Master of Science Thesis

Client-based and Network-based solutions for Distributed Mobility Management

Author: Fabio Giust

Supervisor: Prof. Carlos Jesús Bernardos Cano

Leganés, September 2011
Abstract

Internet traffic has increased steeply in recent years, mainly due to the fruition of video and other streaming contents, social platforms and peer-to-peer networks. In addition, the quick penetration of hand-held devices equipped with multiple radios (e.g., 3G and WiFi), sees to it that wireless access represents an ever-growing portion of current and future demand, thus encouraging operators to investigate and deploy different combinations of wireless access technologies with the purpose of reducing their operational costs (the so-called “4G” architecture).

It is hence necessary to adopt an efficient mobility management technique to meet users’ expectation of an “anywhere, anytime” connectivity. Nevertheless, the use of centralized mobility management approaches – such as Mobile IPv6 and Proxy Mobile IPv6 – is foreseen to bring some difficulties to operators, due to the expected large number of mobile users and their exigent demands.

All this has triggered the need for Distributed Mobility Management (DMM) alternatives, focused on moving the mobility anchors from the core network to the edge, pushing them closer to the users. The purpose of such new research direction is to overcome the limitations imposed by a centralized approach, alleviating operators’ costs by deploying a more efficient network, envisioning also the heterogeneity of the underlying technology.

This work first explores two protocols for mobility support, Mobile IPv6 and Proxy Mobile IPv6, taken as main referents for, respectively, the host-based mobility approach and the network based one. We next elaborate the extensions and the changes to transform them according to the distributed mobility management paradigm, proposing several complete solutions. Finally, we analytically compare the distributed solutions to their centralized counterparts, in order to derive which are the most suitable scenarios for their applicability.
# List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3GPP</td>
<td>3rd Generation Partnership Project</td>
</tr>
<tr>
<td>AP</td>
<td>Access Point</td>
</tr>
<tr>
<td>AAA</td>
<td>Authentication, Authorization and Accounting</td>
</tr>
<tr>
<td>BA</td>
<td>Binding Acknowledgment</td>
</tr>
<tr>
<td>BC</td>
<td>Binding Cache</td>
</tr>
<tr>
<td>BCE</td>
<td>Binding Cache Entry</td>
</tr>
<tr>
<td>BU</td>
<td>Binding Update</td>
</tr>
<tr>
<td>CGA</td>
<td>Cryptographically Generated Address</td>
</tr>
<tr>
<td>CMD</td>
<td>Central Mobility Database</td>
</tr>
<tr>
<td>CN</td>
<td>Correspondent Node</td>
</tr>
<tr>
<td>CoA</td>
<td>Care-of Address</td>
</tr>
<tr>
<td>CoT</td>
<td>Care-of Test</td>
</tr>
<tr>
<td>CoTI</td>
<td>Care-of Test Initialization</td>
</tr>
<tr>
<td>DAD</td>
<td>Duplicate Address Detection</td>
</tr>
<tr>
<td>DAR</td>
<td>Distributed Anchor Router</td>
</tr>
<tr>
<td>DL</td>
<td>Downlink</td>
</tr>
<tr>
<td>DMM</td>
<td>Distributed Mobility Management</td>
</tr>
<tr>
<td>DSMIPv6</td>
<td>Dual-Stack Mobile IPv6</td>
</tr>
<tr>
<td>FAMA</td>
<td>Flat Access and Mobility Architecture</td>
</tr>
<tr>
<td>GTP</td>
<td>GPRS Tunnelling Protocol</td>
</tr>
<tr>
<td>HA</td>
<td>Home Agent</td>
</tr>
<tr>
<td>HMIPv6</td>
<td>Hierarchical Mobile IPv6</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>HNP</td>
<td>Home Network Prefix</td>
</tr>
<tr>
<td>HoA</td>
<td>Home Address</td>
</tr>
<tr>
<td>IEEE</td>
<td>The Institute of Electrical and Electronics Engineers</td>
</tr>
<tr>
<td>IETF</td>
<td>The Internet Engineering Task Force</td>
</tr>
<tr>
<td>LMA</td>
<td>Local Mobility Anchor</td>
</tr>
<tr>
<td>LMD</td>
<td>Localized Mobility Domain</td>
</tr>
<tr>
<td>MAAR</td>
<td>Mobility Anchor and Access Router</td>
</tr>
<tr>
<td>MAG</td>
<td>Mobile Access Gateway</td>
</tr>
<tr>
<td>MH</td>
<td>Mobility Header</td>
</tr>
<tr>
<td>MIHF</td>
<td>Media Independent Handover Functions</td>
</tr>
<tr>
<td>MIIS</td>
<td>Media Independent Information Service</td>
</tr>
<tr>
<td>MIPv4</td>
<td>Mobile IPv4</td>
</tr>
<tr>
<td>MIPv6</td>
<td>Mobile IPv6</td>
</tr>
<tr>
<td>MN</td>
<td>Mobile Node</td>
</tr>
<tr>
<td>MN-ID</td>
<td>Mobile Node Identifier</td>
</tr>
<tr>
<td>ND</td>
<td>Neighbor Discovery</td>
</tr>
<tr>
<td>PBA</td>
<td>Proxy Binding Acknowledgment</td>
</tr>
<tr>
<td>PBU</td>
<td>Proxy Binding Update</td>
</tr>
<tr>
<td>P-CoA</td>
<td>Proxy Care-of Address</td>
</tr>
<tr>
<td>PMIPv6</td>
<td>Proxy Mobile IPv6</td>
</tr>
<tr>
<td>PoA</td>
<td>Point of Attachment</td>
</tr>
<tr>
<td>PoS</td>
<td>Point of Service</td>
</tr>
<tr>
<td>RA</td>
<td>Router Advertisement</td>
</tr>
<tr>
<td>RO</td>
<td>Route Optimization</td>
</tr>
<tr>
<td>RR</td>
<td>Return Routability</td>
</tr>
<tr>
<td>RS</td>
<td>Router Solicitation</td>
</tr>
<tr>
<td>RTT</td>
<td>Round Trip Time</td>
</tr>
<tr>
<td>SLAAC</td>
<td>StateLess Address Auto-Configuration</td>
</tr>
<tr>
<td>UL</td>
<td>Uplink</td>
</tr>
</tbody>
</table>
Table of Contents

1 Introduction 1

2 Centralized vs. Distributed Mobility Management 3
  2.1 Mobile IPv6 3
  2.1.1 Entities 4
  2.1.2 Protocol Overview 4
  2.1.3 Security 5
  2.2 Proxy Mobile IPv6 overview 6
  2.2.1 Entities 7
  2.2.2 Operations 8
  2.2.3 Security 9
  2.3 Limitations of centralized mobility management solutions 9
  2.4 State of the Art 10

3 Client-based solution for DMM: FAMA 12
  3.1 Working scheme 12
  3.2 Security 13

4 Network-based solutions for DMM 16
  4.1 Fully distributed approach 17
    4.1.1 Full distribution with IEEE 802.21 19
  4.2 Partially distributed approach 20
    4.2.1 The CMD is a PBU/PBA relay 22
    4.2.2 The CMD as MAAR locator 23
    4.2.3 The CMD as PBU/PBA proxy 23

5 Evaluation of the DMM solutions 26
  5.1 Overhead Analysis 26
    5.1.1 Client-based scenario 26
    5.1.2 Network-based scenario 27
  5.2 Handover latency 28
  5.3 Communication delay 30
  5.4 Final remarks 32

6 Conclusions 33
References
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Mobile IPv6 overview</td>
<td>5</td>
</tr>
<tr>
<td>2.2</td>
<td>Proxy Mobile IPv6 overview</td>
<td>7</td>
</tr>
<tr>
<td>2.3</td>
<td>Registration to a Proxy Mobile IPv6 domain</td>
<td>8</td>
</tr>
<tr>
<td>3.1</td>
<td>FAMA architecture and example scenarios</td>
<td>13</td>
</tr>
<tr>
<td>3.2</td>
<td>Signaling between the mobile node and the Distribute Access Router</td>
<td>15</td>
</tr>
<tr>
<td>4.1</td>
<td>Fully distributed PMIPv6</td>
<td>17</td>
</tr>
<tr>
<td>4.2</td>
<td>Fully distributed approach with IEEE 802.21</td>
<td>19</td>
</tr>
<tr>
<td>4.3</td>
<td>Partially distributed PMIPv6: initial registration</td>
<td>21</td>
</tr>
<tr>
<td>4.4</td>
<td>Partially distributed PMIPv6: CMD as message relay</td>
<td>22</td>
</tr>
<tr>
<td>4.5</td>
<td>Partially distributed PMIPv6: data flow</td>
<td>23</td>
</tr>
<tr>
<td>4.6</td>
<td>Partially distributed PMIPv6: CMD as MAAR locator</td>
<td>24</td>
</tr>
<tr>
<td>4.7</td>
<td>Partially distributed PMIPv6: CMD as message proxy</td>
<td>25</td>
</tr>
<tr>
<td>5.1</td>
<td>Edge network divided in IP access cells</td>
<td>32</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

The increasing demand of mobile data services from users is no longer a threat to operators, but a reality that needs to be tackled. We are witnessing that the number of wireless mobile subscribers accessing data services does not stop increasing. This is motivated by a variety of different reasons: 3G and WLAN accesses are widely available (combined, coverage reaches almost 100% of dense populated areas in developed countries) and affordable by users (most mobile handsets are 3G and WLAN capable, all laptops and netbooks are equipped with WLAN interfaces, 3G USB modems are quite cheap and operators offer flat rates to their customers). Besides, the number and popularity of applications designed for smart-phones that make use of Internet connectivity is getting larger every day, contributing to an increase of market penetration of such devices (e.g., iPhone, Android, Blackberry and Windows Mobile phones), which results in growing demands for Internet connectivity everywhere.

Additionally, operators are migrating their networks to full IP based networks, as the most recent mobile architectures like the IEEE 802.16 suite (also known as WiMAX)\(^1\) and the 3GPP Evolved Packet System (EPS)\(^2\), for both voice and data, triggering a real need for IP mobility management solutions, which up to now had shown little or no deployment penetration, being proprietary customized solutions used instead.

Most of the currently IP mobility solutions standardized by the IETF\(^3\), like Mobile IPv6 (MIPv6)\(^4\), or Proxy Mobile IPv6 (PMIPv6)\(^5\) rely to a certain extent on a centralized mobility anchor entity. This centralized network node is in charge of i) the control of the network entities involved in the mobility management (i.e., it is a central point for the control signaling), and ii) the user data forwarding (i.e., it is also a central point for the user plane). This makes centralized mobility solutions prone to several problems and limitations, as identified in \(^8\): longer (sub-optimal) routing paths, scalability problems, signaling overhead (and most likely a longer associated handover latencies), more complex network deployment, higher vulnerability due to the existence of a potential single point of failure, and lack of granularity on the mobility management service (i.e., mobility is offered on a per-node basis, not being possible to define finer granularity policies, as for example on a

---

\(^1\)The Working Group’s official webpage can be found at \url{http://grouper.ieee.org/groups/802/16/}

\(^2\)3rd Generation Partnership Project, \url{http://www.3gpp.org/}

\(^3\)Internet Engineering Task Force, \url{www.ietf.org}
per-application basis).

Because of all the aforementioned issues, big operators are now looking for alternative mobility solutions that are more distributed in nature, allowing a cheaper and more efficient network deployment capable to meet their customers’ requirements. In particular, there is an effort in the IETF, called Distributed Mobility Management (DMM), that is currently addressing exactly this particular problem, first starting from a clear definition of the problem statement [8].

There are basically two main approaches being researched now: one aimed at making Mobile IPv6 work in a distributed way, and another one doing the same exercise for Proxy Mobile IPv6. In this work, we explored both approaches producing DMM designs starting from the corresponding IETF protocols and stretching them to work in a distributed way: our objective was to not introduce any new message nor data structure, but to use, instead, what is already standardized, applying the necessary changes in terms small extensions to legacy message formats and, especially, re-ordering nodes’ operations, altering their state-machine.

Hence, we present a complete solution for a mobility client inspired by MIPv6, called “Flat Access and Mobility Architecture” (FAMA) and, besides, we show several proposals for network-based DMM, divided into two categories, the fully distributed approach and the partially distributed one.

The rest of the thesis is organized as follows.

- Chapter 2 summarizes how centralized mobility management works, by describing the operations of Mobile IPv6 and Proxy Mobile IPv6, and highlights the main limitations of this kind of approach, providing some state-of-art DMM schemes, referencing the main output available in literature and in the IETF.

- Chapter 3 presents Flat Access and Mobility Architecture (FAMA), a complete solution for client-mobility based on Mobile IPv6, implementing an alternative security mechanism.

- Chapter 4 introduces partially and fully distributed approaches for network-based DMM, based on Proxy Mobile IPv6, showing several solutions for both approaches.

- Chapter 5 analyzes and compares our solutions with their centralized counterparts in terms of overhead, handover latency and communication delay.

- Chapter 6 concludes the thesis, highlighting the main results and proposing the guidelines for future work.
Chapter 2

Centralized vs. Distributed Mobility Management

The mobility management schemes standardized by the IETF for IPv6 networks are extensions or modifications of the well known Mobile IPv6 protocol (MIPv6) \cite{IETF6275}, such as Proxy Mobile IPv6 (PMIPv6) \cite{IETF6311}, Dual-Stack Mobile IPv6 (DSMIPv6) \cite{IETF6219} and Hierarchical Mobile IPv6 (HMIPv6) \cite{IETF6978}. However, they come at the cost of handling operations at a cardinal point, the mobility anchor, and burdening it with data forwarding and control mechanisms for a great amount of users. This node is usually far away from the edge, deep into the core network, and, although HMIPv6 proposes to split the management hierarchically, this only shifts the problem close to the edge without really addressing the flat IP architecture demand.

Next sections present the two protocols chosen as most relevant examples for the client-based mobility support and the network based one, respectively MIPv6 and PMIPv6. Finally, we highlight the limitations of such approaches, hence the motivations that push for distributed mobility management and some works proposed in literature dealing with DMM.

2.1 Mobile IPv6

A step forward in the design of an efficient mobility protocol was achieved in 2004 with the release of RFC 3775 \cite{IETF6275} by David Johnson, Charlie Perkins and Jari Arkko, named Mobility Support in IPv6\footnote{At the time of writing this report, RFC 6275 is being released that obsoletes the mentioned one. However, the new release is simply an update of the old one, and it does not introduce relevant changes for the purposes of the work}, or simply MIPv6. Mobile IPv6 inherits most of the mechanisms introduced with its predecessor, Mobile IPv4 (MIPv4) \cite{IETF5213}, but, also, it brings several enhancements due to the wise exploitation of IPv6 features.

Next subsections detail the involved entities and the protocol operations. Note that most of the terms introduced in the following paragraphs are common to more mobility protocols, and thus are used throughout the document with that same meaning, except when stated otherwise.
Chapter 2. Centralized vs. Distributed Mobility Management

2.1.1 Entities

Here is the list of nodes comprised in the MIPv6 architecture.

- **Mobile Node (MN).** It is the moving host, usually referred as a terminal that changes point of attachment; the notation of **Mobile Terminal (MT)** is also used in literature\(^2\). The Mobile Node (MN) can configure two types of addresses:
  - **Home Address (HoA).** It is the permanent and globally reachable IP address configured by the MN when at home. The **Home Network** is thus defined as the network where the HoA is topologically valid.
  - **Care-of Address (CoA).** It is the temporary IP address configured by the MN to maintain connectivity when in a foreign network.

- **Home Agent (HA).** This node is in charge of storing an association between the HoA and CoA (a **binding**) per each MN. If the MN is present in the Home Network, the HA is not necessarily involved in the packets delivery to the MN, otherwise, if the MN registered a CoA at the HA, it intercepts the packets destined to the HoA, and it encapsulates them to the CoA to properly route them to the final recipient. This procedure is known also as **IP tunneling**\(^{[31]}\).

- **Correspondent Node (CN).** It is the other endpoint of a communication with the MN. It can be a fixed or moving node.

2.1.2 Protocol Overview

In MIPv6, the HA is an entity located in the home network which enables global reachability and session continuity for the mobile nodes associated to it. The HA anchors the permanent IP address used by the MN, the (HoA), and defends the HoA's reachability when the MN is not at home, by intercepting the traffic addressed to the HoA and redirecting it to the MN's current location.

When away from its home network, the MN acquires a temporal IP address from the visited network – the (CoA) – and informs the HA about its current location by sending a Binding Update (BU) message. The HA, in turn, updates the session information related to the MN stored in a database, called Binding Cache (BC), and sends back the response in a Binding Acknowledgment (BA) message. A bidirectional IP tunnel is established between the two nodes and the data traffic is transmitted through it. The necessity for data packets to traverse the HA lays down a suboptimal path between MN and CN called angular routing, thus an additional operation has been defined to overcome this limitation. Route Optimization (RO) is the procedure standardized to register the MN’s CoA at the CN as well, so that the communication can be moved (i.e., packets are forwarded) through a shorter path to destination. Unfortunately, in order to perform RO, the CN needs to be provided with the MIPv6 module in its TCP/IP stack implementation, otherwise the control messages cannot

\(^2\)For the sake of completeness, in the documents and specifications produced by the 3rd Generation Partnership Project (3GPP) the moving host is called User Equipment (UE). This convention is typical of cellular networks and it refers to the user terminal at all layers of the network stack, while, in the IETF, the term Mobile Node is intended to be technology agnostic, referring in particular at the IP layer.
be processed correctly. Usually, the devices available in the market implement the IPv6 stack without the mobility component. The fact that both MN and CN need to run an extra module in addition to the IPv6 stack is one of the reasons that led to the development of network based mobility solutions (see Section 2.2).

The working scheme of Mobile IPv6 mostly reflects that for IPv4, with minor variations due to the flexibility gained with IPv6 properties. For instance, when a MN is away, it can configure its address in the foreign network using standard Neighbor Discovery (ND) operations and Stateless Address Auto-Configuration (SLAAC) [28]. Hence, there is no need for a foreign agent to take part of the procedure, also for the subsequent signaling. Besides, BU/BA messages are built in a modular way, appending the so-called Mobility Header (MH) to an IPv6 header, and with some additional options as trailers to this new message header.

2.1.3 Security

The security mechanisms adopted in MIPv6 leverage the IPsec suite [1, 12] in Encrypted Security Payload mode [21, 26]. This work does not explore in depth the security issues, thus the scope is limited to present at a very high level the security principles adopted in the protocols, while further details can be found in the cited references and in [19, 29].

Besides the vulnerability intrinsic in the radio channel nature, some malicious attacks are possible when hosts are allowed to change and announce IP addresses. The most common
is known as *redirection attack*. A bogus node registers his address as CoA for a HoA that belongs to another MN. If the attempt is successful, the attacker is able to steal traffic from other users. Conversely, the misbehaving node registers its HoA with a third party’s CoA and starts several IP sessions with the purpose of flooding the victim’s links and/or draining its resources. For this reason, the the IPsec suite is used in MIPv6 to authorize an MN and to authenticate its control communication.

However, such kind of protection cannot be guaranteed for the RO procedure, as no secure associations take place with the CN, which is then particularly prone to redirection attacks. After the RO procedure, a CN basically stores a binding between the MN’s HoA and CoA as an HA does, but BU and BA are secured whilst RO messages are not. Hence, a CN must make sure that:

- the MN is really identified by the HoA it claims;
- it is actually reachable at the CoA.

The security mechanism designed for this scope is called Return Routability (RR), and, for the sake of simplicity, all the details are skipped to highlight the elegance of the basic idea. The MN starts the procedure sending a request through both paths to the CN, i.e., the direct one and that via the HA. The CN replies transmitting two tokens through the distinct paths, i.e., one with destination the HoA and the other with destination the CoA. The two tokens are necessary to generate a key used to encrypt the subsequent registration message (a BU) sent by the MN to the CN. If the key is correct, i.e., if the CN is able to decrypt the message, then it means that the MN has successfully received both tokens, hence it is the legitimate owner of the HoA and it is actually located at the CoA claimed. Hence a BA is sent back to conclude the procedure and enable the optimized path.

### 2.2 Proxy Mobile IPv6 overview

A new concept for mobility management emerged with the network-based approach, for which mobility support is provided to hosts in movement without their involvement, as long as they are roaming inside a Localized Mobility Domain (LMD). The efforts towards this direction resulted in 2007 with the standardization of **Proxy Mobile IPv6** (PMIPv6) [17], developed as an enhancement of MIPv6. In PMIPv6, the home agent is replaced by the Local Mobility Anchor (LMA): it is in charge of routing packets in uplink and downlink containing the IPv6 prefixes assigned uniquely to MNs on a per user basis, the Home Network Prefix (HNP), and it stores the MNs’ mobility sessions information.

The network based scheme is achieved by relocating relevant functionality for mobility management from the MN to the network, in a node called Mobile Access Gateway (MAG). In a Network-based Localized Mobility Domain, the network learns through standard terminal operation, such as ND [28] or by means of link-layer support [25], about an MN’s movement and coordinates routing state updates without any mobility specific support from the terminal. While moving inside the LMD, the MN keeps its IP address, and the network is in charge of updating its location in an efficient manner [37]. The following subsections give an insight of the entities and operations defined in PMIPv6.
2.2.1 Entities

The core functional entities in the PMIPv6 infrastructure are (see Fig. 2.2):

- **Mobile Node (MN)**. It is the moving host.

- **Mobile Access Gateway (MAG)**. This entity performs the mobility related signalling on behalf of an MN that it is attached to one of its access links. The MAG is usually the access router for the MN, i.e., the first hop router and default gateway in the localized mobility management infrastructure. It is responsible for tracking the MN’s movements on the access network. There are multiple MAGs in an LMD.

- **Local Mobility Anchor (LMA)**. This is an entity within the backbone network that maintains a collection of routes for individual MNs within the LMD (i.e., it is the entity that manages the MN’s binding state). The routes point to MAGs managing the links in which the MNs are currently located. Packets for an MN are routed to and
2.2.2 Operations

The sequence of operations in PMIPv6 is quite similar to that drawn in MIPv6, except that those actions performed by the MN in MIPv6 are now responsibility of the MAG.

Once an MN enters an LMD and attaches to an access link, the MAG in that access link, upon identifying the MN, performs the mobility signaling on behalf of the MN. The MAG receives a Router Solicitation (RS) from the MN and sends to the LMA a Proxy Binding Update (PBU), associating its own address with the MN’s identity (e.g., the MN’s MAC address or an ID related with the MN’s authentication in the network). Upon receiving this request, the LMA assigns a prefix – called HNP – to the MN (i.e., it allocates a prefix for the attached interface). The LMA creates a Binding Cache Entry (BCE), which main fields are the Mobile Node Identifier (MN-ID), the prefix assigned and the MAG’s IP address visible from the MN through tunnels between the LMA and the corresponding MAG. The LMA is also responsible for assigning IPv6 prefixes to MNs (e.g., it is the topological anchor point for the prefixes assigned to the MN). There may be more than one LMA in an LMD.
from the LMA (the Proxy Care-of Address (P-CoA)). Then, the LMA establishes on its side a bi-directional tunnel to the MAG for the MN’s traffic forwarding, and it replies to the MAG with a Proxy Binding Acknowledgment (PBA) message, including the prefix assigned to the MN. Once the PBU/PBA handshake is over, the MAG configures the P-CoA as the second end-point of the tunnel with the LMA, and unicasts a Router Advertisement (RA) message to the MN specifying the prefix to be used for the IP connectivity. Now the MN is able to configure one or more addresses from the assigned prefix and the registration procedure is over (see Figure 2.3). Whenever the MN moves, the new MAG updates the MN’s location in the LMA by means of a PBU/PBA handshake, and advertises through a unicast RA the same prefix to the MN. The new MAG shows the same layer-2 and layer-3 identifiers to the MN, thereby making the IP mobility transparent to the MN. Thus, the MN always keeps the address configured when it first entered the LMD, even after changing its point of attachment to the network.

2.2.3 Security

The security mechanisms in PMIPv6 are split into two levels related to i) the accounting of an MN and ii) the authentication of control messages. Indeed, whilst in MIPv6 the secure association between MN and HA converges the procedures into one, as the sender of control messages is the HA or the MN itself, in PMIPv6 the MN is first authenticated and authorized for the service by the MAG, and, next, mobility messages are authenticated between MAG and LMA.

Upon a MN’s attachment to the network, either an authentication mechanism is deployed on the access link, or the MAG performs an Authentication, Authorization and Accounting (AAA) check querying a dedicated infrastructure. RFC 5213 recommends the use of Radius [33] or Diameter [7] for this purpose. In both cases, the MN results to be authorized for the PMIPv6 service and provided with an unique identifier in the domain – the MN-ID. Moreover, the messages between MAG and LMA are secured with IPsec in a similar way as in MIPv6 (see Section 2.1.3).

2.3 Limitations of centralized mobility management solutions

Centralized mobility solutions, such as those described previously, base their operation on the existence of a central entity that anchors the IP address used by the mobile node (e.g., the HA and the LMA). This central anchor point is in charge of tracking the location of the mobile and redirecting its traffic towards its current topological location. While this way of addressing mobility management has been fully developed by the Mobile IP protocol family and its many extensions, there are also several limitations that have been identified [8]:

- **Sub-optimal routing.** Traffic always traverses the HA/LMA, which leads to paths that are, in general, longer than the direct one between the mobile node and its communication peer. This is exacerbated with the current trend in which content providers push their data to the edge of the network, as close as possible to the users. With centralized mobility management approaches, user traffic will always need to go first to the home network and then to the actual content, adding unnecessary delay and wasting operator’s resources. With a distributed mobility architecture, as the anchors are
located at the very edge of the network, close to the user terminal, data paths tend to be shorter.

- **Scalability problems.** With current mobility architectures, networks have to be dimensioned to support all the traffic traversing the central anchors. This poses several scalability and network design problems, as the central mobility anchors need to have enough processing and routing capabilities to be able to deal with all the mobile users’ traffic simultaneously. Besides, the operator’s network also needs to be dimensioned to be able to cope with all the users’ traffic. A distributed approach is inherently more scalable, as the mobility management tasks are distributed and shared among several network entities, which therefore do not need to be as powerful as the centralized alternative.

- **Reliability.** Centralized solutions share the problem of being more prone to reliability problems, as the central entity is a potential single point of failure.

- **Lack of fine granularity on the mobility management service.** With current centralized mobility management solutions, mobility support is offered at a user granularity. This means that the network can just decide if mobility is provided or not to the user, but cannot offer a finer granularity, for example, to allow part of his/her traffic not to be handled by the mobility solution. There are many scenarios in which part or all the traffic of a user does not really need to be mobility enabled, as for example when the user is not mobile (at least during the lifetime of the communication) or the application itself is able to effectively deal with the change of IP address caused by the user movement. In all these situations, it would be more efficient not to enable mobility.

- **Signaling overhead.** This is related to the previous limitation. Any mobility management solution involves certain amount of signaling load. By allowing mobility management to be dynamically enabled and disabled on a per application basis, some signaling can be saved, as well as the associated handover latency. Of course, this depends on the particular scenario, as the use of distributed mobility architectures can also lead to a higher signaling load in case of very dynamic scenarios in which all the traffic is required to be mobility enabled.

## 2.4 State of the Art

The mobility management of users in operator networks has seen in recent times the release of several IPv6 centralized solutions to tackle the issue of continuously providing access and mobility support, as those mentioned in the introduction of this thesis.

However, several researches have already been conducted producing a comprehensive analysis of the comparison between centralized and distributed mobility managing architectures, as reported in the DMM problem statement [8], revealing that a distributed approach already covers most of the issues presented in a centralized approach. The guidelines to develop new systems are given in [38], while [23] summarizes some attempts that have been made to distribute the mobility control and flatten existing architectures, placing mobility entities closer to the edge of the network.
One of the efforts mentioned by [23] is Flat Access and Mobility Architecture (FAMA) [15], our contribution for host-based DMM, described in detail in next chapter. Another proposal is Dynamic Mobile IP (DMI) [20], with the particularity of keeping active and reachable old MN’s addresses (i.e., CoAs) as long as there is a session bound to one of those. Basically it suggests, when the MN is not at home, to start communications with the CoA, so that traversing the HA is avoided for new traffic. However, the draft is not taking into consideration placing the mobility anchors at the MNs’ access routers.

The network-based proposals that are subject of this thesis are also covered in papers and draft submitted by the author himself to international conferences and the IETF community, see [14, 16, 4].

However, other descriptions exist on how to have a PMIPv6-like distributed scheme. In [9], a Proxy Mobile IPv6 domain is envisioned to cover a small area, so that the LMA can be deployed closer to the MAGs, and thus to the users. In this way, many little PMIPv6 domains form a bigger super-domain, and a custom LMA-to-LMA interface is necessary to exchange the control signaling to handle the MN moving from a domain to another. The draft provides also a solution to achieve route optimization in several scenarios, at the cost of an excessive control messages exchange. Also, the architecture deployment requires a big effort, since every small domain is made of the complete Proxy Mobile IPv6 equipment (LMA + MAGs).

Some other projects of distributed architectures inspired by PMIPv6 are Dynamic Mobility Anchoring (DMA) [34] and Dynamic Distributed Mobility Management (DDMM) [5]. According to these documents, mobility support is provided on demand, that is, only for those MNs that change access point with ongoing connections. The MN configures and maintains an IP address for each access network it visits, and the access router in that access network is the anchor for the communications established using the IP address assigned by the router. This means that an access router acts as a standard router when the MN is attached to it, otherwise it tunnels the packets to the access router where the MN is currently attached. Even if the ideas in those papers are similar to what are presented in Chapter 4, in [34] it is not detailed how to build the control signaling, while in [5] a technology-agnostic implementation is used with a more passive approach, greatly reducing signaling on the infrastructure network. It relies on the MN to constantly keep sending packets in order for the network to passively learn where it is located and adequately setup tunnels and routing (this concept is covered again in Section 4.1). In [6] a simulation environment is used to compare behaviors between DMA and MIPv6, proving that although DMA greatly reduces the handover duration, it has an increased end-to-end delay.

Finally, both [13] and [39] propose to use a Distributed Hash Table (DHT) to store the mobility information of the MNs (in MIPv6 this table is called Binding Cache). The former focuses on how to efficiently manage the DHT and other related aspects providing a simulated evaluation, while the latter focuses on a technique to perform handover using bicasting. We grouped these approaches together as they both suffer from requiring a lot of new support and they do not re-use existing legacy standards and solutions.
Chapter 3

Client-based solution for DMM:
FAMA

Distributed Mobility Management approaches try to overcome the limitations of the traditional centralized mobility management, by bringing the mobility anchor closer to the MN. Following this idea, in our proposal, called Flat Access and Mobility Architecture (FAMA), the MIPv6 centralized home agent is moved to the edge of the network, being deployed in the default gateway of the mobile node. That is, the first elements that provide IP connectivity to a set of MNs are also the mobility managers for those MNs. In the following we will call these access routers Distributed Anchor Router (DAR).

3.1 Working scheme

Every time a mobile node attaches to a DAR, it gets an IPv6 address which is topologically anchored at the DAR. That means that while attached to this DAR, the mobile can send and receive traffic using that address without any tunneling nor special packet handling. Every time the mobile node moves to a different DAR, it gets a new IPv6 address from the new access router. In case the MN wants to keep the reachability of the IPv6 address(es) it obtained from the previous DAR (note that this decision is dynamic and can be done on an application basis for example), the mobile has to involve its MIPv6 stack, by sending a Binding Update to the DAR where the IPv6 address is anchored, using the address obtained from the current DAR as care-of-address. In this way, the IPv6 address that the node wants to maintain plays the role of home address, and the DAR from where that address was configured plays the role of home agent (for that particular address). Note that FAMA architecture basically enables a mobile node to simultaneously handle several IPv6 addresses – each of them anchored at a different DAR – ensuring their continuous reachability by using Mobile IPv6 in a distributed fashion (i.e., each access router is a potential home agent for the address it delegates, if required). This distributed address anchoring is enabled on demand and on a per-address granularity, which means that depending on the user needs, it might be the case that all, some or none of the IPv6 addresses that a mobile node configures while moving within a FAMA domain, are kept reachable and used by the mobile.
3.2 Security

In traditional Mobile IPv6, the communication between the MN and the HA is secured through IPsec [12]. Following a similar approach in FAMA is difficult due to the large number of security associations that would be required, since any gateway of the access network can play the role as home agent for any mobile node. In order to overcome this problem and provide authentication between the DAR and the MNs, we propose the use of Cryptographically Generated Address (CGA) [2], as introduced in [24]. Cryptographically Generated Addresses 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. The binding between the public key and the address can be verified by re-computing the hash.

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.
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.

Following the ideas presented above, every time an MN attaches to a DAR, it 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 MN to establish a communication with a remote Correspondent Node (CN) – see Fig. 3.1-(a) – while attached to that particular DAR. If the mobile then moves to a new DAR (nDAR), the following two cases are possible: i) there is no need for the address that was configured at the previous DAR (pDAR) to survive the movement: in this case there is no further action required; ii) the mobile wants to keep the reachability of the address configured at pDAR: in this case Mobile IPv6 is triggered, and the MN sends a Binding Update message to the pDAR, 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 receiving DAR to identify the MN as the legitimate owner of the address. Although the use of CGAs does not impose a heavy burden in terms of performance, depending on the number of MNs handled by the DAR, the processing of the CGAs can be problematic. To reduce the complexity of the proposed solution, we suggest an alternative mechanism to authenticate any subsequent signaling packets exchanged between the MN and the DAR (in case the mobile performs a new attachment to a different DAR). This alternative method relies on the use of a Permanent Home Keygen Token (PHKT), which will be used to generate the Authorization option that the MN has to include in all next Binding Update messages. This token is forwarded to the MN in the Binding Acknowledgment message, sent in reply to the BU. The procedure is depicted in Fig. 3.2. Once the signaling procedure is completed, a bi-directional tunnel is established between the mobile node and the DAR where the IPv6 address is anchored (the “home” DAR – HDAR – for that particular address), so the mobile can continue using the IPv6 address, as shown in Fig. 3.1-(b).

In case the MN performs any subsequent movements and it requires to maintain the reachability of an address for which it has already sent a BU, the following BU messages can be secured using the PHKT exchanged before, reducing the computational load at the receiving DAR.

Although this approach is attractive because it reduces the signaling overhead generated by the mobility support, it can be misused in some particular scenarios by malicious nodes that wish to export an incorrect CoA in the BU message, since it does not provide proof of the MN’s reachability at the visited network. Indeed, the CGA approach assures that the BU message has been sent by the legitimate HoA’s owner but it does not guarantee that the same MN is reachable at the provided CoA. In order to provide a more robust solution, we propose a Return Routability procedure similar to the one defined in MIPv6 Route Optimization to mitigate this security issue. The Return Routability procedure starts after the handoff. Instead of sending the BU message, the MN sends a Care-of Test Initialization (CoTI) message. This message is replied by the DAR with a Care-of Test (CoT) message containing a CoA Keygen Token. The MN 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 MN is the legitimate node who has sent the BU and also is reachable at the CoA indicated. As all security improvements, the one proposed incurs in a performance penalty, in this case an increase in the handover delay. Specifically this enhanced security approach requires four messages to be exchanged between the MN and the 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 two Round Trip Times (RTT) to conclude, instead of one.

Note that on every attachment of a node to a DAR, the terminal also obtains a new IPv6 address which is topologically anchored at that DAR, and that this address can be used for new communications (avoiding in this way the tunneling required when using an address anchored at a different DAR), as shown in Fig. 3.1-(c). A mobile can keep multiple IPv6 addresses active and reachable at a given time, and that requires 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. 3.1-(d), the MN 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.
Chapter 4

Network-based solutions for DMM

The IETF has been working on an analysis document [38] that proposes the guidelines for the development of network-based distributed mobility management solutions, that can be divided into two main categories:

- **Partially distributed**, that basically consists on removing the data path constraint towards the anchor, but maintaining a centralized control plane;
- **Fully distributed**, that consists on eliminating any centralized role in the architecture.

In next sections we address how to develop DMM designs based on PMIPv6 for both approaches, describing in the detail the required procedures. Nevertheless, all solutions share the same data forwarding plane, briefly summarized here: a MAG – that we call Mobility Anchor and Access Router (MAAR) – assigns IPv6 prefixes on a per-MN basis from a prefix pool that belongs to that MAAR only and that are anchored to it (i.e., topologically correct in its access subnet only). In this way, an MN receives a prefix, and therefore configures an IP address, per each visited MAAR. A MAAR anchors the flows started using the prefixes it advertises: it acts as plain IPv6 router (i.e., it does not encapsulate packets) when the MN is still attached to its network, otherwise, it establishes a tunnel with the MAAR currently serving the MN.

We introduce next the terminology related to the roles that the MAAR may play for each IP flow traversing it:

- **Anchor MAAR (A-MAAR)** is referred to the MAAR that advertised the prefix used in the communication/flow.

- **Serving MAAR (S-MAAR)** is referred to the MAAR where the MN involved in the communication/flow is attached to.

It should be noted that this separation is only conceptual and applied on a flow basis: the same MAAR can play simultaneously different roles for the different flows. With respect to the PMIPv6 semantic, an A-MAAR acts as an LMA, and the S-MAAR as a MAG.
Chapter 4. Network-based solutions for DMM

4.1 Fully distributed approach

In the fully distributed approach, the LMA and the MAG are collapsed into the MAAR. In other words, MAARs are not only responsible of the forwarding of flows established with their prefixes in the manner described before, but they also maintain and handle the mobility sessions for those flows. The next use case example clarifies the working scheme of the proposed design.

Figure 4.1: Fully distributed PMIPv6

An MN enters in the domain and attaches to MAAR$_1$, which advertises Pref$_1$. This prefix is used by the MN to configure an IP address and to start a communication with a
remote correspondent node, see Fig. 4.1(a).

When a handover occurs, MAAR\(_2\) learns that the MN has attached and advertises a new prefix, Pref\(_2\). Then, it sends a PBU message to MAAR\(_1\), that is anchoring ongoing MN’s flows, and, upon the corresponding PBA reception, a tunnel is established among them, hence the flows can be redirected through it, as shown in Fig. 4.1(b). It is worth noticing that, in this juncture, MAAR\(_1\) is the A-MAAR for the flow, while the S-MAAR is represented by MAAR\(_2\); in addition, MAAR\(_2\) becomes the A-MAAR and S-MAAR for the flows started by the MN using the new address (see Fig. 4.1(c)). Moreover, upon a new movement of the MN to a new access network, MAAR\(_3\) assigns a third prefix and updates the MN’s location to all its former S-MAARs, i.e. MAAR\(_1\) and MAAR\(_2\), by means of PBU messages. Again, when the PBAs are received, tunnels are set up and the old communications are recovered, as depicted in Fig. 4.1(d).

The main issue in this mechanism is how a MAAR can differentiate between the first attachment to the network and subsequent handovers, that is, the MAAR, upon the MN’s attachment, should be informed of the past MN’s location to be able to contact, if necessary, the formerly visited MAARs. More than one solution is suggested here, each with advantages and disadvantages that should be evaluated looking for the best trade-off:

- **Broadcast PBU.** When a MAAR detects the attachment, it sends a broadcast request to all the MAARs and waits for a reply, that might be a void PBA in case the MN joined the network for the first time. It should be noted that sending the messages only to the neighbor MAARs is not enough to reconstruct the MN’s past history, thus the procedure might result excessively long and introduces unnecessary signaling.

- **Terminal indication.** The MN explicitly sends the prefixes acquired previously when joins the new access network (e.g., as options in the Router Solicitation message). Although this is a fast and light procedure, it requires some capabilities on the host that are not always possible to rely on (or not even desirable), and might pose some security issues if a malicious MN misbehaves.

- **Automatic learning.** In [5] it is proposed that mobility capable access routers learn the previous MN’s location by inspecting the source address of packets sent by the MNs. That is, these routers store in an internal database the prefix pools belonging to the others, and, upon receiving the first uplink packet sent by the MN after a movement, they check which is the access router the prefix belongs to and establish a tunnel with it. This requires little signaling but it may lead to an excessive delay if the MN does not send anything and the router has to explicitly request a packet.

- **Layer-2 handover support.** Handover may follow a Make-Before-Break philosophy and integrate layer-2 and layer-3 mobility procedures within the same framework to assist and drive the handover. IEEE 802.21, Media Independent Handover Services [25], is a suitable protocol for this purpose.

Among the options listed above, we claim that IEEE 802.21 should be preferred as it builds a control plane infrastructure by which a handover is prepared, executed and completed in a controlled and assisted way, according to a Make-Before-Break philosophy. In
addition, the IEEE 802.21 suite is intended to allow inter-technology handovers, giving support to MNs roaming within a heterogeneous environment with multiple radio access technologies. Next paragraphs are devoted to detail this approach.

4.1.1 Full distribution with IEEE 802.21

The IEEE 802.21 services are provided by a cross-layer entity called MIH Function (MIHF). It runs in the terminal and in the network nodes, it interacts with the network interfaces, with the layer 3 (or higher) mobility protocol and with the other MIHF peers, either local in the same node, or remote. A Point of Service (PoS) is a MIHF in the network that

Figure 4.2: Fully distributed approach with IEEE 802.21: message exchange sequence during handover

The IEEE 802.21 services are provided by a cross-layer entity called MIH Function (MIHF). It runs in the terminal and in the network nodes, it interacts with the network interfaces, with the layer 3 (or higher) mobility protocol and with the other MIHF peers, either local in the same node, or remote. A Point of Service (PoS) is a MIHF in the network that
can directly exchange messages with a peer MIHF installed in the terminal. A PoS instance serves one terminal, so multiple instances run in the same network entity that handles the MNs in its access network. In our design, PoSs are implemented in the MAARs. PoSs maintain information about the layer-2 access entities as Base Stations, Access Points and similar nodes connected to the MAAR, by interacting with the MIHF operating in them. These MIHF 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. In Figure 4.2 MAARs and PoSs are represented in the same entities, whilst PoAs are omitted to keep the chart simple. Indeed, the picture highlights the message exchange between the MN and the MAAR, and among MAARs, rather than the complete procedure specified by the IEEE 802.21 standard. According to IEEE 802.21 working scheme, a PoS learns when a handover for its MN is imminent (the details of how this is done are out of scope in this thesis); therefore it queries resources availability in the surrounding PoAs that are not under its direct control, by means of a message called MIH\_N2N\_HO\_Query\_Resources request (highlighted in red in Figure 4.2), sent to the corresponding PoSs connected to the desired PoAs (the latter are discovered by the MN, or by the network, after an information retrieval procedure involving the Media Independent Information Service – MIIS –, and after a scan of the PoAs visible by the MN). This message contains the current S-MAAR’s (to be the A-MAAR) address (denoted in red in the text-box in the picture), that is used later by the new S-MAAR to send the PBU message. Upon the list of candidate PoAs is filled with the requested information, the target for the handover is selected (either by the MN itself or by the PoS) and the corresponding PoS is notified about the decision. The MN can now move to the new PoA: the IP mobility procedure is triggered when the PoA announces the attachment to the PoS (summarized by the “link establishment” text-box), and the conclusion phase takes place, releasing the resources in the old MAAR/PoS.

This procedure works without changing the standard IEEE 802.21 primitives, when only one MAAR needs to be contacted with the PBU message (as in Fig. 4.1(b)). Conversely, when more than one MAAR is anchoring flows, as shown in Fig. 4.1(d), the current PoS needs to send to the candidate PoSs a list of all the past visited MAARs, and this requires a change in the format of the MIH\_N2N\_HO\_Query\_Resources request primitive.

As shown in the above paragraphs, a fully distributed approach requires in all cases many interventions on the terminals, that might be not desirable, and, in the IEEE 802.21 case, also the implementation of a whole control infrastructure. For the purpose of reducing complexity and terminal requirements, we present in next section a partially distributed solution, which come at the cost of keeping the control plane centralized and only distributing the data plane (which is the most critical one).

### 4.2 Partially distributed approach

In this approach, most of the control plane burden is transferred to a centralized network node. According to this principle, in the proposed architecture an additional node is added, called Central Mobility Database (CMD), storing the mobility sessions (bindings) for the MNs. It maintains most of the control functions defined for an LMA but it is relieved of the data forwarding plane since it is not traversed by users’ data traffic. However, as described at beginning of the chapter, the data plane works as in the fully distribution approach, with
the MAARs acting as plain routers when the MN is connected to the MAAR and using the prefix advertised by the MAAR, otherwise the MAAR encapsulates the flow to the serving MAAR when the MN is away, see Figures 4.1 and 4.5.

MAARs leverage on the CMD to access and update information related to the MNs, stored as mobility sessions; hence, this centralized node maintains a global view on the status of the network. The CMD is queried whenever a MN is detected to join the mobility domain. It might be a fresh attachment or a handover, but, as MAARs do not store any mobility session, they contact the CMD to retrieve the data of interest and eventually take the appropriate action. The procedure adopted for the query and the messages exchange sequence might vary to optimize the update latency and/or the signaling overhead. Here are presented three different approaches to update the mobility sessions using PBUs and PBAs. Each approach assigns a different role to the CMD:

- the CMD is a PBU/PBA relay;
- the CMD is a MAAR locator;
- the CMD is a PBU/PBA proxy.

Anyway, the initial registration method is the same in all cases, and it is described in the following paragraph.

Upon MN’s attachment to MAAR₁, an IPv6 global prefix belonging to the MAAR’s prefix pool is reserved for it (Pref₁). The prefix is sent in a PBU with the MN’s Identifier (MN-ID) by the MAAR to the CMD, that, since the session is new, stores a Binding Cache Entry containing as main fields the MN-ID, the MN’s prefix and MAAR₁’s address visible by the CMD (Proxy-CoA). The CMD replies to MAAR₁ with a PBA indicating that the MN’s registration is fresh and no past status is available. MAAR₁ unicasts a Router Advertisement (RA) to the MN including the prefix reserved before, that can be used by the MN to configure an IPv6 address (e.g., with SLAAC). The address is routable at the MAAR, in the sense that the node is on the path of packets addressed to the MN; moreover, the MAAR acts as plain router for those packets, as no encapsulation nor special handling takes place (see Fig. 4.3).
4.2.1 The CMD is a PBU/PBA relay

When the MN moves from its current access, it associates to MAAR\(_2\), which delegates another IPv6 prefix (Pref\(_2\)) and sends it to the CMD for registration. The CMD has already an entry for the MN, binding the MN-ID to its former locations, thus, it forwards the PBU to all the MAARs indicated as Proxy CoAs, in this case only MAAR\(_1\) (as depicted in Fig. 4.4).

Upon PBU reception, MAAR\(_1\) replies to the CMD with a PBA to ensure that the new location has successfully changed, containing the prefix anchored at MAAR\(_1\). The CMD updates the BCE adding the P-MAAR address in the list of old P-CoAs and forwards the PBA to the new S-MAAR, containing the previous Proxy-CoA and the prefix anchored to it, so that a tunnel can be established between the two MAARs and new routes are set appropriately to recover the flow(s).

Now packets destined to Pref\(_1\) are first received by MAAR\(_1\), encapsulated into the tunnel and forwarded to MAAR\(_2\), which finally delivers them to their destination. In uplink, when the MN transmits packets using Pref\(_1\) for the source address, they are sent to MAAR\(_2\), as it is MN’s new default gateway, then tunneled to MAAR\(_1\) which routes them towards the next hop to destination. Conversely, packets carrying Pref\(_2\) are routed by MAAR\(_2\) without any special packet handling both for uplink and downlink (see Figure 4.5).

For next MN’s movements the process is repeated except for the number of P-MAARs involved, that rises accordingly to the number of prefixes that the MN wishes to maintain. Indeed, once the CMD receives the first PBU from the new S-MAAR, it forwards copies of the PBU to all the A-MAARs indicated in the BCE as current P-CoA (i.e., the MAAR prior to handover) and old P-CoAs. They reply with a PBA to the CMD, which aggregates the messages into a single one to notify the S-MAAR, that finally can establish the tunnels with
Chapter 4. Network-based solutions for DMM

Figure 4.5: Partially distributed PMIPv6: data flow

the A-MAARs.

It should be noted that this design separates the mobility management at the prefix granularity, and it can be tuned in order to erase old mobility sessions when not required, while the MN is reachable through the latest prefix acquired. Moreover, the latency associated to the mobility update is bound to the PBA sent by the furthest A-MAAR, that takes the longest time to reach the CMD. The drawback can be mitigated introducing a timeout at the CMD, by which, after its expiration, all the PBAs so far collected are transmitted, and the remaining are sent at a later stage, upon their arrival.

4.2.2 The CMD as MAAR locator

The latency experienced in the approach shown before can be mitigated if the A-MAARs are allowed to signal directly their information to the new S-MAAR. This procedure reflects what was described in Section 4.2.1 up to the moment the A-MAAR receives the PBU. At that point an A-MAAR is aware of the new MN’s location (i.e., S-MAAR) and, besides sending a PBA to the CMD, it also sends a PBA to the S-MAAR including the prefix it is anchoring. The CMD is relieved from forwarding the PBA to the S-MAAR, as the latter receives a copy directly from the A-MAAR with the necessary information to build the tunnel and set the appropriate routes. In Figure 4.6 is illustrated the new messages sequence.

4.2.3 The CMD as PBU/PBA proxy

A further enhancement of previous solutions can be achieved when the CMD sends the PBA to the new S-MAAR before notifying the A-MAARs of the location change. Indeed, when the CMD receives the PBU for the new registration, it is already in possess of all the information that the new S-MAAR requires to set up the tunnel and the routes. Thus the PBA is sent to the S-MAAR immediately after a PBU is received. In parallel, a PBU is sent by the CMD to the A-MAARs to notify them about the new MN’s location, so they receive
the suitable information to establish the tunnel and routes on their side. When A-MAARs complete the update, they send a PBA to the CMD to indicate that the operation is concluded and the information are updated in all network nodes. This scheme is depicted in Figure 4.7.
Figure 4.7: Partially distributed PMIPv6: CMD as message proxy
Chapter 5

Evaluation of the DMM solutions

In this chapter we focus on conducting an analysis of the performance achievable by FAMA and the network-based proposals, comparing them with what would be obtained with plain Mobile IPv6 and Proxy Mobile IPv6.

The comparison is performed considering the three most important characteristics of a mobility protocol:

- Packet and signaling overhead;
- Handover latency;
- Delay between the communication endpoints.

5.1 Overhead Analysis

As explained in Chapters 3 and 4, our proposed Distributed Mobility Management solutions are based on Mobile IPv6 and Proxy Mobile IPv6, pushing, respectively, the HA and LMA functionalities to the edge of the network.

It is common to all DMM approaches to allow the traffic to be forwarded without encapsulation whenever the communication is started using the address configured in the access network where the MN is currently attached to. This is a clear advantage compared to the centralized schemes, as the tunnels are not always required when the MN is away from home. This scenario is worsened in a PMIPv6 domain, in which the concept of “home network” is lost, as the data traffic is always tunneled as long as the MN is connected to the domain. The design of a dynamic mobility management scheme is one of the DMM’s objective; next paragraphs provide the numeric details of how a tunnel impacts on the data transfer.

5.1.1 Client-based scenario

Once an MN moves, the “home” DAR is in charge of tunneling the packets to the new MN’s location, identified by its care-of address. In terms of packet’s overhead, FAMA and Mobile IPv6 (without Route Optimization) share the same overhead since both use a bi-directional tunnel between the MN and an anchor point, hence both incur in a 40-byte
overhead due to the packet encapsulation. However, as mentioned before, when an application is restarted, or a communication refreshed, the MN uses the appropriate IP address configured in that access network, thus the encapsulation is no more required.

Nevertheless, as compared to Mobile IPv6, in FAMA there may be more than one mobility anchor involved, and therefore this introduces a higher signaling load, since the number of BU/BA messages is increased. In particular, in plain Mobile IPv6 there is a single BU/BA exchange at each handoff – as only one home address is maintained by the MN – while in FAMA we have the BU/BA (and the CoTI/CoT in case additional security is required) exchange multiplied by $n$, given that $n$ is the number of IPv6 addresses that need to be kept reachable (or, equivalently, the number of visited access routers). Hence the total amount of messages is $2n$ ($+2n$ in case of additional security).

5.1.2 Network-based scenario

As described in the previous paragraph, when the MN is not at home the traffic is tunneled, carrying 40 extra bytes for the encapsulation. Unlike the host-based scenario, however, the encapsulation is performed among network nodes, saving the packet overhead in the radio link. This can be a desired effect in terms of terminal’s efficiency, as it does not waste energy to send/receive unwanted bytes in the communication. More, it should be noted that, in plain PMIPv6, the encapsulation takes always place, whereas network based DMM saves the tunnel when the MN is using the prefix advertised by the current S-MAAR.

The evaluation of the message number required to update the routing state of the network nodes proceeds considering separately the fully distributed approach and the partial one.

Fully distributed approach

Regardless the method adopted to learn that a handover occurred, the target MAAR has a PBU/PBA handshake with $n$ other MAARs, that is, given that $n$ is the number of IPv6 addresses that need to be kept reachable (or, equivalently, the number of visited access routers so far), a total of $2n$ messages. When the IEEE 802.21 infrastructure is deployed, many more messages are added to the handover phase, but, still, no additional messages are necessary for the IP address continuity and routing state update. Also, the extra message exchanges happen before the actual handover, so no extra delay or packet loss occurs.

Partially distributed approach

The three procedures proposed to access the CMD produce different effects on the proliferation of messages:

- CMD as message relay. Once the CMD receives the first PBU from the new MAAR and forwards copies to the correspondent old MAARs, they reply with a PBA back to the CMD, which in turn sends an aggregated PBA to the new MAAR, containing all the previous Proxy-CoAs. In this case the number of messages transferred grows as $2n + 2$. 
• **CMD as MAAR locator.** In this case the amount of PBU and PBA messages grows as $3n + 1$: the first PBU sent by the new S-MAAR + $n$ more forwarded by the CMD to old MAARs + $2n$ PBAs sent back by old MAARs to the CMD and the S-MAAR.

• **CMD as message proxy.** We have a first handshake with the CMD and the S-MAAR and $n$ subsequent message exchanges with the old MAARs, resulting in $2n + 2$ messages in total.

### 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 Routing 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 as 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.

- **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., HA, DAR, LMA or MAAR) with the new location of the MN (i.e., its CoA or P-CoA) and it highly depends on the distance between the entities participating in the user mobility management: the mobile node on the one side and the HA/DAR on the other side, for the host based approach, the MAAR-to-MAAR and MAAR-to-CMD distances for the network-based solutions.

Hence, we can express the handover latency as follows:

$$T_{handover} = T_{L2\,ho} + T_{MD} + T_{DAD} + T_{auth} + T_{binding}$$
in which the most crucial component is $T_{\text{binding}}$ as it depends on the RTT among the network nodes, whereas the first 4 addends can be considered common to all the scenarios mentioned before, since they do not depend from the IP mobility scheme used. We will see later that this is not totally true for IEEE 802.21 case, as its scope is to cut down the layer-2 operations, but, still, our purpose is to deploy a more efficient scheme at layer-3 level, reducing at most the binding phase.

The term $T_{\text{binding}}$ can be expressed for the different scenarios as follows:

- **MIPv6:**
  $$T_{\text{binding}} = RTT_{MN-HA} \quad (5.1)$$

- **FAMA:**
  $$T_{\text{binding}} = RTT_{MN-HDAR} \quad (5.2)$$

- **PMIPv6:**
  $$T_{\text{binding}} = RTT_{MAG-LMA} \quad (5.3)$$

- **Fully distributed network-based DMM:**
  $$T_{\text{binding}} = RTT_{MAAR_S-MAAR_A} \quad (5.4)$$

- **Partially distributed network based DMM, CMD as message relay, see Figure 4.4:**
  $$T_{\text{binding}} = RTT_{MAAR_S-CMD} + RTT_{MAAR_A-CMD} \approx 2 \cdot RTT_{MAAR-CMD} \quad (5.5)$$

- **Partially distributed network based DMM, CMD as MAAR locator, see Figure 4.6:**
  $$T_{\text{binding}} = \frac{RTT_{MAAR_S-CMD} + RTT_{MAAR_A-CMD} + RTT_{MAAR_S-MAAR_A}}{2} \approx RTT_{MAAR-CMD} + \frac{RTT_{MAAR_S-MAAR_A}}{2} \quad (5.6)$$

- **Partially distributed network based DMM, CMD as message proxy, see Figure 4.7:**
  $$T_{\text{binding}} = \max(\frac{RTT_{MAAR_S-CMD}}{2}; \frac{RTT_{MAAR_S-CMD} + RTT_{MAAR_A-CMD}}{2}) \approx RTT_{MAAR-CMD} \quad (5.7)$$

in which we assume that the CMD is approximately at the same distance to the MAARs.

Comparing (5.1) and (5.2), it is clear that the mean difference between FAMA and MIPv6 in terms of handover delay corresponds to the distance between the MN and the HA/HDAR. This is clearly 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 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 MN gets farther away from its HDAR, the handover delay increases, hence
FAMA 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 (5.3) and (5.4), we can see that a DMM solution produces a shorter latency as long as the A-MAAR is closer to the S-MAAR/MAG than the LMA. This parameter strictly depends on the size of the operator’s network, but, as said in the previous paragraph, we can assume that the LMA is always farther than A-MAARs for short communications with limited user mobility patterns.

Moreover, it can be noted by inspecting (5.5)-(5.7), that the latter solution with the CMD as message proxy outperforms all the others partially distributed proposals. Unfortunately, compared to full distribution, as the CMD needs to be at almost the same distance to the MAARs, it might result quite far from the network’s edge, hence approaching the $RTT_{LMA-MAG}$ distance typical of PMIPv6.

One last word worth mentioning is the experimental study in [27], in which handovers in a PMIPv6 domain are co-ordinated by IEEE 802.21. In the results proposed after the measurement tests, it was shown that, even with the assistance of IEEE 802.21 (i.e., a terminal hands off only when the resources on the target access point are ready), a handover takes a fixed amount of time (around 20 ms) for the attachment/detachment phase, plus a variable interval due to the mobility signaling, that depends on the distance between nodes. Authors of [27] claim that the fixed interval cannot be reduced, as is the time necessary for the network cards to associate to another access point, whilst all the other handover operations are performed before handing off. Therefore, intervening on the IP mobility scheme could be the only solution to decrease the handover latency. For these reasons, the IEEE 802.21 + DMM mobility management technique is supposed to produce the fastest handover.

5.3 Communication delay

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

In Mobile IPv6, user data traffic always traverses the HA, although this path may not be the shortest one between the MN and the CN. This way of forwarding packets is known as angular routing and is characterized by delays that might get to be large, since the packets must go through the MN’s home network, which can be located at a long distance from the MN. Due to the large delays introduced by the angular routing, MIPv6 [19] already includes a procedure called Route Optimization that basically builds a secure direct path between the MN and the CN. Hence packets exchanged between MN and CN flow directly through the shortest path between the two nodes, without passing through the HA. This mechanism needs additional support from the CN, required to enable the route optimization of packets. In the case of FAMA, packets flow between the MN and the CN through the HDAR as in the case of Mobile IPv6 without RO. The difference between both approaches is that in the case of FAMA, DARs are expected to be located near the MN, hence the effect of triangular routing is highly minimized, obtaining delays of the order of RO-enabled Mobile IPv6. In the previous section, it was mentioned that the use of FAMA is better suited for flows with short duration or low mobility MNs. This is due to the fact that as the MN moves away from the HDAR handling a flow, the inefficiency introduced by the angular routing increases.

Likewise, in PMIPv6, packets need to traverse the LMA, whilst in the DMM approach
they are either routed through the shortest path or to another MAAR first, and then to destination. Basically, we experience the same issue described for FAMA and the DARs: as long as the user is not getting too far from the A-MAAR, the effect of angular routing is lower than that introduced by traversing the LMA.

In order to assess how far and how fast an MN can move, we perform the following analysis. Let's suppose a VoIP communication between two peers, being one of them an MN using one of the DMM schemes to handle its mobility. Considering the maximum mouth-to-ear delay as specified in [18] of 150 ms, we can assume that (5.8) holds:

\[ T_{CN \rightarrow HOME - AR} + T_{HOME - AR \rightarrow MN} \leq 150 \text{ms} \]  

in which \( HOME - AR \) stands for HDAR or A-MAAR according to the appropriate situation.

Let's assume the CN and MN to be in the same geographical region or even city. In order to model this delay, we took average values from the PingER project\(^1\), between several client-server pairs located in the same regional area. The average delay obtained corresponds to roughly 20 ms, hence from (5.2) the delay between the HOME-AR and the MN is upper bounded by 130ms. Assuming that the DMM domain has a good internal connectivity and is all managed by the same provider, we can conclude that the delay between two DARs (or two MAARs) is similar to a local delay between two servers located in the same organization from the PingER project (which is on average equal to 5 ms). To simplify, we suppose the access network is deployed in such a way that going farther away from the original HOME-AR 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 (5.8), 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.

The same delay assumptions hold for centralized approach, but we have to consider also the the angular routing intrinsic in MIPv6 or PMIPv6. So, for instance, we can assume that the distance between an MN and CN 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 centralized 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 the case of a WAN technology such as WiMAX or 3G, one access router can serve a cell of about 50 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 [22] and the average call duration is roughly 3 minutes [30]. 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.

---

\(^1\)Ping end-to-end reporting http://www-iepm.slac.stanford.edu/pinger/
5.4 Final remarks

One might argue that a DMM approach produces location updates more often than a centralized one, resulting in an increased signaling overhead. Actually the update frequency is a parameter left to the operator’s choice as it is strictly related to the network conformation. For the sake of clarifying this concept, Figure 5.1 illustrates a possible scenario for network deployment. Let the colored circles be portions of the edge network managed by a single IPv6 access router (or, similarly, by a single MAG, DAR or MAAR). As long as the MN moves within the circle, it may change its layer-2 PoA, but, still, its default gateway remains the same, and therefore the IP address unaltered. Conversely, when the MN during its movement falls into a different circle, it sees a new default gateway and a proper IP address needs to be configured. Hence a location update must take place independently of the mobility management scheme adopted. Thus, the frequency of updates depends on the network configuration, which is related to the IP cells’ size, rather than the mobility protocol used.
Chapter 6

Conclusions

In this thesis we have proposed and analyzed different solutions for distributed mobility management based on a host-based mobility protocol as Mobile IPv6 and also based on a network-based one as Proxy Mobile IPv6.

These solutions bring the advantages of a distributed approach, namely shorter data paths, better scalability and reliability, shorter handover latencies, and better control on the mobility granularity offered by the network. Unfortunately they come at the cost of an increased overhead due to the proliferation of control messages, but it should be noted that, if some intelligence is implemented to determine whether applications need mobility or not, with our DMM approach we can avoid setting up unnecessary mobility session, thus saving the related signaling. If we push this situation to the extreme, we can enable mobility dynamically only when an application that cannot survive an IP address change is running.

Envisioning such smart interaction between applications and mobility stack, we claim that FAMA is a straightforward solution as the two processes run on the same system, whereas a network-based technique is more interesting from a research perspective. Moreover, in the FAMA architecture, two different levels of security protection are proposed, allowing for a faster and easier deployment with respect to the use of IPsec\(^1\).

Nevertheless, client solutions suffer from several limitation due to the hosts’ features: legacy terminals need to be upgraded in order to benefit of DMM technique. In addition, the mobility signaling and the subsequent tunnels are responsibility of the MN, with the undesired effect of introducing extra energy consumption due to the over-the-air control messages and the packet overhead in the data communications.

For these reasons, network-based DMM solutions have been researched as well, trying to stretch PMIPv6 to adapt to a flat network. Indeed, current mobility solutions used in real operator networks are network-based – as the GPRS Tunneling Protocol (GTP) –, hence it is reasonable to go through network-based DMM solutions as well. A preliminary work on the subject was developed in [16]. Here we re-elaborated with more detail the topic, describing the fully distributed approaches, where the data plane and the control plane is handled by the edge routers only, and we researched partially distributed approaches as well, in which the data plane is distributed among the edge routers, but the control plane is co-ordinated.

\(^1\)FAMA design was honored with the Best Paper Award at Mobiworld 2011, a workshop co-located with IEEE INFOCOM 2011, Shanghai, China, April 2011, and was also presented at the 80th IETF Meeting in Prague, Czech Republic, April 2011 [3]
by a central entity. The latter approach is IETF-friendly, as we could re-use as much as possible existing definitions of message formats, data structures and functionalities, by re-arranging the order of the operations and changing the entities’ state machine. This latter part of the work can be found also in [4], an IETF draft submitted in conjunction with the 81st IETF meeting in Quebec City, Canada, July 2011. The fully distributed approach requires heavier intervention on the IETF protocols’ specifications, hence we integrated an existing mobility framework provided by the IEEE 802.21 suite to gain a complete network control over the handover phase. This work is currently under evaluation by the IEEE 802.21 working group [10, 11], with the intent to enhance the standard with the DMM support.

We claim that, given the results in [14], exploiting the IEEE 802.21 infrastructure with a fully distributed IP mobility management scheme is the preferred solution, as it builds an architecture able to control, assist and decide handovers at the IP flow granularity, providing the best performance in terms of handover latency, and offering a transparent support to heterogeneous access technologies. Besides, the architecture proposed in [14] grounds the basis for the design of the mobility management solution within the MEDIEVAL EU project\textsuperscript{2}. The project envisions a future mobile network optimized to enhance video delivery to mobile terminals, and the mobility management scheme adopted exploits the main results presented here, merging in a hybrid architecture a client-based and network-based DMM design co-ordinated by the IEEE 802.21 infrastructure.

Future work includes the experimental evaluation of the DMM proposals via testbed implementation to prove the feasibility of the approach, and to characterize the required changes on the messages formats, state machine and other protocol details. Moreover, further studies would be necessary to infer, by means of simulations, what are the applicability areas of each solution, with respect to the user mobility and traffic statistical patterns.

\textsuperscript{2}MultiMEDia transport for mobIIE Video AppLications, \url{http://www.ict-medicieval.eu/}
References


