A Hybrid Centralized-Distributed Mobility Management for Supporting Highly Mobile Users

Tien-Thinh Nguyen and Christian Bonnet
Department of Mobile Communications
EURECOM
Sophia-Antipolis, France
Email: {Tien-Thinh.Nguyen, Christian.Bonnet}@eurecom.fr

***Abstract*—Distributed Mobility Management (DMM) is a new trend to overcome the limitations of the current IP mobility management protocols raised by the rapid increasing mobile Internet usages. It is based on the idea of flattening the network architecture and providing mobility service when it is necessary. Although DMM outperforms the well known protocol - Proxy Mobile IPv6 (PMIPv6) in terms of optimizing the network resources consumption, the DMM deployment faces several challenges such as complex address and tunnel management, high signaling cost and high handover latency when the number of addresses and tunnels associated with the mobile node increase (e.g., in case of users moving at a high speed). For this reason, we introduce a hybrid centralized-distributed mobility management architecture (H-DMM) for supporting highly mobile users. The numerical results showed that H-DMM can retain the advantages of DMM while limiting its drawback in comparison with PMIPv6 in terms of signaling cost, packet delivery cost, handover latency and end-to-end delay even in case of users with highly mobility features.**

Keywords*—IP Mobility, Distributed Mobility Management, Proxy Mobile IPv6, Network-based Mobility Management.

I. Introduction

Nowadays, the mobile data services have become an essential part of many consumers' daily lives [1], [2]. As a result, the mobile data traffic is expected to grow to 15.9 exabytes per month by 2018, a 11-fold increase over 2013 [1]. Despite the increasing volume of traffic, the average revenue per user is falling fast. On the other hand, as current mobile networks are evolving towards all-IP architecture and the mobile nodes (MNs) may frequently change their point of attachment to the IP network, IP mobility management is a crucial concept to meet the demand of ubiquitous Internet connectivity, for example, providing Internet connection on moving vehicles like cars, buses and subways.

The mobile network operators are being challenged by the increase of mobile data traffic and the new requirements e.g., providing connectivity anywhere and at any time with consistency of user experience, while preserving the economics of their networks and creating new opportunities for revenue growth. Although further dramatic increases in radio capacity of mobile broadband will come with the deployment of new wireless technologies such as Evolved High Speed Packet Access (HSPA+) and Long Term Evolution (LTE), spectrum for operators is however both limited and expensive. Thus, the network operators are looking at different methods to increase the system capacity such as deploying femto and pico cells, together with simplifying the network architecture as well as optimizing the data transmission costs. Accordingly, the mobile network is currently evolving towards flat architecture. The 3GPP[1] also proposes such traffic offloading techniques as Local IP Access/Selected IP Traffic Offload (LIPA/SIPTO) and IP Flow Mobility (IFOM) [3]. Following the same idea, the Internet Engineering Task Force (IETF) has recently chartered the Distributed Mobility Management (DMM) working group[2] which specifies the solutions to address the problems and limitations of the current centralized mobility management (CMM) such as sub-optimal routing, scalability issues, and reliability (the centralized mobility anchor represents a bottleneck and single point of failure) [4], [5].

Since DMM is currently a topic of high interest to both academia and industry, a lot of work has been done considering different DMM approaches [5]. All of them proved that DMM is generally a promising mobility management scheme. The reason is that DMM allows the traffic to be offloaded easily from the core network since the mobility anchor is put at the network edge. In addition, the mobility support is enabled when it is really needed and the traffic is better distributed among the network entities, which as a result helps reduce the network congestion and resources waste. However, DMM deployment also faces several issues such as complex address and tunnel management, high signaling cost and high handover latency as the number of addresses and the number of bi-directional tunnels associated with the MN increase, for example, in case of users moving at a high speed (i.e., smart phone users on vehicles) and/or with long-lasting sessions [6], [7], [8]. Consequently, DMM may not be a suitable scheme for vehicles or users on moving vehicles.

In this document, we propose a hybrid centralized-distributed mobility management architecture (H-DMM) for supporting such highly mobile users. As a network-based approach, it provides the mobility support for all legacy devices. In line with the DMM philosophy, the solution keeps the benefits of DMM while mitigating its limitations compared

This work was performed within SYSTUF project, which is subsidized by the French ministry of Industry in the framework of the AMI ITS program. EURECOM's research is partially supported by its industrial members: BMW Group Research & Technology, IABG, Monaco Telecom, Orange, SAP, SFR, ST Microelectronics, Symantec.

[1]Third Generation Partnership Project, http://www.3gpp.org/
[2]IETF DMM Working Group: https://ietf.org/wg/dmm/

to the centralized approach (including Mobile IPv6 (MIPv6) [9] and Proxy Mobile IPv6 (PMIPv6) [10]). It is done by a smart combination between DMM and PMIPv6. Based on a set of metrics (e.g., the characteristics of the flow and the number of active prefixes) the appropriate approach will be applied. In more details, when moving inside the domain, the MN can obtain two different prefixes, one from the current Mobility Access Router - MAR (as a normal DMM behavior) and the other from the central entity (Local Mobility Anchor (LMA) - following the PMIPv6 operation). The former prefix is changed every time the MN performs a handover, while the latter remains unchanged. The numerical results showed that the solution (H-DMM), as similar to DMM, outperforms PMIPv6 in terms of packet delivery cost and end-to-end delay in a low mobility and/or a short-lived flow scenario. It also helps prevent a high increase of signaling cost, packet delivery cost, and handover latency in DMM compared to PMIPv6 in case of high mobility and/or long-lived flow scenario.

The rest of this paper is organized as follows. Section II presents the background information related to the centralized and distributed mobility management approach as well as highlights the advantages and limitations of each approach. Section III describes the proposed solution regarding its architecture and operations. Section IV provides performance analysis in terms of signaling cost, packet delivery cost, handover latency, and end-to-end delay. Section V shows the numerical results from the given analysis. Finally, Section VI concludes the paper and provides perspectives for future work.

II. Related Work

A. IP Mobility Management: from Centralized to Distributed Approach

Mobile IPv6 (MIPv6) [9], as a global mobility protocol, maintains the mobile node's reachability when it is away from home. It is done by introducing a central entity, namely Home Agent (HA), which is a topological anchor point of the permanent MN's IP address. When the MN is away from home, a bi-directional tunnel is then established between the HA and the MN for redirecting packets from/to the current location of the MN. However, as a host-based protocol the MN needs to perform the mobility-related signaling by means of location update procedure. It is the main obstacle for the deployment of MIPv6 in the reality. For this reason, Proxy Mobile IPv6 (PMIPv6) [10] has been introduced as a network-based protocol to avoid the additional deployment in the MN so that the MN can be kept simple. In other words, mobility can be transparently provided to all legacy devices. While moving inside a PMIPv6 domain, the MN remains its IPv6 address. Thus from IP layer point of view, the MN is unaware of mobility. This is achieved by introducing the network entity called the Mobile Access Gateway (MAG), which performs the mobility-related signaling on behalf of the MNs attached to its access links. In PMIPv6, the LMA, similar to HA in MIPv6, is responsible for maintaining the MN's reachability state and forwarding traffic from/to the current location of the MN. MN's traffic is always encapsulated and tunneled between the LMA and the corresponding MAG.

MIPv6 and PMIPv6 are two typical examples of the current mobility management protocols which are highly centralized and hierarchical. In these protocols, both the mobile context and traffic encapsulation need to be maintained at the mobility anchor. The exponentially increasing number of mobile devices and their traffic demand make the centralized mobility management solutions encounter several problems as inefficient use of network resources, reliability, and scalability issues [4].

To tackle these issues, a new paradigm, the so-called Distributed Mobility Management (DMM), has been introduced from both IETF and the mobile industry. In DMM, instead of having a centralized mobility anchor, the mobility management function is distributed among the network entities at the network edge. Similar to the centralized approach, there are two main groups of solution: the host-based and the network-based [5]. Since the network-based approach does not require any additional deployment at the MN, in this paper, we focus on this approach. In a network-based DMM domain, the MN gets different prefixes when changing its point of attachment. In case of mobility, the MN's flows are anchored (if necessary) at the MAR in which the MN's prefix in use is allocated (called anchor MAR or aMAR). Hence, the packets can be redirected via the tunnel from the anchor to the current MAR (cMAR).

B. Centralized vs. Distributed Mobility Management

As DMM is presently a quite hot topic, a lot of research publications [5] have carried out the analysis on different DMM approaches, and compared them with the conventional mobility management protocols in terms of signaling cost, packet delivery cost, handover delay, packet loss, and end-to-end delay. The results from these analysis showed that DMM helps to save the resources in the network in some scenarios since the mobility support is enabled when it is necessary and the traffic is better distributed among the network entities, thus improving the scalability and reliability of the network. In DMM, the MN obtains a new prefix after each handover, while the old prefixes should be kept as long as their on-going flows are still alive. The on-going flows will be routed via the tunnel between the anchor MAR and the current one. As a result, in the situation where the MN is running the long-lasting flows and/or a node with high mobility features, the DMM deployment encounters several challenges such as complex address and tunnel management, high signaling cost, and long handover latency [6], [7], [8]. Therefore, DMM may not be a good candidate for vehicles or users on a moving vehicle.

In more details, in [6], the authors compared between the network-based DMM approach and PMIPv6 in terms of signaling cost and packet delivery cost. The authors concluded that DMM does not always outperform PMIPv6, especially when the cell residence time is short. In [7], the authors carried the analysis comparing different DMM approaches and the existing centralized ones including MIPv6 and PMIPv6. The results showed that the DMM protocols require an additional registration delay to the centralized approaches. This delay increases significantly when the MN moves far away from its anchor network. In [8], the authors, via an analysis regarding multicast deployment in DMM environment, argued that none of the approaches (PMIPv6 and DMM approaches) is always better than the others. The appropriate approach should be selected dynamically in order to meet a set of requirements in terms of service disruption time, end-to-end delay, packet delivery cost, and tunneling cost.

In [11], the authors introduced a dynamic tunneling scheme for network-based DMM based on PMIPv6 (DT-DMM). The idea is that the prefix allocation is responsible by the MAG instead of the LMA. Thus, the flow, after handover, is routed via the tunnel between the anchor MAG (aMAG) and the current one (cMAG) following the DMM concept. Based on a specific condition, the on-going flows are routed via the tunnel MAG-LMA instead of the tunnel aMAG-cMAG during the handover process. This means that the mobility anchor of the MN's prefix is changed from the current MAG to the LMA. However, it may lead to a significant service disruption (typically in seconds [12]) due to the routing convergence time which reflects the time to update the new anchor location of the prefix used by the on-going flows. This procedure is similar to that in case of changing LMA during the mobility session which is not recommended by the IETF [13], [14].

III. Description of the Solution

In order to keep the advantages of DMM while mitigating its drawback in comparison with a centralized management approach (PMIPv6), we proposed a solution, called H-DMM, that allows intelligently selecting a suitable mobility scheme: PMIPv6 or DMM in an appropriate manner.

The architecture and operations of the solution are illustrated in Fig. 1. The solution relies on the network-based DMM scheme proposed in [15]. The base entity, Mobility Access Router (MAR), basically encompasses the functionality of a plain access router, an MAG, and an LMA. The central mobility database (CMD) is extended to the Central Mobility Anchor (CMA) which also plays the role of an LMA in PMIPv6. When a mobile node (MN1) attaches to an MAR's subnet, it obtains a prefix from this MAR (called MAR-prefix) according to the normal DMM behavior. Also, it can get another prefix from the CMA (called LMA-prefix) as in a PMIPv6 domain. Every time the MN performs a handover, it keeps the LMA-prefix while obtaining a new one from the new MAR. The MN then can select either the prefix allocated at the current MAR or the one from the CMA to start a new communication with a corresponding node (CN). In addition, as a normal DMM behavior, the MN can continue receiving the traffic destined to its previous prefixes. It is noted that the details of how to select the appropriate prefix according to a set of metrics (e.g., type of flow and number of active prefixes) is out of scope of this document. At this stage, the MN should use by default the current MAR-prefix to start a new flow as a normal DMM behavior. However, if the number of active prefixes is greater than a threshold (N_0), then the LMA-prefix should be selected.

A. Initial Registration

The signaling flow when a mobile node (MN1) enters the domain (e.g., attaches to MAR1) is illustrated in Fig. 2. After detecting the presence of an MN by means of a Router Solicitation (RS) which includes the MN's identifier (MN-ID), MAR1 allocates a prefix for the MN (e.g., mpref1::/64). MAR1 then sends a Proxy Binding Update (PBU) message including the MN-ID and mpref1::/64 to the CMA for the

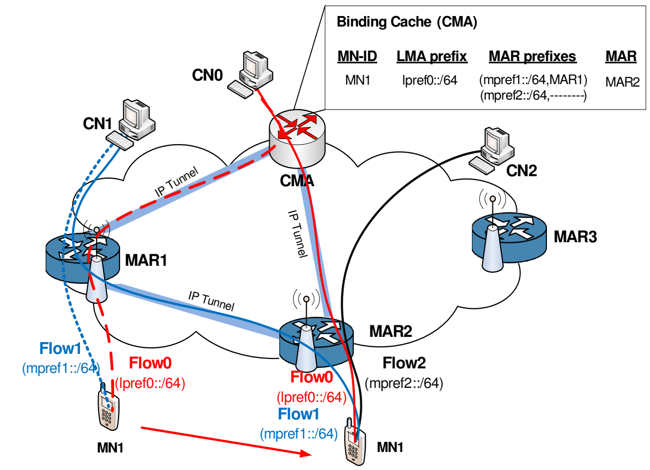

Fig. 1: Architecture and operations of H-DMM.

new prefix registration. In addition, the CMA, acting as a normal LMA, allocates an LMA-prefix (lpref0::/64) for this MN. As the MN enters the domain for the first time, the CMA creates a Binding Cache Entry (BCE) for it. In our solution, the BCE needs to be extended with the information of the LMA-prefix. Thus, the entry consists of the MN-ID, the LMA-prefix, the list of MAR's prefixes and the associated MARs, as well as the address of the current MAR (see Fig. 1). After that, the CMA replies by a Proxy Binding Acknowledgement (PBA) including the MN-ID, lpref0::/64 and mpref1::/64. A bi-directional tunnel is established between the CMA and MAR1 as similarly to that in PMIPv6. MAR1 then unicasts a Router Advertisement (RA) to the MN including the prefixes allocated (lpref0::/64 and mpref1::/64). Based on these prefixes, the MN can configure two IPv6 addresses (lpref0::MN1/64 and mpref1::MN1/64) which can be activated at the same time. The MN then can select one of the two addresses to start a new flow e.g., flow0 using lpref0::/64 and flow1 using mpref1::/64. Note that flow0 is routed via the mobility tunnel between CMA and MAR1 while flow1 is delivered using the normal IP routing without any tunneling mechanism.

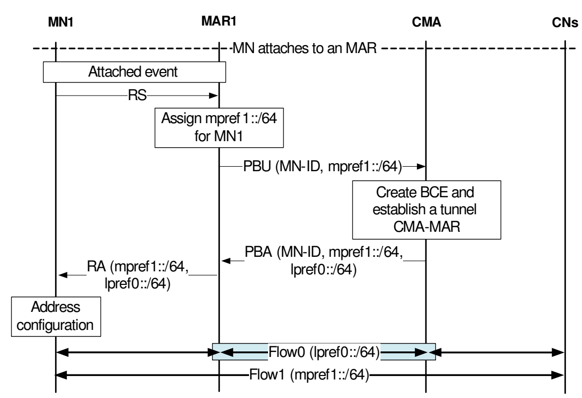

Fig. 2: Signaling flow for initial registration operations.

B. Handover Operation

Fig. 3 describes the signaling flow when the MN performs a handover from MAR1 to MAR2. As a normal DMM behavior, after detecting a new attachment, MAR2 assigns a new prefix (mpref2::/64) for the MN. It then informs the CMA with a PBU

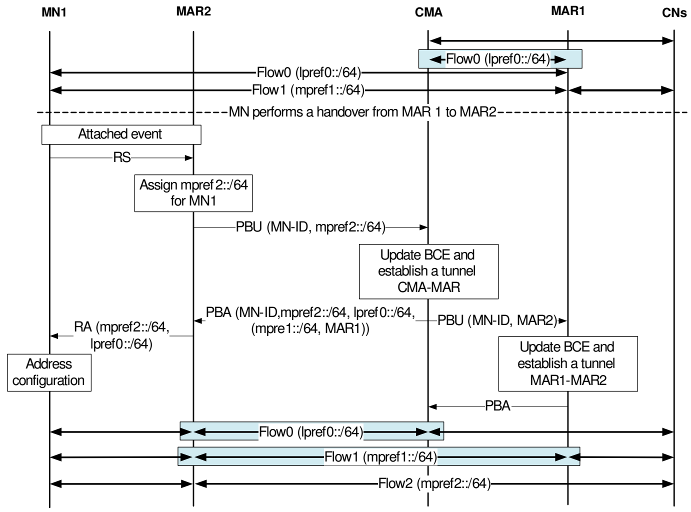

Fig. 3: Signaling flow for handover management operations.

as similar to the initial registration process. After receiving the PBU, the CMA updates the corresponding BCE of the MN with its current location and the new prefix allocated (see Fig. 1). The CMA then replies by a PBA including the new MAR-prefix (mpref2::/64), the LMA-prefix (lpref0::/64), a list of previous prefixes, and the associated MARs (in this case, mpref1::/64 and MAR1). In parallel, the CMA notifies the MN's previous MARs (MAR1) about the current location of the MN by means of a PBU message. MAR1 then updates its BCE and sets up its end-point for the bi-directional tunnel with MAR2 for the on-going flows which are initiated using mpref1::/64 (e.g., flow1). Similarly, MAR2 updates its binding and sets up its end-point for the tunnel with MAR1 (for flow1) and with the CMA (for flow0). It then sends a RA to the MN including the new MAR-prefix (mpref2::/64) and the LMA-prefix (lpref0::/64). As a result, the MN keeps using the LMA-address (lpref0::MN1/64) while configuring a new address based on the new MAR-prefix (mpref2::MN1/64). In more details, the status of mpref1::MN1/64 becomes "Deprecated" while the status of mpref2::MN1/64 and lpref0::MN/64 are "Preferred". Consequently, the MN can start the new flow either using mpref2::MN1/64 (e.g., flow2) or lpref0::MN1/64 as the source address. Regarding the on-going flows, the traffic for flow0 is now routed via the new route CN-CMA-MAR2-MN while the traffic for flow1 is routed from MAR1 via the tunnel MAR1-MAR2 (CN-MAR1-MAR2-MN).

IV. Performance Analysis

This section presents the performance analysis of the proposed solution in comparison with PMIPv6 and DMM (network-based DMM) regarding different metrics: signaling cost, packet delivery cost, handover latency, and end-to-end delay. The details of each protocol is provided in [10], [15].

A. System Models

1) Reference Model: The hop-count distances between the entities for performance analysis are defined as follows: h_{mc} is the average number of hops between the MAR and the CMA; between the MAG and the LMA as well. h_{ac} is the average number of hops between the MAR where the prefix is allocated and the current MAR. h_{cl} is the average number of hops between the CN and the CMA/LMA/MAR. Note that the number of hops between the MN and its MAR/MAG is assumed to be one (wireless link).

As described in Fig. 3, various messages are used in our analysis. The following message sizes in bytes thus are considered: L_{RS} is the size of the RS, L_{RA} - the size of the RA, L_{PBU} - the size of the PBU, L_{PBA} - the size of the PBA, L_P - the size of the packet, L_T - the size of the tunneling header, L_{Addr}/L_{Pref} - the size of the address/prefix options.

Let $\overline{N}_p$ denote the average number of active prefixes (excluding the LMA-prefix). According to [6], $\overline{N}_p$ is calculated as:

$$\overline{N}_p = 1 + \frac{\mu}{\delta}, \tag{1}$$

where μ is the MAR/MAG subnet border crossing rate and $1/\delta$ is the mean value of the active prefix lifetime while the MN is visiting a foreign network. In the context of this document, the low value of $\overline{N}_p$ represents a low mobility and/or a short-lived flow scenario. The higher value of $\overline{N}_p$ corresponds to a high mobility and a long-lived flow scenario. By allowing the MN to select the LMA-prefix or MAR-prefix, our solution keeps the value of $\overline{N}_p$ always lower than a threshold (N_0).

2) Delay Model: We adopt the packet transmission delay model in [16] in which the packet transmission consists of the transmission time and the propagation time. Thus, the transmission delay of a wired link can be calculated as:

$$d_{wd}(l,h) = h(\frac{l}{BW_{wd}} + D_{wd}), \tag{2}$$

where h is the hop-count distances between two nodes, l is the length of the packet, BW_{wd} is the bandwidth of wired link and D_{wd} is the wired link latency.

Unlike the wired transmission which can be considered as reliable, the wireless link is unreliable. The wireless transmission delay is therefore calculated as:

$$d_{wl}(l) = \frac{1}{1-q}(\frac{l}{BW_{wl}} + D_{wl}), \tag{3}$$

where q is the probability of wireless link failure, BW_{wl} is the bandwidth of wireless link and D_{wl} is the wireless link latency.

B. Performance Metrics

1) Signaling Cost: The signaling cost ($SC_{(.)}$) is the signaling overhead for updating the location ($C^u_{(.)}$) as well as for refreshing the bindings ($C^r_{(.)}$) for the MN. It is defined as the total delivery cost of all signaling messages. According to [17], the message delivery cost is calculated as the product of the message size, the hop distance and the unit transmission cost in a wired/wireless link (α for the wired and β for the wireless link). $SC_{(.)}$ is therefore expressed as:

$$SC_{(.)} = \mu\left(C^u_{(.)} + C^r_{(.)}\right). \tag{4}$$

As can be seen in Fig. 3, C^u_{H-DMM} can be given by:

$$\begin{aligned} C^u_{H-DMM} = \beta\left(L_{RS} + L_{RA} + L_{Pref}\right) + \alpha h_{mc}\{L_{PBU} + 3L_{Pref} \\ + L_{PBA} + N_m\left(L_{Pref} + L_{Addr}\right)\} \\ + \alpha N_m h_{mc}\left(L_{PBU} + L_{PBA} + L_{Addr}\right), \end{aligned} \tag{5}$$

where $N_m = min\left(\overline{N}_p, N_0\right)$.

Similarly, C^u_{DMM} and C^u_{PMIP} can be calculated as:

$$C^u_{DMM} = \beta\left(L_{RS} + L_{RA}\right) + \alpha h_{mc}\{L_{PBU} + L_{Pref} + L_{PBA} + \overline{N}_p\left(L_{Pref} + L_{Addr}\right)\} + \alpha\overline{N}_p h_{mc}\left(L_{PBU} + L_{PBA} + L_{Addr}\right), \quad (6)$$

$$C^u_{PMIP} = \beta\left(L_{RS} + L_{RA}\right) + \alpha h_{mc}\left(L_{PBU} + L_{PBA} + L_{Pref}\right). \quad (7)$$

Even the MN is remained at the same subnet, the signaling for refreshing the bindings is sent periodically when the binding timer expires. For a sake of simplicity, we suppose that the binding cache entry lifetime (T_{BCE}) is identical in case of PMIPv6, DMM, and H-DMM (for both MAR-prefix and LMA-prefix). Thus, the refreshing procedure is executed on average $R_r = \lfloor 1/(\mu T_{BCE})\rfloor$ times when the MN attaches to an MAR/MAG. In DMM and H-DMM, the current MAR only needs to exchange PBU/PBA with the CMA (without following by a PBU/PBA exchanged between CMA and the previous MARs). As a result, $C^r_{(.)}$ is given as follows:

$$C^r_{H-DMM} = R_r\alpha h_{mc}\{L_{PBU} + \left(N_m + 1\right)L_{Pref} + L_{PBA}\}, \quad (8)$$

$$C^r_{DMM} = R_r\alpha h_{mc}\left(L_{PBU} + \overline{N}_p L_{Pref} + L_{PBA}\right), \quad (9)$$

$$C^r_{PMIP} = R_r\alpha h_{mc}\left(L_{PBU} + L_{PBA}\right). \quad (10)$$

2) Packet Delivery Cost: The packet delivery cost ($PC_{(.)}$) represents the accumulative cost to deliver the packets between the MN and a CN per unit of time. It is proportional to the distance between the MN and the CN, the size of data packets and the number of packets transmitted. Let λ_{pkt} denote the packet transfer rate. In case of DMM, $\lambda_{pkt} = \lambda_p\overline{N}_p$ where λ_p is the packet rate per active prefix. In PMIPv6, since only one prefix is active, the total packet rate is λ_{pkt}. In H-DMM, the packet rate for the LMA-prefix is $\lambda_{lp} = \left(\overline{N}_p - N_m\right)\lambda_p$. The packet is routed via the route MN-MAG-LMA-CN, MN-cMAR-CN, and MN-cMAR-aMAR-CN in case of PMIPv6, DMM (new flow), and DMM (old flow), respectively. Note that in DMM, the tunneled packet (cMAR-aMAR) belongs to $\overline{N}_p - 1$ prefixes (except the prefix allocated in the current MAR). Thus, $PC_{(.)}$ is calculated as:

$$PC_{DMM} = \lambda_{pkt}\left(\beta + \alpha h_{cl}\right)L_P + \lambda_p\left(\overline{N}_p - 1\right)\alpha h_{ac}\left(L_P + L_T\right), \quad (11)$$

$$PC_{PMIP} = \lambda_{pkt}\{\beta L_P + \alpha h_{mc}\left(L_P + L_T\right) + \alpha h_{cl}L_P\}. \quad (12)$$

As in H-DMM, the MN can select the LMA-prefix or the one at the current MAR to start a new flow, PC_{H-DMM} can be expressed as:

$$PC_{H-DMM} = \frac{N_m}{\overline{N}_p}PC^m_{H-DMM} + \frac{\left(\overline{N}_p - N_m\right)}{\overline{N}_p}PC_{PMIP}, \quad (13)$$

where

$$PC^m_{H-DMM} = \lambda_{pkt}\left(\beta + \alpha h_{cl}\right)L_P + \lambda_p\left(N_m - 1\right)\alpha h_{ac}\left(L_P + L_T\right). \quad (14)$$

3) Handover Latency: Since in a DMM environment, an MN after handover can use the new prefix to start a new flow while the old prefixes for the on-going flows, we consider the handover latency and the service disruption time. The reason is that the handover latency ($T^{HO}_{(.)}$) represents the time needed for the MN after handover to start a new flow, while the service disruption time ($T^{SD}_{(.)}$) is defined as a period from the moment the MN leaves the previous MAR/MAG until the moment the MN continues receiving the packet of the on-going flows from the CN. The handover latency is given by:

$$T^{HO}_{H-DMM} = t_{L2} + d_{wl}(L_{RS}) + d_{wd}(L_{PBU} + L_{Pref}, h_{mc}) + d_{wd}(L_{PBA} + 2L_{Pref} + N_m\left(L_{Pref} + L_{Addr}\right), h_{mc}) + d_{wl}(L_{RA} + L_{Pref}), \quad (15)$$

$$T^{HO}_{DMM} = t_{L2} + d_{wl}(L_{RS}) + d_{wl}(L_{RA}) + d_{wd}(L_{PBU}, h_{mc}) + d_{wd}(L_{PBA} + L_{Pref} + \overline{N}_p\left(L_{Pref} + L_{Addr}\right), h_{mc}), \quad (16)$$

$$T^{HO}_{PMIP} = t_{L2} + d_{wl}(L_{RS}) + d_{wd}(L_{PBU}, h_{mc}) + d_{wd}(L_{PBA} + L_{Pref}, h_{mc}) + d_{wl}(L_{RA}). \quad (17)$$

Regarding the service disruption time, it is calculated as:

$$T^{SD}_{H-DMM} = \frac{N_m}{\overline{N}_p}T^{SD}_{H-DMM-D} + \frac{\left(\overline{N}_p - N_m\right)}{\overline{N}_p}T^{SD}_{H-DMM-P}, \quad (18)$$

where

$$T^{SD}_{H-DMM-D} = t_{L2} + d_{wl}(L_{RS}) + d_{wd}(L_{PBU} + L_{Pref}, h_{mc}) + max\{d_{wd}(L_{PBU} + L_{Pref}, h_{mc}) + d_{wd}(L_{PBA}, h_{mc}), d_{wd}(L_{PBA} + 2L_{Pref} + N_m\left(L_{Pref} + L_{Addr}\right), h_{mc}) + d_{wl}(L_{RA} + L_{Pref})\} + d_{wd}(L_P + L_T, h_{ac}) + d_{wl}(L_P), \quad (19)$$

$$T^{SD}_{H-DMM-P} = T^{HO}_{H-DMM} + d_{wd}(L_P + L_T, h_{mc}) + d_{wl}(L_P), \quad (20)$$

$$T^{SD}_{DMM} = t_{L2} + d_{wl}(L_{RS}) + d_{wd}(L_{PBU}, h_{mc}) + max\{d_{wd}(L_{PBU} + L_{Pref}, h_{mc}) + d_{wd}(L_{PBA}, h_{mc}), d_{wd}(L_{PBA} + L_{Pref} + \overline{N}_p\left(L_{Pref} + L_{Addr}\right), h_{mc}) + d_{wl}(L_{RA} + L_{Pref})\} + d_{wd}(L_P + L_T, h_{ac}) + d_{wl}(L_P), \quad (21)$$

$$T^{SD}_{PMIP} = T^{HO}_{PMIP} + d_{wd}(L_P + L_T, h_{mc}) + d_{wl}(L_P). \quad (22)$$

4) End-to-End Delay: End-to-end delay ($E2E_{(.)}$) is the packet transmission delay from the MN to the CN. In DMM, the new traffic is routed directly from the current MAR (MN-cMAR-CN) while the old traffic is routed via the anchor MAR (MN-cMAR-aMAR-CN). Thus, $E2E_{(.)}$ is defined as:

$$E2E_{DMM} = \frac{1}{\overline{N}_p}E2E^{new}_{DMM} + \frac{\overline{N}_p - 1}{\overline{N}_p}E2E^{old}_{DMM}, \quad (23)$$

where

$$E2E^{new}_{DMM} = d_{wl}(L_P) + d_{wd}(L_P, h_{cl}), \quad (24)$$

$$E2E^{old}_{DMM} = d_{wl}(L_P) + d_{wd}(L_P + L_T, h_{ac}) + d_{wd}(L_P, h_{cl}), \quad (25)$$

$$E2E_{PMIP} = d_{wl}(L_P) + d_{wd}(L_P + L_T, h_{mc}) + d_{wd}(L_P, h_{cl}). \quad (26)$$

Similar to the packet delivery cost, $E2E_{(.)}$ in case of H-DMM is given by:

$$E2E_{H-DMM} = \frac{N_m}{\overline{N}_p}E2E^D_{H-DMM} + \frac{\left(\overline{N}_p - N_m\right)}{\overline{N}_p}E2E^P_{H-DMM}, \quad (27)$$

where

$$E2E^D_{H-DMM} = \frac{1}{N_m}E2E^{new}_{DMM} + \frac{N_m - 1}{N_m}E2E^{old}_{DMM}, \quad (28)$$

$$E2E^P_{H-DMM} = E2E_{PMIP}. \quad (29)$$

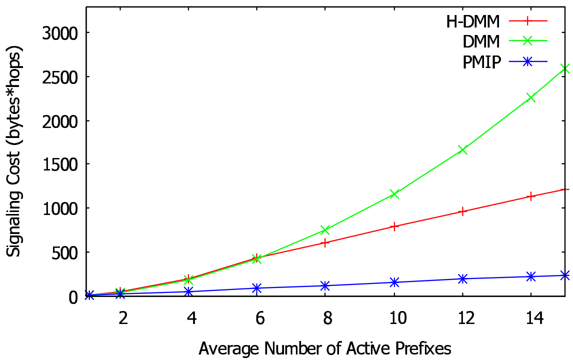

Fig. 4: Signaling cost as a function of $\overline{N}_p$.

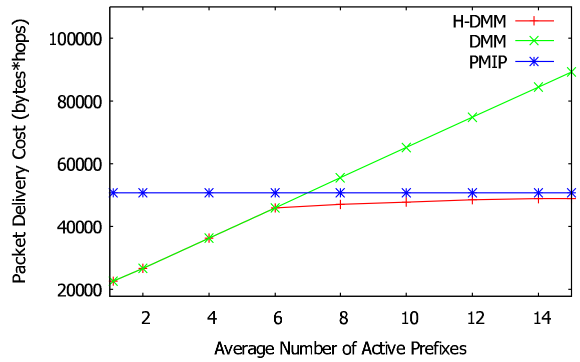

Fig. 5: Packet delivery cost as a function of $\overline{N}_p$.

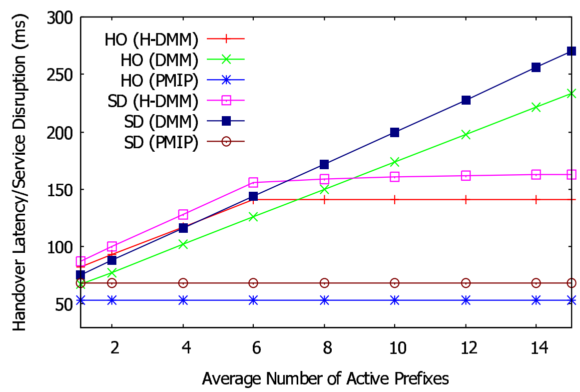

Fig. 6: Handover latency and service disruption time as a function of $\overline{N}_p$.

V. Numerical Results

In this paper, we consider the case where the MN always moves from MAR/MAG to MAR/MAG as if they were linearly deployed (the user is moving further away from the first attached MAR/MAG and never attaches back to a previously visited MAR/MAG). This assumption, while has no impact on PMIPv6, represents the worst-case scenario for DMM. Hence, we have [5]:

$$h_{ac} = \overline{N}_p h_{mm}, \tag{30}$$

where h_{mm} is the average hop distance between two adjacent MARs/MAGs. The default parameter values for the analysis are introduced in Table I in which some of them are taken from [5], [6].

TABLE I: Parameters for the performance analysis.

Parameter	Value	Parameter	Value	Parameter	Value
α	1	β	5	T_{BCE}	300 s
h_{mm}	2 hops	h_{mc}	12 hops	h_{cl}	6 hops
L_{RS}	52 bytes	L_{RA}	80 bytes	L_{PBU}	84 bytes
L_{PBA}	92 bytes	L_T	40 bytes	L_P	200 bytes
L_{Addr}	20 bytes	L_{Pref}	20 bytes	λ_{pkt}	10
$1/\delta$	300s	t_{L2}	29.49 ms	BW_{wd}	100 Mbps
BW_{wd}	11 Mbps	D_{wd}	0.5 ms	D_{wl}	2 ms
q	0.35	N_0	6		

Fig. 4 shows the signaling cost as a function of the average number of active prefixes ($\overline{N}_p$). We can observe that the signaling cost is increased as $\overline{N}_p$ increases. As mentioned earlier, the low value of $\overline{N}_p$ represents a low mobility and/or a short-lived flow scenario, while the higher value corresponds to a high mobility and/or a long-lived flow scenario. In general, PMIPv6 outperforms the others in terms of signaling cost. When $\overline{N}_p$ is small, the difference between the signaling cost in case of PMIPv6 and DMM is small. When $\overline{N}_p$ increases, the difference is getting bigger. Our solution (H-DMM) helps to reduce the difference, thus, mitigating the drawback of DMM (in a high mobility and/or a long-lived flow scenario).

Fig. 5 illustrates the packet delivery cost when $\overline{N}_p$ is varying. The packet delivery cost in DMM and H-DMM is increased when $\overline{N}_p$ increases while it is fixed in PMIPv6 because the distance between the LMA and the MAGs is supposed to be fixed. It appears clearly that when $\overline{N}_p$ is small, DMM outperforms PMIPv6. The reason is that the packet is either routed directly via the current MAR (without tunneling) or via the tunnel cMAR-aMAR which is supposed not too long. When the distance between the cMAR and the aMAR is considerably large (with a high value of $\overline{N}_p$), PMIPv6 is much better than DMM. In both cases, the end-to-end delay in H-DMM is always lower than that in PMIPv6.

Fig. 6 shows the handover latency and the service disruption time as a function of $\overline{N}_p$. They are kept constant in PMIPv6. On the contrary, in DMM and H-DMM they are generally increased when $\overline{N}_p$ increases. DMM causes an additional delay compared to PMIPv6 due to the increasing number of active prefixes associated to the MN as well as the additional time to update the location of the MN at the anchor MAR. As a result, the additional delay is significantly increased as $\overline{N}_p$ increases. As can be seen in this figure, H-DMM helps to keep the additional delay below a certain threshold at a small cost of delay adding to the DMM approach (when $\overline{N}_p$ is small).

Fig. 7 shows the end-to-end delay as a function of $\overline{N}_p$. It is clear that when $\overline{N}_p$ is small, DMM and H-DMM solutions are better than PMIPv6. The reason is that the packets do not need to pass the central entity (LMA) which is quite far away from the current MAR/MAG. When $\overline{N}_p$ increases, the end-to-end delay will be increased in DMM and H-DMM while it is kept the same in PMIPv6. It means the delay in DMM and H-DMM is higher than that in PMIPv6 (when $\overline{N}_p > 7$). The difference between them is getting large when $\overline{N}_p$ increases. Again, our solution mitigates the difference by keeping the end-to-end a bit higher than that in PMIPv6.

Now we investigate the relation between the number of active prefixes ($\overline{N}_p$), the velocity (v) and the active prefix lifetime ($1/\delta$). It is assumed that the subnet residence time

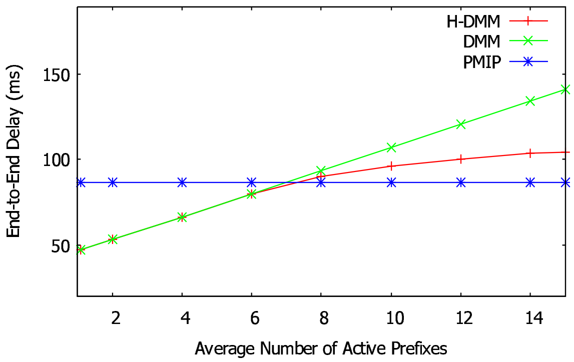

Fig. 7: End-to-End delay as a function of $\overline{N}_p$.

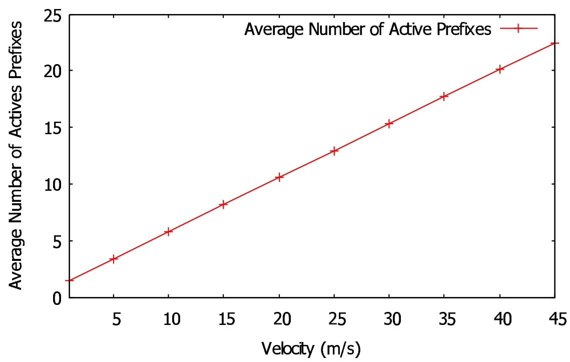

Fig. 8: $\overline{N}_p$ as a function of velocity (υ).

(MAR/MAG subnet) is a random variable which follows an exponential distribution with mean value $1/\mu$ and the MAR/MAG coverage area is circular with radius R. According to [18], the subnet border crossing rate μ is calculated as:

$$\mu = \frac{2\upsilon}{\pi R}, \tag{31}$$

where υ is the average velocity of the MN.

Fig. 8 depicts $\overline{N}_p$ as a function of the velocity when the subnet radius R and $1/\delta$ are fixed to the value of 400m and 300s, respectively. As the velocity increases, $\overline{N}_p$ is increased. According to equation (1), we can obtain a similar curve as in Fig. 8 if the value of υ is fixed while the mean value of the active prefix lifetime ($1/\delta$) is varying. Thus, the low value of $\overline{N}_p$ corresponds to a low mobility and/or a low-lived flow scenario, while the high value of $\overline{N}_p$ represents a high mobility and/or a long-lived flow scenario.

VI. Conclusion

The increasing penetration of the mobile devices is generating a huge number of data traffic over mobile networks. In this context, the concept of DMM aims at overcoming the limitations of the current mobility management protocol created by raising the mobile usage. Although DMM generally helps to save the resources in the network in some scenarios, it does not seem suitable for users with high-mobility features (such as users on board vehicles). In this vein, this paper introduced a hybrid centralized-distributed mobility management (H-DMM) architecture for supporting such highly mobile users. This solution allows the MN to select the appropriate prefix to start a new flow among the prefixes allocated at the current MAR and at the central entity (CMA). By doing so, the number of active prefixes is kept below a threshold value even in a high mobility and/or a long-lived flow scenario. The numerical results showed that H-DMM inherits the advantages of DMM while limiting its drawbacks in comparison with PMIPv6.

In the next step, the network mobility (NEMO) [19] will be considered in our architecture since NEMO is to provide Internet access for a group of users in a moving vehicle in an effective-manner. In addition, to achieve the realistic results, experiments will be conducted based on an existing near-to-real testbed [20].

References

[1] Cisco, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2013-2018", Feb 2014.

[2] Cisco White Paper, "Cisco VNI Service Adoption Forecast, 2013–2018", 2013.

[3] C.B. Sankaran, "Data Offloading Techniques in 3GPP Rel-10 Networks: A Tutorial", IEEE Commun. Mag., vol. 50, no. 6, pp. 46-53, Jun 2012.

[4] H.A. Chan, D. Liu, P. Seite, H. Yokota, and J. Korhonen, "Requirements for Distributed Mobility Management", RFC 7333, Aug 2014.

[5] T.-T. Nguyen, "Optimization of Mobility Mechanisms for IP-based Multicast Flows", Ph.D. thesis, Mobile Communications Dept., EURECOM, Sophia-Antipolis, France, May 2014.

[6] F. Giust, C.J. Bernardos, and A. De La Oliva, "Analytic Evaluation and Experimental Validation of a Network-based IPv6 Distributed Mobility Management Solution", IEEE Trans. Mobi. Comp., no. 99, Feb 2014.

[7] J.-H. Lee, J.-M. Bonnin, P. Seite, and H.A. Chan, "Distributed IP Mobility Management from the Perspective of the IETF: Motivations, Requirements, Approaches, Comparison, and Challenges", IEEE Wireless Communications, vol. 20 , no. 5, pp. 159-168, Oct 2013.

[8] T.-T. Nguyen and C. Bonnet, "On the Efficiency of Dynamic Multicast Mobility Anchor Selection in DMM: Use Cases and Analysis", in Proc. IEEE ICC, Sydney, Australia, Jun 2014.

[9] C. Perkins, D. Johnson, and J. Arkko, "Mobility Support in IPv6", RFC 6275, Jul 2011.

[10] S. Gundavelli, K. Leung, V. Devarapalli, K. Chowdhury, and B. Patil, "Proxy Mobile IPv6", RFC 5213, Aug 2008.

[11] J.-H. Lee, Z. Yan, J.-M. Bonnin, and X. Lagrange, "Dynamic tunneling for network-based distributed mobility management coexisting with PMIPv6", in Proc. PIMRC, London, UK, Sept 2013.

[12] T.-T. Nguyen and C. Bonnet, "Considerations of IP multicast for load balancing in Proxy Mobile IPv6 networks", Journal of Computer Networks, Elsevier, vol. 72, pp. 113-126, Oct 2014.

[13] J. Korhonen, S. Gundavelli, H. Yokota, and X. Cui, "Runtime LMA Assignment Support for Proxy Mobile IPv6", RFC 6463, Feb 2012.

[14] J. Korhonen and V. Devarapalli, "Local Mobility Anchor (LMA) Discovery for Proxy Mobile IPv6", RFC 6097, Feb 2011.

[15] C.J. Bernardos, A. de la Oliva, and F. Giust, "A PMIPv6-based Solution for DMM", IETF Draft (work-in-progress), Jan 2014.

[16] J. McNair, I. Akyildiz, and M. Bender, "An Inter-system Handoff Technique for the IMT-2000 System", in Proc. INFOCOM, San Diego, CA, USA, vol. 1, pp. 208–216, Mar 2000.

[17] J.-H. Lee and T.-M. Chung, "How Much do We Gain by Introducing Route Optimization in Proxy Mobile IPv6 Networks?", Annals of Telecommunications, vol. 65, no. 5-6, pp. 233-246, Jun 2010.

[18] C. Makaya, and P. Samuel, "An Analytical Framework for Performance Evaluation of IPv6-based Mobility Management Protocols", IEEE Trans. Wireless Communications, vol. 7, no. 3, pp. 972-983, Mar 2008.

[19] V. Devarapalli, R. Wakikawa, A. Petrescu, and P. Thubert, "Network Mobility (NEMO) Basic Support Protocol", RFC 3963, Jan 2005.

[20] T-T. Nguyen and C. Bonnet, "Performance Optimization of Multicast Content Delivery in a Mobile Environment based on PMIPv6", in Proc. IEEE WCNC, Apr 2013.