1. Introduction

With the popularization of the Internet and mobile devices, like notebooks and PDAs (Personal Digital Assistants), the concept of portability started to become popular, in the sense that the user could take their device anywhere and start a new connection to the Internet. In the Internet, a node is identified by an IP (Internet Protocol) address that uniquely identifies its point of attachment to the Internet, and packets are routed to the node based on this address. Therefore, a node must be located on the network indicated by its IP address in order to receive datagrams (Akyildiz et al, 2004). However, when moving to different networks, the IP protocol does not allow the current IP address to be valid in the visiting network and the device should ask for a new IP address when entering a new network (Figure 1). It was then necessary to provide a scheme to allow nodes to be reachable and maintain ongoing connections while changing their location within the topology, in a seamless\(^1\) way, when leaving its home (initial) to a visiting (new point of attachment) network.

One of the first solutions proposed to solve this problem for IP networks was the Mobile IP Protocol (Perkins, 1997), also known as MIP. The MIP protocol aims to solve the problem of node mobility by redirecting packets to the mobile node (MN) to its current location. MIP was a scheme suited for interdomain mobility, allowing MN’s movements between different networks from different domains.

But protocols proposed for interdomain mobility are not suited to intrasubnet mobility due to drawbacks such as the need for new protocols on the MNs that exceed their processing power what make these solutions simply undeployable in most mobile devices. The problems are related to the complexity of the proposed solutions which make their

---

\(^1\) Seamless mobility: capability to change the mobile node’s point of attachment to an IP-based network, without losing ongoing connections and without disruptions in the communication.
implementation on small mobile devices such as cell phones and handhelds unfeasible. In fact, manufactures of such devices never considered supporting the present solutions (Zagari et al, 2008).

![Diagram of visiting network](image)

Fig. 1. Current IP address not valid in a visiting network

After MIP, new protocols for mobility between networks in the same domain were proposed. Micromobility protocols aim to improve localized mobility by reducing handover overheads. Other approaches to allow seamless mobility also include mobility provided by transport and session layer schemes.

The objective of this chapter is to provide a major review on mobility protocols and architectures. The protocols and architectures will be classified according their mobility range (intra and inter domain), layer in which it operates (from the link layer to the application layer), and the support required from the mobile node. We will place emphasis on the solutions called “network-centered”, that is, solutions where mobility is handled entirely by the network without the need of installation of mobility protocols on the mobile nodes. The protocols and architectures discussed in this chapter are being proposed by standardization bodies, e.g., IETF, by industry-driven forums, e.g., 3GPP, by academy and by the industry.

This chapter is divided into 10 more sections. Section 2 presents an overview of mobility issues and a classification schema, which will be used in the protocols sections. Mobile IP is presented in Section 3. Section 4 shows link layer based protocols. Section 5 presents mobility solutions based on L2½ protocols. Section 6 presents network layer protocols. Section 7 presents transport layer protocols. Section 8 presents mobility using application layer protocols. Section 9 presents the Mobility Plane Architecture. Section 10 presents a general classification of the mobility solutions seen in this chapter and future work related to mobility in IP networks. Finally, Section 11 concludes this chapter.
2. Mobility Issues

The mobility process starts with the mobile node’s attachment to a local wireless network. The node attachment process happens when a MN enters in the coverage of a wireless access point. In this point a L2 (Layer 2) attachment process is performed. After that, the MN must acquire its IP address from the network. After obtaining its new address, the network is able to route packets to/from the mobile node (Johnson et al, 2008).

When the MN moves away from the current access point, it may detect another wireless access point. Handover is “the process by which an active MN changes its point of attachment to the network, or when such a change is attempted. The access network may provide features to minimize the interruption to sessions in progress” (Manner & Cojo, 2004) by preserving the transport (or higher layers) connections such a way the packets are forwarded to the MN via the new access point.

Mobility management is the key to enable this seamless mobility. It enables wireless or mobile networks to search and locate mobile devices for network communications and to maintain network/applications connections as the MN moves into a new service area. The mobility management is composed of mainly two services: location management and handover management.

Location management consists of two operations: registration or location update and paging, to enable a network to discover the current point of attachment of an MN for information delivery (Saha et al, 2004). Location update is used in support of idle users, and paging is used in support of active communications (Campbell et al, 2002).

Location update procedures need the MN to periodically inform the system to update relevant location databases with its up-to-date location information (Akyildiz et al, 2004). Paging is the ability to track idle mobile hosts. For protocols using this kind of tracking, idle MNs do not have to register if they move within the same paging area, but only if they change paging area (Campbell & Gomez-Castellanos, 2000).

Handover management enables the network to maintain a MN’s connection as it continues to move and change its access point to the network (Saha et al, 2004). There are many types of handover, among which are:

- horizontal and vertical: horizontal handover occurs between wireless cells of the same technology; vertical handover occurs between two different networks of different technologies. This chapter is dedicated to study horizontal handover only.
- mobile and network initiated: in MN initiated handovers the MN is responsible for initiating handover requests, while in network initiated handover the network is responsible for indicating that a handover must occur.
- MN and network controlled: in MN controlled handover, the MN must participate in the handover process, while in the network controlled handover the network handles the entire process.
- fast: fast handover tries to reduce the latency during a handover.
- seamless: change the MN’s point of attachment to an IP-based network, without losing ongoing connections and without disruptions in the communication.

The basic terminology for mobility is (Perkins, 2002; Manner & Kojo, 2004):

- home network: a network having a network prefix matching that of a mobile node's permanent address;
• home address: the IP address acquired when registering in its home network; a stable address that belongs to the mobile node and is used by correspondent nodes to reach mobile nodes;
• home agent (HA): A router located on the home network that acts on behalf of the mobile node while away from the home network;
• correspondent node (CN): Any node that communicates with the mobile node;
• foreign network: Any network (other than the home network) visited by a mobile node;
• foreign agent (FA): A router located on the foreign network that acts on behalf of the MN in this network;
• care-of-address (CoA): An address that is assigned to the mobile node when located in a foreign link.

2.1 Classification parameters
Existing proposals for mobility can be broadly classified into different types, based on many parameters. We will employ a taxonomy based on 4 axis: Mobility Range, Mobility Routing, Mobility Signaling and Mobility Layer. In this section we will briefly introduce each one of them.

2.1.1 Mobility Range
In this work we adopted the definitions by Manner & Kojo (2004) for the mobility scope.

Micromobility
Also called intradomain mobility or local mobility (Kempf, 2007), is the process of mobility over a small area. Usually this means mobility within an IP domain with an emphasis on support for active mode using handover, although it may include idle mode procedures also. Micromobility protocols exploit the locality of movement by confining movement related changes and signaling to the access network.

Macromobility
Also called interdomain mobility or global mobility (Kempf, 2007), is the process of mobility over a large area. This includes mobility support and associated address registration procedures that are needed when a MN moves between IP domains. Interdomain handovers typically involve macromobility protocols. MIP can be seen as a means to provide macro mobility.

2.1.2 Mobility Routing
Defines how the MNs’ location information database is created and maintained.

Routing based
Routing based schemes aim to exploit the robustness of conventional IP forwarding. A distributed mobile host location database is created and maintained within the network domain. The database consists of individual flat mobile-specific address lookup tables and is maintained by all the mobility agents within the domain (Chiussi et al, 2002).

Tunnel based
In tunnel based schemes, the location database is maintained in distributed form by a set of foreign agents in the access network. Each foreign agent reads the incoming packet's
original destination address and searches its visitor list for a corresponding entry. If the entry exists then it contains the address of next lower level foreign agent. The sequence of visitor list entries corresponding to a particular mobile host constitutes the host's location information and determines the route taken by its downlink packets. Entries are created and maintained by registration messages transmitted by mobile hosts (Campbell & Gomez-Castellanos, 2000). Tunnels may be IP-IP (IP over IP) or MPLS (Multi-protocol Label Switching).

2.1.3 Mobility Signaling
Defines if the mobility signaling is carried out by the network alone or also needs the mobile node participation in the signaling process.

Mobile Node Centric
In this approach, the MN must execute an instance of the mobility protocol, thus participating actively in the mobility management process. But the requirement for modification of MNs software may increase their complexity; considering these nodes have less computational capacity, it may lead to performance degradation on the MN.

Network Centric
In network-based mobility management approach, the serving network handles the mobility management on behalf of the MN; thus, the MN is not required to participate in any mobility-related signaling. Contrary to the latter approach, the MN’s performance is not degraded by processing signaling and mobility protocol management.

2.1.4 Mobility Layer
Defines the responsibility of each layer in the mobility management process.

Link Layer
This class includes mobility protocols that use link layer information, when the point of attachment changes, to provide mobility management while the node preserves its network-layer (L3) address. This can fulfill some of the attributes of a micromobility protocol.

Layer 2½
This class uses MPLS to provide mobility management and signaling. MPLS (Rosen et al, 2001) is a technology that substitutes conventional packet forwarding within a network, or part of a network, with a fast operation of label lookup and switching. Each MPLS packet has a label. Label swapping is done by associating labels with routes and using the label value in the packet forwarding process.

In an MPLS cloud, switches are called Label Switching Routers (LSRs) and a connection or tunnel between two endpoints is formed by the union of several LSRs along a route. It is called a Label Switch Path (LSP). When a packet enters into an MPLS cloud, the egress LSR classifies the packet accordingly to the rules defined in its Forwarding Equivalence Class (FEC) and each FEC has an association with a particular LSP.

Through this mapping (FEC - LSP), a label is assigned to the package, which only identifies the LSP to the downstream LSR in the LSP, so that they can continue this procedure until reach the egress (edge) LSR. In core LSRs, the procedure is simpler, since the packet reclassification is no longer required, but just forwarding it to the downstream LSR. Note that the label has only local significance. Before a packet leaves an MPLS domain, its MPLS label is removed (Ren et al, 2001).
Network Layer
All mobility management and signaling is carried out by L3 protocols, based or not in the Mobile IP protocol.

Transport Layer
Mobility on transport layer intends to maintain TCP (Transmission Control Protocol)'s end-to-end reliability and correctness semantics while allowing redirecting the endpoints of an existing transport session (e.g., a TCP connection or a series of UDP - User Datagram Protocol- packets) to arbitrary addresses (Maltz & Bhagwat, 1998).

Application Layer
Mobility provided by application layer protocols intends to allow communication end systems to support mobility, heterogeneity, and multihoming. Terminal mobility also allows a device to move between IP subnets, while continuing to be reachable for incoming requests and maintaining sessions across subnet changes (Schulzrinne & Wedlund, 2000). Session mobility also allows a user to maintain a media session even while changing terminals. For example, a user may want to continue a session initiated on a MN on the desktop PC when entering his/her office. IPv4 or IPv6 mobility does not directly support such session mobility (Nasir & Mah-Rukh, 2006).

3. Mobile IP
The Mobile IP (MIP) (Perkins, 1997; Johnson et al, 2004) uses a stable IP address assigned to mobile nodes. This home address is used to allow the MN to be reachable by having a stable entry in the DNS service, and to hide the IP layer mobility from upper layers. A consequence of keeping a stable address independently of the mobile node's location is that all correspondent nodes try to reach the MN at that address, without knowing the actual location of the mobile node. Therefore, if there are packets forwarded to the home address, and the MN is not at its home network, its home agent is responsible for tunneling packets to the MN’s new location.

MIPv4 (Mobile IP for IPv4 networks) solves the mobility problem by allowing the MN to use a second IP address: the CoA. This address changes at each new point of attachment and it indicates the network prefix, identifying the MN’s point of attachment with respect to the network topology. The CoA is composed of a valid prefix in a foreign network. Thus, the MN will have a home address and one or more CoAs when moving between networks.

MIPv4 works by the cooperation of three separable mechanisms (Perkins, 1998): discovering the CoA, registering the CoA and tunneling to the CoA. The operation of Mobile IP protocol can be briefly described by the following steps (Figure 2):

1. The mobility agents (HA and FA) announces their presence through messages called Agent Advertisement (optionally, these messages can be requested by mobile agents through messages called Agent Solicitation);
2. A MN receives these messages and determines whether it is on its home network or on a foreign network;
3. When a MN detects it moved to a foreign network, it obtains a CoA in that network. The CoA can be allocated by the foreign agent or some other address configuration mechanism, such as DHCP (Dynamic Host Configuration Protocol);
4. When the MN is operating in the new network, it needs to register its CoA with its HA, through the exchange of Registration Request and Registration Reply messages;
5. Datagrams sent to the MN’s home address by a CN are intercepted by the local HA and tunneled to the MN’s CoA. The datagram is received at the exit of the tunnel, and finally delivered to the mobile node in the new network;
6. Datagrams sent by the MN are generally delivered to the destination using standard routing mechanisms, not necessarily through the HA.

The cooperation between MN, HA and CN is called triangular routing, as we can see in Figure 2, which summarizes the MIPv4 operation.

The triangular routing generates a processing overhead on HA, in addition to this being a single point of failure in the network. The MIPv6 solves this problem by optimizing the route.

Mobile IPv6 (Johnson, 2004) is intended to provide mobility support in IPv6 networks. In order to know where the MN is found, an association between home address and care-of address should be performed (binding). This combination of CoA is made by the MN and the HA. This association is achieved by a binding registration where the MN sends messages called Binding Updates (BU) to HA, which responds with a message Binding Acknowledgment (BA) (Figure 3).

The correspondent nodes may carry out route optimization, or they can store bindings between MN’s home address and CoA. Thus, a MN can supply information about its location to the corresponding nodes, through the Correspondent Binding Procedure, which is a mechanism for authorizing the establishment of binding, called the return routability procedure.

![Fig. 2. MIPv4 Operation](image)

Using the Route Optimization process, the CN must support MIPv6 and the MN must register with the CN. In this case, the CN, before sending a package, looks for a cached association between MN’s HA and CoA. If there is an association, the package will be
routed to the CoA of mobile node directly. This eliminates congestion at the home link and the HA.

Fig. 3. MIPv6 Operation

MIPv4 and MIPv6 only define means of managing macromobility but do not address micromobility separately. Indeed, it uses the same mechanism in both cases. So, this protocol is not suited for micro mobility management, because of its high signaling load and long handover delay (Habaebi, 2006), namely movement detection, new CoA configuration, and Binding Update, is often unacceptable to real-time traffic such as Voice over IP (Koodli, 2008).

4. Link Layer related Micro-Mobility

FMIP - The Mobile IPv6 Fast Handovers (FMIPv6) protocol (Koodli, 2008) aims to reduce MN movement detection latency and new MN’s CoA (Care-of Address) configuration latency by providing information to the MN when it is still connected to its current subnet. After discovering available access points, the MN requests subnet information from these APs: prefix, IP address, and L2 address of their associated routers. If the MN eventually attaches to one of the APs, the movement detection delay is reduced because the MN doesn’t need to perform router discovery.

The MN formulates a new CoA (NCoA) based on the prefix of the new subnet and sends a message to its current access router, Previous Access Router (PAR), which communicates with the New Access Router (NAR) to determine whether the NCoA is unique. The PAR also establishes a tunnel to redirect packets arriving for PCoA (Previous CoA) to NCoA. After performing a handover the MN announces its attachment immediately with an Unsolicited Neighbor Advertisement message (Narten et al, 2007) to circumvent the delay associated to neighbor's address resolution. FMIPv6 also defines a different behavior when the MN doesn’t receive acknowledge message prior to its handover. There is also an adaptation of FMIPv6 to IPv4 networks (Koodli & Perkins, 2007).

Access Point (HAP) that is the AP to which the MN was last associated inside its home network, similar to the Home Agent in MIP. When the MN moves to a different network, it sends a modified IEEE 802.11 Reassociation Request (IEEE, 1997) to an AP, the Foreign Access Point (FAP), informing IP addresses of HAP, MN, and Previous FAP (PAP) and this message triggers the mobility management procedure. The FAP communicates with the HAP to establish a bi-directional HAP-FAP tunnel and the HAP starts mapping the MN IP address to the FAP IP address, the Foreign Agent Care of Address (FACOA).

When the MN reassociates with a new FAP (NAP), the same procedure is performed with the addition of a communication between the PAP and NAP to establish a temporary unidirectional tunnel between them. The proposal has also been improved with the provision of more advanced services: secure inter-AP IP-IAPP communications, zero patching on the clients software, and support of clients which use a dynamic IP address (Samprakou et al, 2007). The IEEE 802 Executive Committee approved IAPP withdrawal in 2006, because “the trial use period of 802.11F has expired, there has been no significant deployment of 802.11F implementations and, the functionality provided by 802.11F is being addressed in other standards fora” (IEEE P802.11, 2005).

5. Mobility with MPLS

All of the architectures discussed in this section consider that the MPLS cloud is surrounded by the Internet cloud and that micromobility is to be applied in MPLS cloud while MIP is to be applied in Internet cloud.

Mobile MPLS - The Mobile MPLS is a macro mobility protocol that borrows the mechanisms defined in the MIP standard and applies it to MPLS networks, so that the IP-in-IP tunnels are substituted by MPLS tunnels. The main objective in this migration is to improve the delay time in the tunneling packets from the HA to the FA. Another objective is to facilitate the use of QoS services that are native to MPLS networks (Ren et al, 2001).

In order to track the MN location, an entry in the LIB (Label Information Base) table at the HA is created for each MN that is registered with it. When the MN arrives at a foreign network, it registers itself with this FA and obtains a CoA from it. The FA sends this information to the MN’s HA and it establishes a new LSP for that FA. After the completion of the LSP connection, the HA changes the LIB entry for that MN to reflect the out label and port interface gathered from the previously created LSP. In doing that, whatever packet that is sent to the MN’s home network will be tunneled to this LSP, arriving at the FA in which the MN is actually connected. A lack of an entry about out label and port interface for the MN at the LIB table at the MN’s HA means that the MN returned to its home network.

Three scenarios were discussed. The first one considered was that both FAs and HAs were inside the same administrative MPLS domain. The second one considered was that HAs and FAs were inside different administrative MPLS domains. In order to establish a tunnel between them, Mobile MPLS suggests the use of a border protocol such as BGP (Border Gateway Protocol) (Rekhter & Rosen, 2001). The third scenario considered was that HA and FA were inside a different network tunneling technology, such as MPLS and IP clouds. LER
is responsible for de-tunneling packets from the MPLS cloud and re-tunneling it inside the IP cloud.

**H-MPLS** - Hierarchical Mobile MPLS (Yang & Makrakis, 2001) extends Mobile MPLS, which is a macro mobility protocol, in order to introduce into it micro mobility features. The main objective is to reduce the signaling overhead in creating a LSP from HA to FA, due to MN frequent handover in a small-size cells wireless environment.

To do that, H-MPLS introduces a new element, called FDA (Foreign Domain Agent) whose function is between that defined for HA and FA, as described in the MIP standards. The role of FDA is to track MN local mobility inside a MPLS domain. There is only one FDA per MPLS domain and many FA per subnetworks inside this domain.

The dynamics of the protocol is as follows: whenever a MN enters a foreign MPLS domain and it is its first registration, it acquires a CoA from its FA LSR and registers with it. This FA sends a Registration Request to its FDA, which has an equivalent function of a FA, but its scope is for a domain. This FDA will send back a Label Request message to FA and put as its FEC, the MN's CoA. At meanwhile, FDA will send a Registration Request message to HA, in the same way that was described in Mobile MPLS, but putting its IP address as a FEC for this LSP. So, at this point, there will be two LSPs, one from HA to FDA and another one from FDA to FA.

Now, if a MN does a handover, but stays in the same domain as the previous FDA, only the LSP from the new FA to FDA needs to be established. To avoid FDA sending packets to MN via the old FA, due to an out-of-date entry cache, the new FA sends a Binding Update message to old FA instructing it to create a LSP to the new FA in order to tunnel packets arriving at the MN's old location.

**LEMA** - Label Edge Mobility Agent (Chiussi et al, 2002) is a tunnel-based micro mobility architecture that uses MPLS as a network transport technology. This network is composed of an overlay network whose nodes are called LEMA, an LER that has its function augmented with LEMA features. This overlay network tracks MN location by building a set of LEMAs nodes from the highest to the lowest LEMAs that compound a path.

Highest LEMAs are the ingress node which registers its address in the HA database, and acts as FA in the MIP protocol; while the lowest LEMA is the access router, which remove the MPLS tunnel and delivers messages to the AP which the MN is connected to. This kind of scheme makes a hierarchical network, where only the LEMAs that compose a path to track the MN need to be aware of it.

Others features that can be attributed to it are fast handover capability, scalable design, QoS capability and gradual deployment. It is the MN’s role to define the set of LEMAs that compose its path inside a LEMA network. This path is chosen based on a set of parameters, such as: available bandwidth, mobility patterns, and so on. The algorithm employed to choose a particular set is an open issue, and could be of high complexity. Finally, all LEMAs are connected to themselves by pre-established LSPs.

**MM-MPLS** - Micro Mobile MPLS (Langar et al, 2004) extends mobile MPLS with the principles employed by MIP-RR (L3 protocol) to support micro mobility on MPLS networks. It introduces a new component, called Label Edge Router/Gateway (LER/GW) that resides between HA and FA agents, as defined in the MIP protocol. It acts as a foreign domain agent to HA and it is this address that the MN must register at HA database when it first gets into the domain at which LER/GW is administrating.
So, there is a tunnel/LSP from HA to LER/GW, and an LSP from LER/GW to FA, with MN CoA's address as a FEC. FA here is an AR (Access Router) that remove the MPLS tunnel and delivers the packet to MN that is registered on it. Whenever MN moves to another FA, which is under the same LER/GW, only a regional registration is required that will create a new LSP from LER/GW to the new FA, using the new MN CoA's address as FEC. MM-MPLS uses LDP (Label Distribution Protocol) as signaling protocol in MPLS cloud.

**I-LIB** - Intermediate Label Information Base (Fowler & Zeadally, 2006) maintains the same idea of MM-MPLS architecture in general, where a FDA is placed between HA and FA. FDA has the same role as described early, i.e., its CoA address is registered at HA database and a tunnel connect HA to FDA. On the other side, FDA keeps track of MN by establishing a LSP to FA that is directly connected to it. Whenever a MN does a handover and establishes a new connection to a new FA, this new FA will try to establish a LSP to FDA, sending a Registration Request.

Here is where this proposal differs from the previous ones. Instead of establishing a new LSP from scratch, linking the new FA to the FDA, any segments that are common between the new path and the old path will be preserved. As such, any LSR that already has an entry in its LIB could preserve it and just update it to show the new configuration (the new segment that connects the MN). In order to do that, a new LIB is proposed which augmented the old ones with new fields to contemplate mobility issues. Among the fields that are required, the previous and new MN's CoAs must be accounted for. It is necessary to modify the packet since the FDA and HA know the MN by its old CoA, while the FA knows the MN by its new CoA.

### 6. Network Layer

**Cellular IP** – The CIP (Valko, 1999) architecture is composed of different wireless access networks (CIP access networks) connected to the Internet through a gateway. MIP manages mobility between these CIP access networks, while Cellular IP handles mobility within one domain. The IP address of the gateway is used as the MIP CoA. Thus, packets are first routed to the host’s HA and then tunneled to the gateway, which “detunnels” packets and forwards them toward base stations, using host-specific routing path. Base station (BS) components serve as wireless access points and also route IP packets, but IP routing is replaced by Cellular IP routing and location management. Base stations cache the path taken by uplink packets from MN to gateway for a period of time and use the reverse path to route downlink packets. In order to route packets to idle MNs, Cellular IP employs paging.

**HAWAII** - HAWAII divides the network into a hierarchy of domains. All issues related to mobility management within one domain are handled by a gateway called a domain root router, which uses a specialized path setup scheme which installs host-based forwarding entries in specific routers to support intra-domain micromobility (Ramjee et al, 1999). While moving inside its home domain, the MN maintains its stable IP address. MIP mechanisms are used when the MN moves into a foreign domain. However, if the foreign domain is also based on HAWAII, then the MN is assigned a co-located CoA from its foreign domain to which packets for the MN are tunneled. The domain root router routes the packets to the MN using the host-based routing entries. When the MN moves between different subnets of the same domain, only the route from the domain root router to the BS
serving the MN is modified, and the remaining path remains the same, and connectivity is maintained using dynamically established paths.

The protocol contains three different messages for establishing, updating and refreshing host specific routes for the MN in the domain root router and any intermediate routers on the path towards the mobile host. The protocol also has four different path setup schemes, aiming to reduce disruption to the user traffic during a handoff, classified into two types based on the way packets are delivered to MNs. In the first type, packets are forwarded from the old base station to the new and, in the second type, they are diverted at the crossover router.

**MIP-RR** - MIP-Regional Registration (Fogelstroem et al, 2007) is an optional extension to the Mobile IPv4 protocol, and proposes a mean for mobile nodes to register locally within a visited domain. By registering locally, the number of signaling messages to the home network is kept to a minimum, and the signaling delay is reduced. This protocol introduces a new network node called the Gateway Foreign Agent (GFA). Besides the regular MIP Registration messages, a new pair of registration messages, Regional Registration Requests/Replies, is used between MNs/FAs/GFAs.

There are two models of how the MN uses Regional Registration. In the first model, the FAs in a visited domain advertise the address of the GFA, and, when a mobile node first arrives at this visited domain, it performs a home registration. At this registration, the mobile node registers the address of the GFA as its CoA with its HA. When moving between different foreign agents within the same visited domain, the mobile node only needs to make a regional registration to the GFA. In the second model, the FA can indicate that dynamic assignment of GFA is to be used, if being the FA’s responsibility to choose the GFA after receiving a Registration Request from the MN.

**PMIP** - The Proxy MIP (Gundavelli et al, 2008) is a network-centric micromobility approach that relies on tunnels inside a domain to direct traffic to mobile nodes. PMIPv6 reuses many concepts of MIPv6, like the HA functionality, and defines two new entities, the Mobile Access Gateway (MAG) and the Local Mobility Anchor (LMA).

A MAG typically runs on the Access Router. It is responsible for detecting the mobile node attachments, and, if security policies are fulfilled, establishes tunnels to the LMA for directing traffic to the mobile nodes reached via this MAG. A MAG also emulates (via Router Advertisements messages) the mobile node’s home network in such a way that the mobile node may change the default router in a handover, but preserves the remaining L3 parameters. As the mobile node moves inside the domain, tunnels between MAG and LMA and routes on the LMA are updated.

LMA maintains a binding cache entry for each currently registered MN, providing reachability to the MN’s address. When the LMA receives a packet targeted to the mobile node it forwards the packet via the tunnel ending on the MAG to where the node is attached. PMIPv6 employs local binding update messages between MAG and LMA for signaling purposes and only a single hierarchy of tunnels. Indeed, Proxy MIP considers IPv4 support, but this requires an extension to the original protocol, since it uses some IPv6 features such as auto-configuration and extension headers.

**HMIPv6** - To support local mobility, Hierarchical Mobile IPv6 (HMIPv6) extends Mobile IPv6 and IPv6 Neighbor Discovery (Narten et al, 2007) and introduces Mobility Anchor Point (MAP), a new Mobile IPv6 node. A mobile node entering an HMIP domain receives
Router Advertisements containing information about one or more local MAPs and configures two CoAs: an on-link CoA (LCoA) and a Regional Care-of Address (RCoA). The LCoA is configured on a mobile node's interface based on the prefix advertised by its default router. It is a standard Mobile IP CoA and has a different name just to be distinguished from RCoA. The RCoA is configured on the MAP's link and is obtained by the MN from the MAP employing the address mechanisms described by RFC 4877 (Devarapalli & Dupon, 2007).

After configuration, the MN sends two binding update (BU) messages. The first is a local BU to the MAP to bind the MN's RCoA to its LCoA and to establish a bi-directional tunnel between them. The second is a BU to the home agent to bind the MN’s home address to its RCoA. The MAP receives all packets on behalf of the mobile node it is serving and encapsulates and forwards them directly to the mobile node's LCoA.

When the mobile node moves within the same MAP domain, it only needs to register its new LCoA with its MAP, limiting the amount of Mobile IPv6 signaling outside the local domain. The RCoA remains unchanged and the home agent (HA) or the correspondent nodes (CNs) are not aware of the change in LCoA.

**DMA** - The Dynamic Mobility Agent (Misra et al, 2001) architecture uses the Intra-Domain Mobility Management Protocol (IDMP) (Das et al, 2002) to manage intradomain mobility. The architecture defines two entities to achieve mobility support: Mobility Agent (MA) that acts as a domain-wide point for packet redirection, and the Subnet Agent (SA) that provides subnet-specific mobility services. Two CoAs are associated with a MN: Global CoA (GCoA) and Local CoA (LCoA).

The GCoA is the address used by macromobility protocols to redirect packets and remains unchanged as long as the MN stays in the current domain. The LCoA is an address from the subnet the MN is attached to. IDMP is used in the communication between the MN and the SA, and between the MN and the MA.

When the MN first arrives at the domain, it obtains an LCoA and the SA assigns the MN a MA. The MN registers its LCoA with the MA and obtains a GCoA. After the macromobility updates process is performed by the MN, the packets from remote hosts, tunneled or directly transmitted to the GCoA, are intercepted by the MA and tunneled to the MN's LCoA. When the MN moves to new subnet it obtains a new LCoA and informs its MA of the new LCoA, updating the GCoA-LCoA mapping.

### 7. Transport Layer Mobility

**MSOCKS** is a split-connection proxy-based architecture that uses TCP Splice technique to achieve the same end-to-end semantics as normal TCP connections (Maltz & Bhagwat, 1998). A special host, called a proxy, is placed in the communication path between a mobile node and a correspondent node. An end-to-end TCP connection between a mobile node and a correspondent node are split into two separated connections: one connection between the mobile node and proxy and another between the proxy and the correspondent node. The MSOCKS protocol extends the SOCKS protocol (Leech et al, 1996) to redirect TCP streams to a mobile node’s changing location.

When the mobile node changes the address of its network interface it opens a new connection to the proxy and sends an MSOCK message specifying the connection identifier of the original connection. The proxy unsplces the old mobile-node-to-proxy connection
from the proxy-to-correspondent-node connection, and splices in the new mobile-node-to-proxy connection. Only the proxy is aware of the mobile node migration and the communication between the proxy and the correspondent node remains unchanged. This technique also allows the mobile node to change the network interface used to communicate with the proxy.

**TCP Migrate** – This mobility architecture (Snoeren & Balakrishnan, 2000) (Snoeren et al., 2002) allows an application running on mobile hosts to support transparent connectivity across network address changes. As MNs change their network attachment point, new addresses can be assigned through DHCP, manually or using an auto-configuration protocol. To locate mobile hosts in the new network, Domain Name System (DNS) is used and its ability to support secure dynamic updates.

Because most Internet applications resolve hostnames to an IP address at the beginning of a transaction or connection, this approach is viable for initiating new sessions with mobile hosts. When a host changes its network attachment point (IP address), it sends a secure DNS update to one of the name servers in its home domain updating its current location. The name-to-address mappings for these hosts are un-cacheable by other domains, so stale bindings are eliminated.

Nevertheless, when a MN moves during a previously established connection, it may suspend the open connection and reactivate it from the new address, sending a special packet (Migrate SYN) to the correspondent node, which carries a token that identifies the previous connection. This SYN packet signals the correspondent node to re-synchronize the connection with the MN at the new point of attachment (new address). Thus, it is possible to provide mobility support as an end-to-end service, according to the application’s specific requirements, without changes in the network layer.

### 8. Application Layer Mobility

**Mobility using SIP** - The SIP (Session Initiation Protocol) is a signaling protocol, widely used for setting up and tearing down multimedia communication sessions over the Internet. It can be used in any application where session initiation is necessary.

The SIP registration mechanism is considered the application-layer equivalent of the MIP registration mechanism. However, while mobile IP binds a permanent IP address identifying a host to a temporary CoA, SIP binds a user-level identifier to a temporary IP address or host name (Schulzrinne & Wedlund, 2000). An INVITE message is sent by a MN to its CN to set up a communication session. The mechanism to provide MN mobility during an active session foresees that the MN needs to send another INVITE message to the CN to communicate the information about the new parameters of the communication session after the handover, using the same call identifier as the original call setup.

This solution has some drawbacks (Salsano et al., 2008). The second INVITE is sent end-to-end, and this could lead to high delays. Moreover, the handover procedure relies on the capability of the CN to handle this procedure, thus increasing MN processing needs. An auxiliary mechanism is necessary if the MN and CN move at the same time.
9. Mobility Plane Architecture

The Mobility Plane Architecture (MPA) (Zagari et al, 2008) is an instance of a reference architecture for micromobility support in IP networks (Prado et al, 2008). The goal of MPA is to speed up the handover process in order to minimize communication disruptions when the mobile node changes its network point of attachment. One of the requirements of this architecture is to place the burden demanded by micromobility on the network, not on the mobile nodes. Another requirement is to use, ideally, only well established network protocols. The key point of MPA is to employ an overlay network built above a transport network for directing traffic to the mobile nodes. This overlay network is composed of network elements called Mobility Aware Router (MAR), which are routers enhanced with MPA’s functionalities. MPA employs point-to-multipoint (P2MP) tunnels in order to encapsulate traffic directed to the mobile nodes and allows a gradual deployment once the architecture elements are installed only at the MARs.

MPA addresses the following issues related to mobility in IP networks: tunnel management (tunnel establishment, shutdown, and topology updating); secure mobile node attachment and handover; tracking of mobile node actual point of attachment (location); routing on the overlay network (decoupled from routing on the transport network); and quality and class of service (QoS/CoS) offered to the mobile nodes.

9.1 Functional Description

The architecture defines the following basic elements:

- Transport network - an IP network from which the network operator wishes to offer mobility services.
- Point-to-multipoint (P2MP) tunnel - a tunnel with a topology forming a tree. Nodes in the tree are MARs and arcs are tunnel segments connecting MARs. The tunnel has a single ingress (root) MAR, branch MARs (nodes with branching level greater than one), and egress MARs (leaves of the tree). A packet being forwarded through the tunnel may or may not be replicated at a branch MARs according to the policies enforced by these MARs.
- Access router - a router (usually an egress MAR) connected to a wireless access point.
- Access network - an IP subnetwork formed by the access routers and access points.
- Overlay mobile network - logical network built with one or more P2MP tunnels established through the transport network.

Figure 4 illustrates these basic elements. In addition to the basic elements, four functional blocks (FB) are defined:

- Tunnel Management (TM) Functional Block: TM is the entity responsible for P2MP tunnel establishment, shutdown, and re-routing. It must provide interfaces to the network management system and to the human operator. Tunnel management is a function carried out by MARs.
- Mobile Routing (MR) Functional Block: MR is the entity responsible for tracking the mobile nodes actual point of attachment and for interacting with the MARs forwarding engine in order to route traffic to the mobile nodes correct location. Mobile routing is a function carried out by MARs.
- Address Configuration (AC) Functional Block: AC is the entity responsible for supplying L3 addresses to the mobile nodes when they connects or reconnects to
the access network. Address configuration is a function carried out cooperatively by MARs and mobile nodes.

- Handover Helper (HH) Functional Block: HH is the entity responsible for facilitating the handover process with functions including L2 notification (triggering), L2 re-association, secure node attachment, and handover-related signaling. This function can be spread among MARs, network equipments (e.g., wireless switches), and mobile nodes.

![Fig. 4. MPA overview](image)

### 9.2 Operations Basics

When a packet targeted to a mobile node reaches an ingress MAR, it is tunneled until it reaches an egress MAR able to route the packet to the mobile node. When the mobile node performs a handover, the AC FB presented on the mobile node and on the network interact in order to provide the mobile node with a new L3 address. If the address is identical to the previous one the transport connections are not broken due the handover. The handover also triggers a mobile node location update on the MR FB. The location updating process updates the mobility routing tables on the MARs in such a way that when a packet is targeted to the mobile node the packet is routed to the egress MAR serving the link the mobile node is attached to.

Let us consider the mobile routing and the location update processes. Figure 5 shows a mobile node attached via access router M4. When the mobile node moves to a link served by M5, the entry related to this node on the mobile routing table at M2 must be updated with a different tunnel segment (in this case from segment C to D). If the mobile node roams to a link served by M7, the mobile routing table at M1 and M3 must be updated. Table
updates are performed as soon as the mobile routing protocol messages indicating the new point of attachment are processed by the branch MARs. The entries on the mobile routing table are soft state, meaning that the entries are dropped if location update messages confirming them cease. Soft state is a clear way to drop routes to mobile nodes when they no longer are reached through these routes. This scheme demands that a mobile node perform network attachments periodically in order to generate location update messages that will refresh the routes to it. A way to force the mobile nodes to perform periodic attachments is to provide them L3 address with a short lease time. When the lease time is close to expire, a mobile node performs address renewal that will trigger address location update messages in its behalf.

Fig. 5. MN Routing in MPA

9.3 MPA advantages
The MPA architecture presents the following advantages:

1. The solution is not limited to IPv6, being deployable on both IPv4 and IPv6 networks. Since it relies on tunneling, mixed deployments with IPv4 on the access network and IPv6 on the transport network, and vice-versa are possible.
2. The solution is not affected by middle boxes such as firewall and NAT boxes placed anywhere on the access, transport, or backbone networks.
3. The solution demands no complex protocols such as MIPv6 on the mobile nodes. Since it relies only on the standard IP protocol stack, the solution supports all the commercially available mobile nodes based on, for instance, Windows Mobile, Symbian, and PalmOne operating systems. The architecture does not forbid
enhancements deployed on the mobile nodes in order to improve handover speed and security, for instance. Such enhancements can be installed in user space or in the operating system kernel (e.g., as device drivers).

4. The solution preserves the L3 address of the mobile nodes when they roam inside the access network, causing no disruption of transport connections maintained by the mobile nodes.

5. The solution does not restrict the mobile node to employ macromobility protocols such as MIPv6.

6. The solution complies with security mechanisms related to L2 (e.g. WPA), L3 (e.g. IPSec), and L4+ (e.g., SSL, HTTPS).

7. The solution combines P2MP tunneling management, QoS/CoS management, and mobile node location tracking into the same protocol (RSVP-TE), reducing implementation and operating costs.

8. The solution does not interfere on services already deployed on the transport network such as VPN and VoIP.

9. Only well standardized protocols are employed by the architecture. When extensions to protocols are necessary they are introduced as opaque objects already foreseen by these protocols.

For more details on MPA description, implementation (using IPv4 and IPv6) and performance analysis, see Prado (2008), Zagari (2008), Johnson (2008), and Zagari (2009). Badan et al (2009) details the MPA implementation using MPLS.

### 10. General Analysis and Research Directions

Table 1 summarizes all the protocols and solutions seen in this chapter.

<table>
<thead>
<tr>
<th>Protocol</th>
<th>Mobility Range</th>
<th>Mobility Routing</th>
<th>Mobility Signaling</th>
<th>Mobility Layer</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIP</td>
<td>Micro</td>
<td>Routing</td>
<td>MN-centric</td>
<td>L3</td>
</tr>
<tr>
<td>DMA</td>
<td>Micro</td>
<td>Tunnel</td>
<td>MN-centric</td>
<td>L3</td>
</tr>
<tr>
<td>FMIP</td>
<td>Micro</td>
<td>Tunnel</td>
<td>MN-centric</td>
<td>L2</td>
</tr>
<tr>
<td>Hawaii</td>
<td>Micro</td>
<td>Routing/Tunnel</td>
<td>MN-centric</td>
<td>L3</td>
</tr>
<tr>
<td>HMIP</td>
<td>Micro</td>
<td>Tunnel</td>
<td>MN-centric</td>
<td>L3</td>
</tr>
<tr>
<td>H-MPLS</td>
<td>Micro</td>
<td>Tunnel</td>
<td>MN-centric</td>
<td>L2½</td>
</tr>
<tr>
<td>I-LIB</td>
<td>Micro</td>
<td>Tunnel</td>
<td>MN-centric</td>
<td>L2½</td>
</tr>
<tr>
<td>IP-IAPP</td>
<td>Macro/micro</td>
<td>Tunnel</td>
<td>MN-centric</td>
<td>L2</td>
</tr>
<tr>
<td>LEMA</td>
<td>Micro</td>
<td>Tunnel</td>
<td>MN-centric</td>
<td>L2½</td>
</tr>
<tr>
<td>MPA</td>
<td>Micro</td>
<td>Tunnel</td>
<td>Network-centric</td>
<td>L3</td>
</tr>
<tr>
<td>MIP</td>
<td>Macro</td>
<td>Routing/Tunnel</td>
<td>MN-centric</td>
<td>L3</td>
</tr>
<tr>
<td>MIP-RR</td>
<td>Micro</td>
<td>Tunnel</td>
<td>MN-centric</td>
<td>L3</td>
</tr>
<tr>
<td>MM-MPLS</td>
<td>Micro</td>
<td>Tunnel</td>
<td>MN-centric</td>
<td>L2½</td>
</tr>
<tr>
<td>Mobile MPLS</td>
<td>Macro</td>
<td>Tunnel</td>
<td>MN-centric</td>
<td>L2½</td>
</tr>
<tr>
<td>MSOCKS</td>
<td>Micro</td>
<td>Routing</td>
<td>MN-centric</td>
<td>L4</td>
</tr>
<tr>
<td>PMIP</td>
<td>Micro</td>
<td>Tunnel</td>
<td>Network-centric</td>
<td>L3</td>
</tr>
<tr>
<td>SIP</td>
<td>Macro</td>
<td>Routing</td>
<td>MN-centric</td>
<td>L5</td>
</tr>
<tr>
<td>TCP Migrate</td>
<td>Macro</td>
<td>Routing</td>
<td>MN-centric</td>
<td>L4</td>
</tr>
</tbody>
</table>

Table 1. General Classification
As this work is a non-exhaustive selection of mobility solutions, our comments are limited to these protocols revised in this chapter.

There are a majority of MN-centric solutions, since the beginning of mobility management research, with the Mobile IP. Newer solutions, such as MPA and PMIP, work as network-centric solutions, and it is our belief that this class of mobility signaling will have more research focus and implementations, because of its advantages to the final user: they do not need protocol installation or configuration on the MN, no overheads on devices to handle mobility, and no intervention from the final user to adopt a particular mobility solution.

Another interesting issue is the network layer protocol. About 10 years ago nobody believed in IPv4 for mobility, which is why both MIPv6 and its extensions appeared at that time. Since then, however, the IPv6 protocol has not been massively adopted, so there is reawakened interest in IPv4 (PMIP and MIPv4 raised again, for example). Now, 4G researchers and professionals say IPv6 is inevitable because each cell phone and mobile devices must have a fixed (stable) IP address. Solutions compatible only to IPv4 must foresee this IPv6 adoption and implement a compatible version of their protocol, if intended to be used in the future.

Not mentioned in this chapter, solutions for mobility must adopt security mechanisms for authentication, authorization and general security procedures. For example, for the MPA architecture, the access points can be configured to authenticate mobile nodes based on WPA2 employing Pre-Shared Keys (PSK) or RADIUS. PSK is easy to configure but is not as secure as RADIUS-based authentication. RADIUS authentication can be strengthened by using certificates installed on the mobile nodes. As RADIUS transactions take long time (500ms in our testbed network), RADIUS-based authentication increases considerably the handover overhead.

In order to speed up RADIUS-based authentication, a cache mechanism can be employed such as PMK (Pairwise Master Key) caching (also called proactive key caching). In this mechanism, once a mobile node completes successfully a RADIUS transaction, the access point stores the PMK supplied by the RADIUS server in the cache. When the mobile node connects to a new access point, the access point queries the cache (using the mobile node’s MAC address as a search key) in order to recover the PMK assigned to the node. If an entry is found, the access point accepts the mobile node without the need of a RADIUS transaction. In this case, the PMK found on cache is used to secure the communication between the mobile node and the access point.

As suggestions for future research directions in mobility management, we can point out the development of architectures able to:

- integrate macro and micro mobility into a single mobility solution;
- support both vertical and horizontal handover (e.g., between WiFi and 4G networks);
- support clean slate solution, by designing the network from the mobility requirements (and not to incorporate mobility extensions over the existing networks);
- These solutions decouple host identity from network address as suggested by HIP (Host Identity Protocol) and other related solutions;
- restrict handover within the L2, employing, for instance, tunneling over Ethernet (instead of over IP or MPLS), flat routing (based on MAC – Media Access Control – address), etc.
11. Conclusion

This chapter was intended to present a major review on mobility architectures and protocols. Many solutions were shown, ranging from link layer related solutions to application layer solutions, including network, transport and intermediary layer protocols. We also presented MPA, which is our solution for mobility management, using a network centric paradigm.

Some of the reviewed solutions were abandoned, some were investigated till today. But mobility management solution still needs investigation, to allow massive deployment and use. Although the new mobility architectures and protocols are permanently under investigation, all new solutions are constrained by factors such as: be deployable over existing IPv4 (and future IPv6) fixed networks, operate without upgrading with the mobile nodes already in the marked, comply with current network operation practices, be scalable without degrade quality of service, and allow the introduction of new services with stringent communication requirements such as media streaming (e.g., IP TV), location, and entertainment services. The challenge in mobility for IP networks is to comply with these factors that, unfortunately, are not restricted only to the technical issued commonly addressed by the network architects.

12. References


