**MobiSplit: a scalable approach to emerging mobility networks**

Julien Abeillé, Rui Aguiar, Telemaco Melia, Ignacio Soto, Patrick Stupar

**Abstract**—This paper presents a novel architecture, MobiSplit, to manage mobility in future IP based networks. The proposed architecture separates mobility management in two levels, local and global, that are managed in completely independent ways. The paper describes the flexibility advantages that this architecture brings to operators, and how it is appropriate for the current trend to multiple and very different access providers and operators. Heterogeneity, support for seamless handovers and multihoming, and scalability issues are analyzed in the paper.

I. INTRODUCTION

The increasing complexity being perceived in next generation mobile networks, with multi-mode terminals always best connected, with multiple types of network available, both operator and community supported, has brought mobility issues into a central role for the future networks. In this context, there are large initiatives, both industry and academia led, that address the multiple aspects of mobility. The EU-funded project Daidalos\(^1\) is one such project, addressing architectures for future networks [13]. Starting from current trends discussed in standardization organizations, the Daidalos architecture is a major breakpoint from traditional approaches in IP networks.

In particular, Daidalos is developing a novel mobility architecture, called MobiSplit. The base element of this architecture is the splitting of mobility management in two domains, Local Mobility Domain (LMD) and Global Mobility Domain (GMD), assuming the IP protocol as basic architecture element. This splitting is done according administrative domain considerations. Seamless handovers and multi-technology local domains are also supported.

The paper is structured as follows. Section II describes previous work in the field of localized mobility architectures. Section III describes the Daidalos Network Based Localized Mobility architecture, its motivation, its design principles, main components and associated procedures. Section IV presents the key advantages of our architecture, and section V concludes the paper.

II. LOCALIZED MOBILITY ARCHITECTURES

Discussions on heterogeneous networks agree on the need of a common protocol for communication, the IP protocol. Mobility is also supported at the IP level, with Mobile IP (MIP) becoming intrinsically supported in IPv6 (or with novel proposals such as HIP [1]). MIP has nevertheless well-known deficiencies both in terms of performance and functionalities. Thus most of the research being done recently has been focused in these aspects, in particular along the lines of localized optimization of mobility behavior, separating the local mobility from the global mobility (MIP-based).

The localized mobility proposals aimed initially to reduce signaling outside the local domain, and improve efficiency by managing the “local” mobility closer to the mobile node MN (reducing the time needed for the signaling required by the mobility). Recently, operational exploitation considerations are gaining increasing relevance.

A. Host based localized mobility management

Initial localized mobility techniques were host-based, i.e. hosts had to handle signaling, and to be aware of local and global signaling protocols. Most relevant previous protocols were Hierarchical MIP and Cellular IP.

1) HMIP

Hierarchical Mobile IPv6 ([10]) is a protocol that extends Mobile IPv6 by introducing a dedicated device, named Mobility Anchor Point (MAP), whose task is to handle the movement of a MN within a defined set of Access Routers (AR). Thus the MAP handles local mobility while Home Agent (HA) handles movements among different local domains. The introduction of this hierarchy (i.e. splitting of mobility management between HA and MAPs) aims at optimizing overhead during handovers among ARs of the same local domain: signaling control is reduced since it is exchanged only between Mobile Node (MN) and MAP and therefore handover latency is reduced. HMIPv6 is a host based solution which requires MN to control both local domain and global domain signaling. Another drawback of HMIPv6 is that it introduces an additional tunnel over the air.

2) Cellular IP

Cellular IP (CIP) [1] [3] is slightly different. MN registers its CoA the IP address of a node (called Gateway) in the
LMD. So all the traffic addressed to the MN will reach the Gateway. To send the traffic from the Gateway to the MN, host routes associated with the HoA of the MN are used inside the LMD. To create these host routes, the (non standard) routers of the LMD use the uplink traffic of the MN (or special purpose signaling packets) to update or refresh the route in the reverse direction. Each router will learn from the uplink traffic which is (for the downlink traffic) the next hop to which to send the packets destined to the MN. All the uplink traffic is forwarded hop-by-hop to the Gateway regardless of its destination.

Cellular IP is a host based solution, as the host has to register explicitly in the Gateway when it arrives to the LMD. This allows the MN to discover the CoA that it must register, and it allows the Gateway to learn that it must start forwarding packets to/from the MN. Also route updates may require the MN to send special purpose packets.

**B. Network based localized mobility management**

Aspects of network control and operation have led to the renewed development of localized mobility solutions, including at standardization level in IETF. Unlike host-based mobility where mobile terminals signal a location change to the network to achieve reachability, network based approaches relocate relevant functionality for mobility management from the mobile terminal to the network.

1) **NetLMM**

The NetLMM approach is currently being designed in the IETF NetLMM Working Group. [5] and [6] define the requirements and rationale for NetLMM. The network learns through standard terminal operation, (such as router- and neighbour discovery or by means of link-layer support) about a terminal's movement and coordinates routing state update without any mobility specific support from the terminal. This approach allows hierarchical mobility management: mobile terminals signal location update to a global mobility anchor only when they change the LMD and, in the LMD, mobility is supplied to terminals without any support for mobility management. NetLMM complements host-based global mobility management by means of introducing local edge domains.

Figure 1 shows the entities involved in NetLMM. The Local Mobility Anchor (LMA) is a router defining the edge between the NetLMM domain and the core network. If a global mobility scheme is used, it is the boundary between GMD and LMD. The Mobility Access Gateway (MAG) is the Access Router for the MN. Note that the host routes are only configured in the LMA and the MAG; all intermediate nodes are totally NetLMM unaware, which considerably reduces the signalling in the LMD and avoids the extensive use of resources (routing tables) in the intermediate nodes (which is not the case with Cellular IP, for example).

The NetLMM area of operation is located between the LMA and the MAG. The forwarding method used between the MAG and the LMA (IPv6 in IPv6 tunnelling) can be extended to General Routing Encapsulation ([9]) or Multi Protocol Label Switching ([10]).

The definition of the interface between the MN and the MAG is not in the scope of the NetLMM WG. However, a basic solution was proposed ([8]), which uses standard IPv6 functionalities of the Mobile Node and the MAG.

![Figure 1 NetLMM Architecture](image)

**III. DAIDALOS NETWORK BASED LOCALIZED MOBILITY MANAGEMENT**

Our mobility architecture, called MobiSplit, is being designed to cover the requirements of the operators in the future. The starting point of our architecture is the split between local and global mobility. According to this concept, the network is divided into several local mobility domains connected via a core network in which mobility is handled via local protocol operations which are transparent to the core network i.e. independent of the global mobility protocol. Indeed the GMP only operates in case of terminal mobility across local domains.

This mobility split brings a number of advantages to the operator. On the one hand, performance and scalability is improved by decreasing signalling overhead, both in the network and over the air. Furthermore, running a Local Mobility Protocol (LMP) allows for more optimized and efficient operations (e.g. handovers), as mobility is handled locally and closer to the terminal. Finally, from an operational viewpoint, LMPs give the operator a greater control of mobility operations as they do not depend on functions external to the operator's network: Operators are provided with the flexibility to choose their preferred LMP (either at layer two or at layer three) and Global Mobility Protocol (GMP) or even avoid maintaining mobility.
infrastructure by directly using the GMP for mobility inside their networks. Note that, during the protocol discussion, we use MIPv6 as GMP to simplify the description but that the architecture could be adapted for operation with other GMPs such as HIP.

A. Architecture Requirements

The mobility architecture requirements for the Daidalos networks can be summarized as:

- **R1** Access Network Operators can implement their own mobility solution (within their domains). The solution must be independent of external Mobility Operators (including home).
- **R2** Minimize complexity in the terminal
- **R3** Efficient use of wireless resources
- **R4** Reduce overhead signalling in the network
- **R5** The mobility architecture must be security friendly
- **R6** Seamless handover support
- **R7** Multihoming support
- **R8** Scalability for routing
- **R9** Minimize network side nodes modifications
- **R10** Support for heterogeneous networking
- **R11** Mobility control can be integrated with QoS control
- **R12** Multicast should be supported.

In the following we will address some of the above requirements, and how our solution handles them. Due to space limitations we will avoid QoS, security and multicast considerations.

B. Overall architecture vision

The basis of the MobiSplit solution follows the design described in II.B with additions that we describe below. Note that we describe a L3 protocol for managing the mobility in the LMD, but any other network based Localized Mobility Management (LMM) solution, not in scope of this paper, could be made to work (e.g. L2 clouds, or MPLS rings).

In MobiSplit, a LMD is associated with an access network provider. It can be small or large, depending on the size of the operator. GMDs are associated with home operators, providing the user with subscription and billing capabilities and, from the point of view of mobility, the ability to roam across different access providers. A first advantage of this approach is that mobility is managed closer to the terminals, thus more efficiently and with less overhead (requirement R4). A second advantage is that makes both mobility management solutions independent of each other, which brings a lot of flexibility to operators (R1).

In fact, in our view, both LMD and GMD are fundamentally the domain of different types of operator (and business models). Access network operators do not depend on functions of an external operator to provide their own connectivity (and mobility) services. On the other hand, home operators can focus on costumer support, and rely on multiple access operators with different Localized Mobility Management (LMM) approaches to provide their users with efficient mobility management over large areas.

The architecture can therefore support multiple LMPs. From R2, the terminal should not implement LMP specific functions, but rather implement mechanisms for triggers to be provided to the network. We propose the use of extensions to the basic IEEE 802.21 [4] framework to act as a common interface to the access network.

As the LMD is fundamentally structured as an administrative domain of an access network operator, it is desirable that an LMD can combine different wireless access technologies. Our architecture design supports both homogeneous and heterogeneous LMDs, and retains the same IP address (CoA) within a single LMD. This feature looks appealing from a security point of view (R5) since LMDs can be un-trusted, and keeping the IP address simplifies IPsec tunnel management. It also provides a certain level of location privacy.

In our concept, LMDs can be large: an operator can deploy its whole network inside the same LMD. For this reason efficient routing configuration in needed in the LMDs (this is the rational for R8).

A differentiating aspect in our architecture is the integration of multihoming. Envisioning future environments with multiple access technologies available anywhere and mobile terminals equipped with several network interfaces, the architecture supports improved host multihoming. This feature is provided both at the GMD and the LMD levels. Multihoming support at the GMD is transparent to the LMD, and vice versa. Host multihoming at the LMD level is a novel concept and has been a challenge for the design of the architecture.

C. Mobility

We provide seamless mobility, leaving the choice to users to experience such mobility over the available interfaces (potentially one or more at the same time) according to their preferences. Note that we have chosen a network based LMM solution instead of a host based. The reason is that a host based solution breaks the independence of the Global and Local Mobility management, and impacts the software (complexity) required in the MN. If the MN needs special functions to manage its mobility in the LMD, the operator looses the flexibility to choose how to manage the mobility in its LMD.

The MN is completely unaware of LMM functions. Seamless (proactive) handovers are provided across heterogeneous technologies by using the 802.21 signalling framework as a cross layer solution. To manage the control plane, messages based on the IEEE 802.21 framework trigger the old AR about the handover being undertaken. This generates triggers that the LMD can use to manage efficiently terminal mobility, but how to do this management is an operator’s decision. Another key advantage is that the terminal only needs to implement one single control interface with the outer world.
To manage the data plane, functions for optimal data delivery during handover might be requested to synchronize with the control plane signalling.

The proposed evolutionary L3 solution to achieve mobility uses tunnels between the LMA and the ARs, thus allowing the forwarding of packets from the LMA to the MN inside the LMD, without requiring the MN to change its IP address. Note that tunnels have several advantages compared with route updates: negligible overhead in the infrastructure; avoids extra traffic to refresh routes and additional processing in each forwarding node;

D. Heterogeneous versus Homogeneous LMDs

The operation case for inter technology is similar to the intra technology case: We envision the use of the same IP address on different interfaces being not active at the same time (in this case we are not multi-homed). The 802.21 framework provides methods for IP address renovation. These methods can therefore be extended in order to allow the terminal to keep the same IP address on different interfaces. That is, the inter technology handover becomes transparent to the GMD. From the network point of view both handovers are handled in the same way.

Note that keeping the same IP address while changing interface can bring some difficulties from the point of view of the implementation, because it is not the normal behaviour of the IP stack [6] and different RTTs across technologies may exist.

E. Multihoming: at the GMD

One of the typical features of a MobiSplit terminal is the use of several interfaces concurrently active. Such characteristic implies many benefits such as ubiquitous access (all technologies are not available everywhere), load sharing and load balancing (to distribute several flows to the different active interfaces) and resiliency. Moreover, mobile multi-homed nodes may move from a place where only interfaces A and B are active to a place where only interfaces C and D can be used. In this situation, the mobility protocol is required to accomplish session. Because it is based in the GMD, the multihoming support provided in this way is independent of the LMD, and in fact, the MN can connect its different interfaces to different access providers/LMDs.

In order to support multihoming at GMD, MobiSplit exploits an enhanced version of the MIPv6 stack, allowing i) registration of multiple CoAs associated to the same HoA for the mobility of single flows and ii) assignment of specific flows to interfaces. This implies that single flows are bound to a CoA, which is strictly related to the assigned interface.

When an interface is activated, the MN registers the CoA configured on that interface with its HA. After this operation, the interface can be used to transmit and receive traffic flows that are expected to keep active after an interface handover.

The assignment of a flow to that interface is executed by the algorithm. Such assignment implies several choices depending on the use of bidirectional tunnelling or route optimization, amongst other factors.

If the terminal operates in bidirectional tunnel for the uplink traffic the terminal transmits tunnelled packets through the chosen interface. This is the same operation executed by plain MIPv6. In case of downlink traffic packets are transmitted towards the Home Network and tunnelled by the HA to the interface selected by the MN. Signalling protocol is required between the HA and the MN.

If the terminal operates in Route Optimization, the support of multiple CoA registrations in a CN is necessary. The MN transmits uplink data packets towards the CN through the chosen interface. The CoA of this interface is used as source address of such packets and HoA is inserted in the routing header. This operation is the same as MIPv6. For the downlink traffic the MN can register with the CN the CoA of the chosen interface. This way, CN is not aware that MN is endowed with several active interfaces.

It is worth noting that multihoming at GMD level may be used to handle the case when the several interfaces are connected to different administrative domains.

F. Multihoming: at the LMD

As we have seen in previous section, a multihoming solution based on GMD support provides the MN with the ability to use several interfaces at the same time, and efficiently divide its flows among the different interfaces.

However, managing multihoming in the LMD offers additional advantages. It allows access operators to manage the mobility of visiting MNs inside their domain without depending on an external operator. In particular, an access provider can optimize the use of its resources by using the best interface/technology to provide connectivity to the MN, according to the MN’s preferences, but also to the general situation of the network. The network can for example decide to move a flow to a different interface on the MN in order to make room for additional users.

We are assuming a MN that can have more than one interface active at the same time. When multihoming is handled at the LMD level, the external entities to the LMD (the HA, the CNs) will not be able to distinguish the
situation in which the MN is using more than one interface from the situation in which it is using one. In other words, the MN will only register one CoA per HoA. No signalling will be sent outside the LMD.

There are two aspects to consider. One is the routing of the downlink and uplink traffic in the LMD. The other is the signaling needed so that the access network and the MN agree on the assignment of flows to interfaces.

In the downlink case, the LMA receives the traffic. This traffic is always destined to the CoA registered in the GMD (HA or CNs). The LMA chooses the appropriate interface to send this traffic to the MN (for example, checking the flow), and it sends the traffic so the MN receives it through the chosen interface from the corresponding AR.

In the uplink case, the MN sends the traffic using the chosen interface. The source address of the packets should be the CoA registered in the HA or CNs, at least when the packets leave the LMA towards the GMD.

To achieve this behavior, the same CoA will be configured on each of terminal active interfaces. This CoA is the one registered in the GMD (HA and CNs). So when a MN activates a new interface, it will configure the CoA of its first interface also in the new one.

The downlink traffic destined to this CoA reaches the LMA. The important issue to notice is that in the MobiSplit solution, that CoA is not used for routing inside the LMD, the routing is done using the (eventually implicit) tunnel to the AR. For this reason, it is not a problem if the CoA is duplicated. The LMA can choose the interface in which the MN will receive the traffic just by using the appropriate tunnel.

Note that some problems appear in a situation in which the two interfaces of the MN are attached to the same AR. To overcome this problem, we envision maintaining several routing tables and neighbour caches in the AR and use tagged routing. When a packet arrives at the AR with a particular tag, the routing lookup process will only have access to specific routing tables and neighbour caches.

We have analysed the transport problem. In the following we consider the signalling aspects of the solution.

The MN must actively ask the access provider to manage its multihoming. For this, when it configures an additional interface, it must signal the corresponding AR that this interface is associated to the other one, and the AR will inform the LMA. A MN ID can be used to allow the LMA to know that several interfaces belong to the same MN.

The MN and the LMA must also agree on the mapping between flows and interfaces, so the uplink and downlink traffic belonging to the same flow follow the same path. An easy solution is that the MN chooses the interface for the traffic it initiates, and the LMA chooses the interface for the traffic initiated in a CN. This is not very restrictive, as the normal procedure for intra-LMD network or mobile initiated handovers can be afterwards used to move a flow from one interface to another.

**G. Scalability and large LMDs**

LMDs can be large. This can have an impact on the routing efficiency inside the LMD. If the MN is moving inside the LMD keeping the same LMA, the traffic from the LMA to the AR serving the MN can become quite long. Two solutions are proposed to improve the efficiency of the communication in this scenario. Choosing between them is an operator decision depending on deployment issues. The two solutions are presented below and in Figure 3.

![Figure 3 Two solutions to provide scalability](image)

1) **Inter-LMA handover**

This solution solves the problem by setting up a new and supposedly optimal path between the MN and the HA or CN. It assumes that the ARs have connectivity with several LMAs. The solution consists of two phases (see Figure 3):

In the first phase, a MN performs a standard inter-AR handover: the data path goes from path 1 to path 2. As the path is not optimal and has to be updated, then the network triggers the inter-LMA handover. In a second phase, path 3 is setup between AR and LMA, with the corresponding MN states in these entities. Afterwards, and while both path 2 and path 3 are available, Global Mobility is performed. Once path 3 is used, associations on path 2 can be revoked.

The drawback of this solution is that it forces GMP operations for managing mobility inside the LMD and a change of IP inside the LMD, loosing some of the advantages of the design. For this reason, the triggering of the inter-LMA handover should not be too frequent. The decision for this trigger may be based on the difference between the prefix belonging to the LMA and the prefix belonging to the AR or on the algorithm used at bootstrap by an AR to discover the closest LMA.

2) **Inter-LMA communication**

This solution may be less efficient than the previous one, but avoids the drawbacks presented above: global mobility is not involved and the MN keeps its IP address.

It requires that the physical links between the LMAs are high performance links, which can be a reasonable deployment scenario. It is composed of two phases: in the first phase, a MN performs a standard inter-AR handover.
The data path goes from path 1 to path 2. A specific algorithm to be designed detects that the path is not optimal and has to be updated and the network triggers the inter-LMA communication. An association is created between LMA1 and LMA2, and between LMA2 and AR2; states for the MN are also created in these entities. Once the new path (path 4) is used, the states and routes corresponding to the old path (path 2) can be revoked.

Though the new path may be geographically longer than the old one, the high performance of the link between the LMAs should sustain the performance of the communication.

IV. PROTOCOL COMPARISON

Many protocols have been developed for handling mobility, as we have seen. Over these protocols, many variants have been developed. For instance, CIP principles could be also used in a network based localized mobility solution. Without being exhaustive, Table 1 shortly summarizes some characteristics of different protocols.

<table>
<thead>
<tr>
<th>Protocol</th>
<th>CIP</th>
<th>H</th>
<th>MIP</th>
<th>Net LMM</th>
<th>Mobi split</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local/global</td>
<td>L</td>
<td>L</td>
<td>G</td>
<td>L+G</td>
<td>L+G</td>
</tr>
<tr>
<td>Multihoming</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Network overhead</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Y</td>
</tr>
<tr>
<td>Seamless handover</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Terminal modifications</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
</tbody>
</table>

Nevertheless, the key differentiating aspect of MobiSplit is not clear from the table. In MobiSplit, domains are associated with administrative ownership. With the above splitting, interconnecting different operators becomes easier. While retaining the overall interoperation, network operations can be managed according to access provider’s or home operator’s preferences. This is achieved through the synergistic way we incorporated network-based mobility, multihoming and choose a cross-layer approach (802.21) in the terminal to handle signalling across multiple technologies.

V. CONCLUSION

In this article we have presented a novel architecture for future mobile operators. The architecture recognizes the current trend in networks to a heterogeneous landscape of access providers. In this environment it is important to give the access providers the flexibility of managing the mobility inside their domains according to their needs, technologies, and requirements, without being conditioned by how mobility is managed in other domains.

To cope with this concept, the architecture proposed in this paper splits the mobility management in two levels: the local domain and the global domain; and the management of the mobility in these two levels is kept completely independent. The architecture proposal also supports features that are essential to make this proposal useful for operators, such as heterogeneous (multi-technology) local domains, seamless handovers, support for multihoming, and scalability and routing efficiency in local domains considering that, depending on access operator needs, local domains can range from very small to really large ones.

REFERENCES