II. Mobility Architectures: Background and State Of the Art

This section introduces the background on host mobility management for mobile multimedia services, reviews the principal architectural solutions proposed in the literature that address issues on host mobility management, describes a taxonomy that organizes these solutions based on their design principles, in particular the aspects that impact in the deployment in a real scenario and limit their applicability. Our solution also (ABPS/MAGELLANO) in introduced in this taxonomy, so as to provide a comparison among the different architectures.

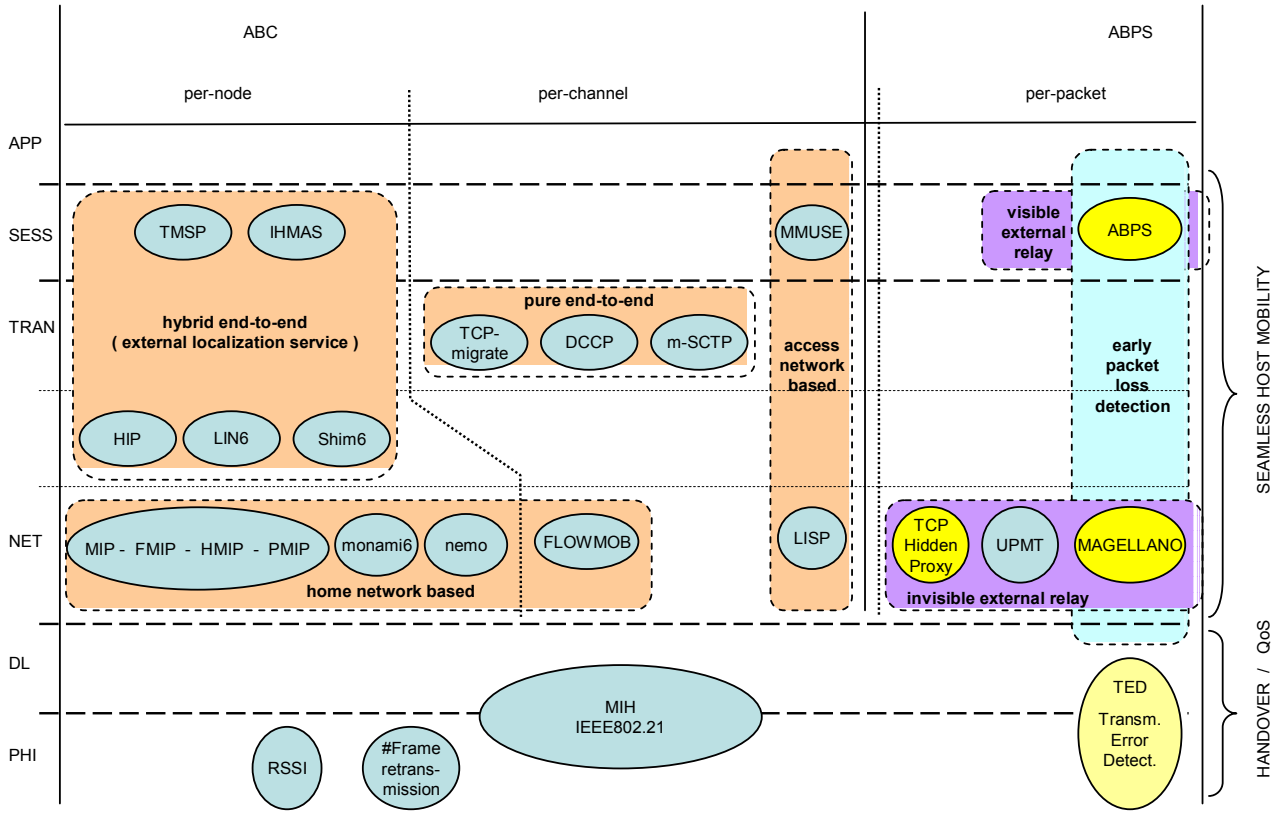

Figure 1: Architectures for mobility management.

II.1 Seamless Host Mobility Architectures

The aim of an host mobility architecture is to ensure that a MN can move across different access networks and seamlessly gain access to network services. Handover is the process of maintaining a user's active session when a mobile terminal changes its connection point to the access network (called "points of attachment"). Seamless handover aims in performs handover with no user perceivable interruptions. A horizontal handover takes place between points of attachment supporting the same datalink-layer network technology. On the other hand, vertical handover occurs between points of attachment supporting different network technologies. Various sophisticated approaches, aiming at providing seamless communications, exploit multihoming as a basic solution to increase reliability, i.e. they assume that a MN is equipped with multiple NICs (and corresponding IP addresses) that it can use. Ideally, a seamless mobility architecture is responsible for: i) identifying univocally each given MN; ii) allowing each MN to be reachable from its correspondent nodes (CNs, for the purposes of this discussion, a CN is either a fixed or mobile node communicating with the MN); iii) monitoring the QoS provided by different networks to predict the need of a handoff and to select a new preferred NIC and its connected network access point; iv) performing the handoff seamlessly, ensuring the continuity of the communications without use-perceivable interruptions.

In IP-based communications, the MN's IP address plays the twofold role of MN's identifier and MN locator, as it distinguishes uniquely the MN, and identifies its position in the network. When a MN joins a network, it is assigned an IP address that is valid within that network only. When that MN moves in a different network, it acquires a new IP address from the new network, loosing its identity. Hence it needs to inform CNs of its new identity and location. More generally, a mobility management architecture must ensure that two hosts can communicate even when they both move and change simultaneously their addresses.

To this aim, all the mobility management architectures adopt some mechanism that i) defines uniquely a MN identifier, independent from the location of the MN, and ii) provides a **localization service** that maintains a mapping between the MN identifier and the MN current location, even when the MN moves. The localization service is composed of a location registry, a service assumed to be always available. When a MN changes its IP address, it communicates with such registry, informing it of this change (**registration phase**). Once a CN wants to initiate a new communication with the MN, or when it wants to continue a communication with the MN that has changed its address, the CN starts a **lookup phase**, asking the location registry for the MN's current address, and then uses the obtained MN's address to contact directly the MN. This is the general principle adopted by different architectures, and may be developed inside different network entities and using very different algorithms and protocols. The architectural solutions examined below adopt this principle; however, they are implemented at different levels of the ISO-OSI reference model, as illustrated in Figure 1.

In addition, these architectures may be classified depending on where the localization service is deployed: **pure end-to-end solutions** distribute the localization service on both the end system involved in a communication; **home network -based solutions** deploy the service inside the network from which the mobile node belongs to; **border gateway -based solutions** instead deploy the service inside border gateway at the edge of each network domain that a given mobile node exploits as its access point to the Internet; **hybrid end-to-end solutions** manages the reconfiguration of the nodes on both the end system involved in a communication but place the localization service in a separate server; finally, **external relay solutions** integrate both the localization service and the packet relay service at separate servers independent from both home and access networks so as to be deployed with no impact on the network infrastructures and moreover to overcome the presence of firewall and NAT systems.

Another important classification relates to the granularity of the solution. Granularity defines the target to be assigned to a selected NIC. When triggered, **per-node** switchers migrate every active flow, by adopting a coarse-grained approach and exploiting one NIC only according to the requirements of the whole node. On the contrary, **per-channel** switchers can also migrate only one flow at a time, in a finer-grained manner, thus enabling each flow to exploit its most suitable NIC. Finally, **per-packet** switcher allows each single IP datagram be routed through the most suitable NIC allowing fine-grained load-balancing and recovery policies.

In the rest of this Section, we shall examine these solutions in isolation, according to the ISO-OSI reference layer to which they belong.

II.2 Solutions at the Network Layer

Among the architectures working at the network layer, it is worth citing the efforts in the protocol Mobile IP version 6 (MIPv6) [15] and its optimizations, e.g. the Fast Handover Mobile IPv6 (FMIP) [18], Hierarchical Mobile IPv6 (HMIP) [28] and Proxy Mobile IPv6 (PMIP) [12]. All these approaches employ an Home Agent, i.e. an additional entity working inside the access network to which the MN belongs. This Home Agent plays the role of the location registry mentioned above, and routes datagrams towards the MN when such node is outside its "home network". . In addition, if the mobile host wishes to register its binding with a correspondent node, so as to communicate directly with it without the interposition of the home agent, that mobile host must perform return routability operations as described in [RFC-3775]. Unfortunately, if the correspondent node in protected by a firewall or NAT system the return routability operation fails. In order to work properly, the MIPv6-based approaches require that all the end-systems have IPv6 capabilities so as to insert in the IP datagrams some extension headers that transport both the MN's identifier (the home address) and the current MN address. A clear limitation is that these architectures only work on infrastructures with IPv6 capabilities. Moreover, the MIPv6 specification does not allow the simultaneous use of the multiple MN's NIC. For each given MN, the address of a single NIC is registered at the Home Agent. In addition, as demonstrated in [20], the handover latency results very high due to the numerous authentication messages and this causes a service disruption time not compatible with strongly interactive services such as Internet Telephony. Finally, the MIPv6-based approaches do not overcome the presence of firewall and NAT systems that make unreachable the mobile host from outside the firewall. We will explain in more details the limitation imposed by firewalls and NAT systems in the following subsection II.6.

A latest extension of MIPv6, called multiple care of address registration (monami6) [draft-ietf-monami6-multiplecoa-14.txt] has been recently proposed that allows host mobility and multihoming. If a mobile node configures several IPv6 global addresses on one or more of its interfaces, it can register these addresses with its home agent as care-of addresses. This extension also does not overcome the presence of firewall and NAT systems. In addition, the return routability operations cannot be easily extended to verify multiple care-of-addresses and, as usual, if the correspondent node in protected by a firewall or NAT system the return routability operation fails.

Network mobility Basic Support Protocol (NEMO BSP or simply NEMO) is concerned with managing the mobility of an entire network. Network mobility aims providing seamless Internet connectivity of the whole mobile network that consists of mobile routers (MRs) and mobile network nodes (MNNs). The network moves around along with vehicles as a whole. The application scenario is public transportation, such as trains and buses [Per05]. NEMO BSP is based on MIPv6 with minimal extensions. Therefore, the handover mechanism of a mobile router (MR) is essentially the same as that of mobile node (MN) in MIPv6. In NEMO BSP, Mobile Router serves as a gateway; a permanent address called

home address (HoA) is obtained on the home link as an identifier of MR. When MR moves away, it acquires a care of address (CoA) from the access router (AR) in foreign network. MR sends binding update (BU) message to its home agent (HA) located in the home network, binding the CoA with the HoA. After the binding process, a bi-directional tunnel is established between MR and HA. Packets from corresponding node (CN) with the destination of MR's HoA are directly routed to HA, and HA are in charge of rerouting all packets to the CoA of MR through tunnel. Mobile network nodes (MNN) in the mobile network have permanent addresses taken from mobile network prefix (MNP) advertised on MR's ingress interface, and packets intended to or originated from the MNNs are encapsulated in the tunnel. Recently, various multihoming MR scheme have been proposed that allow multiple care-of-addresses registration, and in [Che10] an experimental evaluation has been conducted that proves that NEMO schemes shares the same strengths and weaknesses of the monami6 approach previously described.

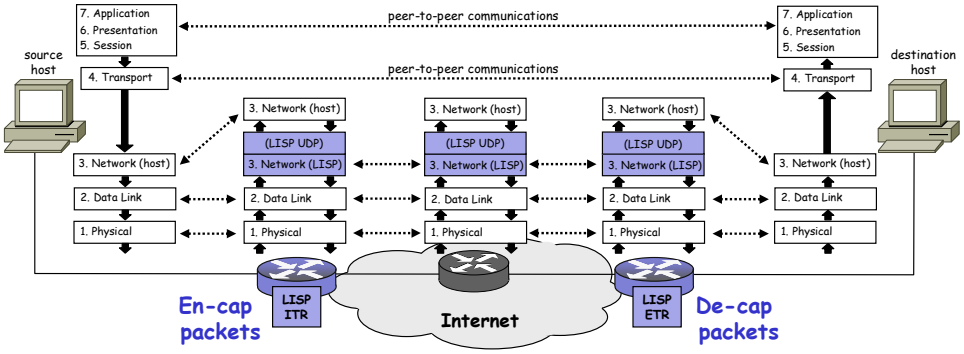

Figure XXXXXX: The LISP architecture.

Finally, another notable architecture is Location/ID Separation protocol (LISP) [Mey08], [Mey10], [Far11] proposed by Cisco, that belongs to the border-gateway class. As depicted in the previous figure XXXXXX, LISP makes use of an overlay network of LISP routers, located at the edge of the network domains and classified as Ingress Tunnel Router (ITR) and Egress Tunnel Router (ETR). The ITR intercept the IP datagram belonging from the mobile nodes, map the mobile node sender address (the location) to the sender ID, encapsulate these datagram in LISP packet and route them through the LISP routers towards the destination. When the LISP packet reaches the ETR at the edge of the destination domain, the ETR extracts the IP datagram from the LISP packet, maps the destination ID in the destination locator and route the IP datagram to the destination. In this way, LISP does not require changes on both the end systems. The main drawback of LISP is shared with all the other architectures that introduce functionalities at the edge of the network domains, they requires that all the domains be LISP enabled by deploying Ingress/Egress tunnel routers at their edges. In fact, these architectures require that all the paths between each mobile node and the Internet flows through at least a LISP-enabled border router. If a mobile node NIC connects to a network domain that have no LISP-enabled router at its edge, the LISP architecture fails in providing network continuity to the mobile node. In other words, LISP requires that all the network domains deploy LISP-enabled border routers.

II.3 Solution Between the Network and Transport Layers

Approaches have been presented in the literature that insert an intermediate layer between the network and transport layers of the protocol stack. This layer works on all the end-nodes involved in the communication, i.e., in a one-to-one communication as in VoIP, between the MN and its CN. Notable examples include the Host Identity Protocol (HIP) [23], Location Independent Addressing for IPv6 (LIN6) [31] and Level 3 Multihoming Shim Protocol for IPv6 (Shim6) [RFC5533]. Based on these solutions, the location registry is a DNS-like mapping function that operates as a service outside the access networks and associates host identifiers to host locations. In particular, using the Reachability Protocol (REAP) [REAP] in case of outage, Shim6 may perform a end-to-end exploration of the peer available addresses set and update the locators, but this approach is not suitable for highly dynamic environment as it is related to timer expirancy and not to movement detection.

A limitation of these approaches is the requirement to modify the protocol stack at all end-nodes involved in a communication. While it is reasonable to ask a MN to install additional software to properly work while moving, its CN can be a fixed node that may not be interested in supporting the mobility of the MN.

II.4 Solutions at the Transport Layer

The common approach of the protocols working at the transport-layer, such as the datagram-oriented Datagram Congestion Control Protocol (DCCP) [19], the stream-oriented Mobile Stream Control Transport Protocol (m-SCTP) [24] and the TCP enhancement TCP-migrate, is very different: each given end-system plays the role of proactive location registry that directly informs the CN whenever its configuration changes. Unfortunately, this approach fails

when both the end-systems change simultaneously their IP configuration, because the two end-systems become mutually unreachable. Such a situation is not unusual, because it happens when both the end-systems are mobile ones and leave simultaneously their current network access point.

As in the previous approaches, solutions working at the transport-layer require modifications of the applications on both the MN and CN to invoke the services of the novel transport layer and to implement suitable recovery policies. This fact prevents the reuse of the existing applications.

II.5 Solutions at the Session Layer

When solutions are devised to let nodes communicate through different networks, a key role might be played by session protocols that control the dialogue between end-points and incorporate functionalities used by the localization service. Today, the Session Initiation Protocol (SIP) is the main protocol employed for controlling multimedia, multi-homed communication sessions [25, 30]. Since our approach employs SIP, as well as other solutions outlined in the rest of this section, we review now its basic properties and functioning.

SIP is a session-layer text protocol that uses a message/response handshake for signaling purposes. In particular, it is used to establish or change communication parameters such as IP addresses, protocol ports and audio/video codecs between the end-systems. The SIP specification is extensible and allows application-defined fields to be added to the SIP messages.

The SIP messages worthy of mention, in view of the discussion that follows, are *REGISTER*, *INVITE* and *re-INVITE*. The *REGISTER* message allows a given node to declare that it is available for communications; it is usually sent to a SIP server that works as the localization service. The *INVITE* message is used to establish a communication session between two nodes. Typically this message is sent from the user to the SIP server that replies with the address of the other end-node, together with some communication parameters. Then, the two end-nodes can communicate directly. A *re-INVITE* message may be used when communication parameters (such as the IP address) change.

Another important aspect is that the SIP protocol allows the presence of SIP proxies that can be transparent to the application (proxy agent) or can masquerade the end systems (back-to-back (B2B) user agent) working as an opaque relay.

Several proposals exist that employ SIP to control the session of a (multimedia) multi-homed communication. For instance, the Terminal Mobility Support Protocol (TMSP) [21] exploits an auxiliary SIP server, located outside the access networks, as location registry that maps a user identifier (e.g. ghini@cs.unibo.it) to the current user's location (the IP address of the user's MN). Each MN has a SIP user agent that sends REGISTER messages to the SIP server in order to update its current location. INVITE messages are sent to establish communications with the other nodes.

Similarly, [33] presents an architecture able to manage vertical handoffs, by using a SIP-based approach. The scheme complies to the IP Multimedia Subsystem (IMS), a standardized overlay architecture for session control, authentication, authorization and accountability in all-IP networks [8]. Another related proposal is that presented in [16], that only supports vertical handoffs from 3G networks to a Wi-Fi.

The session-layer solutions seem to be not efficient as they invoke an external localization service when an IP reconfiguration occurs. In particular, the SIP-based services introduce an additional delay due to their message/response behavior; in case of reconfiguration the MN interrupts the communication, sends a SIP signaling message to the CN and waits for the response before resuming the transmission. With this in view, the IHMAS work presented in [6] provides a solution to minimize handoff delays by exploiting a SIP-based, IMS compliant proactive mechanism that performs registration and renegotiation phases for novel connections while keeping the media flows active over old connections, if these are available.

II.6 Coping with NAT and Firewall Systems

Most of the previously described mobility management solutions do not take into account the possible presence of firewall and NAT systems [METTERE RIFERIMENTO CHE CLASSIFICA FW e NAT] in between end-nodes involved in a (multimedia) communication. Usually, in order to overcome this limitation, at the communication setup the applications resort to services offered by external STUN [26] or TURN [27] servers, but these protocols do not work correctly when both the end-systems (MN and CN) are protected by the most common and restrictive symmetric firewalls. These most restrictive firewalls require an intermediate application-layer relay server, outside any firewall and NAT systems, that receives and forwards the packet exchanged between the MN and its CN for the entire duration of the communication. The relay rewrites the transport and network datagram header so as to appear as the peer node for both MN and CN.

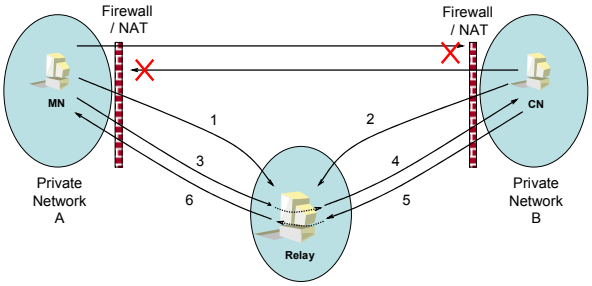

Figure 2: The data relay allows the undirected communication between two nodes behind different firewalls or NAT systems.

An example of system taking into account the possible presence of firewall and NAT systems is MMUSE [29]. Such system requires that an auxiliary SIP server (namely, the Session Border Controller, SBC) be located at the edge of the autonomous system inside which the MN will move. This autonomous system may be composed of several subnets using heterogeneous network technologies. While the MN moves across the subnets, each subnet provides the MN with a different IP address. The SBC aggregates the functionalities of SIP and RTP proxy, firewall and NAT systems, and intercepts the communications that enter and leave the network, in particular the SIP messages between the MN and its CN outside the network edge. Based on the outgoing SIP messages, the SBC sets up the firewall rules that allow the subsequent SIP, RTP and RTCP communications between MN and CN. Moreover, when the MN moves to a different subnet and changes its IP address, the SBC modifies the outgoing datagram in order to hide to the CN the current location of the MN.

The main limitation of the MMUSE solution is that the MN traffic needs to flow always through a given SBC that resides in the edge of the network. This implies that the MN may move only inside a given autonomous system, but it cannot move across different networks administered by different organizations.

II.7 Simultaneous Use of Different Network Interface Cards

It is worth to point out that, except that for monami6, all the previously described approaches do not allow the simultaneous utilization of all the NIC available on the MN. In other words, all the packets are transmitted using a NIC only.

The monami6 extension of MIPv6, namely Multiple Care of Address (MCoA), allows a MN to register its multiple IP addresses with its Home Agent. The Flow Mobility technique (FlowMob) described in [32] uses this extension to allow the MN to separate its outgoing traffic in different flows based on the protocols, port numbers and IP addresses, and forward each given flow using a selected NIC. This technique allows a flow-based switching through the MN's NICs, and in [V.E. Zafeiris, E.A. Giakoumakis, "Optimized traffic flow assignment in multi-homed, multi-radio mobile hosts", Elsevier Computer Networks, issues 55 (2011), pp. 1114–1131.] the problem of assigning traffic flows to available interfaces is optimized using heuristic algorithm. This approach shares the same limitations as the other MIPv6-based approaches: it needs that the network infrastructures be modified so as to add IPv6 capabilities and, moreover, the handover latency results very high owing to the large number of authentication messages. Finally, this approach does not provide solutions for symmetric firewall and thus requires external relay to be used, increasing the overall network latency between MN and CN.

II.8 Zero impact on Existing Network Infrastructures

To limit the need of network infrastructure upgrade, the **external relay solutions** locate the localization service at separate server independent from both home and access networks so as to be deployed with no impact on the network infrastructures. This external server incorporates also the functionality of data relay to overcome firewalls. This class of solutions requires support at the MN and splits the communications in two consecutives paths, from the MN to the relay and from the relay to the CN. It is important to point out that all the previously presented solutions requires an additional relay service to overcome firewall presence, thus this two-parts path does not represent a weaknesses with respect the other solutions. The external relay class of host mobility solutions may be separated in two sub-classes: the **visible** class of solutions supports only those applications and protocols (such as Web and VoIP) that define the concept of proxy, and operate as a pair of explicit proxies, one at the MN and another at the relay server. The application at the MN needs only to be configured so as to select the proxy at the localhost. On the other hand, the **invisible** class groups those solutions that operate as a tunnel between MN and relay without to be seen from the applications at the MN. Typically, this last sub-class is independent from the application but manages differently IP datagram that delivers different transport protocols. It is worth to point out that the application at the MN does not see the two proxies but on the contrary, from the standpoint of the CN, the proxy server seems to be the MN.

II.8.1 Visible Relay Service and Early Packet Loss Detection

The most enhanced representative of the visible relay class is our Always Best Packet Switching (ABPS-SIP/RTP) architecture, designed to support application based on SIP/RTP such as VoIP and VOD, that, in contrast with the previously presented approaches, enables the transmission of each datagram through the most appropriate NIC. Our ABPS-SIP/RTP architecture operates at the session-layer using a proxy client on the MN and an auxiliary proxy server in an external server, and allowing the MN to move across different autonomous systems. A cross-layer technique is employed to monitor all the concurrent NICs which are available, their performances and those that become active (inactive); based on their current status, automatic reconfiguration is performed in the MN.

Our solution uses SIP-compliant proxy servers which interact with additional software modules, installed on the proxies, working between the network and the transport layer; these modules are in charge of managing the application-layer data flows enabling their transmission through different NICs. In practice, our proxy decides on a per-packet basis which is the best NIC to use to transmit the data. Moreover, each proxy adds a digital signature in the packet so as to the proxy server may identify the sender in spite of changing of IP address. The use of such a signature technique avoids in a transparent way the typical delays introduced by the two way message/response handshake of the SIP signaling messages.

It is important to notice that all the approaches we have introduced earlier usually consider the problem of changing a NIC in use as soon as it becomes unavailable. Thus, the decision to change communication technology is not taken based on some particular QoS metric; rather, it is taken based on the failure of the currently adopted NIC. Vice versa, our approach takes into consideration QoS metrics to identify the best NIC to use at any given moment. In particular, a cross-layer mechanism (namely, Transmission Error Detector, TED) has been developed, that provides the applications with information about the successful (unsuccessful) datagram transmissions through a given wireless access point. At the moment, the implementation of TED is for WiFi NIC only.

Currently, to determine the "best" NIC to be used to transmit a given packet, the system adopts a simple "WiFi first" hybrid metric that takes into account both the energy consumption and the packet loss rate. In particular, at the MN, if all the available NICs are correctly configured and working, the proxy client selects the WiFi NIC, that consumes expensive than other (cellular) technologies, it sends the packet and waits for a few milliseconds to receive from the local TED a notification of the successful or unsuccessful packet transmission to the access point. In case of transmission error the packet is sent again using another (e.g. UMTS) NIC. The proxy server, instead, saves the sender IP address of the last-sent datagram received from the proxy client, and sends each packet directed to the proxy client exactly to that last saved address. If there is no traffic from the proxy client to the proxy server, the proxy client periodically sends to the proxy server an empty ABPS-SIP message in order to: i) keep open the path through the firewalls between proxy client and proxy server, and ii) implicitly give to the proxy server the IP address of the selected NIC.

The software module is extensible, in the sense that different metrics can be employed to decide which is the best NIC to use. Hence, the approach can be configured so as to take into consideration classic QoS criteria of network performance, such as costs, bandwidth, quality of the signal (i.e. Received Signal Strength Indicator - RSSI), indicators measured using the IEEE 802.21 Media Independent Handover [3] standard, or other solutions [17].

In essence, the advantages provided by our ABPS-SIP/RTP system are the following:

- our approach perfectly works over all IP based networks and does not require any modifications of the actual infrastructures;
- it is SIP compliant;
- it avoids delays occurring in classic SIP-based approaches when the MN changes its preferred NIC or its configuration. In this case, in fact, other schemes employ reconfiguration phases based on the exchange of INVITE messages, while our approach avoids this additional message exchange;
- it supports RTP-based applications such as VoIP and VOD services;
- it can cope with vertical handoffs, without introducing any additional delay during the passage from the use of a NIC to the next one, since as soon as this becomes available it is promptly configured to work;
- it optimizes the use of NICs, by deciding on a per-packet basis which is the best interface to be used, based on the monitored performances of the available networks and on the employed QoS metrics;
- it overcomes the presence of NATs and firewalls.

Obviously, the main weakness is that the ABPS-SIP/RTP architecture is strictly dedicated to SIP/RTP-based applications and cannot be exploited for other applications. The visible relay class of solutions aims in overcoming this weakness.

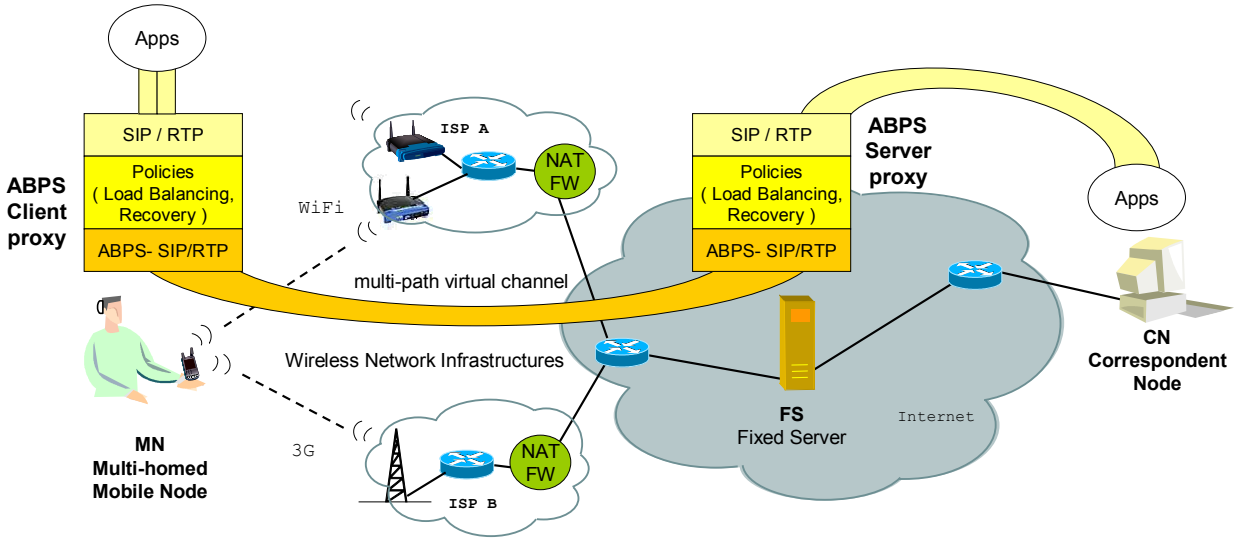

Figure 3: The ABPS architecture.

II.8.2 Invisible Relay Service

The Visible Relay Service solutions are not suitable for applications that are not designed to work with proxies or, more generally for legacy applications. To overcome this limitation, the Invisible Relay Service class operates transparently from the application standpoint and intercepts and redirects to a local proxy the messages sent by the applications on the MN. The local proxy delivers each message to the external relay proxy using all the available NICs of the MN. In the MN, a virtual Ethernet interface, (i.e. an Ethernet interface not linked to a physical one) and a particular routing table configuration causes the applications bind their outgoing connections to the virtual interface that is independent from the physical interfaces (and their faults). The earliest representative of this class has been the "hidden proxy" [Ghi06], developed on Linux systems and based on the iptables/netfilter mechanism and on the tun/tap virtual interface. This proxy was dedicated to TCP-based applications. More recently, the UPMT architecture [Bon09] provides support for UDP-based applications also. Both these architectures consider the problem of changing a NIC in use as soon as it becomes unavailable. Our proposal, Magellano, extends the hidden proxy so as to manage both TCP and UDP protocols but, with respect to the UPMT architecture, integrates the Transmission Error Detector of ABPS so as to recognize packets lost quickly and retransmit them through a different NICs.

II.9 Comparison among Host Mobility Architectures

The following Table HHHHH concentrates the discussion presented in this chapter providing a concise comparison among the different host mobility architectures we have presented. The comparison takes into account criteria that belong to three classes: requirements, deployment and performance.

The requirements class describes what protocols and protocol features are necessary for each given architecture.

The deployment class describes what kind of modifications need to be introduced in the network entities (e.g. infrastructures, protocols, terminals, applications) so as each given architecture may be deployed. These criteria are particularly important because they measure the applicability of a solution. In this way, the necessity of modifying a given entity avoids the simple deployment of the solution. The necessity of modifying the protocol stack or the applications in the MN is not a particular limitation because the MN takes advantage of the mobility and thus its user is very motivated in deploying the solution. On the contrary, modifying the CN to allow MN mobility is not convenient for the CN and thus this changing will be difficulty deployed. Analogous consideration refers to changing necessary in applications on the MN and on the CN; in addition, modifications on application are not possible for legacy applications. In this way, pure or hybrid end-to-end host mobility architectures should be difficultly deployed. Modifications on the network infrastructures are generally heavy but particularly when a solution needs to modify all the access networks or border gateways. Thus the deployment of home network –based and access network –based host mobility architectures pays a tremendous inertia. On the contrary, solutions that rely on an external relay server only are simply deployable because they do not impact on the existing infrastructures.

The performance class describes the quality of service provided by the architectures, in particular the performance criteria include both the ability of provide support to different applications (based on SIP/RTP, UDP, TCP or legacy applications) and the results provided in terms of service continuity, service unavailability, end-to-end latency, security, ability on exploiting simultaneously all the available NICs and switching granularity.

The unavailability interval measures the length of the interval time in which the MN cannot communicate with the CN due to a handover and thus it is very important for interactive applications such as VoIP. This unavailability interval

includes the time necessary to re-configure all the network entities involved in the end-to-end MN-to-CN communication. For this reason, all the solutions based on MIPv6 pay an high unavailability interval (1-3 seconds) because they require an intense packet exchange between MN, home network and CN to register the new care-of-address on the home-agent and to perform the return routability operation. However, if during the re-configuration of the preferred NIC the MN may exploit a different NIC, the unavailability interval decreases. For this reason, all the solution that allows multi-homing on a packet basis limits the unavailability interval length. In addition, it is very important notice that the unavailability interval includes the period immediately before the handover starts, period in which the NIC appears to be functioning but looses packets. If the MN detects these losses and may retransmit immediately the packet exploiting a different NIC the unavailability interval decreases. For this reason, the solutions that adopt the "early packet loss detector" such as ABPS and MAGELLANO reduce strongly the unavailability interval. For the sake of completeness, in the row 21 the symbol ∞ means that the architecture may fail in completing the handover: for the pure end-to-end solutions this happens when both MN and CN change their IP address contemporaneously; for the access network –based solutions this happen when the MN enters a access network without the border gateway able to support the host mobility.

The continuity interval measures the time during which the MN may communicate with the CN, in other words it measures the distance between two consecutive unavailability intervals. This metric reward the ability in switching, with no service interruption, from the NIC that performs the handover to another NIC already working, thus all the host mobility solutions that do not allow the simultaneous utilization of multiple NIC causes a number of service interruptions and thus a low level of continuity. Moreover, also the architectures that may fail the handover (see the discussion of the unavailability interval) provide not enough continuity. Obviously, the maximum continuity interval is provided by the solution based on the early packet loss detection because they may recover from packet losses by switching to and retransmitting through a different NIC.

The end-to-end latency measures the time necessary to deliver a packet from the MN application to the CN application and is very important for interactive applications. This minimum (low) latency occurs in a direct communication between MN and CN as it happens in case of pure and hybrid end-to-end host mobility solutions. The access network –based solution does not add delay because the packet transit through a entity (the border gateway) that is already in the path between MN and CN. In home network -based solutions, instead, the situation is more complex because if the return routability operation success the communication between MN and CN becomes practically direct (low latency) but if the operation fails the packets transit through the home network and adds latency depending on the distance between CN and home network and between home network and MN. The solutions based on an external relay obviously add a delay depending on the distance of the relay server. However, it is worth to point out that in the real scenario there are often firewall and NAT systems that avoid a direct communication and impose the interposition of a relay. Thus, in this common situation the latency caused by all the solutions increases dramatically, except that for the external relay –based solutions that utilize already a relay. The rows 23 and 24 compare the end-to-end latency in the best scenario (with no firewalls) and in the most common scenario (with symmetric firewalls), respectively, and highlight how in the most common scenario all the solutions are equivalent from the standpoint of the end-to-end latency.

	classes	Pure End-to-End			Hybrid End-to-End					Home network - based				Access network - based		Invisible external relay			Visible external relay
	Architectures / **Criteria**	**TCP-migrate**	**DCCP**	**m-SCTP**	**HIP**	**LIN6**	**Shim6**	**TMSP**	**HIMAS**	**MIPv6…**	**Monami6**	**Nemo**	**FlowMob**	**LISP**	**MMUSE**	**Hidden Proxy**	**UPMT**	**Magellano**	**ABPS**
Requirements	Requires IPv6					Y	Y			Y	Y	Y	Y	Y					
	Requires MIPv6									Y	Y	Y	Y						
	Requires IPSec									Y	Y	Y	Y				Y		
Deployment - needs modification in	Protocol stack in MN	Y	Y	Y	Y	Y	Y			Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	Protocol stack in CN	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y						
	Access Networks									Y	Y	Y							
	Border Gateways													Y	Y				
	Requires External Relay															Y	Y	Y	Y
	Application in MN		Y	Y				Y	Y										
	Application in CN		Y	Y				Y	Y										
Performance	Support to SIP/RTP				Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y		Y	Y	Y
	Support to TCP	Y			Y	Y	Y		Y	Y	Y	Y	Y	Y		Y	Y	Y	
	Support to UDP				Y	Y	Y		Y	Y	Y	Y	Y	Y			Y	Y	
	Support to Legacy App.								Y	Y	Y	Y		Y		Y	Y	Y	
	Sender Identification									Y	Y	Y		Y			Y	Y	Y
	Exploit simultaneously multiple NICs										Y	Y	Y	Y		Y	Y	Y	Y
	Granularity	Per-chan	Per-chan	Per-chan	Per-node	Per-node	Per-node	Per-node	Per-node	Per-node	Per-node	Per-node	Per-chan	Per-chan	Per-chan	Per-pkt	Per-pkt	Per-pkt	Per-pkt
	Overcome FW & NAT														Y	Y	Y	Y	Y
	Unavailability Interval due to handover	med ∞	med ∞	med ∞	med	med	med	high	high	high	high	high	high	med ∞	high ∞	med	med	low	low
	Continuity Interval	low	low	low	low	low	low	low	low	low	high	high	high	med	med	high	high	high	high
	E2E Latency (best case)	low (**)	low (**)	low (**)	low (**)	low (**)	low (**)	low (**)	low (**)	low-high (*) (**)	low-high (*) (**)	low-high (*) (**)	low (**)	med (**)	med-high (+)	med-high (+)	med-high (+)	med-high (+)	med-high (+)
	E2E Latency with symmetric Firewall (common case)	high	high	high	high	high	high	high	high	very high	very high	very high	very high	very high	med-high	med-high	med-high (+)	med-high (+)	med-high (+)

(*) the E2E latency becomes high if the return routability fails due to firewall presence
(**) the E2E latency becomes high when it is necessary a external relay to overcome firewall
(+) the E2E latency becomes high if the relay server is far from the MN

Table HHHHHH: Comparison among host mobility architectures

References

[1] Boing wireless, http://www.boingo.com/, 2008.
[2] FONERA, http://www.fon.com/, 2008.
[3] IEEE, “Media Independent Handover”, IEEE Draft Standard 802.21, 2008.
[4] Linux wireless extensions api. http://www.hpl.hp.com/personal/Jean_Tourrilhes/Linux/Tools.html. 2008.
[5] M. Baugher, D. McGrew, M. Naslund, E. Carrara, K. Norrman, “The Secure Real-time Transport Protocol (SRTP)”, RFC 3711, March 2004.
[6] Bellavista, P.; Corradi, A.; Foschini, L.; “IMS-Compliant management of vertical handoffs for mobile multimedia session continuity”, *Communications Magazine, IEEE*, vol.48, no.4, pp.114-121, April 2010.
[7] G. Bolch, S. Greiner, H. de Meer and K. Trivedi, Queueing Networks and Markov Chains: Modeling and Performance Evaluation with Computer Science Applications, Wiley, 1998, pp. 284-287.
[8] G. Camarillo, M.A. Garcia-Martin, M.A. Garcia-Martin, The 3G IP Multimedia Subsystem (IMS): Merging the Internet and the Cellular Worlds, 2nd Ed. Wiley 2006.
[9] L. Chen, “Recommendation for Key Derivation Using Pseudorandom Functions”, NIST Special Publication 800-108, Nov. 2004, available at http://csrc.nist.gov/publications/nistpubs/800-108/sp800-108.pdf
[10] V. Ghini, G. Lodi, F. Panzieri, “Always Best Packet Switching: the mobile VoIP case study”, Journal of Communications, Academy Publishers Ed., accepted for publication, May, 2009, also available at http://www.cs.unibo.it/~ghini/accepted/Ghini_JCM17.pdf
[11] V. Ghini, S. Ferretti, F. Panzieri, "Mobile Games Through the Nets: a Cross-Layer Architecture for Seamless Playing", in Proceedings of the International Workshop on DIstributed SImulation & Online gaming (DISIO 2010) - ICST Conference on Simulation Tools and Techniques (SIMUTools 2010), Torremolinos (Spain), ICST, March 2010.
[12] S. Gundavalli et al., “Proxy Mobile IPv6”, IETF Internet Draft, draft-ietf-netlmm-proxymip6-01.txt, June 2007.
[13] E. Gustafsson and A. Jonsson, “Always Best Connected", IEEE Comm. Mag., vol. 10, no. 1, Feb. 2003, pp. 49–55.
[14] S. Ferretti, V. Ghini, F. Panzieri, E. Turrini, “Seamless Support of Multimedia Distributed Applications Through a Cloud”, in Proceedings of the 3rd International Conference on Cloud Computing (IEEE Cloud 2010), Miami (USA), IEEE, July 2010.
[15] D. Johnson, C. Perkins, J. Arkko, “Mobility support in IPv6”, RFC 3775, June 2004.
[16] Kalmanek, C.; Murray, J.; Rice, C.; Gessel, B.; Kabre, R.; Moskal, A.; "A network-based architecture for seamless mobility services," Communications Magazine, IEEE , vol.44, no.6, pp.103-109, June 2006.
[17] S. Kashihara, K. Tsukamoto, “Service-oriented mobility management architecture for seamless handover in ubiquitous networks”, IEEE Wireless Communications, pp. 28-34, Apr. 2007.
[18] R. Koodli, “Fast Handover for Mobile IPv6”, IETF RFC 4068, July 2005.
[19] E. Kooler et al., “Datagram Congestion Control Protocol (DCCP)”, IETF RFC 4340, March 2006.
[20] K. Kong et al, “Mobility management for All-IP mobile networks: Mobile IPv6 vs. Proxy Mobile IPv6”, *IEEE Wireless Communications*, April 2008.
[Per05] E. Perera, V. Sivaraman, A. Seneviratne, Survey on network mobility support, Mobile Computing and Communications Review 8 (2) (2005).
[Che10] X. Chen, H. Zhang, Y. Chang, H. Chao, Experimentation and performance analysis of multi-interfaced mobile router scheme, Simulation Modelling Practice and Theory, Elsevier, 18 (2010) 407–415.
[Mey08] The Locator/Identifier Separation Protocol (LISP), D. Meyer, February. 2008, available at http://www.lisp4.net/documentation/extensive-lisp-overview/
[Mey10] D. Meyer, D. Lewis, D. Farinacci, LISP Mobile Node, Internet-Draft, draft-meyer-lisp-mn-04.txt, October 2010, available at http://tools.ietf.org/html/draft-meyer-lisp-mn-04#page-11
[Far11] D. Farinacci, V. Fuller, D. Meyer, D. Lewis, "Locator/ID Separation Protocol (LISP)", Internet-Draft, draft-ietf-lisp-11, March 2011, available at http://tools.ietf.org/html/draft-ietf-lisp-11.
[21] T.M. Lim, Chai Kiat Yeo, Francis Bu Sung Lee, Quang Vinh Le, "TMSP: Terminal Mobility Support Protocol," IEEE Transactions on Mobile Computing, vol. 8, no. 6, pp. 849-863, June 2009.
[22 M. Marzolla, The qnetworks Toolbox: A Software Package for Queueing Networks Analysis, Proceedings Analytical and Stochastic Modeling Techniques and Applications, 17th International Conference (ASMTA 2010), Cardiff, UK, June 14—16, 2010, volume 6148 of Lecture Notes in Computer Science, Springer, pp. 102-116.
[23] R. Moskowitz, P. Nikander, “Host Identity Protocol (HIP) Architecture,” IETF RFC 4423, May 2006.
[24] M. Riegel, M. Tuexen, “Mobile SCTP,” IETF Internet draft, draft-riegel-tuexen-mobile-sctp-07.txt, Oct. 2006.
[25] J. Rosemberg et al., “SIP: session initiation protocol”, IETF RFC 3261, June 2002.
[26] J. Rosenberg, J. Weinberger, C. Huitema, R. Mahy, “STUN - Simple Traversal of User Datagram Protocol (UDP) Through Network Address Translators (NATs)”, RFC 3489, March 2003.
[27] J. Rosenberg, R. Mahy, C. Huitema, “Traversal Using Relay NAT (TURN)”, Internet-Draft, draft-rosenberg-midcom-turn-08, September 2005.
[28] H. Soliman et al., “Hierachical Mobile IPv6 Mobility Management (HMIPv6)”, IETF RFC 4140, Aug. 2005.
[29] S. Salsano et al., “SIP-based mobility management in next generation networks”, IEEE Wireless Communications, pp. 92-99, April 2008.
[30] H. Schulzrinne, E. Wedlund. 2000. Application-layer mobility using SIP. SIGMOBILE Mob. Comput. Commun. Rev. 4, 3 (July 2000), 47-57.
[31] F. Teraoka, “LIN6: A Solution to Multihoming and Mobility in IPv6”, IETF Internet Draft, draft-teraoka-multi6-lin6-00.txt, 2006.
[RFC5533] E. Nordmark, M. Bagnulo, RFC 5533 "Shim6: Level 3 Multihoming Shim Protocol for IPv6", RFC, June 2009.
[REAP] [REAP] J. Arkko, I. van Beijnum, "Failure Detection and Locator Pair Exploration Protocol for IPv6 Multihoming; IETF

draft, draft-ietf-shim6-failure-detection-06, September 2006.
[32] U. Toseef et al., “Realization of multiple access interface management and flow mobility in IPv6”, ACM Mobilware, Innsbruck (Aut), Feb. 08.
[33] Udugama, A.; Kuladinithi, K.; Gorg, C.; Pittmann, F.; Tionardi, L.; "NetCAPE: Enabling Seamless IMS Service Delivery across Heterogeneous Mobile Networks," Communications Magazine, IEEE , vol.45, no.7, pp.84-91, July 2007.
[Ghi06] V. Ghini, S. Cacciaguerra, F. Panzieri, P. Salomoni, ``An hidden proxy for seamless & ABC multimedia mobile blogging". Proc. of IEEE Consumer Communications & Networking (CCNC 2006), 2nd IEEE International Workshop on Networking Issues in Multimedia Entertainment (NIME2006), Las Vegas, NV (USA), January 2006.
[Bon09] M. Bonola, S. Salsano, A. Polidoro, “UPMT: Universal Per-application Mobility management using Tunnels”, IEEE GLOBECOM 2009, Dec 2009, Honolulu, Hawaii.
[34] P. Zimmermann, A. Johnston, J. Callas, "ZRTP: Media Path Key Agreement for Secure RTP", Internet-Draft, draft-zimmermann-avt-zrtp-15, March 2009.