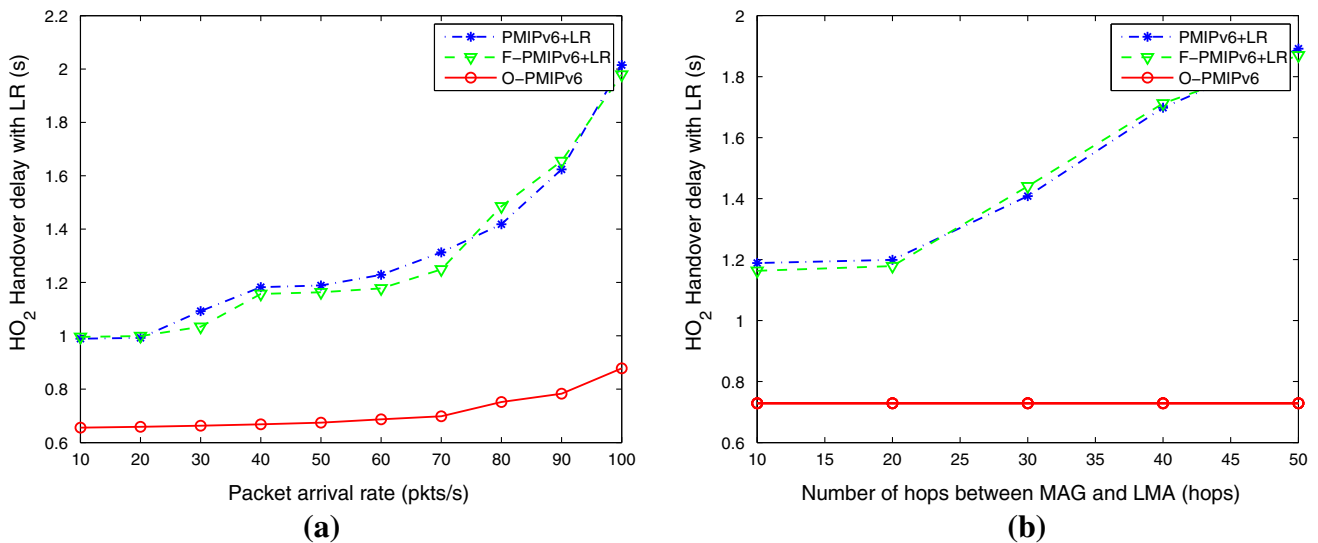


Fig. 9 LR handover latency simulation. **a** Network congestion versus HO_2 delay, **b** distance between MAG and LMA versus HO_2 delay

overlap, however they are very close to each other. This is just due to the simulation environment where delays are random and based on the condition of the network.

5.2.2 Case 2: distance between MAG and LMA

The handover performance is measured against the distance between each MAG and LMA in terms of hops. The distance is increased by adding more nodes between MAG and LMA that do basic routing. The distance between MAG and LMA is varied from 10 hops to 50 hops. The congestion level of the network is set by external traffic sending at 50 pkt/s. The distance between the MAGs is kept at five hops. Also, the packet sending rate of MN is kept at 5 pkt/s where each packet has the size of 1000 bytes. The result of this test case is illustrated in Fig. 9b. As can be seen, both PMIPv6 and F-PMIPv6 handover delay with LR increases as the distance between MAG and LMA. This is for the same reasons discussed in the previous case. As the distance between the MAG and LMA increases, the handover delay increases significantly because the exchange of PBU/PBA and LRI/LRA takes longer to be exchanged between MAG and LMA. Also, it can be seen that there is a sharp increase starting from when the number of hops is equal 20. This is due to the increased possibility of having higher packet drop rate and retransmissions when the distance starts to get big resulting in the increased delay. In the case of O-PMIPv6, the handover is relatively lower and is constant. The reason is that the LR is established prior to handover and the data packet going from MN to CN sits on the optimal path which is not affected by the distance between MAG and LMA. It can be seen that the PMIPv6 and F-PMIPv6 curves don't exactly overlap, however they are very close to each other.

5.3 Signaling cost simulation

The signaling cost was calculated by monitoring the number of signaling packets needed to be exchanged for the purpose of handover of MN and establishing its LR. The results acquired were focused on the mobility management protocol related signaling (i.e., IP layer signaling). Therefore, the signaling required to do L2 handover, which is 802.11 specific, is not considered as it is outside the scope of this paper and is common to all the protocols. The signaling cost is measured against the number of nodes performing handover at the same time. The number of MNs performing handover is varied from 1 to 10 MNs in which all the MNs are communicating with their corresponding CNs. The results of this test case are illustrated in Fig. 10 and are consistent with the num-

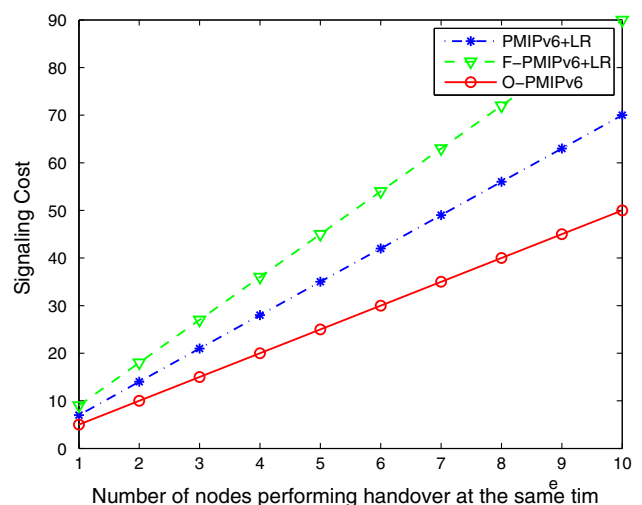


Fig. 10 Number of nodes performing HO versus Signaling cost

ber of signaling messages mentioned before (see Table 4). F-PMIPv6 has the highest signaling cost as it involves the extra signaling due to the exchange of HI/HACK packets. PMIPv6 has a lower signaling cost than F-PMIPv6 as it does not need to exchange the HI/HACK packets. However, O-PMIPv6 is the lowest in terms of signaling cost, since it encapsulates the LR information in the HI/HACK packets which allows considerable savings of network resources and performance improvement. As the number of nodes performing handovers increases, the performance gain becomes more apparent. This is due to the fact that O-PMIPv6 has the lowest number of signaling messages that are required to perform handover and establish LR. This benefit is magnified since the number of nodes increases as the number of signaling messages adds up.

5.4 LMA utilization simulation

The LMA utilization during MN handover is measured by dividing the average data packets arrival rate at LMA over the LMA average processing time. Portion of the data packets sent by MN when it is attached to the new network follows the non-optimal path (i.e., through LMA) until the handover and LR establishment is completed. We are specifically interested in the LMA utilization over this period to analyze the influence of different mobility protocols on LMA utilization in different cases. The following cases have been tested:

5.4.1 Case 1: number of MNs performing handover

The LMA utilization is measured against the number of MN performing handover at the same time. The number of MN is varied from 1 to 10. Each MN is connected to a different MAG (as well as their CNs) as shown in Fig. 11a. All the MAGs are connected to the same LMA using the same distances. The introduction of new MAGs is done to minimize the packets collision over the wireless medium that are sent by MNs and to reduce any other L2 technology specific effect. In addition, since MAGs are generally slower, we wanted to minimize the effect of this node being a bottleneck to reduce the effect on the number of data packets forwarded to LMA and therefore its utilization. The congestion level of the network is set by external traffic sending at 30 pkt/s. The distance between the MAGs is kept at 5 hops while the distance between any MAG and LMA is kept at ten hops. Also, the packet sending rate of MN is kept at 5 pkt/s where each packet has the size of 1000 bytes. The LMA utilization is assumed to start at 20% due to other processing that is not related to this test or any mobility protocol, i.e. some of the external traffic is passing through LMA.

It can be seen that in both PMIPv6 and F-PMIPv6, LMA utilization is high and increases dramatically given that number of MNs performing handover increases. This is due to

the fact that portion of the data packets after handover passes through LMA until LR is established. It can be seen also that when the number of MN performing handovers is 6 then we have a relatively high utilization. This is due to the fact that there was not a lot of dropped packets and few retransmissions which resulted in most of the packets being received by the LMA. In case of O-PMIPv6, it can be seen that LMA utilization stays at the default value of 20% independently of the number of nodes performing handover. This is due to the fact that LR is established prior to handover, therefore all the packets sent by MN after being attached to the new MAG are forwarded on the optimal path. It is worth noting that one may see that PMIPv6 and F-PMIPv6 curves don't overlap each other exactly as expected. The reason is that in the case of F-PMIPv6, packets sent from MN are forwarded immediately from MAG 1 to MAG 2 to go on the non-optimal path to LMA. This process gives high probability of dropping packets and more steady traffic received by the LMA resulting in the above utilization. While in the case of PMIPv6, packets are buffered at MAG1 until PBA is received, then all packets are forwarded to LMA. This has less number of nodes to go through and high arrival rate in small period of time at LMA. This is the reason why utilization is a bit higher than F-PMIPv6.

5.4.2 Case 2: number of MNs performing handover

The LMA utilization is measured against the host (i.e. MN) packet sending rate. The packet sending rate is varied from 5 to 30 pkts/s. The congestion level of the network is set by external traffic sending at 30 pkt/s. The size of any data packet is 1000 bytes. The distance between the MAGs is kept at 5 hops while the distance between any MAG and LMA is kept at 10 hops. The LMA utilization is assumed to start at 20% due to other processing that is not related to this test or any mobility protocol, i.e. some of the external traffic is passing through LMA.

The result of this test case is illustrated in Fig. 11b. It can be seen that in both PMIPv6 and F-PMIPv6, LMA utilization is high and increases dramatically as the host packet sending rate increases. This is due to the fact that portion of the data packets after handover passes through LMA until LR is established and the higher the data packet rate is, the higher the portion of packets received by LMA is and consequently, the higher the LMA utilization is. In case of O-PMIPv6, it can be seen that LMA utilization stays at the default value of 20% independently of the host packet sending rate. This is due to the fact that LR is established prior to handover, therefore all the packets sent by MN after being attached to the new MAG are forwarded on the optimal path. It can be noticed here also that PMIPv6 utilization is slightly higher than F-PMIPv6. This is due to the same reason mentioned in the previous test case.

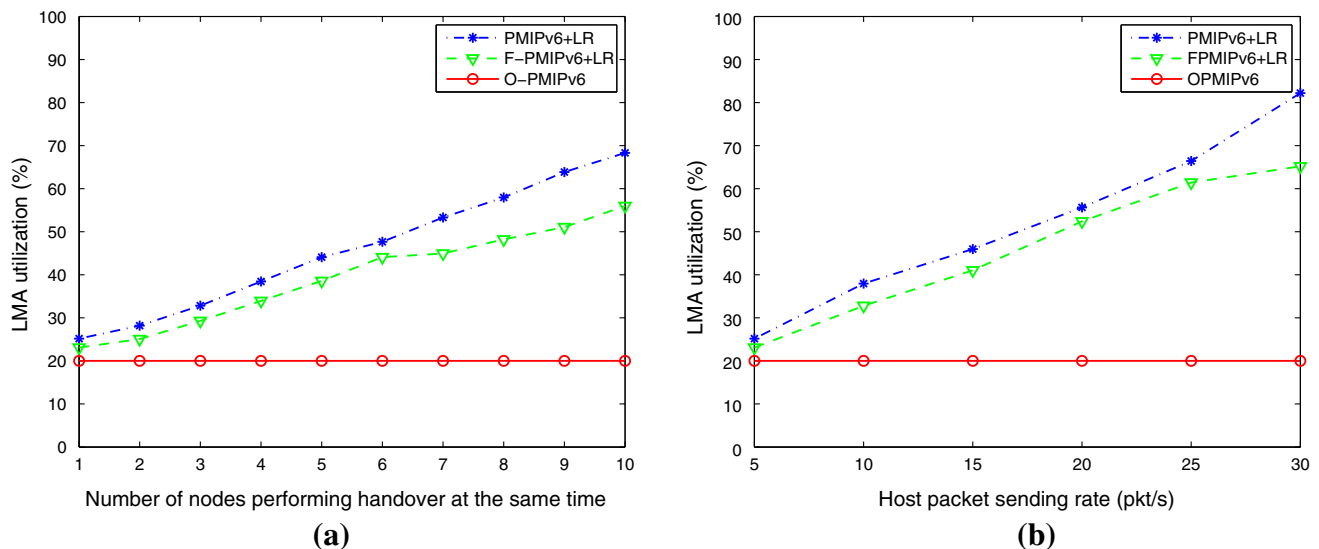


Fig. 11 LMA utilization simulation. **a** Number of nodes performing HO versus LMA utilization. **b** MN sending packet rate versus LMA utilization

6 Conclusion and future work

Since the design of IP-based mobility management protocols such as MIPv4, MIPv6, and PMIPv6, their optimization has attracted a lot of interests in the networking community. Our work shares the same goal. However, in PMIPv6, the optimization is based on two separate procedures, HO optimization and LR. By performing these two procedures independently, some redundancies are introduced with signaling message exchange. Our solution, called O-PMIPv6, performs these two procedures simultaneously.

As shown by the mathematical and simulation results, implementing O-PMIPv6 gives network operators a huge improvement to multiple performance factors such as localized routing handover delay, signaling cost, and LMA utilization. This is extremely crucial in real network deployment in which highly mobile nodes have localized routing established with their remote CN. In addition, O-PMIPv6 exhibits higher performance benefits over F-PMIPv6 and PMIPv6 in term of handover delay and packet loss rate. This is due to the fact that O-PMIPv6 extends F-PMIPv6 by anticipating LR re-establishment when the need of handover is initiated. It was proved mathematically and by simulation that O-PMIPv6 performs better than F-PMIPv6 and PMIPv6 in terms of LR handover latency, signaling cost, and core network utilization.

Our future work is to study the performance of O-PMIPv6 for a scenario where multiple-LMA PMIPv6 domains are involved. Another future work is to study the performance of O-PMIPv6 when the mobile node is performing inter-domain handover. This is important as the mobile node may decide to handover to another domain and its LR will need to be re-

established with minimum performance impact. In the case of inter-domain handover, the LR setup is involving more than one LMA and these LMAs may have different policies for managing their LR as they can be controlled by different operators.

References

- Choi, H. Proxy Mobile IPv6 in NS-2.29, Aug 2009. <http://communi.net/pmip6ns>.
- Choi, Y., & Chung, T. (2009). Enhanced light weight route optimization in Proxy Mobile IPv6. In *Fifth international joint conference on INC, IMS and IDC* (pp. 501–504).
- Gundavelli, S, Leung, K., Devarapalli, V., Chowdhury, K., & Patil, B. (2008). Proxy Mobile IPv6. IETF RFC 5213.
- Han, B., Lee, J., Lee, J., & Chung, T. (2008). PMIPv6 route optimization mechanism using the routing table of MAG. In *Third international conference on systems and networks communications* (pp. 274–279).
- Han, B.-J., Lee, J.-M., Lee, J.-H., & Chung, T.-M. (2008). PMIPv6 route optimization mechanism using the routing table of MAG. In *3rd international conference on systems and networks communications* (pp. 274–279).
- Kim, S., Lee, J., Chung, T. (2009). Performance analysis of fast handover schemes for proxy mobile IPv6. In *The fourth international conference on systems and networks communications* (pp. 37–48).
- Koodli, R. (2009). Fast Handovers for Mobile IPv6. IETF RFC 5568, July 2009.
- Krishnan, S., Koodli, R., Loureiro, P., Wu, Q., & Dutta, A. (2012). Localized Routing for Proxy Mobile IPv6. IETF RFC 6705.
- Lach, H., & Janneteau, C. (2003). Network mobility in beyond-3G systems. *IEEE Communications Magazine*, 41(7), 52–57.
- Lee, J., & Chung, T. (2010). How much do we gain by introducing route optimization in proxy mobile IPv6 networks? *Annals of Telecommunications*, 65(5–6), 233–246.
- Lee, J., & Park, J. (2008). Fast Handover for Proxy Mobile IPv6 based on 802.11 Networks. In *IEEE 10th international conference on advanced communication technology* (pp. 1051–1054).

12. Lee, J.-H., Bonnin, J.-M., You, I., & Chung, T.-M. (2013). Comparative handover performance analysis of IPv6 mobility management protocols. *IEEE Transactions on Industrial Electronics*, 60(3), 1077–1088.
13. Lee Sri, J., & Chung, T. (2009). A performance analysis on route optimization for Proxy Mobile IPv6. In *IEEE international conference on communications (ICC)* (pp. 1–6).
14. Lei, J., & Fu, X. (2008). Evaluating the benefits of introducing PMIPv6 for localized mobility management. In *Wireless communications and mobile computing conference* (pp. 74–80).
15. Li, Y., Su, H., Su, L., Jin, D., & Zeng, L. (2009). A comprehensive performance evaluation of PMIPv6 over IP-based cellular networks. In *IEEE 69th vehicular technology conference* (pp. 1–6).
16. Liebsch, M., Jeong, S., & Wu, Q. (2011). Proxy Mobile IPv6 localized routing problem statement. IETF RFC 6279.
17. Liu, Y., Xu, X., Ding, X., & Cui, Y. (2014). Localized routing optimization for multi-access mobile nodes in PMIPv6. In *23rd international conference on computer communication and networks (ICCCN)* (pp. 1–5).
18. Makaya, C., & Pierre, S. (2008). An analytical framework for performance evaluation of IPv6-based mobility management protocols. *IEEE transactions on wireless communications*, 7(3), 972–983.
19. Modares, H., Moravejosharieh, A., Lloret, J., & Salleh, R. B. (2014). A survey on proxy mobile IPv6 handover. *IEEE Systems Journal*, 9, 1–10. PP.
20. National Institute of Standards and Technology (NIST). Mobility Extension for ns-2, Aug. 2009. http://www.nist.gov/itl/antd/emntg/ssm_tools.cfm.
21. ns-2. The network simulator, version 2.29, Nov. 2005. <http://www.isi.edu/nsnam/ns/>.
22. Perkins, C. (2010). IP Mobility Support for IPv4. Revised. IETF RFC 5944 Nov. 2010
23. Perkins, C., Johnson, D., & Arkko, J. (2011). Mobility Support in IPv6. IETF RFC 6275, 2011.
24. Quoc, A. K., Kim, D.S., & Choo, H. (2014) A novel scheme for preventing out-of-order packets in fast handover for Proxy Mobile IPv6. In *2014 International conference on information networking* (pp. 422–427).
25. Rasem, A., Makaya, C., & St-Hilaire, M. (2012) Efficient handover with route optimization in proxy mobile IPv6 domain. In *8th IEEE international conference on wireless and mobile computing, networking and communications (WiMob)* (pp. 47–54).
26. Rasem, A., St-Hilaire, M., & Makaya, C. (2012). A comparative analysis of predictive and reactive mode of optimized PMIPv6. In *8th International wireless communications and mobile computing conference (IWCMC)* (pp. 722–727).
27. Soliman, H. (2008). Hierarchical mobile IPv6 (HMIPv6) mobility management. IETF RFC 5380.
28. Yan, Z., & Lee, J.-H. (2013). State-aware pointer forwarding scheme with fast handover support in a PMIPv6 domain. *IEEE Systems Journal*, 7(1), 92–101.
29. Yokota, H., Chowdhur, K., Koodli, R., Patil, B., & Xia, F. (2010). Fast Handovers for Proxy Mobile IPv6. IETF RFC 5949, 2010.
30. Zhang, Q., & Shan, Z. (2013). Using fast handover to quicken the re-establishment procedure of localized routing for Proxy Mobile IPv6. In *22nd Wireless and optical communication conference (WOCC)* (pp. 349–354).



Ahmad Rasem completed his MASc in Systems and Computer Engineering from Carleton University in 2011. Prior to that, he graduated with a Bachelor of Communications Engineering with high distinction and received the Senate Medal for Academic Excellence from Carleton University in 2006. In addition, he has recently completed a Master Certification in Project Management. Ahmad was a teaching assistant for programming courses for over 5

terms at Carleton University and was a recipient of the Teaching Excellence award. Later, he worked in Nortel as a software designer as part of R&D on various network concepts including but not limited L2 Ethernet, LANS/VLANS, ATM, and PseudoWire in addition to the working knowledge on O&M. Currently, Ahmad works in the Government of Canada as an Agile Team Leader for IT design and development delivering multi-million dollars solutions. In addition, Ahmad does mobile development by developing Android mobile applications and has currently multiple apps on the market. Ahmad's main interests are in technology, networking, mobile development and Agile project management.



Marc St-Hilaire joined Carleton University in 2006 upon completion of his PhD in Computer Engineering, from École Polytechnique de Montréal. He is currently an associate professor with the School of Information Technology with a cross appointment with the Department of Systems and Computer Engineering at Carleton University. Dr. St-Hilaire is conducting research on various aspects of wireline and wireless communication systems. More precisely,

he is interested in network planning and infrastructure, network protocols, network interconnection, and performance analysis. With more than 100 publications, his work has been published in several journals and international conferences. Finally, Dr. St-Hilaire is involved in the organisation of several national and international conferences and workshops. He is also a senior member of the IEEE.



Christian Makaya is currently a researcher at IBM T.J. Watson Research Center. He received his Ph.D. in Computer Engineering from the Ecole Polytechnique de Montreal, University of Montreal, Canada in 2007. Prior to joining IBM, he was a senior research scientist at Telcordia Technologies and a visiting researcher at Ericsson. Dr. Makaya leads several research and development programs in the area of networking and distributed systems. The focus of his

research interests is on cloud computing, network functions virtualization (NFV), software-defined networking (SDN), big data and analytics, wireless and mobile networking, Internet of things (IoT), and network security. Dr. Makaya has authored numerous technical papers in peer-reviewed journals, conferences, and book chapters. He is an active contributor to IEEE NGSON, IEEE 802.21, OMA DM, 3GPP, and IETF as well as to several open source communities. He received the M.Sc. degree in Computer Science from the University of Montreal and the M.Sc. degree in Telecommunications from INRS-EMT, University of Quebec, Montreal, Canada in 2003 and 2004, respectively. Christian is a member of the IEEE and the IEEE Communication Society.