Noname manuscript No.
(will be inserted by the editor)

Combining client and network-based Distributed Mobility Management

A hybrid approach

Fabio Giust · Carlos J. Bernardos · Antonio de la Oliva

the date of receipt and acceptance should be inserted later

Abstract The introduction of modern portable devices has exacerbated the increase of mobile data demand from users. These new mobile terminals exhibit a great variety of applications, some of which require active sessions to be maintained, while others are just short-lived. This introduces additional challenges to mobility management, which traditionally assumed that all traffic had to be mobility-enabled. Altogether, mobile network operators are now facing the challenges posed by a huge data demand generated by users that are able to connect from different access networks and establish several active sessions simultaneously, while being mobile. This triggered the introduction of a new paradigm: the distributed mobility management (DMM) which aims at flattening the network and distributing the entities in charge of managing users' mobility.

This article describes a novel hybrid DMM solution which benefits from combining a network-based mobility approach, based on Proxy Mobile IPv6, with a client-based one, based on Mobile IPv6. This combination provides additional flexibility to the mobile network operators, which can decide when and how to combine these two approaches. An analytic evaluation of the solution is also provided, comparing the obtained results to the classical centralised mobility approaches. Also, a real implementation for two network-based DMM solutions has been developed, proving the feasibility of the design and showing

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7-ICT-2009-5) under grant agreement n. 258053 (MEDIEVAL project) and from the Spanish Government, MICINN, under research grant TIN2010-20136-C03.

Fabio Giust
Institute IMDEA Networks
University Carlos III of Madrid,
E-mail: fabio.giust@imdea.org

Carlos J. Bernardos · Antonio de la Oliva
University Carlos III of Madrid,
E-mail: {cjbc, aoliva}@it.uc3m.es

that the overall performance for the two solutions highly depends on the underlying network topology. This study provides the trade-offs that an operator should consider when deploying a distributed mobility management architecture.

Keywords Distributed Mobility Management · IP mobility · PMIPv6 · Wireless systems · cellular architecture · handover mechanisms · experimental evaluation

1 Introduction

Mobile connectivity is now far from being a luxury service. Users demand Internet access while on the move, and the volume of traffic generated by mobile subscribers has been exponentially increasing during the last few years. This has been motivated by the incredible success on the development and wider introduction in the market of smart-phones, tablets and netbooks, such as Android or iOS-based terminals, which have changed not only the way users consume data services, but also the place they do it from. This ubiquitous Internet connectivity has also changed the catalog of available services, which is no longer limited to a small well known set of services such as web browsing and email, but it is now composed of a huge plethora of applications, some of which are even based on capabilities residing in the cloud. As a consequence of this paradigm shift, mobile network operators are witnessing how their networks need to cope with an increasing volume of data, saturating their access links, and triggering the need for additional access technologies to be made available to the users. In addition, as radio accesses with more capacity are deployed, and operators migrate their networks to full IP based architectures, such as the WiMAX [1] related standards or the 3GPP Evolved Packet System (EPS)[2] – which can also benefit from offloading to WiFi available networks – the load will spread between the different access networks, but the congestion in the core will tend to increase due to current centralised network design. Additionally, the use of a full IP-based core for both voice and data triggers the need for standardised IP mobility management solutions, which up to now had shown little or no deployment penetration, being proprietary customised solutions used instead.

Because of the new requirements imposed by mobile users' traffic, operators with a large number of mobile subscribers are now looking for alternative mobility solutions that are more distributed in nature, allowing a cheaper and more efficient network deployments capable of meeting their customers' requirements. In particular, there is an effort within the IETF, called *Distributed Mobility Management*[3] *(DMM)*, that is currently addressing exactly this particular problem. After defining the problem statement [1], the group is

[1] `http://www.wimaxforum.org/`

[2] 3rd Generation Partnership Project, `http://www.3gpp.org/`

[3] `http://datatracker.ietf.org/wg/dmm/charter/`

currently analysing the limitations of current standardised mobility management protocols, identifying the gaps that need to be covered with new DMM protocols [2]. We review the motivations of DMM in Section 2. A thorough review of the state of the art, including the most relevant approaches for distributing the mobility management is performed in Section 3.

Most of the research work being performed on distributed mobility management focuses basically in two main approaches: solutions aimed at making Mobile IPv6 (MIPv6) work in a distributed way, and proposals doing the same exercise for Proxy Mobile IPv6 (PMIPv6). In this article, we propose HDMM, a hybrid DMM solution composed of two independent components following each of those approaches (Section 4): *i)* a MIPv6-based solution, which evolves the MIPv6 architecture to tackle flat network deployments (Section 4.1); *ii)* a PMIPv6-based approach, extending the standard PMIPv6 operation to operate in a distributed way (Section 4.2) and *iii)* the combination of both approaches (Section 4.3). This article represents many refinements and extensions to our original works [3–5], extending the scope of those solutions and providing significant contributions to the previous work, not only in the design and scope of the protocols, but also performing an important analytic evaluation and practical validation, based on a Linux implementation.

The evaluation of our solution is divided in two parts. We first present in Section 5 an analysis of the overhead an handover latency, comparing each component of our hybrid approach with Mobile IPv6 and Proxy Mobile IPv6. Then we report on the results obtained from an experimental evaluation based on a partial implementation of the solution (Section 6). To the best of our knowledge, this is the first working implementation of a DMM solution. Finally, we conclude this work in Section 7.

2 Background and Motivation

Recent mobile architectures, such as the Evolved Packet System (EPS), are intended to be IP-based both for data and voice communications, triggering a real need for the optimisation of IP protocols for mobile networks. In this scenario, IP mobility management plays a key-role in providing the *always-on* and ubiquitous service envisioned by future technologies.

Most of current mobility management solutions derive from Mobile IPv6 (MIPv6) [6], the first mobility protocol standardised by the IETF for IPv6. MIPv6 (see Fig. 1) enables global reachability and session continuity by introducing the home agent (HA), an entity located at the *home network* of the mobile node (MN) which anchors the permanent IP address used by the mobile node, called home address (HoA). The home agent is in charge of defending the HoA's reachability when the mobile node is not at home (i.e., where the HoA is not topologically valid), and redirecting received traffic to the node's current location. When away from its home network, the MN acquires a temporal IP address from the visited network – called care-of address (CoA) – and informs the home agent about its current location by sending a Binding

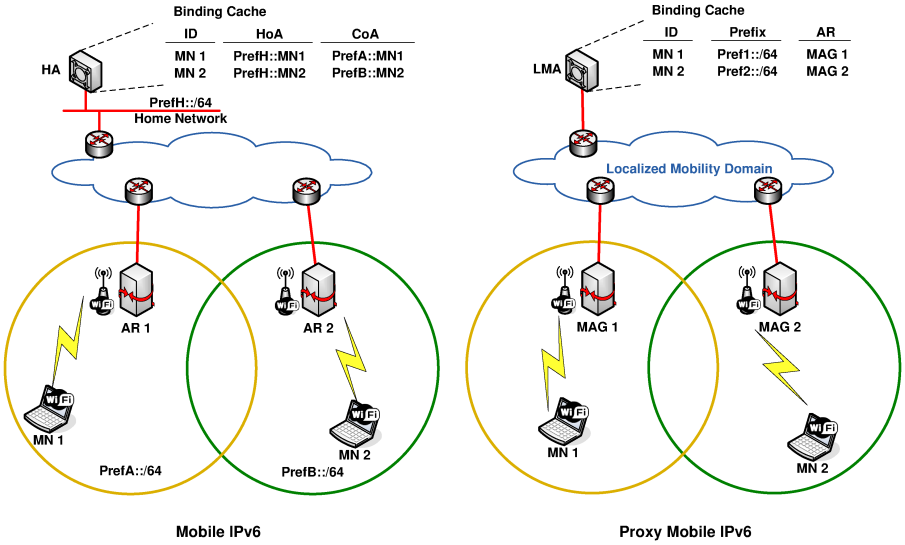

Fig. 1 Centralized IP mobility protocols operation.

Update (BU) message. An IP bi-directional tunnel between the mobile node and the home agent is then used to redirect traffic from and to the MN.

While MIPv6 requires the explicit participation of the mobile node in the signalling procedures (this is referred to as *client-based* mobility), there is also a family of protocols that provide mobility support without the active involvement of the mobile node (the so-called *network-based* mobility). The effort towards network-based mobility management resulted in the standardised Proxy Mobile IPv6 (PMIPv6) [7], developed as an enhancement of MIPv6. In PMIPv6, the home agent is replaced by the local mobility anchor (LMA), the network entity in charge of routing data packets in uplink and downlink containing the IPv6 prefixes assigned uniquely to MNs on a per user basis, the home network prefix (HNP), and also storing the MNs' mobility sessions information (see Fig. 1).

PMIPv6 evolved from MIPv6 by relocating relevant functionalities for mobility management from the MN to a network node, called the mobile access gateway (MAG), which is the first IP hop and default gateway seen by the terminal. In PMIPv6 indeed, mobility is transparent for MNs: the network learns through standard terminal operation, such as router and neighbour discovery [8], about MN's movements and coordinates routing state information using Proxy Binding Update (PBU) and Proxy Binding Acknowledgment (PBA) messages. The former is sent by the mobile access gateways to the local mobility anchor to indicate the mobile node's location, and the latter is sent as a response to ensure that the procedure succeeded. The LMA stores a binding cache entry (BCE) containing the MN's identifier, its prefix and the serving MAG's address, called proxy care-of address (Proxy-CoA). Users' traffic is encapsulated between the LMA and the Proxy-CoAs. The set of deployed

MAGs and the corresponding LMA forms the *localised mobility domain*, in which mobility is completely transparent to the IP stack of the mobile nodes.

As described above, currently standardised IP mobility solutions come at the cost of handling operations at a central point – the mobility anchor – and burdening it with data forwarding and control mechanisms for a great amount of users. This central anchor point is in charge of tracking the location of the mobile and redirecting traffic towards its current topological location. While this way of addressing mobility management has been fully developed by the Mobile IP protocol's family and its many extensions, it brings several limitations: *a)* sub-optimal routing, as traffic always traverses the central anchor, leading to paths that are, in general, longer than the direct one between the mobile node and its communication peer; *b)* scalability problems, as existing mobile networks have to be dimensioned to support all the traffic traversing the central anchors, and the anchor itself has to be powerful enough, and; *c)* reliability, as the central entity is a potential single point of failure.

In order to address these issues – which will soon start to become an operational problem for large-scale mobile network operator – a new paradigm has gained momentum recently: the so-called Distributed Mobility Management [1, 9]. DMM basically develops the concept of a flatter system, in which the mobility anchors are placed closer to the user, distributing the control and data infrastructures among the entities located at the edge of the access network [10].

Most of existing DMM proposals are based on extending or modifying already existing IETF protocols, due to the benefits inherent to extending an already accepted protocol such as PMIPv6 or MIPv6 in 3GPP standards. In the next subsection, we provide an overview of existing DMM proposals.

3 State of art on DMM

It was already mentioned in Section 2 that a key characteristic for a mobility management protocol is the entity in charge of the mobility management; hence, the solutions are distinguished into client and network-based. These categories are preserved when designing a DMM protocol, but, in addition, the latter is further split according to the level of distribution of the control plane [1, 32]:

- *Partially distributed solutions* are characterised by completely distributing the data path, but keeping the control plane centralised. In this way, a central entity is in charge of handling the users' mobility sessions (or other relevant mobility parameters), while the data forwarding role is responsibility of the anchors only.
- *Fully distributed solutions* are the ones that completely distribute both the data and control planes. In this case, there is no longer a centralised control entity.

Besides this first conceptual division, that applies to all mobility protocols, the design of a DMM solution may follow different approaches, namely: *i)*

clean-slate approaches, proposing novel network architectures tackling from the foundation the problems inherent to current IP mobility architectures, *ii)* architecture-dependent solutions, such as the different efforts initiated in the 3GPP (e.g., LIPA, SIPTO and LIMONET), *iii)* peer-to-peer approaches, distributing the mobility management functionality across a P2P network, and *iv)* solutions based on or extending existing IETF protocols (see an overview on this concept in [33]).

3.1 Clean slate approaches

The more relevant *clean slate approach* that has been discussed in the DMM related forums is the one presented in [11]. This solution presents a novel approach that breaks with current trends on mobility management. It proposes the use of routing updates between routers to manage the mobility of the nodes within the mobility domain. It relies on DNS updates and lookups to detect the prefix assigned to the node and BGP update messages to renew the information in the routing within the domain. Global roaming is also supported by issuing BGP route updates between several ASs. Although the proposal has been subject to deep discussion within the IETF, there is still a lack of a deep analysis of its expected performance. This protocol can be accounted within the client-based category as the MN has the responsibility to update its location in the DNS server.

The Dynamic Mobility Anchoring (DMA) presented in [12] is a generic solution for mobility management in flat IP networks that can be included in the clean-slate family. Mobility management is offered on a per IP flow basis. Indeed, the design encompasses two roles for an access node, depending on the service offered to the data flows generated by an MN: first, the access node can behave as a visited access node (VAN) when the functionality provided to the MN includes only the provision of IP connectivity. Second, an access node can become an anchor access node (AAN) when it is in charge of anchoring MN's IP flows after it has moved to a different VAN.

Packets arriving at the AAN are forwarded to the correct VAN by means of an IP tunnel. This tunnel is established without requiring any extra signalling with the access nodes. A VAN learns the corresponding AAN through packet inspection of uplink traffic. Similarly, an AAN learns the current VAN when receiving encapsulated traffic. In order to address the situation in which there is no uplink traffic, the mobile node is required to send uplink void packets to timely recover connectivity with an AAN. The side effect of this approach is the introduction of unnecessarily latencies at handover execution. This proposal is evaluated in [34] through simulations, but the lack of a real implementation does not reveal any feasibility evidence of such solution. This design is extended in [35] to support a prefix relocation mechanism, capable of relocating the prefixes used by the mobile node to prefixes allocated to the serving access router. This requires mobile node modifications to indicate to the network the best moment to perform the relocation.

Other existing clean slate approaches leverage the concept of identifier/locator split to provide flatter architectures. In [13], the authors present a novel approach called HIMALIS (Heterogeneity Inclusion and Mobility Adaptation through Locator ID Separation) that advocates for mobility management built on top of the concept of locator and ID separation. End host traffic is routed through the optimal direct path by the swapping of the locators used in the communication, while the connection is not closed as the identifiers are kept constant. The functionality provided by HIMALIS resemble classical approaches such as HIP [36] or SHIM6 [**?**], being not only as a protocol for mobility management but rather a new architecture for networking. In the same way as the HIP/SHIM6 approaches, its main drawback lays on the difficulties to deploy it, given that the hosts' IP stack is considerably changed.

A similar approach is followed in [14], where the Locator/Identifier split is obtained through the use of the Locator Identifier Separation Protocol (LISP) [38] . The draft indeed proposes a solution called DMM-LISP. As mentioned for the previous work, these concepts do not just apply to mobility, but they address a wider problem space, therefore they encounter tough obstacles for their deployment.

3.2 Architecture dependent solutions

Regarding the second category, *architecture dependent solutions*, it is worth mentioning the relevance of the work being performed in the 3GPP along the lines of flattening the network and distributing the anchors. The 3GPP is currently looking for solutions specifically focused on providing enhanced mechanisms for local breakout, offloading and, simply put, reducing the volume of users data traffic that transits through the operator core network, hence alleviating their overloaded networks. There are several standardisation efforts, such as Selected IP Traffic Offload (SIPTO) and Local IP Access (LIPA) [39], or its extension to improve their mobility capabilities: LIPA Mobility and SIPTO at the local Network (LIMONET) [40].

These subjects are taken into consideration by Hahn in two different articles: [15] and [16]. Both works provide complementary solutions for packet data network gateway (P-GW) relocation within the 3GPP Release 10 specification. The main idea proposed by both works is the definition of new mechanisms for application aware non-optimal path detection.

Additionally, the work in [17] explores the deployment of client and network based DMM solutions in the EPS architecture, providing a detailed description of the required operations and the re-use of the architectural elements and interfaces to support also non-3GPP access. Authors of such document describe how to adapt DSMIPv6 [41] for the client-based solution, while for the network-based approach two variants are presented, extending the GPRS Tunnelling Protocol (GTP) [**?**] and PMIPv6.

3.3 Peer-to-peer approaches

One of the key aspects of the DMM concept is the distribution of the mobility management functionality across multiple entities. *Peer-to-Peer (P2P)* paradigms are naturally prone to devise the interaction of such entities. Relevant works using this approach are [18], [20], [21] and [19].

In [18] the authors present m-Chord, a protocol used to distribute the home agent and foreign agent functionality of Mobile IPv4. Their performance analysis concludes that in some cases their solution performs even better than standard Mobile IP, although in the general case, there is a performance drawback from the use of the P2P technology.

Similar to the previous work, [19] presents a solution for mobility management that distributes the functionality of the home agent across multiple nodes through the use of a P2P approach. The protocol selected for the distribution of the information is Chord. In this solution, MNs and CNs are enabled with a MIPv6-capable module. During handover the MN sends BUs to all the CNs to timely inform them about the new mobility parameters. The authors argue that one of the main drawbacks of using P2P overlays for mobility management is the lack of coherence between the overlay and the actual physical topology of the nodes. Hence they propose to extend the P2P protocol to consider physical information through a Markov decision process, optimising the update and query performance.

In [20], a new mobility management protocol based on Distributed Hash Tables (DHT), called Distributed IP Mobility Approach (DIMA) is presented. The protocol is similar to Mobile IP but the home agent functionality is split and spread across different nodes that share a common binding distributed database. The data traffic towards the mobile node is intercepted by one of these nodes, which acts as home agent, anchoring the mobile node's home address. The distributed mobility is achieved by relocating the nodes acting as distributed home agents, closer to the mobile nodes. Differently from MIPv6, the MN does not take active part in handling location updates, as the set of home agents are in charge of transmitting the Binding Update and Acknowledgment messages.

Finally, the article [21] also describes a DMM solution that leverages a DHT storing the MNs' ID/location pairs. Nevertheless this can be accounted as a client-based solution, because the entities located at the edge of the network are responsible to handle the MN's mobility context coordinated by the messages exchanged with the MN itself, which employs a dedicated handoff module. The session continuity during handover is granted by the bi-casting mechanism. Authors claim that their DMM solution is less demanding than Fast Handover MIPv6 (FMIPv6) [42] in terms of signalling cost.

3.4 Extension of existing protocols

There are several benefits inherent to the *extension of already established protocols* to support DMM. In fact, the 3GPP adopted the use of Proxy Mobile IPv6 as an alternative to the well-known GTP and DSMIPv6 [41] as the client-based mobility management protocol of choice. Additionally, this is the preferred approach within the IETF, which is trying to avoid standardising yet another mobility protocol without first studying if the DMM requirements might be achieved by extending existing protocols.

An intermediate step towards DMM has been studying techniques to offload MNs' traffic, as those proposed for MIPv6 in [43] and for PMIPv6 in [44] and [45]. Both approaches are intended to alleviate congestion of networks nodes by moving traffic to different access networks if the MN is capable to be simultaneously connected through multiple interfaces (e.g., 3G and WiFi). These solutions, however, still rely on a single anchor, the home agent or the local mobility anchor, thus not solving the limitations inherent to centralised mobility management protocols.

Therefore, DMM works focusing on Mobile IPv6 based solutions try to reduce the impact of the triangular routing on the overall performance. In [22], the ADA (Asymmetric Double Agents) extension to Mobile IP is presented to optimise handover latency and communication delays. These improvements come at the cost of introducing two new entities in the network, the local mobile proxy (LMP), that takes care of the functionality of the home agent in Mobile IP, but is located closer to the mobile node; and the correspondent mobile proxy (CMP), which is located near the correspondent node to provide an optimised route towards the LMP.

A different approach for reducing the HA-MN delay is taken in [23]. This work proposes a solution that enables the use of multiple home agents distributed through the Internet, interconnected by high speed links and communicated through anycast routing. Hence these nodes can be always placed near the mobile node, in this way reducing all the problems of centralised deployments.

Last, but not least, works based on Proxy Mobile IPv6 are mainly focused on providing route optimisation mechanisms between mobile access gateways. In [24], authors perform an analysis of the different mobility functionalities required in PMIPv6, to then propose a solution splitting these functionalities across several nodes in the network. Nevertheless, the proposed solution uses for the actual routing of the flows a centralised approach, not providing local breakout of the connections, hence no real distributed mobility is achieved.

In [25], route optimisation is proposed in PMIPv6 domains. In this solution, the MAGs serving the MN and CN leverage on the information stored at the LMA to establish a direct tunnel between them, so a better path can be used for the communication. This mechanism still suffers from using a tunnel for the whole length of the data session. Also, the CN is required to be connected to the PMIPv6 domain too.

These drawbacks are partly mitigated in the draft [26], where a different Route Optimisation technique for PMIPv6 is discussed. Authors provide different operation modes, but beyond the specific sequence of messages and operations, the protocol either needs the CN to be connected to the PMIPv6 domain or to be able to interpret some modified Proxy Binding messages.

The proposal described in [27] suggests to split the localised mobility anchor (LMA) of PMIPv6 into two distinct nodes, a control plane LMA (CLMA) and a data plane LMA (DLMA). The former maintains the mobility sessions for the MNs, whereas the second is the anchor for the MNs traffic. The CLMA also assigns the most suitable DLMA to the MNs. This proposal relieve the LMA's burden, but, in general, it does not fit for flat architectures, as the DLMA/MAG hierarchy is preserved, along with the tunnels, which are established for the whole duration of a data session. The solution, however, envisions an operating mode by which, if the MN and CN are under the same CLMA's administration, route optimisation can be set up between the corresponding MAGs.

In [28], three mobility schemes are proposed: signal-driven PMIP, data-driven distributed PMIP and signal-driven distributed PMIP which explore partially and fully distributed solutions. The three mechanisms rely on control/data split (for the partially distributed solution) and multicast or peer to peer communication (for the fully distributed one) to route the data towards the mobile node through the optimal path.

In the article [29], the authors present an extension for Proxy Mobile IP that enables the local mobility anchor to select an entity to handle a given mobile node's flow. The anchoring function will follow the mobile node as it roams across the mobility domain. The new entity in charge of performing route optimisation between the MAGs is called intermediate anchors (IA). This entity is in charge of establishing tunnels with old and new MAGs, hence providing connectivity between them. The main problem of this solution is that it cannot provide the optimal path, but just an approximation to it.

Finally, the DMA proposal in [12] has been modified taking into account the PMIPv6 legacy in [30]. The modified DMA solution relies on mobility capable access routers (MAR) that exchange PBU and PBA messages to update the MN's location and addresses used. MARs also interact with a central database to retrieve the mobility sessions and coordinate the routing state for the MNs. An analytical evaluation of such protocol is provided in [46].

The paper [31] is somehow different to all the previous because it tackles DMM for a moving network (NEMO). Authors propose to use a PMIPv6-based DMM solution similar to [30] to provide mobility support to a network moving around the mobility domain, for instance for a automotive scenario.

To finalise this section, Table 1 presents the summary of the different related works, highlighting their main characteristics and classifying each of them following the proposed taxonomy.

Table 1 Summary chart of main DMM solutions

Solution	Client-based	Network-based	
		Partially	Fully
Clean slate approaches			
McCann [11]	BGP/IBGP/DNS based		
DMA-Bertin [12]			Automatic learning
HIMALIS [13]	Loc./ID split		
Zhang [14]		LISP-based	
Architecture dependent solutions			
Hahn [15, 16]			P-GW relocation in 3GPP EPC
Bernardos [17]	DSMIPv6-based for 3GPP EPS		GTP/PMIPv6-based for 3GPP EPS
P2P appraches			
m-Chord [18]			multiple HAs and FAs interact through Chord
Zhai [19]	MIPv6 and Chord based		
DIMA [20]			DHT updated with BU and BA messages
Yu [21]	Loc./ID pairs stored in DHT		
Extension of existing protocols			
Liu [22]	MIPv6 based		
Wakikawa [23]	MIPv6 based		
DMA-Chan [24]			PMIPv6 based
Ernst [25]		RO for PMIPv6	
Xue [26]		RO for PMIPv6	
D-PMIPv6 [27]		LMA split into CLMA and DLMA	
Jung [28]		PMIPv6 based	PMIPv6 based
Anchor PMIPv6 [29]		LMA for control plane IAs for data plane	
DMA-Seite [30]		PMIPv6 based	
Do [31]		PMIPv6 based for NEMO	

4 HDMM: hybrid DMM for future mobile network operators

In this section we describe our proposal, called HDMM: Hybrid Distributed Mobility Management for future mobile network operators. It is based on extending current IP mobility standards, namely Mobile IPv6 and Proxy Mobile

IPv6. The rationale behind this is the following: we believe that the difficulties posed by deploying a new architecture are in contrast with how networks are operated. We have seen that the Mobile IP family of protocols is already present in most of current standards for cellular/metropolitan mobile networks (e.g., 3GPP, WiMAX), and the main efforts are focused on evolving current infrastructures due to the huge cost of its deployment. Also, we believe operators prefer to have as much flexibility as possible in terms of the mobility solutions they can put in place, as different scenarios impose different requirements, which can be more easily met by complementary mobility approaches:

- *Network-based mobility approaches* do not require any specific IP mobility support on the mobile node, which allows for an easier deployment in some situations. On the negative side, this kind of approach makes more challenging inter-technology mobility and inter-domain roaming, as some security associations have to be in place, and this is not always possible when crossing operator boundaries. HDMM extends Proxy Mobile IPv6 to operate in a distributed way, which can include control and data planes, or just data plane, as will be described later in this article.
- *Client-based mobility approaches* do require specific IP mobility support on the mobile node, as well as potentially complex security configurations. However, if the support is available, and the mobile node can be properly provisioned, this approach provides more flexibility, as it is easier to perform mobility management when the mobile node plays an active role. Besides, inter-domain mobility becomes easier, as there is no need for deploying security associations between network entities belonging to different operators, just between the mobile node and the home agent. In this case, HDMM extends Mobile IPv6 to support its distributed operation, as well as its combination with the network-based operation mode, for those cases in which this feature is required.

It is important to note that HDMM comprises two solution components that can also be used independently. We next describe each of this components in Sections 4.1 and 4.2, and then explain how their combination can be used to provide a unified inter- and intra-domain mobility solution (Section 4.3).

Before describing how each solution component works, we introduce some common terminology that is used throughout the article:

- *Distributed Anchor Router (DAR).* It corresponds to the first IP router (with mobility functionality) which a mobile node attaches to. Upon attachment, the distributed anchor router provides the mobile node with a topologically correct IPv6 address/prefix. In case the mobile node later moves to a different location, this DAR is in charge of ensuring the reachability of the previously delegated address/prefix. In this way, the DAR can be actually considered as a distributed version of the anchors defined by the classical centralised mobility protocols: the home agent and the local mobility anchor.
- *Serving DAR (S-DAR).* This term is used to refer to the distributed anchor router where the mobile node is currently connected. Note that the mobile

node may have visited different DARs before, and might still be using addresses configured from some of them. As described later, this entity can be considered as a modified version of the mobile access gateway (for the case of the network-based component of HDMM). In this article, we consider, for simplicity, that a mobile node can only be attached to a single serving distributed anchor router at a time.
– *Previous DAR (P-DAR).* This term denotes a distributed anchor router previously visited by a mobile node, and which is still anchoring one or more active IP flows of the mobile node. For a given mobile node, there might be multiple P-DARs.

4.1 Client-based HDMM component

This section presents the client-based component of HDMM, which is basically a distributed version of Mobile IPv6. Following this idea, the functionality of the Mobile IPv6 centralised anchor – the home agent – is distributed and moved to the edge of the network, so an instance of it is deployed in each default gateway the mobile node attaches to, which we refer to as distributed anchor router (DAR), as introduced before. In the following we assume the presence of at least one DAR per access network.

4.1.1 Solution overview

The client-based component of HDMM operation can be summarised as follows. At every mobile node attachment to a new access network, served by a distributed anchor router, the MN configures an IPv6 address delegated and locally anchored by the S-DAR: this address can be seen as a home address for the home network managed by the S-DAR. If the mobile node previously visited other access networks in which there was a distributed anchor router deployed, and there are active flows using addresses delegated by some of them, the mobile node can maintain the reachability of these addresses. This is done by sending a regular Mobile IPv6 Binding Update message to each of the previous DARs anchoring an address used by an active flow (distributed anchor routers are effectively playing the role of home agents), using the address configured at the serving DAR as care-of address. A bi-directional tunnel is established between the mobile node and the anchoring P-DAR for each of the home addresses, which is then used to forward the respective data traffic. In this way, active connections requiring mobility service are maintained, while new sessions can make use of the last configured IPv6 address (i.e., the one delegated by the S-DAR), hence using an IPv6 address that is topologically correct at the current mobile node's location. Compared with regular Mobile IPv6, the client-based component of HDMM introduces the use of several (distributed) home agents, and the additional intelligence on the mobile node to be able to simultaneously manage several home addresses and tunnels, as well as to effectively select the best possible source address for new connections.

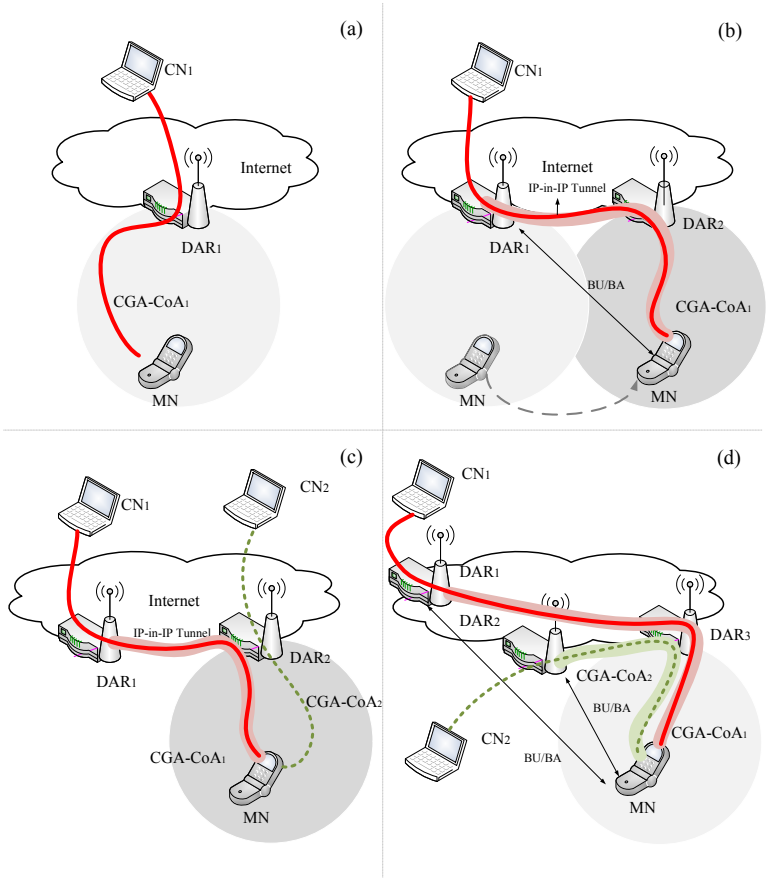

Fig. 2 HDMM client-based component operation.

Note that the operation of HDMM is fully compatible with legacy centralised home agents, as it might be required for some traffic to traverse the mobile network operator's core (e.g., because of service agreements, location privacy or simply for sessions that are known in advance to be long-lived and it is more efficient to anchor them centrally).

Although the operation of client-based HDMM and Mobile IPv6 are very similar, the distributed operation of HDMM poses additional requirements in terms of security. In traditional Mobile IPv6, communication between the mobile node and the home agent are secured through IPsec [47]. Following a similar approach in HDMM would be challenging due to the large number of security associations that would be required, since any distributed anchor router deployed in an access network can play the role of home agent for any mobile node. In order to overcome this problem and provide authentication between the distributed anchor routers and mobile nodes, we propose the use of cryptographically generated addresses (CGAs) [48], as introduced in [49].

CGAs are basically IPv6 addresses for which the interface identifier is generated by computing a cryptographic one-way hash function from a public key and the IPv6 prefix[4]. The binding between the public key and the address can be verified by re-computing the hash function and comparing the result with the interface identifier. To authenticate a message, the packet is signed with the corresponding private key, hence the receiver is able to authenticate the message with the knowledge of the address and the public key. CGAs are a powerful mechanism allowing packet authentication without requiring any public-key infrastructure, and hence it is well-suited for this application.

We next make use of an example, shown in Fig. 2, to complete the explanation of how the client-based component of HDMM works. When a mobile node attaches to a distributed anchor router, the mobile node configures a CGA from a prefix anchored at the DAR (e.g., by using stateless address auto-configuration mechanisms). This address can then be used by the mobile node to establish a communication with a remote correspondent node (CN) – see Fig. 2-(a) – while attached to the S-DAR. If the mobile node moves to a new distributed anchor router, and in case maintaining reachability of the addresses configured at P-DARs is required (e.g., a VoIP call is in place), then Mobile IPv6 procedures are triggered. Following the Mobile IPv6 standard operation, the mobile node sends a Binding Update message to the P-DAR, using the address configured at the previous DAR as home address, and the address configured at the new DAR as care-of address. This BU includes the CGA parameters and signature, which are used by the P-DAR to identify the mobile node as the legitimate owner of the address. Once the signalling procedure is completed, a bi-directional tunnel is established between the mobile node and the P-DAR where the IPv6 address is anchored, so the mobile can continue using that IPv6 address, as shown in Fig. 2-(b).

Any P-DAR serving the user may start behaving as its home agent as soon as the mobile node hands off to a different network, if session continuity is required for an active flow using an address anchored by the P-DAR, and it will keep providing this functionality as long as there are active sessions using that address. Therefore, it is possible that a P-DAR receives multiple Binding Update messages while the mobile node is roaming through different networks. Upon reception of a BU message, the P-DAR has to process the CGAs being carried on the message. Although the use of CGAs does not impose a heavy burden in terms of performance, depending on the number of mobile node sessions handled by the P-DAR, the processing of the CGAs can be troublesome. To reduce the complexity of the proposed solution, we suggest an alternative mechanism to authenticate any subsequent signalling packets exchanged between a mobile node and the P-DAR. This alternative method relies on the use of a Permanent Home Keygen Token (PHKT), which is then used to generate the Authorisation option that the mobile node includes in all subsequent Binding Update messages. This token is forwarded to the mobile

[4] There are additional parameters that are also used to build a CGA, in order to enhance privacy, recover from address collision and make brute-force attacks unfeasible.

node in the Binding Acknowledgment message sent in reply to the first BU. For any subsequent movement requiring to maintain the reachability of an address for which the MN has already sent a BU, the following BU messages can be secured using the PHKT exchanged before, reducing the computational load at the receiving P-DAR.

Another security threat that is specific to HDMM is the possibility of a redirection attack, where a malicious node tries to use an incorrect care-of address in a Binding Update message. By doing so, the attacker could achieve a DoS attack by redirecting any session established using the *spoofed* CoA. Indeed, this is a threat that must be tackled separately since the CGA approach only provides proof of message authenticity (e.g., it assures that the BU message is sent by the legitimate HoA's owner) but it does not provide proof of reachability at the CoA. In order to provide a more robust solution, we propose a return routability (RR) procedure similar to the one defined for the Mobile IPv6 route optimisation mechanism. A return routability procedure is initiated after a handover, so instead of directly sending a BU message, the mobile node first sends a Care-of Test Init (CoTI) message to the respective P-DAR. This message is replied by the P-DAR with a Care-of Test (CoT) message containing a CoA Keygen Token. The mobile node can now send a BU using both Home and CoA Keygen tokens to proof its reachability at both the HoA and the CoA. The message and the knowledge of both tokens is a proof that the mobile node is the legitimate node who has sent the Binding Update message and also is reachable at the CoA indicated. This last security improvement incurs in a performance penalty, namely an increase in the handover delay. The enhanced security approach requires four messages to be exchanged between the mobile node and the P-DAR, instead of the two messages of the original solution. In terms of handover delay, it increases it by a factor of two, as the new solution requires an amount of time equal to two MN-to-P-DAR round trip times (RTTs) to conclude, instead of just one. The performance of the solution is analysed in detail in Section (Section 5).

To conclude the explanation of the protocol, it is worth highlighting that at every attachment to a distributed anchor router, the terminal obtains a new IPv6 address which is topologically anchored at that serving DAR. This address can be used for new communications (avoiding in this way the tunnelling required when using an address anchored at a different DAR), as shown in Fig. 2-(c). A mobile node can keep multiple IPv6 addresses active and reachable at a given time, but this requires the MN to send – every time the MN moves – a BU message to all the previous DARs that are anchoring the IP flows that the MN wish to maintain. For instance, in the example depicted in Fig. 2-(d), the mobile node sends a BU to the first DAR containing CGA-HoA as home address, while the BU it sends to the second DAR contains CGA-CoA1 as home address.

4.2 Network-based HDMM component

This section describes the network-based component of HDMM, which is basically a distributed version of Proxy Mobile IPv6. Both the MAG and LMA functionalities are implemented by the distributed anchor router (DAR) entity introduced before.

The network-based HDMM component is characterised by the split between the control and data plane. In the following sections we present two approaches for the control plane implementation, differentiated by their degree of centralisation. First, a partially distributed approach is presented in Section 4.2.2, which relies on a central entity to keep track of the movement of the users and the previous anchor points. Second, in Section 4.2.3 a completely distributed version, in which the control plane does not use any central entity but messages between the involved anchors, is explained. It is important to highlight, that both approaches for the distribution of the control path share the same data forwarding scheme, which is illustrated next.

4.2.1 Data plane management

A serving DAR provides IP connectivity to the mobile node through a locally anchored IPv6 address. Packets using that address are forwarded by the S-DAR without encapsulation, as a plain IPv6 access router, both in downstream and upstream directions. If the mobile node moves, a new IPv6 address is obtained from the new S-DAR, which is (in general) preferred by the MN to start new IP flows, so packets benefit from optimal routing. However, ongoing data sessions still need reachability of the old address. Hence a bi-directional tunnel is setup between the S-DAR and the previous DAR to not disrupt the communication. Borrowing PMIPv6's definition, the S-DAR behaves as a MAG, and the P-DAR as an LMA. The MN may have hence a number of flows directly routed by the S-DAR to and from the global Internet without encapsulation, and another set of streams anchored to the P-DAR. Depending on the previous MN's movement history and the active sessions, this situation might be replicated for multiple P-DARs.

At the control plane level, the key element to achieve this traffic configuration is to let the S-DAR interact with the P-DARs so that the correct routing state can be set up. This concept leads to the definition of a partially distributed scheme first.

4.2.2 Partially distributed approach

This solution leverages a central entity to store the mobility sessions and maintaining the state about S-DAR and P-DARs for all the MNs in the domain. We name this entity as central mobility database (CMD), and basically it implements all the tasks related to keeping the Binding Cache up to date, as a regular PMIPv6 LMA does, updating its entries with the information received from the DARs. However, its operation differs from the one of a legacy LMA

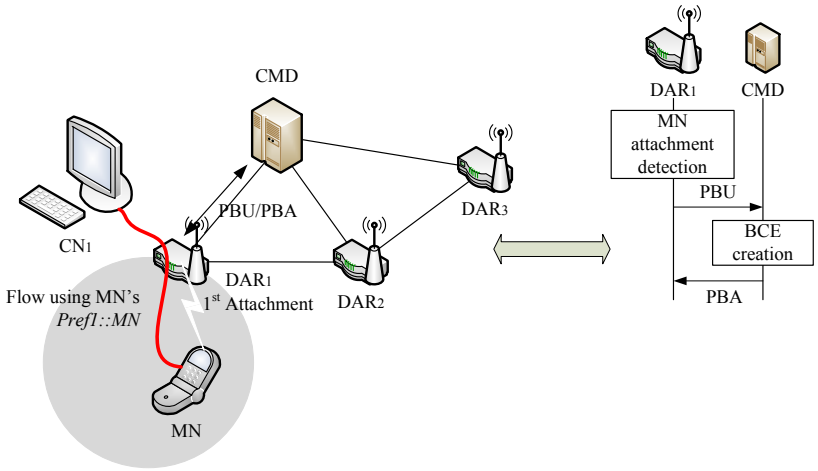

Fig. 3 Partially distributed network-based HDMM: initial registration.

in that the CMD does not perform any data forwarding task, therefore users' data traffic does not traverse it. Similar concepts can be found in the related work [30,31]; a server acting as mobility/policy store is queried by the serving anchor, which interacts with the anchors indicated in the response to set up the proper routing configuration. A similar functionality can be found in our scheme, although with a key difference. In our proposal, the central server (the CMD) does not passively provide the response, but it rather takes an active role forwarding the messages to the MN's P-DAR(s), since it is the entity in possession of the whole picture in terms of P-DARs and prefixes allocated.

The following paragraphs describe how the initial registration to the domain is performed and how the handover is handled.

Initial registration: Upon mobile node's attachment to a DAR (see Fig. 3), say DAR_1, the MN's unique identifier in the domain (MN-ID) is retrieved, and an IPv6 global prefix belonging to the S-DAR's prefix pool is reserved for it ($Pref_1$). The pair MN-ID and the prefix are stored locally as part of a temporal binding cache entry (BCE) at the DAR.

These parameters are conveyed to the CMD in a PBU message. Since the MN is attaching to the domain for the first time, the CMD has no previous entry for it. Hence a fresh BCE is created, containing as main fields the MN-ID, the MN's prefix and DAR_1's address (the proxy-CoA in the PMIPv6 terminology). The CMD then replies to DAR_1 with a Proxy Binding Acknowledgment (PBA) message, which is mainly a copy of the PBU message received before, meaning that the mobile node's registration is new and no additional information was available at the CMD. DAR_1 finalises the registration for the temporal BCE previously created and unicasts a router advertisement (RA) to the mobile node, including the IPv6 prefix reserved before, that is used by the MN to configure an IPv6 address (e.g., with stateless auto-configuration).

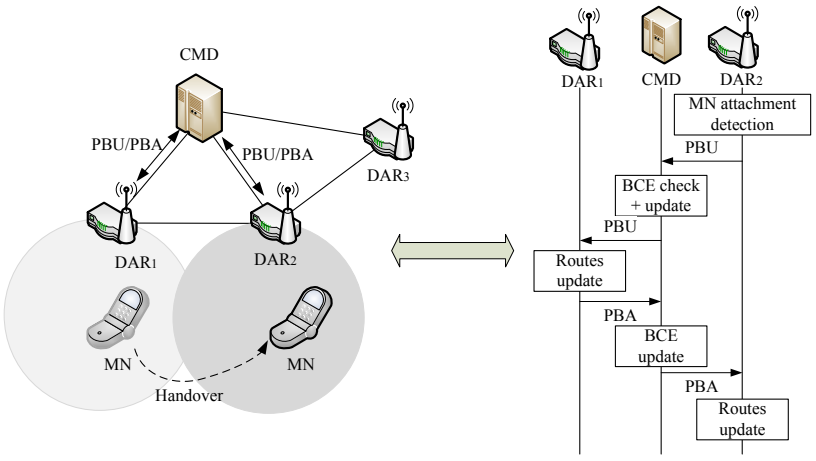

Fig. 4 Partially distributed network-based HDMM: CMD behaves as PBU/PBA relay.

Since this address is locally anchored at the S-DAR, no encapsulation nor special handling is required to route packets of IP flows started there.

When a handover occurs, there are several possible signalling schemes that can actually be used by the DARs to interact with the CMD and set up all the required state in the network. Each approach assigns a different role to the central mobility database, with different pros and cons associated, in terms of handover latency and signalling overhead:

- the CMD behaves as a PBU/PBA relay,
- the CMD behaves as a DAR locator,
- the CMD behaves as a PBU/PBA proxy.

The CMD behaves as a PBU/PBA relay: When the MN moves from its current access and associates to another DAR (see Fig. 4), say DAR_2 (now the S-DAR), the L3 handover is handled in 4 phases:

1. DAR_2 reserves an IPv6 prefix ($Pref_2$) from its local pool, storing it in a temporal BCE, and sends a plain PBU to the CMD for registration (as the initial registration phase).
2. Upon PBU reception and binding cache lookup, the CMD retrieves an already existing BCE for the MN. The BCE indicates a DAR's address in the P-CoA field, so the CMD forwards the received PBU message to it (in our example DAR_1), appending to the message the S-DAR's global address (DAR_2). The P-CoA is changed indicating the new S-DAR's address.
3. Upon reception of the PBU from the CMD, DAR_1 sets up its end-point for the bi-directional tunnel towards DAR_2 and adds the required routing entries for $Pref_1$. DAR_1 informs the CMD that these steps have been successfully performed by sending a PBA message.
4. The CMD, after receiving the PBA, adds an item in the BCE called P-DARs list. An entry of the P-DARs list is composed by the pair P-DAR's

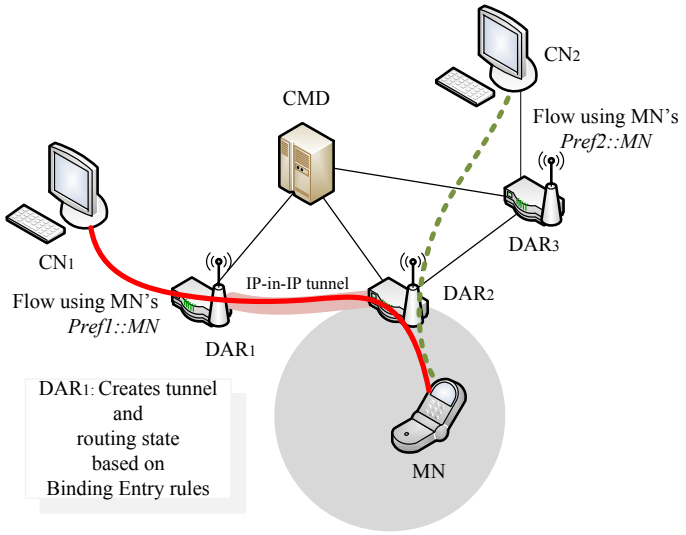

Fig. 5 Partially distributed network-based HDMM: data flow.

address and the prefix it allocated to the MN (in our example DAR_1's address and $Pref_1$). Finally, the CMD sends a PBA to the current S-DAR, which includes the P-DAR's address and the associated anchored prefix. This message enables the S-DAR to finally establish the correct routing state, i.e., the bi-directional tunnel with the P-DAR (DAR_1) and the routing entries for $Pref_1$.

5. The S-DAR advertises the local anchored prefix to the MN, and sends an additional RA including the old prefix but indicating a non zero valid lifetime and a zero preferred lifetime. In this way the old address can be correctly used to terminate old data sessions, whilst it is deprecated for new ones, forcing the MN to pick the address advertised by the S-MAR.

Fig. 5 illustrates how old and new IP flows are routed in the domain.

Any subsequent mobile node's handover follows the same procedure, involving all the P-DARs that are anchoring active flows incrementally. Indeed, when the CMD receives the first PBU message from the S-DAR, it forwards a copy of the message to the P-CoA and to all the P-DARs indicated in the P-DAR list. All these DARs reply back with a PBA message to the CMD, which then aggregates all the messages into a single PBA sent to the new S-DAR, hence the routing state has been re-configured in the whole domain.

It should be noted that this design separates the mobility management at the prefix granularity, and it can be tuned in order to remove old mobility sessions when not required, while the mobile node is always reachable through the address configured from the latest IPv6 prefix acquired. Moreover, the latency associated to the mobility update is bound to the PBA sent by the farthest P-DAR, that is, the one that takes the longest time to reach the central mobility database. This drawback can be mitigated introducing a timeout at

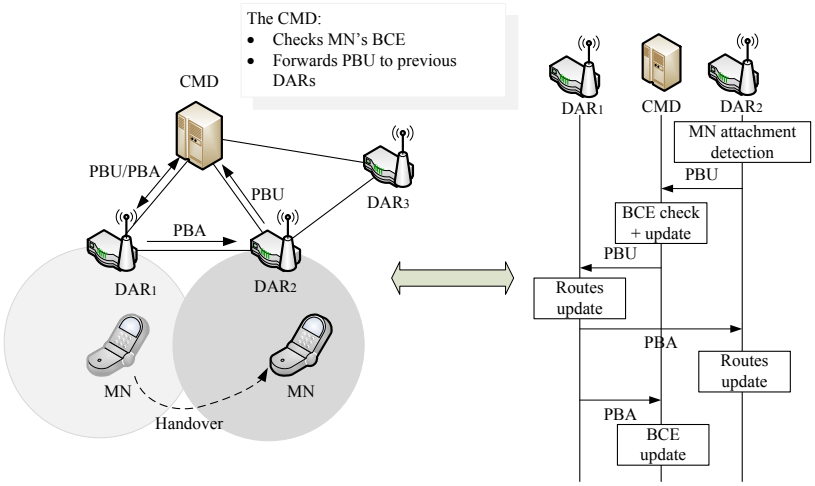

Fig. 6 Partially distributed network-based HDMM: CMD behaves as DAR locator.

the CMD, by which, after its expiration, all the PBAs so far collected are transmitted, and the remaining are sent at a later stage, once they are received.

The CMD behaves as DAR locator: This mobility update procedure follows the same steps defined before up to step 2, the moment when the P-DAR receives the PBU message from the CMD. At that point, the P-DAR is aware of the new mobile node's location (because the S-DAR address is contained in the PBU message). Therefore, the P-DAR signals with a PBA message directly to the S-DAR the prefix it is anchoring for the MN. A similar message is sent to the CMD too, to maintain the consistency in the database. The routing state can be recovered and the procedure is expected to terminate quicker than the previous scheme. Fig. 6 illustrates the new signalling sequence, while the data forwarding remains unaltered.

The CMD behaves as PBU/PBA proxy: Previous mechanism can be further sped up if the CMD simultaneously replies to the new S-DAR with a PBA message and notifies the P-DARs with a PBU. Indeed, the CMD possesses the whole MN's picture, so the serving DAR is notified immediately with a PBA message, including the necessary parameters. In parallel, a PBU message is sent by the CMD to the P-DARs notifying them about the new mobile node's location, so they can establish the required tunnels and routing entries on their side. Every P-DAR, after completing the update, sends a PBA message to the central mobility database to indicate that the operation is concluded and the state has been updated. This scheme is depicted in Fig. 7, where, again, the data forwarding remains the same.

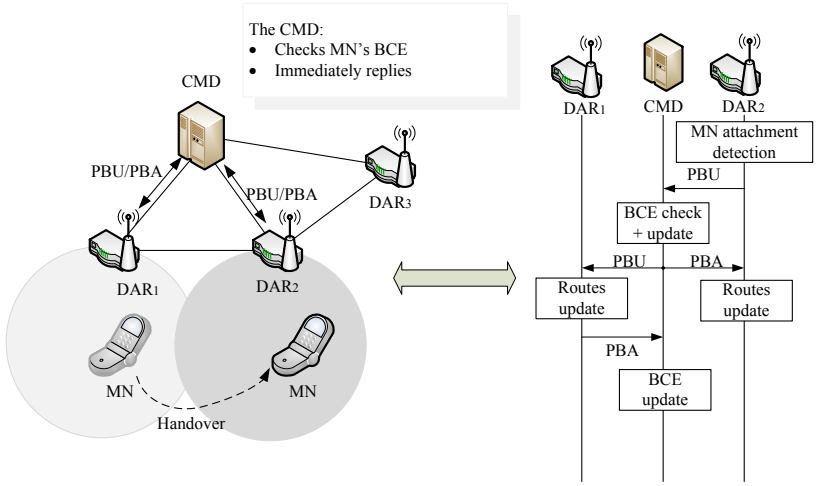

Fig. 7 Partially distributed network-based HDMM: CMD behaves as PBU/PBA proxy.

4.2.3 Fully distributed approach

As introduced at the beginning of the section, we develop in the following the network-based DMM solution in which both control and data planes are distributed. In order to provide a full functional distribution, where there is no central entity in charge of keeping mobile node's state, DARs only interact with each to maintain the mobility sessions up-to-date.

The example used next to provide an overview of the solution is depicted in Fig. 8, where the prefix assignment and routing configuration concepts are identical to the scheme seen in previous paragraphs. The key difference is that the PBU/PBA handshake takes place between the new S-DAR and the P-DAR(s) without the intermediation of other entities. The illustrations in Fig. 8 show how an IP flow is handled when generated at the initial DAR (Fig. 8-(a)), how the flow is routed after a handover (Fig. 8-(b)) and how a second flow started at the new S-DAR is routed in the network as compared with previous flows (Fig. 8-(c)). Finally, Fig. 8-(d) exhibits the handover to a third (and in general, to all subsequent) DAR.

The critical point of this mechanism is how an S-DAR finds out if the attached mobile node has any active flow anchored at a previously visited P-DAR, and, if so, which P-DARs are and what IPv6 prefixes they are anchoring. According to [12] a S-DAR should learn automatically by inspecting uplink MN's packets that the address used by the MN is not anchored to it, hence a tunnel-based re-direction is required to the P-DAR corresponding to that prefix. We argue that this solution does not timely react to the handover event. One might argue that the mobile node itself could provide the serving DAR with the list of active P-DARs and corresponding prefixes, for example by defining new options in the router solicitation message or introducing new notifications. Although this approach would be faster and introduces low over-

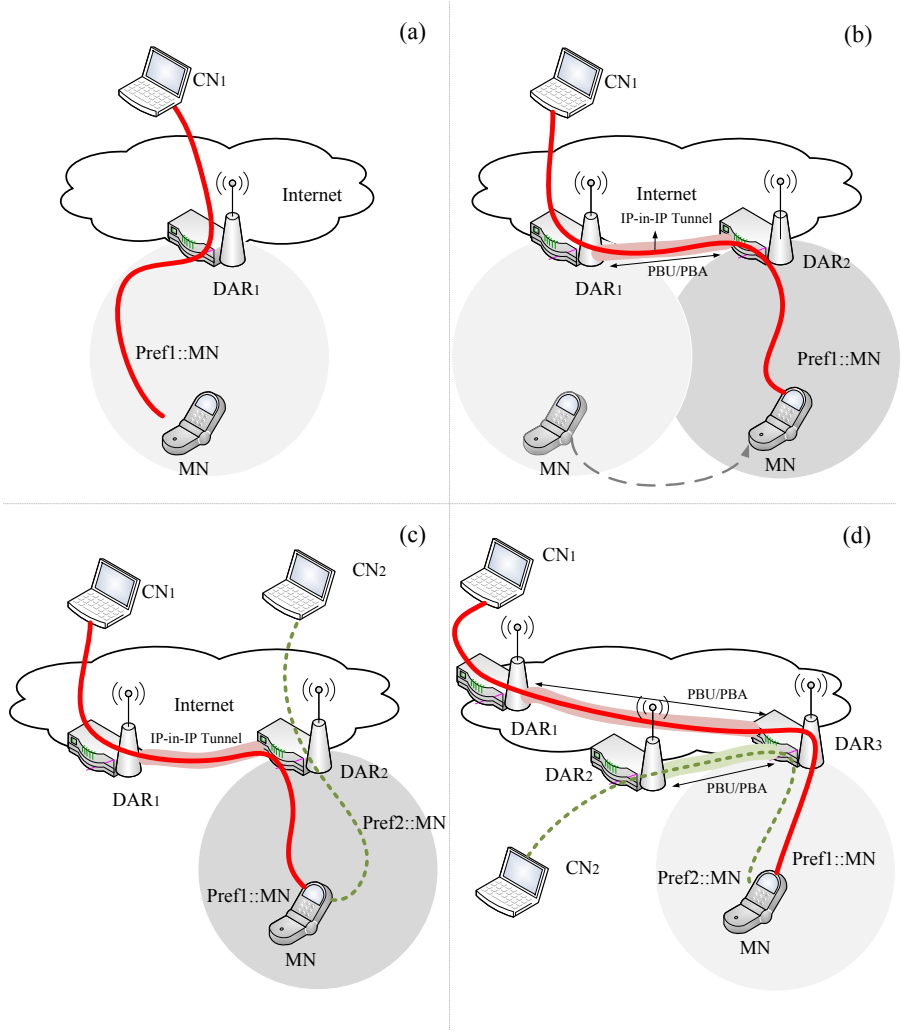

Fig. 8 Fully distributed network-based HDMM: protocol operation.

head, it requires some capabilities on the host that are not always possible to rely on (or not even desirable, as one of the main benefits of network-based approaches is that they do not require any special IP support on the mobile nodes), and might pose some security issues if a malicious mobile node misbehaves.

We here propose the following:

– Multicasting the PBU by the S-DAR to the group formed by all the DARs of the domain. In case no answer (PBA) is received within a timeout interval, the S-DAR may assume this is the first time the MN is joining the network. Unfortunately this approach might not provide a good perfor-

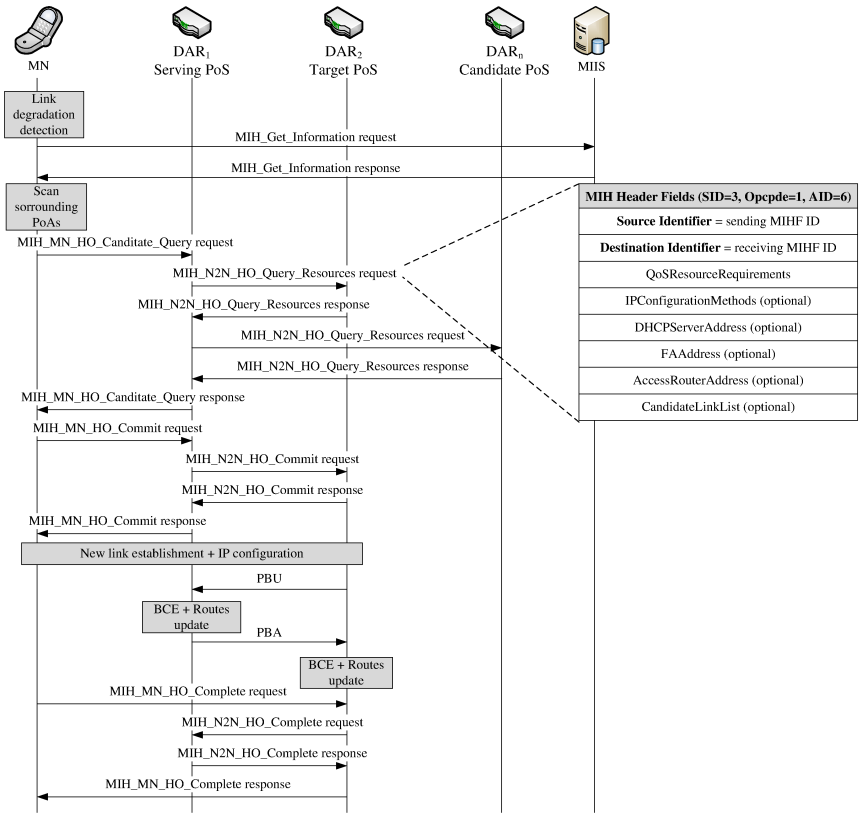

Fig. 9 Fully distributed network-based HDMM: IEEE 802.21-aided message exchange sequence during handover.

mance in terms of handover delay and adds unnecessary signalling in the network.

– Layer-2 handover support through Media Independent Handover Services specification (IEEE 802.21) [50]. The latest revisions of the most used wireless technologies such as IEEE 802.11 or IEEE 802.16, already provide support to the so-called *Link Layer Events*. Through these mechanisms, a network interface is able to indicate changes in e.g., point of attachment or reconnection. Therefore, a handover is handled by a dedicated control plane infrastructure by which the movement is prepared, executed and completed in a controlled and assisted way, according to the *make-before-break* philosophy. Additionally, the IEEE 802.21 suite is intended to allow inter-technology handovers, providing support to mobile nodes roaming within a heterogeneous environment.

For the first approach there is no need for further details, next paragraph is devoted to give some more insights about how to shape IEEE 802.21 to become the control plane of a fully DMM solution.

IEEE 802.21 services are provided by a cross-layer entity called Media Independent Handover Function (MIHF). Its role is to act as an interface that provides communication between lower and higher layers. The services offered by the IEEE 802.21 specification are classified in: *i)* Event Services, where information pertinent to the status of the layer-2 and physical interfaces is forwarded to higher layers (e.g., Mobile IPv6 stack or an application) on a local or remote node; *ii)* Command services, enabling higher layers to command actions and configure some of the interface parameters; and *iii)* Information services, enabling the terminal and network entities to gather information about neighboring networks to help in network selection. The handover of an MIH enabled terminal is in all cases controlled by an entity called Point of Service (PoS). This network entity is able to exchange MIH messages with a peer MIHF installed in the terminal. In a similar way, points of service maintain information about the layer-2 access devices as base stations, access points and similar nodes connected to the DAR, by interacting with the MIHF operating in them. These MIHFs are called Point of Attachments (PoAs), and provide the access link to MNs but they do not establish a direct control communication with the MIHF in the terminals.

Fig. 9 presents the detailed procedure including the IEEE 802.21 signaling required to perform a fully distributed network-based handover. In the figure, DARs and PoSs are co-located, whilst PoAs are omitted to keep the chart simple. Indeed, the diagram highlights the message exchange between the MN and the S-DAR, and among DARs.

According to IEEE 802.21 operations, a point of service learns when a handover for its mobile node is imminent (e.g., through the use of a remote `Link_Going_Down` event). Therefore, the PoS queries the resources availability in the surrounding points of attachment, by means of a message called `MIH_N2N_HO_Query_Resources request`, sent to the corresponding points of service connected to the target PoAs (the latter are discovered by the mobile node, or by the network, after an information retrieval procedure involving the media independent information service – MIIS, and after scanning the PoAs visible by the mobile node). This message contains the current serving DAR's address, that is used later by the new serving DAR to send the PBU message. Once the list of candidate PoAs is filled with the requested information, the target for the handover is selected (either by the mobile node itself or by the point of service) and the corresponding PoS is notified about the decision. The mobile node can now move to the new PoA: the IP mobility procedure is triggered when the PoA announces the attachment to the new serving distributed anchor router (summarized by the "link establishment" text-box in Fig. 9), and then the conclusion phase takes place, releasing the resources in the old DAR/PoS.

No changes are required to the standard IEEE 802.21 primitive messages when only one DAR needs to be contacted with the PBU message (as in Fig. 8-(b)). However, if more than one P-DAR is anchoring active flows of the mobile node, as shown in Fig. 8-(d), then a change in the format of the

`MIH_N2N_HO_Query_Resources request` primitive is required to allow current PoS to send to the candidate points of service the list of all the past P-DARs.

As described above, a fully distributed approach, although perfectly feasible, requires in all cases some support from the mobile nodes, and even the deployment of a whole control infrastructure (as in the case of IEEE 802.21). This might be not desirable, but the deployment of such an architecture would yield to a more scalable and bottleneck-free operator infrastructure, where no single point of failure could bring the network down.

4.3 Hybrid DMM: combining network- and client-based mobility components

Previous sections have described how the individual components of our hybrid distributed mobility management solution work. Although these network- and client-based components can be used as standalone solutions, we argue that future mobile network operators can benefit from a framework allowing a seamless integration of both solutions. We mentioned in Section 1 some example scenarios supporting our claim for the need of a hybrid DMM solution. In this section we pick one of those as driving use-case to explain how our HDMM approach works: inter-domain mobility.

Fig. 10 shows an example of hybrid DMM operation. A mobile node first attaches to a mobile operator and benefits from network-based distributed mobility management (by using either the partially or the fully distributed flavor described in Section 4.2).

While roaming within the same operator network, the security associations required among the involved distributed anchor routers can be easily set up (on demand or can be already pre-configured). If the mobile node moves to an access network managed by a different operator, the new operator might not even support DMM (i.e., there are no distributed anchor routers deployed) or, if DMM is supported, setting up security associations that cross operator boundaries might not be possible. In both cases, using a client-based DMM approach appears as the best possible solution to provide those sessions anchored at the previous domain with session continuity. HDMM supports this by activating the client-based component of the solution, and using the IP address configured on the new domain as care-of address where active sessions anchored elsewhere can be redirected. This operation can be executed exploiting enhanced features of the terminal's connection manager. This latter is a software construct widely available in most of today's portable devices. Thus, the connection manager can be extended including mechanisms to detect that the surrounding point of attachment do not belong to the home operator (for instance, 3GPP access technologies and IEEE802.11u provide the operator's name when the MN scans the radio environment) or extensions to standard EAP authentication between UE and operator core can also provide hints about the support of DMM capabilities. Then, if no DMM support is assumed, the connection manager can activate the mobility client at the UE. Taking into account the different levels of intelligence that my be implemented

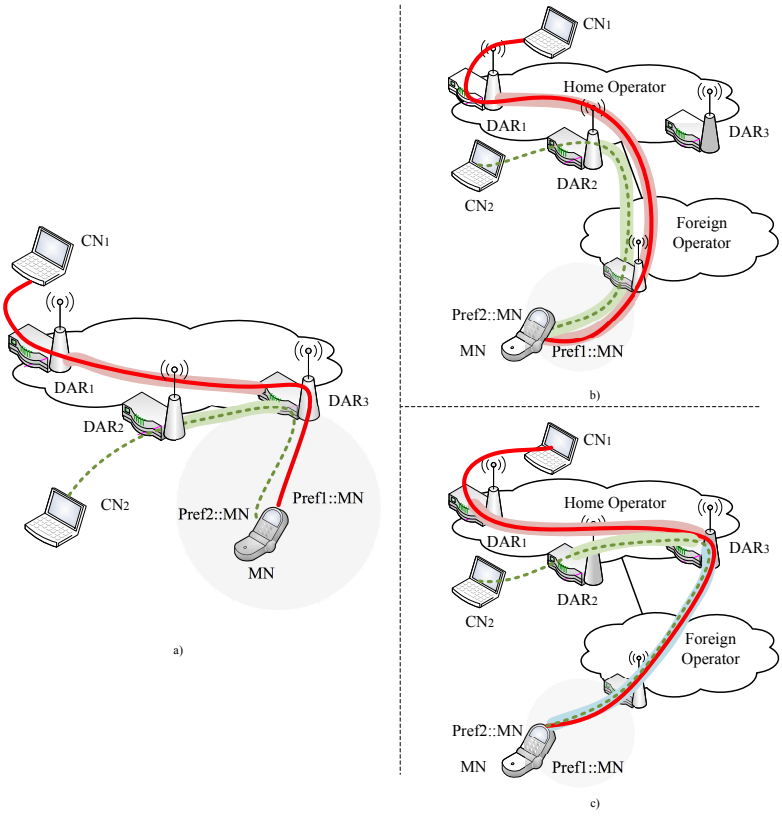

Fig. 10 An example use-case of HDMM: inter-operator mobility.

in the connection manager, two situations may be considered: First, the case where the connection manager has taken track of all the DARs visited by the MN when in the home domain, then the MN can send a BU message to such P-DARs and completely optimise the path between the UE and the DARs in the home network, by establishing tunnels between the MN and the P-DARs to keep the ongoing data sessions (see Fig. 10-a)). Second, taking advantage of the fact that all flows are already tunnelled from the previously visited DARs to the last visited DAR in the home domain, then the MN can simply notify the latest visited DAR and establish a tunnel only with it. Data sessions with old P-DARs will traverse a tunnel segment within the home domain established previously and then another tunnel from the latest P-DAR to the MN (see Fig. 10-b)). Unfortunately, even if the connection manager is a user space software that can be easily installed in modern devices, the client for mobility management cannot be applied to a legacy host's IP stack.

Note that if the new domain also supports HDMM, then subsequent handovers within that domain could be transparently managed by the network-

based mobility solution in place, without requiring any action on the client-mobility stack running on the mobile node[5].

This example clearly shows how a given distributed anchor router can be simultaneously playing – on a delegated prefix basis – the roles of plain IPv6 access router (for prefixes locally anchored used by attached mobile nodes), as local mobility anchor (for prefixes locally anchored that are in use by a mobile node which is no longer attached), as mobile access gateway (to enable address continuity for prefixes anchored at a different DAR) and as home agent (for prefixes locally anchored by a mobile node which is no longer attached and is using the client-based HDMM component).

Next sections are devoted to report on *i)* a performance analysis of the solution, in terms of handover latency and signaling overhead, comparing obtained results to Mobile IPv6 and Proxy Mobile IPv6, and *ii)* an experimental evaluation of the solution, based on real experiments conducted with a prototype of the solution.

5 Analytic evaluation

This section provides an analytic evaluation of HDMM, comparing obtained results with those of Mobile IPv6 and Proxy Mobile IPv6. The analysis is conducted considering the three key performance metrics of a mobility protocol:

- Packet and signalling overhead.
- Handover latency.
- End-to-end delay.

5.1 Overhead Analysis

A common characteristic for both HDMM components is that a mobile node can use the locally anchored address provided by the serving DAR for new communications, benefiting from no additional encapsulation. This is a clear performance advantage compared to centralised schemes, where tunnelling is always used if the mobile node is not at home (with regular Proxy Mobile IPv6, data traffic is tunnelled for the whole MN permanence in the domain).

Note that the locally anchored IPv6 address might also be used not only by new communications, but also by already established ones in which the application is able to cope with an IPv6 address change (e.g., progressive HTTP download). In this way, IP mobility support might be provided only to those applications that require it, feature known as *dynamic* mobility management, which is inherently supported by our HDMM approach.

[5] In this case, the mobile node could actually decide if it prefers to update the care-of address used in the bi-directional tunnels established with P-DARs located at the other domain, or just let the network-based distributed mobility support deployed in the new domain provide address continuity to the care-of address used to set-up the tunnels.

An IPv6-in-IPv6 tunnel adds 40 extra bytes to each packet. More, it consumes processing resources for the encapsulation/de-encapsulation operations and for its management. In network-based solutions, the encapsulation is performed among network nodes, so the extra packet overhead is not present in the last radio link and the processing is done by powerful dedicated equipment. This is actually good in terms of terminal's efficiency, as it does not waste energy to send/receive unwanted bytes in the communication.

We next develop the signalling overhead analysis. As suggested in [51], a general expression to compute the average signalling cost for a mobility scheme is given by the following:

$$C_{SIGNAL} = \frac{1}{SMR \cdot \sqrt{M}} \left[(\sqrt{M} - 1) C_{INTRA} + C_{INTER} \right] \tag{1}$$

where SMR is the *Session-to-mobility ratio*, M is the number of subnets for a single domain, C_{INTRA} and C_{INTER} are the binding update signalling costs for the intra- and inter-domain handover respectively. These costs are proportional to $d(X, Y)$, distance in number of hops from a node X to a node Y[6], multiplied by the link factors τ and ω, for a wired and a wireless link respectively. Therefore, the cost for transferring a packet from the MN to the S-DAR is $C_{MN,S\text{-}DAR} = \omega\, d(MN, S\text{-}DAR) = \omega$, whereas from a S-DAR to a P-DAR it is $C_{S\text{-}DAR,P\text{-}DAR} = \tau\, d(S\text{-}DAR, P\text{-}DAR)$, value that depends on the size and configuration of the network.

We next examine how C_{INTRA} and C_{INTER} are computed, in accordance with the protocol chosen for the purpose.

5.1.1 INTRA domain handover

In HDMM, the intra domain scenario is managed by a network based protocol, to be picked among one of those presented in Section 4.2. This scenario is hence compared with PMIPv6.[7]

PMIPv6 The signaling consists in a PBU/PBA handshake between the new MAG and the LMA:

$$C_{INTRA}^{PMIPv6} = 2\,\tau\, d(LMA, MAG) \tag{2}$$

Partially distributed approach: Depending on the actual procedure used to update the central mobility database, the total signalling load varies:

[6] We assume that the links are symmetric, $d(X, Y) = d(Y, X)$.

[7] However, we remark that the two parts constituting our HDMM design can be used independently. Hence one might deploy a client solution even for intra domain mobility, and the comparison with MIPv6 will be discussed as well.

– CMD behaves as PBU/PBA relay. Besides the handshake between the CMD and the S-DAR, there is an additional PBU/PBA exchange with n active P-DARs. This accounts for a total number of $2n + 2$ messages:

$$\begin{aligned} C_{INTRA}^{Partially-relay} &= 2\,\tau\,d(CMD, S\text{-}DAR) + 2n\,\tau d(CMD, P\text{-}DAR) \\ &= (2n + 2)\,\tau\,d(CMD, DAR), \end{aligned} \tag{3}$$

where $d(CMD, DAR)$ is the average distance between the CMD and the DARs in the domain.

– CMD behaves as DAR locator. In this case, the amount of PBU and PBA messages is $3n + 1$: a first PBU message sent by the new S-DAR, plus n copies sent by the CMD to the active P-DARs, and $2n$ PBA messages sent back by the P-DARs to the CMD and the S-DAR:

$$\begin{aligned} C_{INTRA}^{Partially-locator} &= \tau\,d(CMD, S\text{-}DAR) + 2n\,\tau\,d(CMD, P\text{-}DAR) + \\ &+ n\,\tau\,d(S\text{-}DAR, P\text{-}DAR). \end{aligned} \tag{4}$$

– CMD behaves as a PBU/PBA proxy. Apart from the re-ordering, the number of messages sent is identical to the *relay* case, $2n+2$, thus Eq. (3) holds in this case as well:

$$C_{INTRA}^{Partially-proxy} = C_{INTRA}^{Partially-relay} \tag{5}$$

Fully distributed approach: Regardless the method adopted to learn that a handover occurred, the S-DAR has to perform a PBU/PBA handshake with n active P-DARs. Being n the number of IPv6 addresses that need to be kept reachable and $d_{S-DAR,P-DAR}$ the average number of hops, the result is a total of $2n$ control messages:

$$C_{INTRA}^{Fully} = 2n\,\tau\,d(S\text{-}DAR, P\text{-}DAR). \tag{6}$$

From the above equations, we deduce that, in general, the DMM solutions introduce more messages than PMIPv6, because there are more anchors to update. However, the cost for a fully distributed scheme may be close to the PMIPv6's one, in a scenario in which there are very few P-DARs to be updated and they are much closer to the S-DAR than how the LMA is to the MAG.

Client based approach: In plain Mobile IPv6 there is a single BU/BA exchange per handover (the mobile node uses a single home address and we omit route optimisation), whereas in HDMM we have n BU/BA exchanges (plus the CoTI/CoT ones in case of additional security), where n is the number of IPv6 addresses that need to be kept reachable. This accounts for a total of $2n$ ($+2n$ in case of additional security) control messages. MIPv6 messages traverse a wireless link from the MN to S-DAR, and the wired path from the

S-DAR to the home agent. In client DMM, after the wireless segment, message packets are delivered by the S-DAR to the P-DAR(s). In total we have:

$$C_{INTRA}^{MIPv6} = 2\,[\omega + \tau\, d(S\text{-}DAR, HA)] \tag{7}$$

$$C_{INTRA}^{Client} = 2n\,[\omega + \tau\, d(S\text{-}DAR, P\text{-}DAR)] \quad \text{w/o add. security} \tag{8}$$

$$C_{INTRA}^{Client} = 4n\,[\omega + \tau\, d(S\text{-}DAR, P\text{-}DAR)] \quad \text{w/ add. security}$$

The trade-offs for this scenarios are straightforward: on the one hand, client DMM introduces more traffic at the control plane level, but it allows using optimal or close to optimal routes for data traffic. On the other hand, MIPv6 requires less signaling but all the user's data need to traverse the home agent. The route optimization procedure enables the MN to use an optimal path with the CNs, but all the CNs need to be notified with some signaling, leading to an equal or larger number of control messages than client DMM.

5.1.2 INTER domain handover

For the inter domain scenario, a client solution is better suited to handle mobility. This situation is similar to the client approach mentioned above, except that now the MN is not connected to a S-DAR, but to a generic Access Router (AR). Thus the nodes involved are different:

$$C_{INTER}^{MIPv6} = 2\,[\omega + \tau\, d(AR, HA)] \tag{9}$$

$$C_{INTER}^{Client} = 2n\,[\omega + \tau\, d(AR, P\text{-}DAR)] \quad \text{w/o add. security}$$

$$C_{INTER}^{Client} = 4n\,[\omega + \tau\, d(AR, P\text{-}DAR)] \quad \text{w/ add. security} \tag{10}$$

The same considerations seen before hold in this scenario as well.

5.2 Handover latency

The handover latency corresponds to the time during which an IPv6 address is not usable because of a change of the point of attachment. During this process there are multiple operations performed like the layer-2 attachment, the movement detection, the address configuration and duplicate address detection, and the mobility signaling. In the following we explain the different components of the handover delay:

- Layer-2 handover time (T_{L2}^{ho}). This is defined as the time required by the layer-2 technology to perform a handover (i.e., disconnecting from its current point of attachment and connecting to a new one).
- Movement detection time (T_{MD}). This delay corresponds to the time required by the terminal to detect that it has moved to a different layer-3 point of attachment. In IPv6 this can be done in different ways. The most simple (and the most widely supported) consists in the appropriate use of the router advertisement (RA) messages. An access router periodically

multicasts unsolicited RA messages. Movement detection can also be assisted by the use of layer-2 triggers, such the ones implemented by IEEE 802.21. In this case, the movement detection delay can be extremely low.
- IP address configuration and duplicate address detection (T_{DAD}). This time corresponds to the configuration of the IP address based on the prefix received in the RA (i.e., the MN uses stateless auto-configuration) and the address uniqueness test in the network. Note that DAD is only used for new prefixes in the network-based approach, since old prefixes are maintained from previous allocations and do not require of new DAD processes.
- Network authentication delay (T_{auth}). The handover delay also depends on the particular authentication method used in the network being accessed by the user terminal.
- Mobility signaling delay ($T_{binding}$). This is the time required to update the mobility anchor (i.e., the home agent, the localized mobility anchor or the distributed anchor router) with the new location of the mobile node (denoted by its care-of address or the associated proxy care-of address). It is highly dependent on the distance between the entities participating in the user mobility management. For client-based approaches this is the distance between the mobile node and the home agent/distributed anchor router, while for network-based approaches, this is the distance between the mobile access gateway and the local mobility anchor, or the distance between the serving DAR and the previous DAR, or the the distance between the central mobility database and the involved DARs, depending on the solution flavor.

The handover latency is thus expressed as follows:

$$T_{handover} = T_{L2}^{ho} + T_{MD} + T_{DAD} + T_{auth} + T_{binding}, \tag{11}$$

in which the most relevant component for the comparison is $T_{binding}$. The other delay components can be considered common to any of the analyzed mobility solutions[8]. The term $T_{binding}$ can be expressed, for each of the different scenarios, as follows:

- Mobile IPv6:

$$T_{binding} = RTT_{MN \leftrightarrow HA}. \tag{12}$$

- Client-based HDMM component:

$$T_{binding} = RTT_{MN \leftrightarrow P\text{-}DAR}. \tag{13}$$

- Proxy Mobile IPv6:

$$T_{binding} = RTT_{MAG \leftrightarrow LMA}. \tag{14}$$

[8] Actually T_{auth} has a different form in the client DMM solution when the additional security procedure is in place. We omit this procedure in the analysis.

– Partially distributed network-based HDMM component, CMD behaves as PBU/PBA relay:

$$\begin{aligned} T_{binding} &= RTT_{S\text{-}DAR\leftrightarrow CMD} + RTT_{P\text{-}DAR\leftrightarrow CMD} \\ &\approx 2 \cdot RTT_{DAR\leftrightarrow CMD}. \end{aligned} \tag{15}$$

– Partially distributed network-based HDMM component, CMD behaves as DAR locator:

$$\begin{aligned} T_{binding} &= \frac{RTT_{S\text{-}DAR\leftrightarrow CMD} + RTT_{P\text{-}DAR\leftrightarrow CMD} + RTT_{S\text{-}DAR\leftrightarrow P\text{-}DAR}}{2} \\ &\approx RTT_{DAR\leftrightarrow CMD} + \frac{RTT_{S\text{-}DAR\leftrightarrow P\text{-}DAR}}{2}. \end{aligned} \tag{16}$$

– Partially distributed network-based HDMM component, CMD behaves as PBU/PBA proxy:

$$\begin{aligned} T_{binding} &= \max(RTT_{S\text{-}DAR\leftrightarrow CMD}; \frac{RTT_{S\text{-}DAR\leftrightarrow CMD} + RTT_{P\text{-}DAR\leftrightarrow CMD}}{2}) \\ &\approx RTT_{DAR\leftrightarrow CMD}. \end{aligned} \tag{17}$$

– Fully distributed network-based HDMM component:

$$T_{binding} = RTT_{S\text{-}DAR\leftrightarrow P\text{-}DAR}. \tag{18}$$

For the cases of partially distributed network-based HDMM, we assume that the central mobility database is approximately at the same distance to all the distributed anchor routers ($RTT_{DAR\leftrightarrow CMD}$).

Comparing Eqs. (12) and (13), it is clear that the mean difference between the client-based HDMM component and Mobile IPv6 in terms of handover delay corresponds to the distance between the mobile node and the home agent/distributed anchor router. This is the main advantage of a distributed mobility management approach as compared with classical centralized mobility solutions, because the delay between the mobile node and its anchor is likely lower in the distributed approach as the anchor in this case resides at the edge of the network, instead of at the core of the operator. It is also worth noting how as the mobile node gets farther away from an active previous DAR, the handover delay increases, thus HDMM is better suited for flows with short duration or mobile nodes with low mobility. This characteristic is explored in more detail in the next section.

Similarly, from Eqs. (14) and (18), we can see that the network-based HDMM solution produces a shorter latency as long as the distance between the serving and previous DARs is shorter than the one between the MAG and LMA for the case of Proxy Mobile IPv6. This parameter strictly depends on the size of the operator's network, but, we can safely assume that an LMA would always be always farther than active previous DARs for the case of short communications with limited user mobility patterns.

Moreover, it can be noted by inspecting Eqs. (15)-(17), that the network-based HDMM solution with the CMD behaving as message proxy outperforms

all the others partially distributed proposals. For all partially distributed solutions, as the central mobility database is pushed into the core of the operator, so its distance to all DARs is similar, the handover delay approaches to the one of Proxy Mobile IPv6: $RTT_{LMA \leftrightarrow MAG}$.

5.3 End-to-end delay

We next analyze the delay experienced by packets exchanged between the mobile node and its communication peer (i.e., a CN).

In Mobile IPv6, user data traffic always traverses the home agent, although this path may not be the shortest one between the mobile node and the correspondent node. This way of forwarding packets is known as angular routing and is characterized by delays that might be large, since the packets must go through the MN's home network, which can be located at a long distance from the mobile node. Due to the large delays introduced by the angular routing, Mobile IPv6 [6] already includes a procedure called route optimization (RO) that basically builds a secure direct path between the mobile node and the correspondent node. Thanks to the use of route optimization, packets exchanged between the mobile and the correspondent node can flow directly through the shortest path between the two nodes, without passing through the home agent. This mechanism needs additional support from the correspondent node, required to enable the optimization of the path. In the case of our HDMM approach, packets flow between the mobile node and the correspondent node through the serving DAR as in the case of Mobile IPv6 without RO. The difference between both approaches is that in our case, DARs are expected to be located near the mobile node, hence the effect of angular routing is highly minimised, obtaining delays of the order of RO-enabled Mobile IPv6. In the previous section, it was mentioned that the use of DMM is better suited for flows with short duration or low mobility MNs. The reason for this is the fact that as the mobile node moves away from the serving DAR handling a flow, the inefficiency introduced by the angular routing increases.

In order to assess how far and how fast a mobile node can move, we performed the following analysis. Lets suppose a VoIP communication between two peers, being one of them a mobile node using one of the HDMM schemes to handle its mobility. Considering the maximum mouth-to-ear delay as specified in [52] of 150 ms, we can assume that Eq. (19) holds:

$$T_{CN \rightarrow HOME-AR} + T_{HOME-AR \rightarrow MN} \leq 150ms, \tag{19}$$

in which $HOME-AR$ stands for the serving DAR or HA/LMA according to the solution in place.

Let's assume the correspondent node and the mobile node are in the same geographical region or even city. In order to model this delay, we took average values from the PingER project[9], between several client-server pairs located

[9] Ping end-to-end reporting: `http://www-iepm.slac.stanford.edu/pinger/`

in the same regional area. The average delay obtained corresponds to roughly 20 ms, so Eq. (19) indicates the delay between the HOME-AR and the MN is upper bounded by 130ms. If we consider the network-based HDMM solution, we can assume that the DMM domain has a good internal connectivity. In this way, we can also assume that the delay between two distributed anchor routers is similar to a local delay between two servers located in the same organisation (from the PingER project this delay is on average equal to 5 ms). To simplify, we suppose that the access network is deployed in such a way that going farther away from the first DAR to which the mobile node attached to increases the delay in a linear way (note that this is a worst case scenario). The maximum number of hops allowed for the VoIP communication can then be derived from Eq. (19), resulting in a maximum distance of 26 hops. This number represents a limit on the diameter of the DMM domain, which depends on the access technology used.

In the case of the client-based HDMM component, we could repeat a similar analysis but considering the distance between the distributed anchor routers is longer than in the network-based case. If we assume a inter-DAR delay of roughly 10ms (intermediate value between a regional and local delay), our solution allows approximately 10-13 hops before degrading the VoIP call.

The same delay assumptions hold for centralised approach, but we have to consider also the the angular routing intrinsic to Mobile IPv6 and Proxy Mobile IPv6. For instance, we can assume that the distance between a mobile node and a correspondent node is twice the client-server distance mentioned before: one to get to the HA/LMA, and another to reach the recipient (we can safely assume that the anchor is equidistant from the communication endpoints, as they are all located in the same region). With these assumptions, after 4/5 hops DMM degrades as a centralised scheme. However, the advantage of DMM is that when the delay becomes not tolerable, the application might be restarted, or the communication refreshed, so that the most suitable IP address can be picked, thus leading to traverse a shorter (direct) path with better delay.

In order to evaluate the advantages and disadvantages of DMM, it would be desirable to understand what are the constraints in terms of mobility due to the number of hops previously calculated. In the case of a WAN technology such as WiMAX or 3G, one access router can serve a cell of few Km of radius, while in the case of a LAN technology such as IEEE 802.11, the cell radius is reduced to less than 100m. Now let's look at a typical use case, where a user starts a VoIP conversation and walks across a DMM domain using IEEE 802.11. The typical speed for pedestrians is 4-5 Km/h [53] and the average call duration is roughly 3 minutes [54]. This means that during the call, the user will walk around 250m, hence performing two handovers and adding a delay of roughly 10ms more than the direct path between the CN and MN. This simple example shows two of the benefits of DMM: simplicity and low added end-to-end communications delay.

6 Experimental evaluation

This section provides an experimental evaluation based on real tests conducted with a prototype of our proposal. The goal is two-fold: on the one hand to show that the designed solution is feasible in a real testbed. On the other hand, to assess some performance metrics.

From the two main components of our HDMM solution, the client-based mobility one is conceptually very similar to Mobile IPv6, being the main difference the fact that the mobile node is able to simultaneously operate with multiple home addresses handled by different home agents. Therefore, there is little value in developing a prototype of the client-based HDMM component, as the results would not differ from those already available for Mobile IPv6 (of course, using the same mobile node - anchor delay). Because of this, we preferred to focus or implementation and evaluation efforts on the network-based components of HDMM, which do present significant differences as compared to legacy Proxy Mobile IPv6. One of the main contributions of this study is to compare the partially and fully distributed approaches when delivering real traffic.

The prototype implementing the network-based components of HDMM is written in C and runs in Linux-operated machines. It is based on the OAI PMIPv6 implementation[10], extended with the new characteristics explained in Section 4.2. The testbed is composed of five Linux Ubuntu 10.04 boxes (running a Linux-2.6.32 kernel): four desktop PCs playing the role of three DARs and one CMD, plus one laptop playing the role of mobile node. In terms of connectivity between the different entities, both the central mobility database and the distributed anchor routers are connected to the same Ethernet switch, while the mobile nodes obtains connectivity using IEEE 802.11g as wireless technology.

The partially distributed approach is implemented following the *CMD as Proxy* flavour because it provides the quickest reaction to the handover event in terms of routing state re-configuration. Also, it yields to the least number of signalling messages exchanged.

Regarding the fully distributed approach, we decided to not implement for this tests the complete IEEE 802.21 chain of messages, since all of them are performed before the actual L2 handover and they do not impact on the performance metrics collected on this section. Hence, regarding the IEEE 802.21 support, we only implemented a custom Layer-2 attachment and detachment detection mechanism. Nevertheless, it is worth highlighting there is an ongoing joint effort with authors of [55] to integrate their IEEE 802.21 implementation (called ODTONE[11]) and our DMM code within the EU project MEDIEVAL[12], but this work is very complexity and has not yet been completed. Since the IEEE 802.21 framework is not available at the current status of the platform,

[10] OpenAir Interface PMIPv6: `http://www.openairinterface.org/components/page1103.en.htm`

[11] `http://helios.av.it.pt/projects/odtone`

[12] `http://www.ict-medieval.eu`

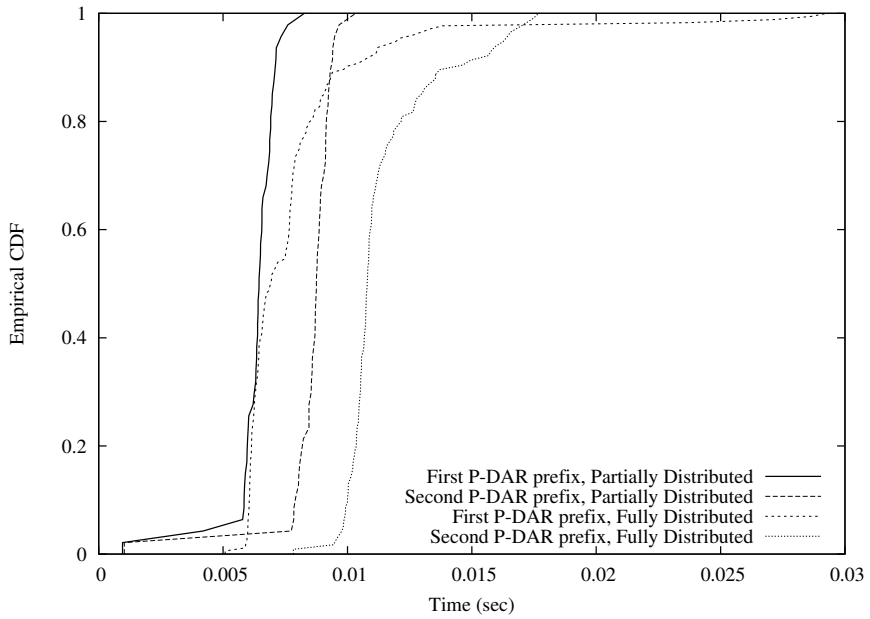

Fig. 11 Comparison of the CDF of the handover latency with different number of active prefixes.

in the experiments the P-DARs' addresses to be used for the signalling, are statically configured at each node.

In order to compare and understand the performance of the partially and the fully distributed solutions, Fig. 11 presents the empirical cumulative distribution function (eCDF) of the handover latency for both approaches and different number of active prefixes. In the experiments we repeated several times a cycle in which the MN connects to DAR_1, next moves to DAR_2 and DAR_3, and finally disconnects. Hence after the first handover only one prefix is updated, whereas after the second movement two prefixes are updated. To better understand the contribution of the IP mobility operations to the overall handover latency, we have set timers in the code to extract the timestamps when the PBU, PBA, RS and RA messages are sent and received. With respect to the handover analysis conducted in Section 5.2, this is equivalent to measuring the interval $T_{binding}$. These results do not include the Layer-2 handover delays ${T_{L2}}^{ho}$, the times T_{auth}, T_{DAD} and T_{MD} because they are identical for both set of experiments, thus not relevant.

A first observation is the difference to maintain the first prefix (observed during both handovers) and the second one (observed in the last handover only). The reason is inherent to how the implementation handles multiple prefix updates, because each prefix is handled sequentially within a single execution thread, rather than simultaneously with parallel threads. Therefore, with each additional active prefix, the handover latency of that prefix is in-

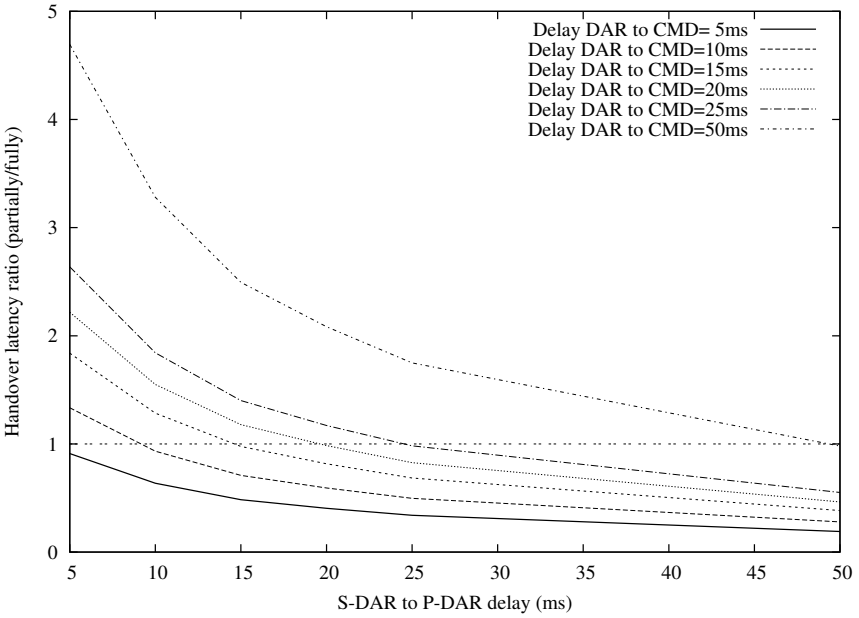

Fig. 12 Comparison of the handover latency ratio between the partially and fully distributed solutions versus the DAR-DAR and DAR-CMD delays.

creased, as shown in Fig. 11. We can observe how this latency is shorter for the case of the partially distributed approach than for the fully distributed one. The reason lies in how many messages are required to convey the same infromation. Indeed, in the partially distributed scenario, the CMD provides to the S-DAR the list of P-DARs' addresses and associated anchored prefixes in one single packet. On the contrary, in the other case, the serving DAR obtains the information about each prefix through a different message. This results in a difference in the processing time required for both operations.

The results for the fully distributed approach also show a slightly higher dispersion. This is because there are more machines and more links involved in parallel in the fully distributed mechanism than in the partial one, each producing slight differences at each repetition, thus adding more random variation effects. These facts lead to a more evident heterogeneity in the measurements observed for the fully distributed scheme, while they have a lower impact on the partially distributed solution, where all the heavy processing is performed at the CMD.

However, taking all previous comments into account, we can safely conclude from Fig. 11, that there is no significant difference in terms of handover delay between the partially and fully distributed solutions due to processing, as in a real deployment the most relevant contribution to the overall latency would be given by the distance between the involved network entities, which in our case is negligible, being the machines in the same network segment.

For the reason mentioned above, we run an experiment aiming at assessing the impact of the delay between the serving DAR and previous DARs for the fully distributed solution, and of the delay between the DARs and the CMD for the partially distributed one. Using different delays, we measured the total handover latency for both the partially and the fully distributed approaches. We then plot in Fig. 12 the ratio between the handover latency in the two protocols, versus the S-DAR to P-DAR delay, and for different DAR to CMD delays. As it can be observed from the figure, both approaches behave as expected. For S-DAR to P-DAR delays smaller than S-DAR to CMD, the fully distributed approach offers less delay that the partially distributed. A line for the ratio equal to one is also plotted, for an easier performance comparison. The obtained results show that both approaches perform quite similarly for comparable delays between the involved network entities, similarly to what stated for the previous experiment. Hence, we argue that the solution selection must be performed in function of the network infrastructure characteristics where the solution is going to be deployed. As an example, if the architecture of the operator is already distributed in nature and IEEE 802.21 is deployed, then the fully distributed approach seems the best candidate. In contrast, if the operator is evolving a mature network, where the underlying network was dimensioned for use with a centralized solution, then the partially distributed solution is better suited.

7 Conclusions

The unexpected success of smart-phones, tablets and netbooks has fostered an incredible increase in mobile data volumes. Large-scale mobile operators are very much concerned about how their networks are going to tackle this exponentially growing users' traffic demand in the near future. Current mobility architectures are heavily centralized, making the network dimensioning extremely challenging, as the core has to be able to cope with all this traffic load. This has triggered a special interest on a new mobility paradigm, the so-called Distributed Mobility Management (DMM), where the network architecture is flattened and the mobility task is no longer performed by a centralized entity.

This article discusses a novel solution that proposes the combination of a network-based DMM approach with a client-based one. The resulting Hybrid DMM (HDMM) solution aims at providing mobile network operators with a powerful, yet flexible, framework that could lead them towards effectively flattening their networks and distributing the mobility management. HDMM is composed of two main components: a distributed version of Proxy Mobile IPv6, and a distributed version of Mobile IPv6. For the former case, different signaling schemes are proposed and analyzed. An analytic and experimental evaluation has been conducted, showing that HDMM solutions are comparable in terms of overhead and handover delay to existing centralized approaches

(Mobile IPv6 or Proxy Mobile IPv6), while the use of HDMM solutions would heavily alleviate the mobile operator's core.

References

1. H. Chan, "Requirements for Distributed Mobility Management," Internet-Draft (work in progress), draft-ietf-dmm-requirements-03.txt, 2012.
2. J. C. Zuniga, C. J. Bernardos, T. Melia, and C. Perkins, "Mobility Practices and DMM Gap Analysis," Internet-Draft (work in progress), draft-zuniga-dmm-gap-analysis-03.txt, December 2012.
3. F. Giust, A. de la Oliva, and C. J. Bernardos, "Flat Access and Mobility Architecture: an IPv6 Distributed Client Mobility Management solution," in *2011 IEEE INFOCOM MobiWorld Workshop*, April 2011, pp. 361–366.
4. F. Giust, A. de la Oliva, C. Bernardos, and R. Da Costa, "A network-based localized mobility solution for Distributed Mobility Management," in *Wireless Personal Multimedia Communications (WPMC), 2011 14th International Symposium on*, October 2011, pp. 1 –5.
5. F. Giust, C. Bernardos, S. Figueiredo, P. Neves, and T. Melia, "A hybrid MIPv6 and PMIPv6 distributed mobility management: The MEDIEVAL approach," in *Computers and Communications (ISCC), 2011 IEEE Symposium on*, July 2011, pp. 25–30.
6. C. Perkins, D. Johnson, and J. Arkko, "Mobility Support in IPv6," RFC 6275, Internet Engineering Task Force, July 2011.
7. S. Gundavelli, K. Leung, V. Devarapalli, K. Chowdhury, and B. Patil, "Proxy Mobile IPv6," RFC 5213, August 2008.
8. T. Narten, E. Nordmark, W. Simpson, and H. Soliman, "Neighbor Discovery for IP version 6 (IPv6)," RFC 4861, September 2007.
9. H. A. Chan, H. Yokota, J. Xie, P. Seite, and D. Liu, "Distributed and dynamic mobility management in mobile internet: Current approaches and issues," *Journal of Communications*, vol. 6, no. 1, pp. 4–15, 2011.
10. P. Bertin, S. Bonjour, and J. Bonnin, "Distributed or centralized mobility?" in *Global Telecommunications Conference, IEEE GLOBECOM 2009*, 2009, pp. 1–6.
11. P. McCann, "Design of a flat wireless Internet Service Provider network," in *Wireless Personal Multimedia Communications (WPMC), 2011 14th International Symposium on*, October 2011, pp. 1 –5.
12. P. Bertin, S. Bonjour, and J.-M. Bonnin, "A Distributed Dynamic Mobility Management Scheme Designed for Flat IP Architectures," in *New Technologies, Mobility and Security, 2008. NTMS '08.*, November 2008, pp. 1–5.
13. V. Kafle, Y. Kobari, and M. Inoue, "A Distributed Mobility Management scheme for future networks," in *Kaleidoscope 2011: The Fully Networked Human? - Innovations for Future Networks and Services (K-2011), Proceedings of ITU*, December 2011, pp. 1 –7.
14. H. Zhang, F. Qiu, H. Zhou, X. Li, and F. Song, "A Distributed Mobility Management Solution in LISP networks," Internet-Draft (work in progress), draft-zhang-dmm-lisp-00.txt, September 2012.
15. W. Hahn, "3GPP Evolved Packet Core support for distributed mobility anchors: Control enhancements for GW relocation," in *ITS Telecommunications (ITST), 2011 11th International Conference on*. IEEE, 2011, pp. 264–267.
16. ——, "Flat 3GPP Evolved Packet Core," in *Wireless Personal Multimedia Communications (WPMC), 2011 14th International Symposium on*, October 2011, pp. 1–5.
17. C. Bernardos, J. Zuniga, and A. Reznik, "Towards Flat and Distributed Mobility Management: a 3GPP Evolved Network Design," in *Communications (ICC), 2012 IEEE International Conference on*, June 2012, pp. 6855 –6861.
18. R. Farha, K. Khavari, N. Abji, and A. Leon-Garcia, "Peer-to-Peer mobility management for all-IP networks," in *Communications, 2006. ICC'06. IEEE International Conference on*, vol. 5. IEEE, 2006, pp. 1946–1952.

19. Y. Zhai, Y. Wang, I. You, J. Yuan, Y. Ren, and X. Shan, "A DHT and MDP-based Mobility Management Scheme for Large-Scale Mobile Internet," in *2011 IEEE INFOCOM MobiWorld Workshop*, 2011, pp. 379–384.
20. M. Fischer, F.-U. Andersen, A. Kopsel, G. Schafer, and M. Schlager, "A Distributed IP Mobility Approach for 3G SAE," in *Personal, Indoor and Mobile Radio Communications, 2008. PIMRC 2008. IEEE 19th International Symposium on*, September 2008, pp. 1 –6.
21. L. Yu, Z. Zhijun, L. Tao, and T. Hui, "Distributed Mobility Management Based on Flat Network Architecture," in *Wireless Internet Conference (WICON), 2010 The 5th Annual ICST*, March 2010, pp. 1–6.
22. M. Liu, X. Guo, A. Zhou, S. Wang, Z. Li, and E. Dutkiewicz, "Low latency IP mobility management: protocol and analysis," *EURASIP Journal on Wireless Communications and Networking*, vol. 2011, no. 1, pp. 1–16, 2011.
23. R. Wakikawa, G. Valadon, and J. Murai, "Migrating home agents towards internet-scale mobility deployments," in *Proceedings of the 2006 ACM CoNEXT conference*.
24. H. Chan, "Proxy Mobile IP with distributed mobility anchors," in *GLOBECOM Workshops (GC Wkshps), 2010 IEEE*, December 2010, pp. 16 –20.
25. P. Ernest and H. Chan, "Enhanced handover support and routing path optimization with distributed mobility management in flattened wireless networks," in *Wireless Personal Multimedia Communications (WPMC), 2011 14th International Symposium on*, oct. 2011, pp. 1 –5.
26. K. Xue, L. Li, P. Hong, and P. McCann, "Routing optimization in DMM," Internet-Draft (work in progress), draft-xue-dmm-routing-optimization-01.txt, December 2012.
27. L. Yi, H. Zhou, D. Huang, and H. Zhang, "D-pmipv6: A distributed mobility management scheme supported by data and control plane separation," *Mathematical and Computer Modelling*, no. 0, pp. –, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0895717712003445
28. H. Jung, M. Gohar, J. Kim, and S. Koh, "Distributed mobility control in Proxy Mobile IPv6 networks," *IEICE Transactions on Communications*, vol. 94, no. 8, p. 2216, 2011.
29. M. Boc, A. Petrescu, and C. Janneteau, "Anchor-based routing optimization extension for Proxy Mobile IPv6 in flat architectures," in *Wireless Personal Multimedia Communications (WPMC), 2011 14th International Symposium on*, October 2011, pp. 1 –5.
30. P. Seite and P. Bertin, "Distributed Mobility Anchoring," Internet-Draft (work in progress), draft-seite-dmm-dma-06.txt, January 2013.
31. T.-X. Do and Y. Kim, "Distributed network mobility management," in *Advanced Technologies for Communications (ATC), 2012 International Conference on*, oct. 2012, pp. 319 –322.
32. H. Yokota, P. Seite, E. Demaria, and Z. Cao, "Use case scenarios for Distributed Mobility Management," Internet-Draft (work in progress), draft-yokota-dmm-scenario-00.txt, October 2010.
33. H. Chan, "Distributed Mobility Management with Mobile IP," in *Communications (ICC), 2012 IEEE International Conference on*, June 2012, pp. 6850 –6854.
34. P. Bertin, S. Bonjour, and J.-M. Bonnin, "An Evaluation of Dynamic Mobility Anchoring," in *Vehicular Technology Conference Fall (VTC 2009-Fall), 2009 IEEE 70th*, September 2009, pp. 1–5.
35. P. Louin and P. Bertin, "Network and host based distributed mobility," in *Wireless Personal Multimedia Communications (WPMC), 2011 14th International Symposium on*, October 2011, pp. 1 –5.
36. A. Gurtov, M. Komu, and R. Moskowitz, "Host identity protocol: identifier/locator split for host mobility and multihoming," *Internet Protocol J*, vol. 12, no. 1, pp. 27–32, 2009.
37. E. Nordmark and M. Bagnulo, "Shim6: Level 3 multihoming shim protocol for ipv6," RFC 5533, June, Tech. Rep., 2009.
38. D. Farinacci, V. Fuller, D. Meyer, and D. Lewis, "The Locator/ID Separation Protocol (LISP)," RFC 6830, Internet Engineering Task Force, January 2013.
39. 3GPP, "General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access," 3rd Generation Partnership Project (3GPP), TS 23.401, September 2011.

40. ——, "LIPA Mobility and SIPTO at the Local Network," 3rd Generation Partnership Project (3GPP), TR 23.859, July 2011.
41. H. Soliman, "Mobile IPv6 Support for Dual Stack Hosts and Routers," RFC 5555, Internet Engineering Task Force, June 2009.
42. R. Koodli, "Mobile IPv6 Fast Handovers," RFC 5268, Internet Engineering Task Force, June 2008.
43. G. Tsirtsis, H. Soliman, N. Montavont, G. Giaretta, and K. Kuladinithi, "Flow Bindings in Mobile IPv6 and Network Mobility (NEMO) Basic Support," RFC 6089, Internet Engineering Task Force, January 2011.
44. C. J. Bernardos, "Proxy Mobile IPv6 Extensions to Support Flow Mobility," Internet-Draft (work in progress), draft-ietf-netext-pmipv6-flowmob-05.txt, October 2012.
45. S. Gundavelli, X. Zhou, J. Korhonen, G. Fegie, and R. Koodli, "IPv4 Traffic Offload Selector Option for Proxy Mobile IPv6," Internet-Draft (work in progress), draft-ietf-netext-pmipv6-sipto-option-10.txt, February 2013.
46. H. Ali-Ahmad, M. Ouzzif, P. Bertin, and X. Lagrange, "Comparative performance analysis on dynamic mobility anchoring and proxy mobile IPv6," in *Wireless Personal Multimedia Communications (WPMC), 2012 15th International Symposium on*, sept. 2012, pp. 653 –657.
47. V. Devarapalli and F. Dupont, "Mobile IPv6 Operation with IKEv2 and the Revised IPsec Architecture," RFC 4877, April 2007.
48. T. Aura, "Cryptographically Generated Addresses (CGA)," RFC 3972, Internet Engineering Task Force, March 2005.
49. J. Laganier, "Authorizing Mobile IPv6 Binding Update with Cryptographically Generated Addresses," Internet-Draft (work in progress), draft-laganier-mext-cga-01.txt, October 2010.
50. LAN/MAN Committee of the IEEE Computer Society, "IEEE Std 802.21-2008, Standards for Local and Metropolitan Area - Part 21: Media Independent Handover Services," 2008.
51. M. Skorepa and R. Klugl, "Enhanced analytical method for ip mobility handover schemes cost evaluation," *Telecommunication Systems*, pp. 1–10, 2011. [Online]. Available: http://dx.doi.org/10.1007/s11235-011-9524-2
52. R. ITU-T and I. Recommend, "G. 114," *One-way transmission time*, vol. 18, 2000.
53. R. Knoblauch, M. Pietrucha, and M. Nitzburg, "Field studies of pedestrian walking speed and start-up time," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1538, pp. 27–38, 1996.
54. A. Noll, "Cybernetwork technology: Issues and uncertainties," *Communications of the ACM*, vol. 39, no. 12, pp. 27–31, 1996.
55. D. Corujo, C. Guimaraes, B. Santos, and R. Aguiar, "Using an open-source IEEE 802.21 implementation for network-based localized mobility management," *Communications Magazine, IEEE*, vol. 49, no. 9, pp. 114 –123, september 2011.