Computer Networks 94 (2016) 129–144

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

DMMS: A flexible architecture for multicast listener support in a distributed mobility management environment

Tien-Thinh Nguyen*, Christian Bonnet

Department of Mobile Communications, EURECOM, 450 Route des Chappes, Biot 06410, France

ARTICLE INFO

Article history:
Received 4 February 2015
Revised 26 May 2015
Accepted 11 November 2015
Available online 22 November 2015

Keywords:
IP multicast
IP mobility
Network-based mobility management
Distributed mobility management
Multicast listener mobility

ABSTRACT

The mobile network operators are being challenged by the explosion of mobile data traffic in terms of network performance and generated revenue. On one hand, efficient mobility management plays a crucial role to support mobile users, however, the current mobility protocols have several major limitations from their centralized and hierarchical nature (e.g., sub-optimal routing, scalability and reliability issues). Distributed mobility management (DMM) is a new, very promising trend to overcome these limitations by flattening the network architecture and dynamically providing the mobility support. Based on the fact that the mobile Internet traffic will be dominated by the mobile video, the scalability and bandwidth efficiency from multicast routing makes the IP multicast play a crucial role. However, one of the main challenges for multicast support is the mobility of a multicast node, leading to several issues for both multicast service and network operator such as long service disruption, high end-to-end delay, non-optimal routing and traffic replication. Driven from the fact that different multicast flows have very different characteristics and each network operator has different policies for multicast support, we propose a dynamic multicast support scheme (DMMS), taking into account both the user and network operator point of view. DMMS allows to dynamically provide the appropriate multicast support mode based on a set of contexts such as the service's characteristics, mobility of the node and network context to adapt to the service's requirements as well as operator's policies.

1. Introduction

Technology has now driven us in the mobile era in which mobile traffic generated by mobile phones has exceeded that from mobile PCs, tablets and mobile routers. Estimates say that global mobile broadband subscriptions grew by around 30% year-on-year and reached 2.5 billion in the third quarter of 2014 [1]. Additionally, this trend does not show any sign of slowing down. Global mobile broadband subscriptions are predicted to reach 8.4 billion by 2020, accounting for more than 90% of all mobile subscriptions. The increasing number of subscriptions has been driven by a variety of reasons such as the increasing number of mobile devices which become more and more powerful and intelligent (especially, in the low- and mid-range price); the enhancement of wireless access technology in terms of coverage, speed and quality; as well as the explosion of mobile applications [1–3]. As a result, mobile data traffic will increase around 8 times between 2014 and 2020, to reach 17 exabytes per month by 2020. However, despite the increasing volume of traffic, mobile data revenue per user is falling down [4]. On the other hand, the current mobile networks are evolving toward all-IP architecture. In this context, the mobility nature of the mobile nodes (MNs) makes IP mobility management play a crucial role in the mobile network. In fact, today's mobility management protocols (e.g., Mobile IPv6 (MIPv6) and Proxy

* Corresponding author. Tel.: (+33) 4 93 00 82 15.
E-mail addresses: tien-thinh.nguyen@eurecom.fr (T.-T. Nguyen), christian.bonnet@eurecom.fr (C. Bonnet).

http://dx.doi.org/10.1016/j.comnet.2015.11.015
1389-1286/

Mobile IPv6 (PMIPv6)) have several major limitations from their centralized and hierarchical nature. Centralizing both the control and data plane functions at the central mobility anchor introduces scalability and reliability issues [5,6]. Also, it leads to sub-optimal paths between the MNs and their corresponding nodes (CNs).

The mobile network operators are being challenged by the explosion of mobile data traffic (especially the video 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. Faced with these challenges, the operators are seeking for innovative solutions to improve their network performance and efficiency, as well as to reduce the costs expended on network operation and maintenance. One possible solution is to increase the radio capacity of mobile broadband by deploying new wireless technologies such as Evolved High Speed Packet Access (HSPA+), Long Term Evolution (LTE) and LTE-Advanced (LTE-A). However, the radio spectrum for operators is both limited and expensive. Consequently, 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 topologies are currently evolving toward a flat architecture. The Third Generation Partnership Project (3GPP)[1] also proposes such traffic offloading techniques as Local IP Access/Selected IP Traffic Offload (LIPA/SIPTO) and IP Flow Mobility (IFOM) [7]. 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, and reliability issues (the anchor represents a bottleneck and single point of failure) [5,6].

On the other hand, IP multicast is regarded as a key technology for improving traffic delivery efficiency when data is sent to several receivers at the same time, for example, in such areas as multimedia distribution, gaming and software update. The reason is that IP multicast can provide significant advantages compared to unicast communication regarding overall resources consumption (e.g., bandwidth, server load and network load) and deployment cost to deliver the traffic [8–10]. However, after more than a decade of researches and development efforts, IP multicast deployment, in general, has been sluggish on the global Internet, mainly due to the practical, security and business concerns [11,12]. However, the new business models, a huge traffic demand (especially multimedia traffic), the revenue per data reducing phenomenon in the mobile operator networks, as well as the advantages of new multicast model (source-specific multicast—SSM) bring again the strong interest of IP multicast from both academia and industry [13]. As a result, IP multicast is expected to play more important role in the future networks.

Altogether, to deal with a huge number of devices and traffic, while DMM is expected to be an effective solution in terms of IP mobility management, IP multicast can be considered as a valuable solution from service point of view. But one of the major challenges for multicast support is when mobility is considered. It is due to the fact that the multicast protocols were designed for fixed networks. As a result, the interaction between the IP multicast and IP mobility protocols raises several issues from both multicast service and network operator point of view such as service disruption (and packet loss), routing non-optimization, packet duplication, and waste of network resource [9,12].

However, a relatively limited work has been done considering IP multicast in a network-based DMM environment. At this stage, following the DMM concept, the multicast traffic is routed directly from the native multicast infrastructure via the current access router (mobility access router or MAR) for the new flow. For the flow after handover, the multicast traffic is tunneled from the MAR where the flow is initiated to the current one via the mobility tunnel between them, like unicast traffic [14,15]. Thus, the MAR where the flow is initiated plays the role of the multicast mobility anchor (MMA) which is a logical entity from where the MN receives the ongoing multicast flow while on the move; and is identical with the unicast mobility anchor. The multicast flow will be anchored at the initially assigned MMA during its lifetime. Therefore, even when the MN moves far away from its anchor, the multicast traffic still traverses the anchor. As a result, it not only causes several issues to the ongoing multicast flow such as service disruption and end-to-end delay, but also leads to the non-optimal routing and tunnel convergence problem. These problems become serious when considering the interruption- and delay-sensitive services. Besides, to address the tunnel convergence and non-optimal routing problem, the MAR can obtain the multicast traffic of the ongoing flow directly from its upstream MR. However, it may cause a significant service disruption due to the time needed to join the multicast delivery tree and deliver the traffic to the current MAR [16]. In addition, the distribution of mobility anchors will not help to balance the traffic between them [17]. On the other hand, the operator may not want to deploy the tunnel between the access routers, or in some situations, desires applying a specific policy for a particular multicast channel. Altogether, a multicast support mode may not be always good regarding different requirements. That is the reason why a flexible manner to support multicast mobility in DMM should be provided.

In this document we propose a solution, namely dynamic multicast mobility support—DMMS, which allows to provide a suitable multicast support mechanism in DMM from both user and network operator point of view. The multicast support scheme can be provided in a flexible manner based on the collected contexts to meet a set of requirements. From user perspective, it helps satisfy the requirements in terms of service disruption and delay, especially when considering real-time services. From network operator perspective, it provides a mechanism to better distribute the load among the network entities and save network resource by preventing the packet duplication and shortening the leave latency. This mechanism not only takes into account the multicast service context (e.g., interruption- and delay-sensitive services) but also the mobile node's mobility and the network context (such as the load of MARs and the multicast channel

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

policy), thus enabling a per-flow multicast support. DMMS performance was then validated not only by the numerical results, but also via a proof-of-concept with preliminary results.

The rest of this document is organized as follows. Section 2 introduces background information related to IP mobility management, the issues and different approaches for the multicast support in DMM. A detailed description of DMMS is given in Section 3, while Section 4 presents the performance analysis regarding different metrics as service disruption, end-to-end delay, signaling cost and packet loss. The numerical results are given in Section 5, while Section 6 shows the current status of the implementation and the preliminary results. Section 7 discusses the scenario in which the multicast router is deployed at the access router and the multicast source mobility support. Finally, Section 8 concludes this document.

2. Related work

2.1. IP mobility management

2.1.1. Centralized mobility management

Today's mobility management protocols rely on a central entity such as Home Agent (HA) in Mobile IPv6 (MIPv6) [18], and Local Mobility Anchor (LMA) in Proxy Mobile IPv6 (PMIPv6) [19] to maintain the mobile node's reachability when it is away from home. In MIPv6, it is done by using a tunneling mechanism between the HA and the MN for redirecting packets from/to the current location of the MN. However, as a host-based protocol, an additional component is required at the MN to perform the mobility-related signaling. The additional component is considered as the main obstacle for the deployment of MIPv6 in reality. For this reason, PMIPv6 has been introduced as a network-based protocol. PMIPv6 introduces a new entity called MAG to perform mobility related signaling on behalf of MNs. Accordingly, MN's traffic is always encapsulated and tunneled between the LMA and the corresponding MAG. Additionally, while moving inside a PMIPv6 domain, the MN remains its IPv6 address. Consequently, the MN can be kept simple and is unaware of mobility as well. In other words, mobility can be transparently provided to all legacy devices.

On the other hand, MIPv6 and PMIPv6 are two typical examples of the current mobility management protocols which have several major limitations from their centralized and hierarchical nature. Centralizing both the control and data plane functions at the central mobility anchor introduces scalability and reliability issues (the central entity represents a bottleneck and single point of failure) [5]. Centralized approach also suffers from triangle routing problem. Therefore, it affects network performance in terms of routing efficiency and end-to-end delay transmission [5].

To address the limitations of the current mobility management protocols, distributed mobility management (DMM) solutions have been proposed [5]. The key concept of DMM is that instead of having a centralized mobility anchor, the mobility anchors are distributed among network entities and placed as close as possible to the MN e.g., at the router edge of the access network. In other words, the MAR where a flow is initiated plays the role of the mobility anchor for this flow. DMM also offers dynamic mobility features (per prefix granularity). As DMM is presently an active topic, the performance evaluation of DMM has been extensively analyzed using different approaches and network metrics [6,14]. The results from these analysis showed that DMM helps to save the resources in the network 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. Among the proposals, the network-based solution proposed in [20] appears to be the most promising scheme.

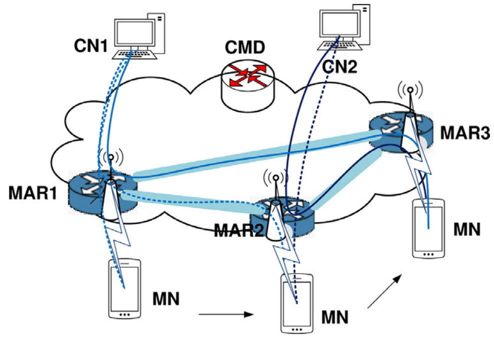

Fig. 1. DMM architecture.

2.1.2. Network-based distributed mobility management

This paper follows the concept of the network-based DMM [20] proposed by the IETF DMM Working Group. The central entity, namely Central Mobility Database (CMD), still exists but for the control plane only. The CMD stores the information of mobility sessions of all mobile nodes in the domain. The base entity, mobility access router (MAR), basically encompasses the functionality of a plain access router, an MAG, and an LMA. In a DMM domain, each MAR owns a pool of IPv6 prefix. An MN obtains different prefixes when it changes its attachment points (MARs). In case of mobility, the MN's flows are anchored (if necessary) at the MAR in which its prefix in use is allocated (called anchor MAR or aMAR). Hence, the packets can be redirected via the mobility tunnel from the anchor to the current MAR (cMAR).

Fig. 1 represents an example of how DMM works. Once the MN enters a DMM domain (attaches to MAR1), it configures an IPv6 address (pref1::MN1/64) based on the prefix allocated at the current MAR (pref1::/64) and can use its address to initiate a flow with a CN (say flow1). Flow1 is routed in a standard routing manner without any tunneling mechanism. After moving to MAR2, the MN can start a new flow (say flow2) using the new allocated prefix (pref2::/64). If flow1 is kept alive, it will be redirected via the mobility tunnel between MAR1 and MAR2. The tunnel establishment is done thanks to the coordination between CMD, MAR1, and MAR2. Similarly, when the MN moves to MAR3, flow1 and flow2 are anchored at MAR1 and MAR2, respectively. From flow1 point of view, MAR1, MAR2, and MAR3 are the anchor (aMAR), the previous (pMAR) and the current MAR (cMAR), respectively. For flow2, MAR2 is both the anchor and the previous MAR.

2.2. IP multicast mobility: from centralized to distributed mobility management

As the multicast protocols (including the group management and routing protocols) are originally designed for a fixed network, considering a multicast node in a mobile environment brings several challenges to both the multicast service and the network operator. The mobility of the multicast node has different impacts on the multicast service upon such factors as: (i) the role of the node in the multicast session (source or listener); (ii) the considered multicast model (any-source multicast (ASM) or source-specific multicast (SSM)); (iii) the multicast routing and multicast group management protocols in use; (iv) the mobility management protocol; and (v) the wireless access technology [14]. Accordingly, the mobility-related issues can be divided into 4 main groups as the general multicast problems (due to multicast protocols), the specific mobile listener problems, the specific mobile source problems and the deployment issues [9,12,21]. In the context of this document, we focus on the issues caused by the mobility of a multicast listener including long service disruption and high number of lost packets, packet duplication (or tunnel convergence problem), sub-optimal routing, high end-to-end delay, and waste of network resource [9,12].

2.2.1. Multicast listener mobility in a network-based DMM paradigm

Since DMM is still in the preliminary stage of standardization, there is no complete solution for multicast in place. Typically, all major aspects for multicast support in a network-based DMM environment are inherited from that in PMIPv6, while an additional complexity is added [16]. Compared to PMIPv6, DMM introduces also two basic scenarios to enable multicast regarding the multicast functionality deployed at the MAR i.e., multicast router (MR) and multicast listener discovery (MLD) proxy [15]. However, the operators may not want to support the multicast routing function at the MAR due to its implementation and operational costs. Therefore, this paper assumes that the MAR and MLD proxy are co-located (the scenario in which MAR acts as a multicast router is briefly discussed in Section 7). Also, only multicast listener mobility in the network-based DMM is further studied.

In network-based DMM, when a multicast flow is initiated, the traffic is routed directly from the multicast infrastructure via the current MAR. In case of handover, the traffic is routed from the anchor to the current MAR via the mobility tunnel between them, just like unicast traffic. The operations in details are illustrated in Fig. 2 and described as follows.

After detecting the presence of a new mobile node by means of a router solicitation (RS) message, the current MAR allocates a prefix for the MN (e.g., pref1::/64). According to the normal DMM behavior, a proxy binding update (PBU) including the allocated prefix is sent to the CMD for the prefix registration. Upon receiving the PBU, the CMD creates a binding cache entry (BCE) for this MN and replies by a proxy binding acknowledgement (PBA) message. The MAR then unicasts a router advertisement(RA) including the allocated prefix to the MN. In parallel, the MLD proxy instance at MAR adds the MN to its downstream interface, and configures its upstream interface toward its default upstream MR. The MN, after

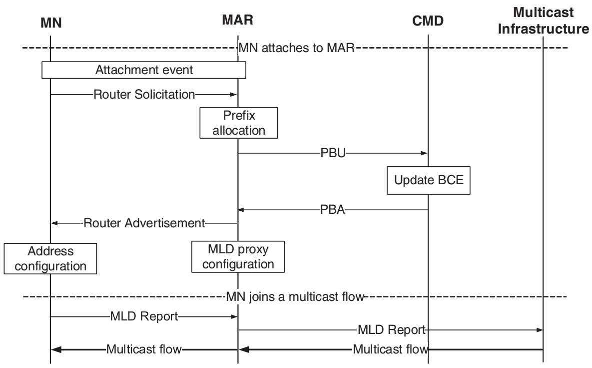

Fig. 2. Initial operations.

configuring its IPv6 address (pref1::MN1/64), can join a multicast flow via the current MAR by means of an MLD report. As a result, the current MAR receives the multicast packets from the native multicast infrastructure and forward them to the MN. It is noted that instead of using RS/RA message, another mechanism can be used to allow the MN obtain an IPv6 address e.g., by using Dynamic Host Configuration Protocol (DHCPv6). In this case, the DHCP request and DHCP reply messages will play the role of RS and RA, respectively.

In case of handover (see Fig. 3), following the standard DMM operations, the MLD proxy instance at the cMAR adds the MN to its downstream interface, and configures its upstream interface toward the aMAR. The cMAR, after obtaining the MN's subscription information by means of the normal MLD operation, sends an aggregated MLD report to the aMAR to join the ongoing flows. Finally, the multicast traffic is transmitted from aMAR to cMAR via the mobility tunnel between them and reaches the MN. This approach is similar to the base deployment for multicast listener support in PMIPv6 [22]. In this paper, it is called the default multicast support mode (DF) [15,23,24]. However, this mode does not address any multicast-related issues caused by the movement of listener such as long service disruption (and high number of packet losses), non-optimal routing, high end-to-end delay, and tunnel convergence problem.

For more details, when a listener moves to a new MAR (cMAR), it may cause a noticeable disruption for the ongoing flows due to multicast subscription acquisition time (based on the normal MLD query/report mechanism). It is about 5 s in the normal case, and 1 s in the best case [25]. This delay is much longer than the maximum tolerated time for normal services, as specified in [26] is 500 ms. As a result, for an interruption sensitive flow, the multicast context transfer [27,28] may be required to avoid a large delay. However, even with the multicast context transfer, it is difficult to meet the requirement in terms of service disruption for the interruption-sensitive service when the delay between the anchor and the current MAR is large [16,29]. It is because the multicast traffic has to pass through the aMAR, which plays the role of multicast mobility anchor (MMA). Also, the traffic follows a non-optimal path via the aMAR. In particular, when considering a large domain, it can cause a high end-to-end delay. This issue becomes serious in case of end-to-end delay sensitive service.

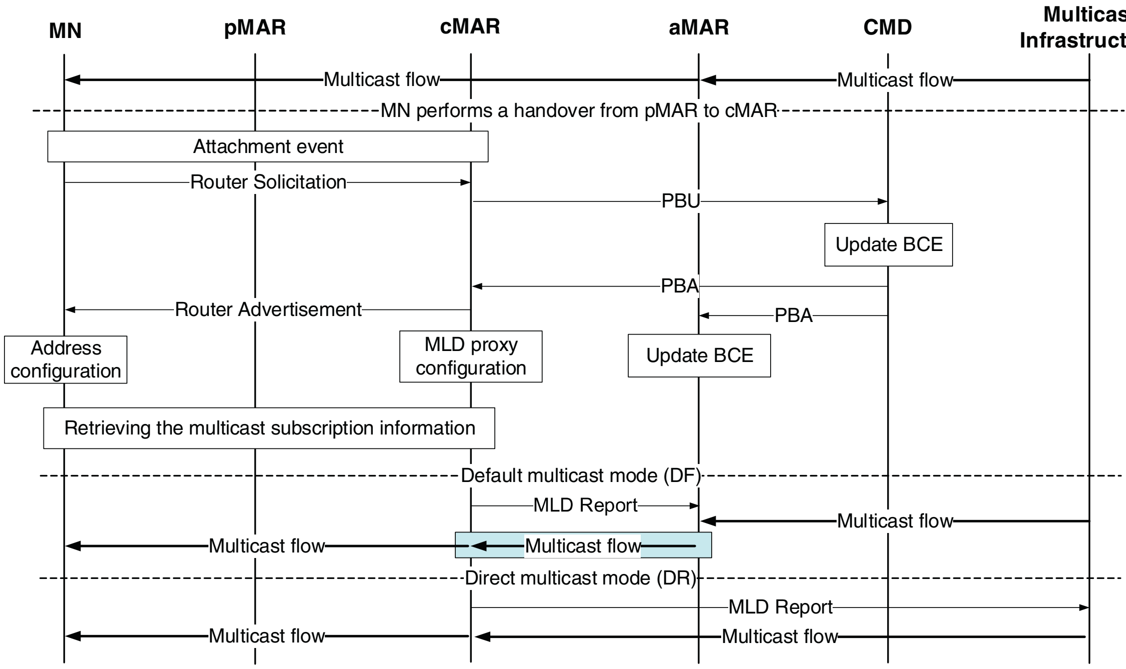

Fig. 3. Multicast mobility support in the default and direct mode (with and without context transfer).

Also, since the listener is unaware of mobility, it will not send an MLD report for explicitly leaving the group in the previous MAR. As a result, if the last member of a multicast group moves to another MAR, the previous one will continue to deliver the multicast traffic until it updates its membership information. Thus, it causes waste of network resource. Using the explicit tracking function and the context transfer, in this case, could help.

On the other hand, the mobility of the node may result in the tunnel convergence problem (or packet duplication). It occurs when multiple instances of the same multicast packets converge to an MAR (from different MARs via the mobility tunnel between them), leading to duplicated multicast packets. Since the purpose of DMM is moving the mobility anchors from the core to the edge of the networks, the number of mobility anchors (the mobility tunnel as well) in a DMM domain will be much more than that in a PMIPv6 domain. As a consequence, the tunnel convergence problem is supposed to be much more severe than that in PMIPv6.

To avoid the tunnel convergence problem, the cMAR can obtain the multicast traffic of the ongoing flow directly from its upstream MR (thus using the native multicast infrastructure for delivering the traffic, as similar as in [30]). This mode is called direct routing (DR). It also allows the multicast traffic to be routed in a better manner. However, although it is the simplest way to support multicast in DMM, it may cause a significant service disruption due to the time needed to join the multicast delivery tree and deliver the traffic to the current MAR [16].

Our previous work in [16] argued the need for a dynamic multicast mobility anchor selection to adapt to the service's requirements as well as operator's policy. This article elaborates the idea proposed in [16] with a full functionality of the system. Also, the performance comparison between our proposal and the DF and DR modes (with and without the multicast context transfer) is conducted.

3. Dynamic multicast mobility support

This section proposes a solution, namely dynamic multicast mobility anchor support (DMMS), which provides a suitable multicast support modes in DMM in a flexible manner depending on different contexts. Our method decides dynamically from which MAR the multicast traffic should be retrieved based on different mobility scenarios (with or without using mobility tunnel). DMMS aims at addressing the issues caused by the mobility of the listener as described in the previous section from both the service and the network operator point of view.

In this section, we first highlight how the solution reflects the considered contexts and its benefits. We then present in more details the architecture as well as the operations of the proposed solution.

3.1. Considered context

To provide multicast support mode in a flexible manner, different contexts are taken into account including the multicast service context, the node's mobility context, and the network context. Then upon these contexts, an appropriate behavior should be followed. The multicast service context refers to the service's requirement in terms of service disruption, end-to-end delay, packet loss, and its feature (e.g., long-lived, short-lived). If the multicast service does not require any requirements in term of service disruption, and end-to-end delay; the simplest support mode such as DR and DF should be used. On the other hand, when services are sensitive to interruption or packet loss, the service disruption time should be minimized. For instance, it should be less than 300 ms for a real-time service, while 500 ms for a normal one [26]. For the end-to-end delay-sensitive service, the long mobility tunnel, which can result in a high end-to-end delay, should be avoided. ITU-T Recommendation G.114 [31] suggests that if one-way transmission time for connection delays can be kept below 150 ms, most applications will experience a transparent interactivity. If the on-going flow is long-lived, the MN may perform many handovers and move far away from its anchor during the flow's lifetime, thus affecting the service disruption and end-to-end delay.

In terms of mobility context, a mobile node with high mobility performs frequent handovers. In this case, almost all ongoing multicast flows are the handover ones. If the multicast traffic is always routed through the aMAR utilizing the mobility tunnel (DF mode), the longer the flow's lifetime is, the more serious the impact will be. Also, the number of anchors and tunnels may be increased. On the contrary, for the low mobility node, the MN is expected to stay at one or several MARs most of the time. It is noted that the mobility features can be defined based on the average number of handovers per unit of time.

Besides, several network contexts such as current load of the MARs, geographical proximity of the MAR to the MN as well as the multicast channel policy[3] also are considered. For example, when the load of MAR is high, it may cause a long delay and a high number of packet losses if it serves as the MMA. In this case, the least loaded MAR (for example, among the MARs having the multicast forwarding state for this channel) can be a potential candidate. The reason lies in the fact that if the channel is already available at the selected MAR, the service disruption time can be minimized (no need of extra time to join the multicast channel). Also, with a negligible increase of load, this MAR can forward the traffic to the cMAR [8].

3.2. Benefits of the solution

The proposed architecture is designed from both user and network operator perspectives, thus it can offer such benefits as:

- *A complete solution* for all the major issues related to multicast listener mobility including service disruption, tunnel convergence problem, network resource waste (due to leave latency), sub-optimal routing, and packet loss.
- *A per-flow multicast support*: depending on the contexts, each multicast flow can be treated differently.
- *Route optimization*: the multicast flows will be routed in a better route since they do not always pass through their mobility anchor.
- *Tunnel convergence problem avoidance*: by using an extension to MLD proxy to support multiple upstream interfaces [32], the tunnel convergence problem can be avoided. In this case, only one proxy instance will be deployed at MAR with its upstream interfaces being configured toward different aMARs and its upstream MR. Accordingly, the MAR will receive only one instance of the multicast packet.
- *Dynamic utilization of mobility tunnel*: the utilization of mobility tunnel for the ongoing multicast flows is enabled in appropriate cases e.g., for remote content, or for a flow with strict delay requirements.
- *Effective tunnel management*: in a DMM environment, it is infeasible to pre-establish all the tunnels between MARs since the number of MARs is supposed to be large. Also, by enabling the multiple upstream interfaces in DMM, it may cause the complex tunnel management (e.g., maintenance of the tunnel and keep alive signaling). Thus, the proposed solution, which is based on a combination between the multicast and unicast mobility management module, can help to solve this issue.
- *Multicast flow load distribution*: since the mode selection takes the current load of the MARs into account, it helps in better distribution of the multicast traffic load among MARs.
- *Centralized channel management*: the central entity (multicast control entity or MCE) collects and manages the considered contexts (e.g., the multicast channels and their scope (local or remote), thus enhancing the control of network providers.

In addition, DMMS can also be applied in case of source mobility and is compatible with the unicast mobility.

3.3. Architecture of DMMS

In order to collect and manage the considered contexts, a logical entity, called multicast control entity (MCE), is introduced. This entity can be collocated with the CMD to become multicast mobility controller (MMC). This entity, beside playing the role of a mobility database as in a normal DMM, can act as a central entity for managing multicast channel in the domain. It therefore stores such multicast-related information as service context (e.g., based on QoS class), node's mobility context, network policy configuration as well as current load of the mobility access router.

Residing in the MAR, the multicast mobility management module (MUMO) takes responsibility for all actions related to the multicast mobility. The structure of this module is depicted in Fig. 4 and briefly described as follows:

- The multicast group management module (MGM) refers to the multicast group management operations and information storage, which is developed based on the MLD proxy with multiple upstream interfaces.[4] This module also supports the multicast explicit tracking function in order to keep a per-host multicast membership state [33] thanks to its database. Besides, it holds a counter structure for the number of listeners per IP multicast channel, allowing it to identify when a node is the last subscriber of a group.

[3] The network operator can define the channel policy in which some channels should be received directly from the native multicast infrastructure (to gain benefit from local content) while the others from their anchor MAR [23].

[4] It is noted that in case of MAR acting as a multicast router, this module is relied on the multicast router function e.g., MRDv6.

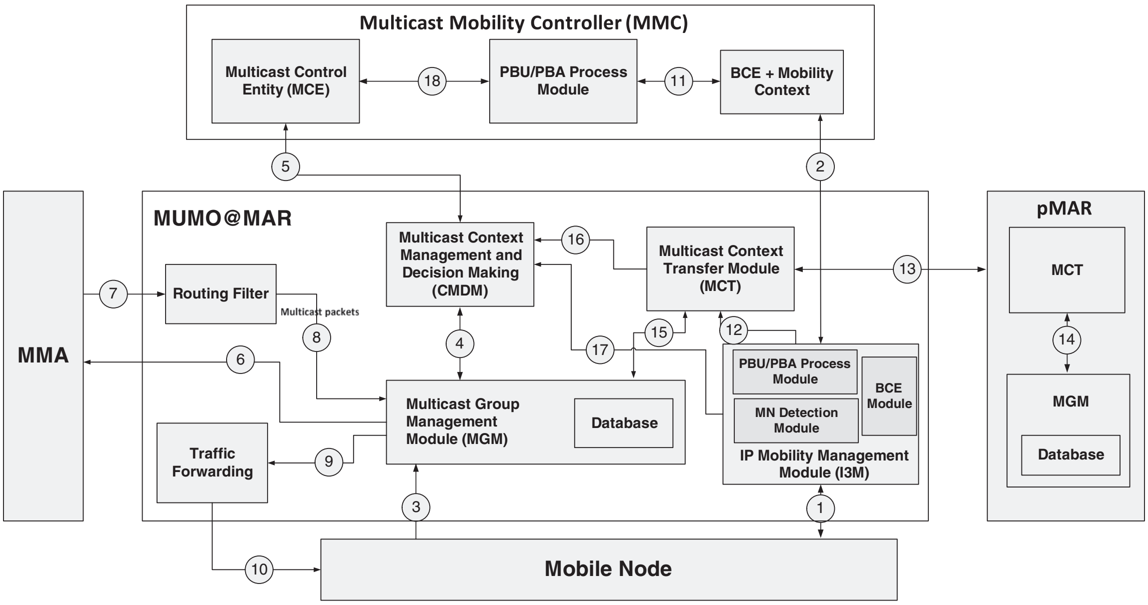

Fig. 4. Multicast mobility management module (MUMO) in the MAR and the multicast-related operations.

- The multicast context management and decision making module (CMDM) collects and manages the considered context including the channel policy, service context and the mobility context. Based on the collected context, it decides the suitable mode among the candidates. The CMDM then indicates the MGM to configure an appropriate upstream interface and join the corresponding MMA.
- The multicast context transfer module (MCT) is in charge of exchanging the MN's multicast subscription information between MARs. So that the new MAR can join the on-going flows in advance to minimize the service disruption.
- The IP mobility management module (I3M) resembles the mobility protocol stack (DMM). It is responsible for detecting MN's attachment, assigning and maintaining the IP connectivity of an MN roaming inside the DMM domain. In other words, it is responsible for all the mobility management-related actions.

3.4. Operations of DMMS

Figs. 4 and 5 represent the components as well as operations of the proposed solution. It is noted that these figures mainly focus on the multicast-related operations. Basically, our solution introduces two levels of intelligence. First, the MMC, for example, based on the network policy, can force the MAR to follow a specific multicast support mode. Besides, MMC can also provide additional information to the MAR (MUMO) which then can decide the most suitable support mode to follow. The detailed operations are described as follows.

When an MN attaches to a new MAR, at first, the typical DMM operations are executed to update the MN's location and configure its address (pref1::MN/64) as specified in the previous section (steps ①, ②). In parallel, MGM module at MAR adds the MN to its downstream interface. The MN then can join a multicast flow via the current MAR by means of an MLD report (③). If the MGM had the multicast information of the requested flow (in other words, the MGM already subscribed to the requested flow), it simply forward this flow to the MN (⑨, ⑩). Otherwise, the MGM triggers the channel configuration acquisition procedure at the MMC by the CMDM module (④, ⑤). The CMDM, based on the acquisition information, decides and returns the appropriate multicast support mode to the MGM. The MGM, acting as an MLD proxy, sends an MLD report to the corresponding MMA to join the flow on behalf of the MN (⑥). After receiving the multicast packets from the MMA, the MAR forward them to the corresponding interface to send to the MN (⑦, ⑧, ⑨, ⑩).

In case of handover to a new MAR (nMAR), similarly, the I3M at the nMAR allocates a new prefix for this MN (e.g., pref2::/64). It then sends a PBU to the MMC for the new prefix registration (②). By looking up the BCE table, the MMC updates the entry corresponding to the MN with its new prefix and location. It is noted that the BCE is extended to store the mobility characteristic of the MN (for example, with 3 different levels including low, medium, and high mobility) (⑪). The mobility feature, the list of anchor MARs' address, the corresponding prefixes, and the address of the previous MAR are then conveyed in the PBA message sent to the I3M at the cMAR (②). The MMC also notifies the anchor MARs the current location of the MN via a PBA message. The mobility tunnel is then established between cMAR and each aMAR to redirect the on-going unicast traffic destined to the MN's active prefixes (e.g., pref1::/64). The I3M then unicasts a RA including the new prefix (pref2::/64) to the MN (①). Based on this prefix, the MN can configure a new IPv6 address (pref2::MN/64) while keeping using the old one for the ongoing flows. The I3M also updates its BCE and triggers the MCT to perform the context transfer (CXT request/response) exchanged with the pMAR (⑫, ⑬) to obtain the MN's subscription information. It is noted that the pMAR retrieves the MN's

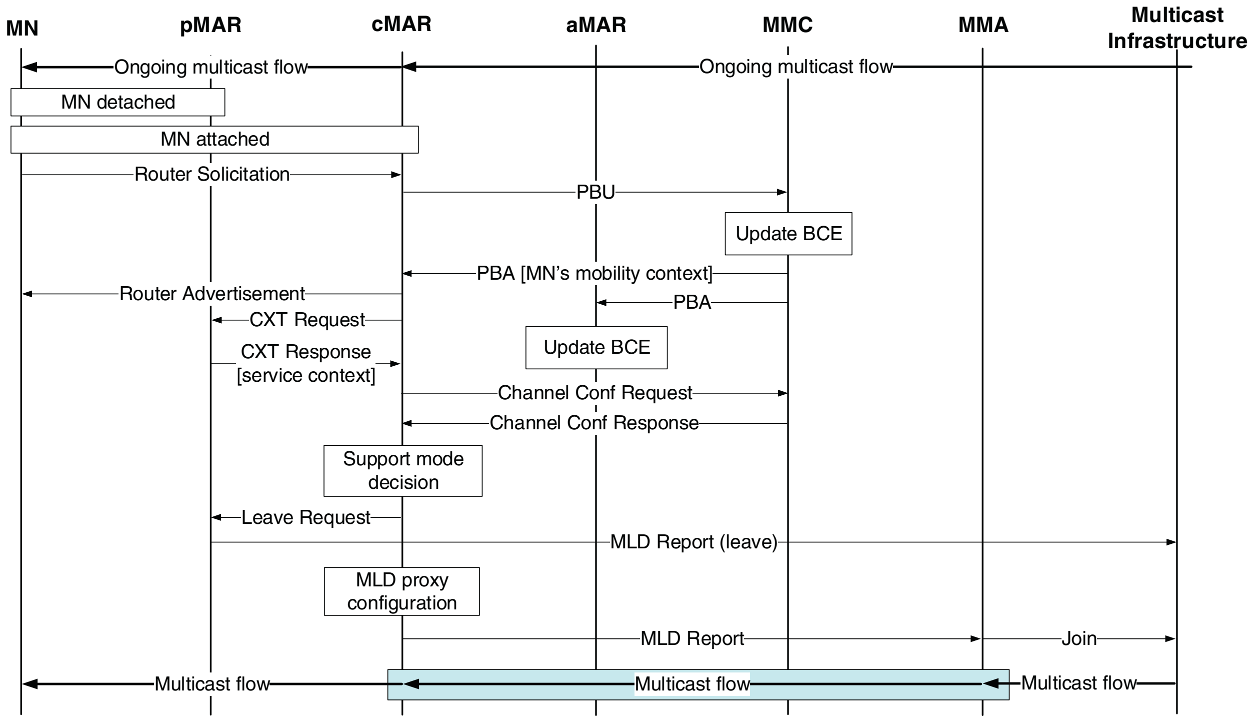

Fig. 5. Multicast-related handover signaling with the multicast context transfer.

subscription information from its database thanks to the explicit tracking function (⑭). MCT then updates these informations in the database (⑮) and sends them to the CMDM for the decision making (⑯). In parallel, I3M sends the mobility features and triggers the CMDM to get the multicast channel configuration from the MMC (⑰, ⑤). Again, based on the channel configuration, the mobility context as well as the multicast service context, the CMDM decides to select the most appropriate support scheme. The CMDM then sends a join request to the MGM including the multicast group and the corresponding MMA address. MGM, after joining and getting the multicast packet for the corresponding flow, forward them to the MN. In addition, if the cMAR does not get the traffic from the pMAR, it will request the pMAR to stop forwarding the flow (if the MN is the last member of the channel at the pMAR) to lower the leave latency, thus reducing the waste of resource.

For a handover flow, the multicast traffic can be received from the aMAR, the pMAR, the cMAR, a common MMA (COMMA) which serves as only one MMA for the domain (as similar in [30]) or even an MAR in which the multicast channel is already available, or a less loaded MAR so as to meet a set of requirements. At this stage, we consider only 4 candidates including the cMAR, pMAR, aMAR, and COMMA corresponding to different multicast support modes: MMA(a) – similar to the default mode, the multicast traffic is routed from the anchor to the current MAR via the mobility tunnel between them; MMA(p) – similar to MMA(a), however, the traffic is received from the previous MAR instead of aMAR; MMA(c) – using the native multicast infrastructure for delivering multicast traffic (without using mobility tunneling); and MMA(co) – the traffic is received from a common multicast router in the domain.

To reduce the complexity and the signaling cost for the context collection process, the MMC can store the MN's subscription information, however, only for the privileged users and the channels with strict requirement in terms of service disruption and end-to-end delay. For those channels, the mode decision will be made by the MMC while for the normal ones, it is done by the MUMO at the cMAR. As a result, for the channels with the strict requirement, the corresponding MMAs will be conveyed via the extended PBA from MMC to the cMAR. Accordingly, the time for the multicast context exchanged as well as the channel configuration acquisition can be ignored.

In addition, DMMS can work in a "simple mode" (SIMP) as DF and DR. It is done by adding an option, namely multicast mode (MM) to the PBA sent from MMC to cMAR (step ② as in Fig. 4) to indicate the multicast mode that cMAR should follow. For example, the value 0 indicates that the normal MMA operations should be executed. On the other hand, the value 1 shows that the default mode without context transfer (DF-C) should be applied. That means the context transfer and channel configuration acquisition processes should not be activated. Consequently, the cMAR simply joins the ongoing flow from the aMAR via the mobility tunnel. Similar procedure happens for other approaches including DF with context transfer (DF-C, MM = 2), DR (MM = 3), and DR-C (MM = 4).

4. Performance analysis

In wireless mobile networks, the mobility anchor is responsible for tracking the location of the mobile node to provide the mobility support. Thus, location management is crucial for the effective operation of wireless networks. In this context, the signaling cost is defined as the cost to update the location of the node which can be considered as a function of different metrics as the hop distance between the entities, the unit transmission cost over wired/wireless link, and the handover rates. Signaling cost is an important factor since it influences the scalability of the system as well as the cost for data delivery. This metric becomes even more critical with the presence of wireless links who have a limited capacity. As the data and control plane are no longer coupled, in case where a huge amount of traffic is generated in the network, the packet delivery cost and the tunneling cost could play a very important role.

On the other hand, from an application point of view, service disruption, end-to-end delay, and number of packet losses are the most important metrics. The service disruption time is defined as a period when a node cannot receive/send the packets while performing a handover. During this period, the packets will be lost. Thus, it may result in noticeable service disruption, especially in case of interruption sensitive applications like video and Voice over IP (VoIP). The number of lost packets typically is proportional to the service disruption, packet arrival rate, network condition, etc. However, since this paper focuses on the impact of mobility to the performance metrics, thus, only two main factors are considered i.e., service disruption and packet arrival rate. In IPv6-based networks, QoS may be defined by packet loss, handover latency and signaling overhead. As a result, a long service disruption time and a large number of lost packets may degrade the quality of service. Besides, end-to-end delay is also one important metric, especially in case of delay-sensitive service.

Accordingly, this section presents the performance analysis of the proposed solution in comparison with different approaches (DF, DF-C, DR, and DR-C) regarding different metrics as multicast service disruption, end-to-end delay, signaling cost, packet delivery cost, tunneling cost and packet loss.

At this stage, when a listener subscribes to a new multicast flow, this flow will be received directly from the native multicast infrastructure. This means the mode selection in the initial phase will be left for future works. For a handover flow, the traffic can be received from the aMAR, the pMAR, the cMAR, a common MMA (COMMA), corresponding to 4 different modes: MMA(a), MMA(p), MMA(c), and MMA(co). We also do not consider the network operator's policies. It is noted that COMMA generally reflects the multicast deployment in PMIPv6.

4.1. Reference model

Fig. 6 shows a reference topology for the performance analysis. The hop-count distances between the entities are defined as follows:

- h_{ac}: the average number of hops between the aMAR and the cMAR.
- h_{ap}: the average number of hops between the aMAR and the pMAR.
- h_{pc}: the average number of hops between the pMAR and the cMAR.

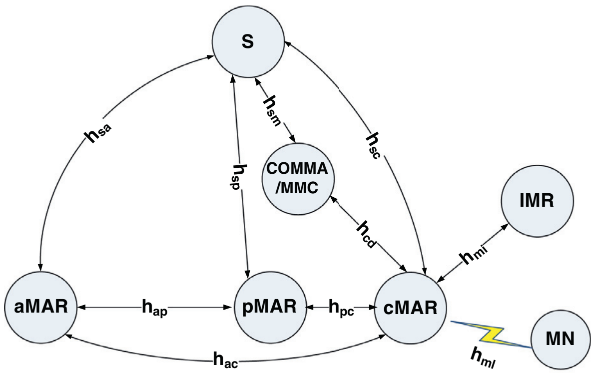

Fig. 6. Reference network topology.

- h_{cd}: the average number of hops between the MAR and the MMC/COMMA.
- h_{sa}, h_{sp}, h_{sc}, h_{sm}: the average number of hops between the source S and the aMAR, the pMAR, the cMAR, the COMMA, respectively.
- h_{mi}: the average number of hops between the cMAR and the intersection MR (IMR) which already has a multicast forwarding state for the group. In the context of this document, h_{mi} represents the popularity of users subscribing to the same flow.

It is noted that the average numbers of hops between the MAR and the listener (wireless link) and between the MAR and its upstream MR are assumed to be one.

4.2. Analytical modeling

4.2.1. Multicast service disruption time analysis

The multicast service disruption time ($SD_{(.)}$) is defined as a period when a listener is unable to receive the multicast packets. Assuming that the delay associated with the processing of the messages in the network entities (e.g., time for PBU/PBA processing and updating binding cache in MAR) is included in the total value of each variable. Then $SD_{(.)}$ in the proposed solution is expressed as:

$$SD_{MMA}^{(.)} = T_{L2} + T_{RA-RS} + T_{MMC} + T_{CXT} + T_{Conf} + T_M^{(.)}, \quad (1)$$

where T_{L2} is the layer 2 (L2) handover duration, T_{RA-RS} is the time for RS/RA exchanged between MAR and MN, T_{MMC} is the time needed to get the address of the anchor/previous MARs from the MMC and update the current location of the MN (at the aMAR), T_{CXT} is the time for the context transfer messages exchanged, T_{Conf} is the time to get the multicast configuration information from the MMC, $T_M^{(.)}$ is the time needed for the cMAR to join, get the first multicast packet and deliver it to the MN after handover.

Although different signaling messages have different sizes, we assume that they have the same size for simplicity. Also, the delay for transmitting a signaling message is supposed to be proportional to the distance between the source and the destination. The proportion is τ for wired and κ for wireless link. As a result, T_{RA-RS}, T_{MMC}, T_{CXT}, and T_{Conf} are given by:

$$T_{RA-RS} = 2\kappa,$$

$$T_{MMC} = 2\tau h_{cd},$$

$$T_{CXT} = 2\tau h_{pc},$$

$$T_{Conf} = 2\tau h_{cd}.$$

Regarding $T_M^{(.)}$, in case of MMA(c), the cMAR has to get the multicast traffic from the IMR which already has a multicast forwarding state for this group. Thus,

$$T_M^c = (h_{mi} + 1)\omega + 2\tau h_{mi} + \kappa,$$

where ω is the delay time in which an MR (and an MLD proxy) needs to join a multicast flow at each router (proxy) in the Internet [34].

In case of MMA(p), the pMAR already had the multicast state for this flow. We have:

$$T_M^p = 2\omega + 2\tau h_{pc} + \kappa.$$

In case of MMA(a), the aMAR may need to re-join the multicast channel, leading to an extra delay. It happens, for example, in case the multicast traffic was received from the multicast infrastructure in the pMAR and the aMAR has left the channel as a result of using MLD proxy with multiple upstream interfaces (called worst-case scenario, or wc). Let p_a denote the probability that this situation happens. As a result, $T_M^{(.)}$ is calculated as:

$$T_M^a = (1 - p_a)T_M^{DF} + p_a T_M^{a-wc},$$

where T_M^{DF} is the time the cMAR needs to get the multicast packet from the aMAR, as similar to in the default mode (DF). We have:

$$T_M^{DF} = 2\omega + 2\tau h_{ac} + \kappa,$$

$$T_M^{a-wc} = (2 + h_{mi})\omega + 2\tau h_{ac} + 2\tau h_{mi} + \kappa.$$

In case of MMA(co), we have:

$$T_M^{co} = 2\omega + 2\tau h_{cd} + \kappa.$$

Similarly, the service disruption time in the default mode (DF—without context transfer, DF-C using context transfer) is given by:

$$SD_{DF} = T_{L2} + T_{RS-RA} + T_{MMC} + T_{Sub} + T_M^{DF}, \quad (2)$$

$$SD_{DF-c} = T_{L2} + T_{RS-RA} + T_{MMC} + T_{CXT} + T_M^{DF}, \quad (3)$$

where T_{Sub} is the time needed for the MAR to obtain the MN's subscription information based on the normal MLD process. We have:

$$T_{Sub} = T_{MSA} + T_{QRD} + 2\kappa,$$

where T_{MSA}, T_{QRD} represent the multicast service activation time and the query response delay, respectively [25]. We suppose that MLD Queries are followed immediately the link-up event or the auto-configuration of IPv6 link-local address of an MN [22]. As a consequence, the multicast service activation time can be ignored ($T_{MSA} = 0$).

The service disruption time in the direct mode (DR) is given by:

$$SD_{DR} = T_{L2} + T_{RS-RA} + T_{MMC} + T_{Sub} + T_M^c, \quad (4)$$

$$SD_{DF-c} = T_{L2} + T_{RS-RA} + T_{MMC} + T_{CXT} + T_M^c. \quad (5)$$

4.2.2. End-to-end delay

End-to-end delay ($E2E_{(.)}$) is the packet transmission delay from the source to the listener. In the MMA(c), the cMAR receives the multicast traffic directly from the multicast infrastructure similar to the direct mode. Hence, we have:

$$E2E_{MMA}^c = E2E_{DR} = \tau h_{sc} + \kappa. \quad (6)$$

In case of MMA(a), the multicast packet is routed from the source to the cMAR via the aMAR, representing the default mode. We have:

$$E2E_{MMA}^a = E2E_{DF} = \tau h_{sa} + \tau h_{ac} + \kappa. \quad (7)$$

In MMA(p) mode, if the traffic for the multicast flow is already available at the current MAR, this MAR will simply forward it to the MN. After each handover, the current MAR can obtain the multicast traffic either from its upstream interface (with probability p_p) or from the previous one (with probability $1-p_p$). Since the flow is started at the anchor MAR, the delay in case of MMA(p) is therefore given by:

$$E2E_{MMA}^p = \kappa + (\tau h_{sa} + \tau \overline{N}_{mar} h_{mm}) p_p^{\overline{N}_{mar}-1} + \sum_{i=1}^{\lfloor \overline{N}_{mar} \rfloor - 1} [\tau h_i + \tau (i+1) h_{mm}] p_p^i (1 - p_p), \quad (8)$$

where $\overline{N}_{mar}$ denotes the average number of MARs involved in the data traffic forwarding between aMAR and cMAR; and h_i is the hop-count distances from the source to the ith MAR in the moving path of the MN (from the aMAR to the cMAR).

Considering the MMA(co), $E2E$ is expressed as:

$$E2E_{MMA}^{co} = \tau h_{sm} + \tau h_{cd} + \kappa. \quad (9)$$

4.2.3. Cost analysis

4.2.3.1. Signaling cost. The signaling cost ($SC_{(.)}$) is the signaling overhead for supporting the handover including multicast-related procedures. It is noted that we consider only the signaling cost per handover, thus the costs for refreshing and de-registration are not taken into account. It therefore can be calculated as:

$$SC_{(.)} = \mu \big(LU_{(.)} + MC_{(.)} \big), \quad (10)$$

where μ is the MAR subnet border crossing rate, $LU_{(.)}$, $MC_{(.)}$ is the signaling cost for the location update and the multicast-related procedures, respectively. According to [35], the signaling 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). We obtain:

$$LU_{MMA} = LU_{DF} = LU_{DR} = 2\beta + 2\alpha h_{cd} + 2\alpha h_{ac}, \quad (11)$$

$$MC_{MMA}^c = MC_{CXT} + MC_{conf} + \alpha(1 + h_{mi}), \quad (12)$$

where

$$MC_{CXT} = 2\alpha h_{pc},$$

$$MC_{conf} = 2\alpha h_{cd}.$$

Similarly, we have:

$$MC_{MMA}^p = MC_{CXT} + MC_{conf} + \alpha h_{pc}, \quad (13)$$

$$MC^{a}_{MMA} = (1 - p_a)MC^{a-df}_{MMA} + p_a MC^{a-wc}_{MMA}, \quad (14)$$

where

$$MC^{a-df}_{MMA} = MC_{CXT} + MC_{conf} + \alpha h_{ac},$$

$$MC^{a-wc}_{MMA} = MC_{CXT} + MC_{conf} + \alpha (h_{ac} + 1 + h_{mi}),$$

$$MC^{co}_{MMA} = MC_{CXT} + MC_{conf} + \alpha h_{cd}. \quad (15)$$

Regarding DF and DR mode, MC_{DF} and MC_{DR} are expressed as:

$$MC_{DF} = \alpha h_{ac} + 2\beta, \quad (16)$$

$$MC_{DF-c} = MC_{CXT} + \alpha h_{ac}, \quad (17)$$

$$MC_{DR} = \alpha + \alpha h_{mi} + 2\beta, \quad (18)$$

$$MC_{DR-c} = MC_{CXT} + \alpha + \alpha h_{mi}. \quad (19)$$

4.2.3.2. Packet delivery cost. The packet delivery cost ($PC_{(.)}$) represents the cost of delivering multicast packets to the MN per unit of time. Let S_c, λ_p denote the average session length at the cMAR and the packet arrival rate, respectively. Again, the packet delivery cost in the MMA(a) corresponds to the default multicast mode. The packet delivery cost is expressed as:

$$PC^{c}_{MMA} = PC_{DR} = S_c \lambda_p (\alpha h_{sc} + \beta), \quad (20)$$

$$PC^{a}_{MMA} = PC_{DF} = S_c \lambda_p (\alpha h_{sa} + \alpha h_{ac} + \beta). \quad (21)$$

Similar to $E2E^{p}_{MMA}$, PC^{p}_{MMA} can be calculated as:

$$PC^{p}_{MMA} = S_c \lambda_p \beta + S_c \lambda_p (\alpha h_{sa} + \alpha \overline{N}_{mar} h_{mm}) p_p^{\overline{N}_{mar}-1} + S_c \lambda_p \alpha \sum_{i=1}^{\lfloor \overline{N}_{mar} \rfloor -1} [h_i + (i+1)h_{mm}] p_p^{i} (1 - p_p). \quad (22)$$

$$PC^{co}_{MMA} = S_c \lambda_p (\alpha h_{sm} + \alpha h_{cd} + \beta). \quad (23)$$

4.2.3.3. Tunneling cost. Regarding the packet tunneling cost ($TC_{(.)}$), it is defined as the additional cost from the tunneling overhead. In MMA(c), the multicast traffic is received directly from the multicast infrastructure, thus, there is no tunneling cost. On the contrary, in MMA(a), MMA(p), and MMA(co) the traffic is routed via the tunnel aMAR-cMAR, pMAR-cMAR, and cMAR-COMMA, respectively. Note that the tunneling cost in the MMA(a), MMA(c) corresponds to the default mode, and the direct mode, respectively. The tunneling cost is therefore computed as:

$$TC^{c}_{MMA} = TC_{DR} = 0. \quad (24)$$

$$TC^{a}_{MMA} = TC_{DF} = \alpha S_c \lambda_p h_{ac}. \quad (25)$$

$$TC^{p}_{MMA} = \alpha S_c \lambda_p h_{mm} \overline{N}_{mar} p_p^{\overline{N}_{mar}-1} + \alpha S_c \lambda_p h_{mm} \sum_{i=1}^{\lfloor \overline{N}_{mar} \rfloor -1} (i+1) p_p^{i} (1 - p_p), \quad (26)$$

$$TC^{co}_{MMA} = \alpha S_c \lambda_p h_{cd}. \quad (27)$$

Table 1
Parameters for the performance analysis.

Parameter	Value	Parameter	Value
T_{L2}	29.9ms	τ	2
λ_p	10 packets/s	κ	15
α	1	β	5
ω	10 ms	h_{mm}	1 hop
h_{cd}	8 hops	h_{mi}	2 hops
h_{sa}	14 hops	h_{sp}	14 hops
h_{sc}	14 hops	h_{sm}	14 hops
S_c	60 s	p_p	0.9
p_a	0.5	T_{QRD}	1 s

4.2.4. Packet loss

During the handover, packets may be lost. For sake of simplicity, we assume that the number of lost packets is proportional to the service disruption time and the packet arrival rate. As a result, the number of lost packets is given by:

$$PL_{(.)} = \lambda_p SD_{(.)}. \quad (28)$$

5. Numerical results

This section presents the numerical results for the proposed solution in comparison with the DR, DR-C, DF, and DF-C mode. It is worth to note that all the 4 different schemes in our proposed solution (including MMA(a), MMA(c), MMA(p), MMA(co)) are presented to highlight the need of a dynamic multicast support mechanism. The default parameter values for the analysis are introduced in Table 1, in which some parameters are taken from [25,27,34].

Also, we consider the case where the MN always moves from MAR to MAR as if they were linearly deployed (the user is moving further away from the first attached MAR and never attaches back to a previously visited MAR). It represents the worst-case scenario. Thus, we have $h_{ac} = \overline{N}_{mar} h_{mm}$, $h_{ap} = (\overline{N}_{mar} - 1)h_{mm}$, and $h_{pc} = h_{mm}$. According to [36], $\overline{N}_{mar}$ is calculated as:

$$\overline{N}_{mar} = 1 + \frac{\mu}{\delta}, \quad (29)$$

where $1/\delta$ is the mean value of the active prefix lifetime while the MN is visiting a foreign network.

To give the idea of the relation between the number of MARs involved in the data traffic forwarding to/from an MN ($\overline{N}_{mar}$) and the velocity (υ), we assume that the subnet residence time (MAR subnet) is a random variable which follows an exponential distribution with mean value $1/\mu$ and the MAR coverage area is circular with radius R. According to [37], the subnet border crossing rate μ is calculated as:

$$\mu = \frac{2\upsilon}{\pi R}, \quad (30)$$

where υ is the average velocity of the MN.

Fig. 7 depicts the value of $\overline{N}_{mar}$ as a function of the velocity when the subnet radius R and $1/\delta$ are fixed to the value of 400 m and 300 s, respectively. As the velocity increases, $\overline{N}_{mar}$ is increased. According to Eq. (29), we can obtain a similar curve as in Fig. 7 if the value of υ is fixed while the mean value of the active prefix lifetime ($1/\delta$) is varying. Thus, the figure for this case is not shown here. As we can observe, the

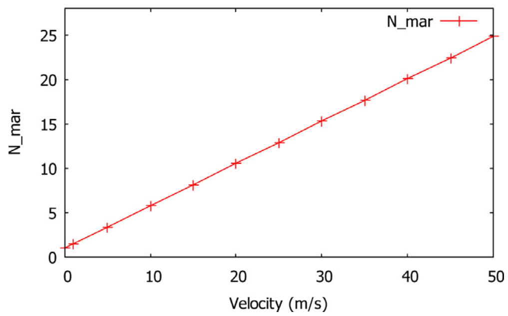

Fig. 7. $\overline{N}_{mar}$ as a function of velocity (υ).

low value of $\overline{N}_{mar}$ represents a low mobility and/or a low-lived flow scenario. The higher value of $\overline{N}_{mar}$ corresponds to a high mobility and/or a long-lived flow scenario.

5.1. Multicast service disruption

Fig. 8 shows the multicast service disruption time as a function of $\overline{N}_{mar}$ and h_{mi}. In Fig. 8(a), the value of $\overline{N}_{mar}$ is varied over a range from 1 to 15. It appears clearly that without the multicast context transfer, the service disruption in the DF and DR mode (about 1200 ms) is definitely higher than that in the other cases and leading to such an unacceptable service disruption. When $\overline{N}_{mar}$ is small, DF-C gives a better performance than the others. As $\overline{N}_{mar}$ increases, while the disruption time in DR-C, MMA(p), MMA(c), MMA(co) is kept constant, that in DF-C, MMA(a) is significantly increased. As a result, DR-C outperforms the others, while MMA(p) introduces a minor additional disruption compared to that in DR-C. Similarly, as can be seen in Fig. 8(b), when the h_{mi} is increased, the service disruption time in MMA(p), MMA(co) and DF-C is fixed while that in MMA(c), MMA(a) and DR-C is notably increased. Altogether, the service disruption time cannot be guaranteed in both cases DR-C and DF-C. Besides, the service disruption in MMA(p) is slightly higher compared to the lower value among DF-C and DR-C. It is also stable in both figures. In conclusion, the difference between the service disruption in MMA(p), DF-C, and DR-C can be considered as a cost of our solution to obtain the required context to keep the service disruption time stable and as low as possible as well.

5.2. End-to-end delay

Regarding the end-to-end delay, two different scenarios are considered. In the first scenario, the source is supposed to be outside of the DMM domain. Thus, the distance between the source and the MAR is supposed to be the same. The results of this scenario are shown in Fig. 9(a). The second scenario is used to illustrate the case where the source (inside the domain) is extremely close to the aMAR (on the right side of Fig. 9(b)) or extremely close to the cMAR (on the left side of Fig. 9(b)). It is done by varying the value of h_{sc} while fixing the value of $h_{sa} + h_{sc}$ (for example, 16 hops).

In the first scenario (Fig. 9(a)), as $\overline{N}_{mar}$ increases, the end-to-end delay in case of MMA(a)/DF is dramatically increased, while that in MMA(c)/DR and MMA(co) is kept constant. On the other hand, the delay in MMA(p) is relatively small increased. Note that the delay in MMA(c)/DR and MMA(p) is kept below the value 50 ms, which means that these modes satisfy the strict requirement in terms of end-to-end delay (for real-time gaming as specified in [31]).

In the second scenario, as can be observed in Fig. 9(b), even when the source is very close to the aMAR ($h_{sa} = 1$, $h_{sc} = 15$), the delay in the MMA(c)/DR is lower than that in the other cases. Therefore, the impact of the mobility tunnel (cMAR-aMAR and cMAR-pMAR) on the end-to-end delay is obvious. In conclusion, the MMA(c)/DR is generally well suited for the delay-sensitive flows, while MMA(p) introduces a minor additional delay compared to that in MMA(c)/DR.

5.3. Signaling cost

Fig. 10 shows the signaling cost of the proposed solution in comparison with the DF and DR mode. As can be seen in this figure, the DR and DF modes give a better performance compared with our solution. It is obvious since in our solution, more signaling messages are needed to collect the context. In Fig. 10(a), the signaling cost is decreased as the subnet residence time ($1/\mu$) increases (μ decreases). However, the signaling overhead is negligible. In Fig. 10(b), when the value

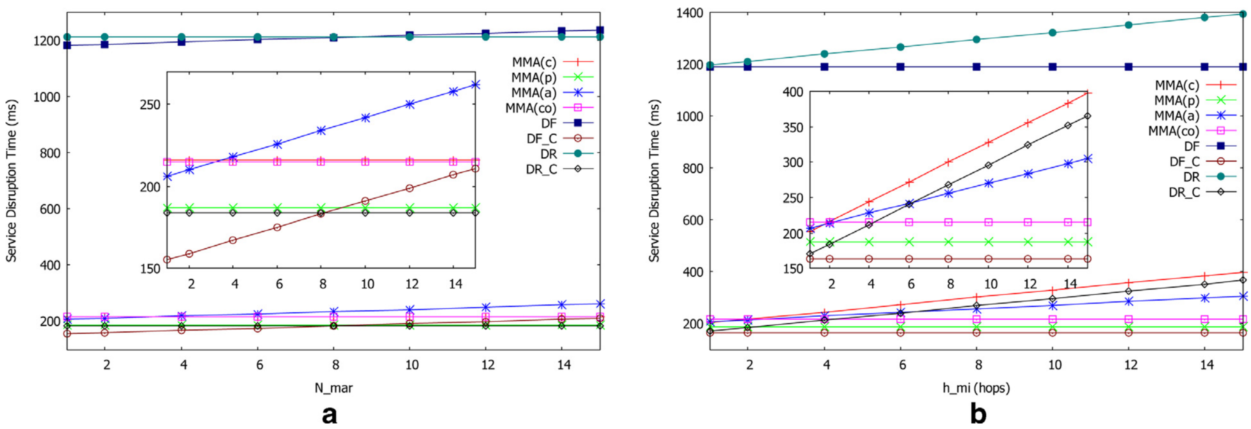

Fig. 8. Service disruption as a function of: (a) $\overline{N}_{mar}$, (b) h_{mi}.

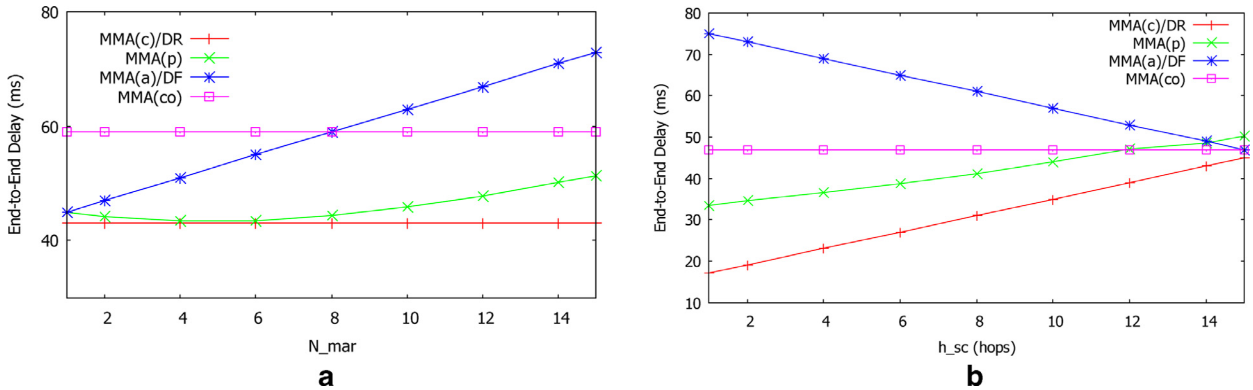

Fig. 9. End-to-End delay as a function of: (a) $\overline{N}_{mar}$ and (b) h_{sc}.

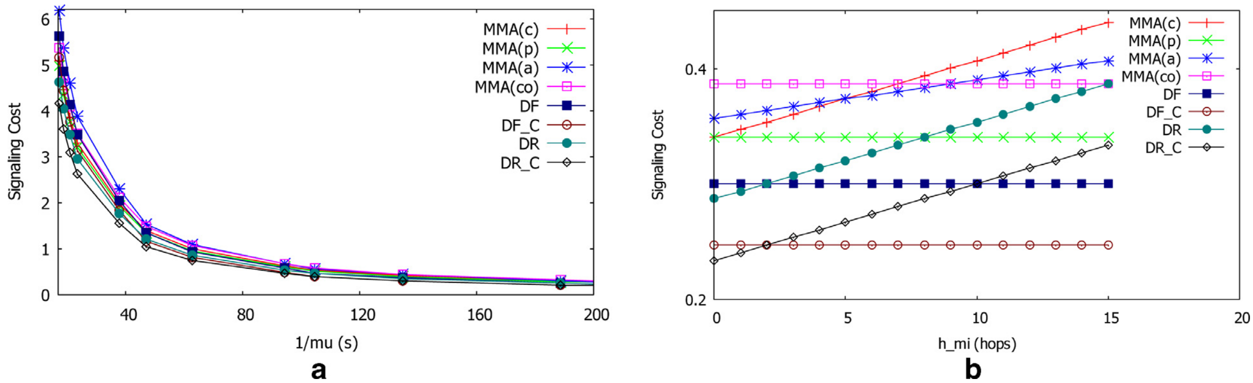

Fig. 10. Signaling cost as a function of: (a) $1/\mu$ and (b) h_{mi}.

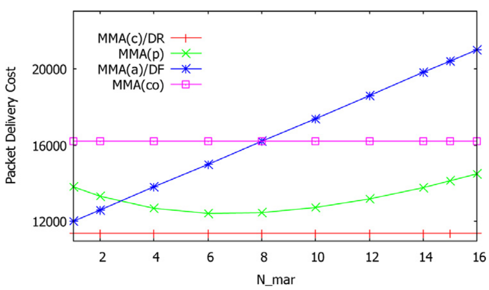

Fig. 11. Packet delivery cost versus $\overline{N}_{mar}$.

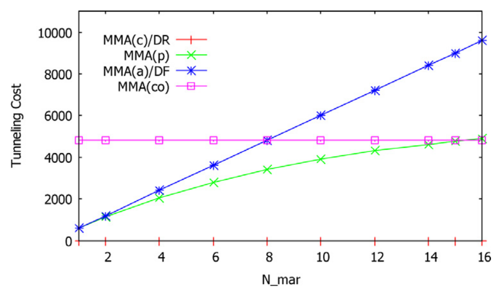

Fig. 12. Tunneling cost as a function of $\overline{N}_{mar}$.

of h_{mi} increases, the signaling cost is fixed in case of DF, DF-C and MMA(p), while it is increased in the other cases. The reason is that in case of DF, DF-C, and MMA(p) the current MAR obtains the multicast traffic from the MAR which had the subscription information of the flow. In general, our solution introduces an acceptable signaling overhead compared to the DF and DR modes, as a cost of the context collection process. Additionally, this overhead can be avoided if our solution works in the SIMP mode.

5.4. Packet delivery cost

Regarding the packet delivery cost (see Fig. 11), MMA(c) and DR mode give the best performance compared to the others. When $\overline{N}_{mar}$ increases, the packet delivery cost in MMA(c)/DR and MMA(co) is kept constant while that in MMA(a)/DF is notably increased. In the middle, MMA(p) introduces a minor increase in compared to MMC(c)/DR.

5.5. Tunneling cost

Fig. 12 depicts the tunneling cost as a function of $\overline{N}_{mar}$. As can be seen in this figure, MMA(c) and DR mode do not introduce any tunneling overhead since the traffic is routed directly from the multicast infrastructure without using the mobility tunnel. On the contrary, the tunneling cost in case of MMA(a) and DF is significantly increased as $\overline{N}_{mar}$ increases. It is due to the fact that the traffic is routed via the tunnel aMAR-cMAR which is supposed to be long. The tunneling overhead in case of MMA(p) is increased, however, still lower than that of MMA(co) which is kept constant.

5.6. Packet loss

With a fixed value of packet arrive rate, the packet loss is directly proportional to the service disruption time as given in Eq. (28). Accordingly, the figure depicting the packet loss (as a function of $\overline{N}_{mar}$ and h_{mi}) has a similar curve as in Fig. 8, thus, is not shown here. Again, the DF and DR represent a large number of packet losses, while MMA(p) can be considered as a good option to guarantee the low number of packet losses.

6. Implementation and preliminary results

6.1. Current status of the implementation

The proposed solution, DMMS, is under development [24,27]. At this stage, MGM, MCT, and I3M modules are already available. In detail, MGM was developed based on an open-source for MLD Proxy, namely ECMH.[5] It is extended to support the explicit tracking function and the multiple upstream interfaces. There is another possibility in which MGM can work as an MR e.g., by using MRD6.[6] MCT module (written in C), which was developed as a separate component, can be easily applied to the other solutions in PMIPv6 as well as in DMM [24,27]. On the other hand, in our laboratory, a Linux-based DMM was recently implemented on top of the UMIP[7] and OAI PMIP implementation [38]. Our DMM implementation, which follows a partially distributed approach, is perfectly fit with I3M module. The CMD is extended to play the role of MMC. At this stage, CMDM module is executed in a simple way: (i) with a low mobility node, MMA(a) is always used; (ii) with a high-mobility node, for a disruption-sensitive service, MMA(p) is applied while for a delay sensitive service, MMA(c) is selected. In addition, when the MN plays the role of a source, MMA(a) is applied.

6.2. Testbed deployment and scenario description

Following the idea of a near-to-real testbed described in [25], the proof-of-concept architecture, as indicated in Fig. 13, is deployed. The testbed is a combination of virtualized environment which consists of multiple virtual machines (e.g., using User-Mode Linux[8]) and the Network Simulator NS-3.[9] It is composed of a central entity playing the role of MMC, 16 MARs with MUMO functionality, 16 access points (APs), 15 multicast routers (MRs) supporting Protocol Independent Multicast—Sparse Mode (PIM-SM),[10] one multicast source and one MN acting as a multicast listener. The architecture of the PIM-SM domain is hierarchically formed as a tree structure with 4 layers in which the source's MR acts as

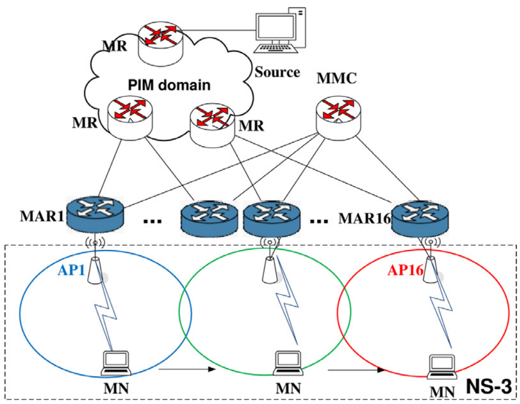

Fig. 13. Testbed deployment.

the root of the tree [39]. The MARs are then connected to the PIM-SM domain in a similar manner, as the leafs of the binary tree. Thus, this structure allows to measure the impact of h_{mi} to the performance metrics. For instance, the testbed is deployed on a single physical machine running Ubuntu 14.04 LTS. The MMC, MARs, MRs and the source are the virtual machines, while the MN and the APs are NS-3 nodes. By using NS-3, the mobile node can easily move between MARs.

To generate the multicast traffic, several tools can be used e.g., Iperf[11] and MINT.[12] During the experiments, a network analyzer tool (e.g., Wireshark[13]) is used to capture the packets exchanged between the entities. At this step, two performance metrics are considered including service disruption and end-to-end delay. It is noted that for the improvement of the credibility, we performed the experiment in a large amount of times. Based on the collected results, we calculated the average value and the standard deviation to improve the degree of confidence.

Then two different experimental scenarios are defined as follows.

- Scenario 1 (S1): This simple scenario is to explain the situation in which the flow will be terminated after one handover and the value of h_{mi} is fixed to a value of one. In this scenario, the MN will first attach to the MAR1. It subscribes to a multicast flow which is being broadcasted by the source. The MN will then perform a handover from MAR1 to MAR2.
- Scenario 2 (S2): This general scenario is used to illustrate the situation when both the number of active prefixes and the value of h_{mi} are varied. In this case, the MN will first attach to the MAR1. It subscribes to a multicast flow which is being broadcasted by the source. It then moves from MAR1 to MAR16 passing the MARi ($i = 2, \ldots, 15$). The flow is kept alive during the movement. We then calculate the average service disruption as well as end-to-end delay for all handovers.

[5] Easy Cast du Multi Hub, http://unfix.org/projects/ecmh/.
[6] MRD6, http://fivebits.net/proj/mrd6/.
[7] UMIP - Mobile IPv6 and NEMO for Linux, http://umip.org.
[8] User-Mode Linux, http://user-mode-linux.sourceforge.net.
[9] The Network Simulator NS-3, https://www.nsnam.org.
[10] The multicast router functions are deployed by using MRD6 implementation.
[11] Iperf, https://iperf.fr.
[12] MINT, http://mc-mint.sourceforge.net.
[13] Wireshark, https://www.wireshark.org.

Table 2
Preliminary results: the average and standard deviation values of service disruption and end-to-end delay in milliseconds.

Metric/method	MMA	DF-C	DR-C
SD (S1)	(203.6, 60.3)	(170.4, 45.5)	(199.3, 49.4)
SD (S2)	(213.3, 65.2)	(311.3, 160.6)	(327.3, 150.2)
E2E (S1)	(43.2, 11.3)	(46.1, 14.5)	(42.2, 13.4)
E2E (S2)	(48.2, 13.2)	(67.3, 20.5)	(40.3, 12.2)

6.3. Preliminary results

Table 2 shows the average and standard deviation values of service disruption time and end-to-end delay of our solution in comparison with DF-C and DR-C in two different scenarios (S1 and S2).

Regarding the multicast service disruption, in the scenario 1 (S1) the disruption in case of MMA is slightly longer than that in DF-C and DR-C as a cost of additional signaling messages exchanged between the entities for collecting the contexts. However, in the scenario 2, while the service disruption in case of MMA is slightly increased, those in DF-C and DR-C are significantly increased. Thus, it is clear that by dynamically providing a suitable multicast support mode, our solution helps to keep the value of service disruption stable. Moreover, this value is much lower than the one in case of DR-C and DF-C in the scenario 2. The reason is that by moving the MN from MAR1 to MAR16, we varied the value of two parameters i.e., number of active prefixes and h_{mi} at the same time. The similar thing happens in terms of end-to-end delay. In detail, in the first scenario, three modes give the similar results. However, while the delay in case of MMA and DR-C is almost stable, the one in case of DF-C is notably increased. In the next steps, more experiments will be considered to validate the behavior of our proposed solution.

7. Discussions

7.1. Multicast router function deployment at the MAR

Our analysis can also be applied when the multicast router function is deployed at the MAR. As in the Medieval project, the MGM represents the functionality of a PIM-SM multicast router (e.g., based on MRD6 implementation). In this case, the Multicast Routing Information Base (MRIB) not only can be based on the unicast RIB, but also on the information from the CMDM. For example, in order to apply MMA(p) for a specific channel (say C1), the cMAR uses an explicit PIM join message to join the C1 at pMAR. In other words, pMAR becomes a reverse path forwarding (RPF) neighbor router of the cMAR regarding the specific flow C1.

7.2. Multicast source mobility support

At this stage, our solution can also support source mobility in DMM. However, the MMA(a) will always be applied to avoid the potential impact on the service disruption. In case of ASM, an extension of PIM-SM [40] can be used to route the multicast traffic directly from the cMAR to the rendez-vous point (RP) bypassing the aMAR. Thus, the multicast traffic is routed in a better way. In more detail, the explicit RPF mechanism is used to build the multicast delivery tree via an explicitly configured path included in the PIM join messages. After receiving the unicast-encapsulation packets from the current MAR, the RP will send a Join message including the address of the sender (cMAR's address) in a new type-length-vector (TLV). It allows the RP to establish the shortest path tree toward the current location of the source. The native multicast traffic then will be sent via the new delivery tree from the cMAR and reaches the listeners (PIM phase two).

8. Conclusions and perspectives

As a variety of multicast service with different features is expected to be widely used in the future networks, a flexible architecture to support multicast is required. In this paper, we introduced an architecture to support multicast in a flexible manner (namely DMMS) from both the service and the operator point of views. Depending on a set of contexts, an appropriate support mode will be enabled, thus, providing a per-flow multicast support. The numerical and experimental results showed that DMMS guaranties the service requirements in term of service disruption and end-to-end delay while keeping the low value of signaling, packet delivery and tunneling cost compared with other proposals. Also, DMMS helps to avoid the traffic replication issue while better distribute the load among the access routers. In the next step, to achieve a more realistic result, more experiments will be conducted based on the under-deployment real testbed.

Acknowledgments

EURECOM acknowledges the support of its industrial members: BMW Group Research & Technology, IABG, Monaco Telecom, Orange, SAP, SFR, ST Microelectronics, Symantec.

References

[1] Ericsson, Ericson Mobility Report: On the Pulse of the Networked Society, November 2014.
[2] Cisco, Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2013–2018, February 2014.
[3] Cisco White Paper, Cisco VNI Service Adoption Forecast, 2013–2018, 2013.
[4] Telefonica Sees ARPU Decline as Data Grows and Voice Falls, June 2009, http://www.mobileeurope.co.uk/News-Analysis/telefonica-sees-arpu-decline-as-data-grows-and-voice-falls.
[5] H.A. Chan, D. Liu, P. Seite, H. Yokota, J. Korhonen, Requirements for Distributed Mobility Management, RFC 7333, August 2014.
[6] H.-A. Chan, H. Yokota, J. Xie, P. Seite, D. Liu, Distributed and dynamic mobility management in mobile internet: current approaches and issues, J. Commun. 6 (1) (2011) 4–15.
[7] C.B. Sankaran, Data offloading techniques in 3GPP Rel-10 networks: a tutorial, IEEE Commun. Mag. 50 (June (6)) (2012) 46–53.
[8] B. Williamson, Developing IP Multicast Networks, Cisco Press, 1999.
[9] T. Schmidt, M. Waehlisch, G. Fairhurst, Multicast mobility in mobile IP Version 6 (MIPv6): problem statement and brief survey, RFC 5757, February 2010.
[10] E.W. Paper, LTE Broadcast: A Revenue Enabler in the Mobile Media Era, 2013.
[11] A. El-Sayed, V. Roca, L. Mathy, A survey of proposals for an alternative group communication service, IEEE Netw. 17 (1) (2003) 46–51.
[12] I. Romdhani, M. Kellil, H.-Y. Lach, A. Bouabdallah, H. Bettahar, IP mobile multicast: challenges and solutions, IEEE Commun. Surv. Tutorials 6 (1) (2004) 18–41.
[13] Ericsson, LTE Broadcast: A Revenue Enabler in the Mobile Media Era, White Paper, 2013.

[14] T.-T. Nguyen, Optimization of mobility mechanisms for IP-based multicast flows, Mobile Communications Dept., EURECOM, Sophia-Antipolis, France, May 2014 (Ph.D. thesis).
[15] S. Figueiredo, S. Jeon, R.L. Aguiar, IP Multicast Use Cases and Analysis over DMM, IETF Draft (expired), 2012.
[16] T.-T. Nguyen, C. Bonnet, On the efficiency of dynamic multicast mobility anchor selection in DMM: use cases and analysis, IEEE Int. Conf. Commun. (ICC) (2014) 2828–2834.
[17] H. Ali-Ahmad, D. Moses, H. Moustafa, P. Seite, T. Condexia, Mobility Anchor Selection in DMM: Use-case Scenarios, IETF-Draft (work-in-progress), 2013.
[18] C. Perkins, D. Johnson, J. Arkko, Mobility Support in IPv6, RFC 6275, July 2011.
[19] S. Gundavelli, K. Leung, V. Devarapalli, K. Chowdhury, B. Patil, Proxy Mobile IPv6, RFC 5213, 2008.
[20] C.J. Bernardos, A. de la Oliva, F. Giust, A PMIPv6-based Solution for DMM, IETF Draft (work-in-progress) (2015).
[21] H. Asaeda, H. Liu, Q. Wu, Tuning the Behavior of the Internet Group Management Protocol (IGMP) and Multicast Listener Discovery (MLD) for Routers in Mobile and Wireless Networks, RFC 6636, 2012.
[22] T. Schmidt, M. Waehlisch, S. Krishnan, Base Deployment for Multicast Listener Support in Proxy Mobile IPv6 (PMIPv6) Domains, RFC 6224, 2011.
[23] T.-T. Nguyen, C. Bonnet, Efficient multicast content delivery over a distributed mobility management environment, in: Proceedings of VTC Fall, Las Vegas, USA, 2013.
[24] S. Figueiredo, C. Guimaraes, R.L. Aguiar, T.-T. Nguyen, L. Yadin, N. Carapeto, C. Parada, Broadcasting user content over novel mobile networks, in: ICC 2013, 2013.
[25] T.-T. Nguyen, C. Bonnet, Performance Optimization of Multicast Content Delivery in a Mobile Environment based on PMIPv6, WCNC, 2013.
[26] 3GPP, TR 25.913., Requirements for Evolved UTRA (E-UTRA) and Evolved UTRAN (E-UTRAN), Release 9, 2009.
[27] MEDIEVAL Project, Deliverable D4.4, Final Operational Mobility Architecture, 2012.
[28] L.M. Contreras, C.J. Bernardos, I. Soto, Proxy Mobile IPv6 (PMIPv6) Multicast Handover Optimization by the Subscription Information Acquisition through the LMA (SIAL), RFC 7161, 2014.
[29] S. Figueiredo, S. Jeon, R.L. Aguiar, Use-cases Analysis for Multicast Listener Support in Network-based DMM, PIMRC 2012, 2012.
[30] J.C. Zuniga, L.M. Contreras, C.J. Bernardos, S. Jeon, Y. Kim, Multicast Mobility Routing Optimizations for Proxy Mobile IPv6, RFC 7028, 2013.
[31] G. ITU-T Recommendation, General Characteristics of International Telephone Connections and International Telephone Circuits: One-Way Transmission Time, 2003.
[32] H.-K. Zhang, S. Gao, T.C. Schmidt, B.-H. Feng, L.-L. Wang, Multi-Upstream Interfaces IGMP/MLD Proxy, IETF Draft (expired), 2013.
[33] H. Asaeda, IGMP/MLD-Based Explicit Membership Tracking Function for Multicast Routers, IETF Draft (work-in-progress), 2015.
[34] S.J. Yang, S.H. Park, A dynamic service ranged-based multicast routing scheme using RSVP in mobile networks, in: GLOBECOM, 2001.
[35] J.-H. Lee, T.-M. Chung, How much do we gain by introducing route optimization in proxy mobile IPv6 networks? Ann. Telecommun. 65 (5–6) (2010) 233–246.
[36] F. Giust, C.J. Bernardos, A. De La Oliva, Analytic evaluation and experimental validation of a network-based IPv6 distributed mobility management solution, IEEE Trans. Mobi. Comp. 13 (11) (2014) 2484–2497.
[37] C. Makaya, P. Samuel, An analytical framework for performance evaluation of IPv6-based mobility management protocols, IEEE Trans. Wireless Commun. 7 (3) (2008) 972–983.
[38] OpenAir Interface Proxy Mobile IPv6, http://www.openairinterface.org/openairinterface-proxy-mobile-ipv6-oai-pmipv6.
[39] Y. Min-hua, Y. Lv-yun, L. Yu, Z. Hui-min, The implementation of multicast in mobile IP, in: Proceedings of WCNC, 2003.
[40] J. Asghar, I.J. Wijnands, S. Krishnaswamy, A. Karan, V. Arya, Explicit RPF Vector, IETF-Draft(work-in-progress), 2015.

Tien-Thinh Nguyen is a Post-doc researcher at EURECOM, Sophia Antipolis, France. He received his B.S. degree in Computer Engineering from Hanoi University of Science and Technology, Vietnam in 2006. He received the M.S. degree in Networks and Communication Systems from University of Claude Bernard Lyon 1, France in 2010. He received his Ph.D. degree in Computer Science and Networking from Telecom Paristech, France in 2014. His research interests are in the areas of wireless networks and mobile networks with emphasis on IP multicast, mobility management, and Software-defined Networking.

Christian Bonnet received his M.S. degree in 1978 from Ecole Nationale des Mines de Nancy, France. He is currently a Professor in the Department of Mobile Communications at EURECOM, Sophia Antipolis, France. His teaching activities are real-time and distributed systems, mobile communication systems, wireless LANs and protocols for mobility management. His main areas of research are wireless protocols, wireless access to IP networks and data communications in mobile networks including mobile ad hoc networks.