IP mobility.
Macromobility, micromobility, quality of service and security.*

Authors: Josep Mangues-Bafalluy[1], Albert Cabellos-Aparicio[2], René Serral-Gracià[2], Jordi Domingo-Pascual[2], Antonio Gómez-Skarmeta[3], Tomás P. de Miguel[4], Marcelo Bagnulo[5], Alberto García-Martínez[5]

[1] Telecommunications Technological Center of Catalonia (CTTC), [2] Technical University of Catalonia (UPC), [3] University of Murcia (UMU), [4] Technical University of Madrid (UPM), [5] University Carlos III of Madrid (UC3M)

Abstract. The current trend towards offering seamless connectivity no matter the place, time, application in use, or access technology served to coin the expression *Always Best Connected (ABC)* for describing such a framework. A key issue in accomplishing this goal is the provisioning of mobility to users and/or terminals. This paper overviews some of the solutions for offering mobility at the network layer as well as other relevant issues to solve towards this goal, namely quality of service and security. Both of them pose challenging research problems due to the variability of conditions found in mobile environments and their increased security threats.

Keywords: IP Mobility, QoS, Security, AAA

1. Introduction

The current trend towards offering seamless connectivity no matter the place, time, application in use, or access technology served to coin the expression *Always Best Connected (ABC)* [GuJo03]for describing such a framework allowing a user to choose the best available access network and device at any point in time. The definition of *best* depends on multiple parameters, like personal preferences, size and capabilities of device, applications requirements, available network resources, security. A key issue in accomplishing the ABC goal is the provisioning of mobility to users and/or terminals. However, mobility may be understood in different ways, and solutions to offer mobility at the subnetwork, network, transport, and application layers have appeared in the literature. Often, mobility is differentiated from portability. Whilst in the former the connection is not lost when changing the point of attachment to the network, this is not the case for the latter, which just guarantees that communications may be established, but not necessarily using always the same address, and thus, not maintaining the ongoing communications. In this paper, we mainly focus on mobility offered at the network layer to the terminal, which is the type of mobility that has received most attention from the research community, particularly with the development of MobileIP in the IETF.

Mobility support has been designed for IPv4, but also for IPv6. However, from the design point of view, the situation of both protocols is not the same. As the core of IPv4 was developed well before mobility scenarios were conceived,

* This research work has been partially funded by the Spanish Ministry of Science and Technology and European Funds for Regional Development (FEDER) under contract TIC2002-04531-C04 (Project: Advanced Mobile Services (SAM)).

mobility mechanisms were incorporated as extensions to the protocol. As a consequence, some of these extensions, though presenting technical advantages, were difficult to deploy in the wide scale. On the other hand, mobility has been considered from the very beginning in the design of IPv6. The same applies for other features that might be of interest in the path towards the ABC scenario, like quality of service (QoS) and security.

With respect to QoS, the same set of services available to the fixed user should be offered to the mobile user. But, as mobility is usually associated to wireless links, their variability makes accomplishing such goal very difficult. Besides, as the mobile terminal moves, it is expected to change its point of attachment to the network, a potentially disruptive process.

Security is also a challenging field of research in a mobility framework, as potential threads increase due to mobility often being associated to wireless media, but also due to mobility schemes requiring interactions between nodes that in a fixed Internet are often considered unusual. Furthermore, in an ABC environment, there is the need for a coordinated infrastructure for Authentication, Authorization, and Accounting (AAA) due to the variety of access technologies and potential users with diverse requirements.

In this paper, all the above issues will be dealt with by briefly explaining the main operational issues and reviewing some of the options found in the literature for providing macromobility and micromobility (section 2), QoS (section 3), and security (section 4) in mobile environments.

2. Mobility

Mobility management architectures are divided into two main parts, location management and handoff management. The former entails registering changes in the position of the mobile node (MN) and also the localization of an idle MN when an outside client wants to contact it. The other important point is handoff management, which tries to maintain all the connections of the MN alive despite the frequent changes of its point of attachment to the network. The process by which such change takes place is called handoff, during which communication may be interrupted and delay increased. Depending on the type of handoff, the process is more complex, as it may entail changes in the access point, the access router, the access gateway, the access technology, and/or the administrative domain.

From the network point of view, mobility management is seen from two different perspectives. On the one hand, there is the mobility inside a single administrative domain confined to a localized geographical region, which is called micromobility, and on the other hand, macromobility deals with mobility across larger regions, which often comprise various networks, with potentially different access technologies, which may belong to different administrative domains. Micromobility protocols try to solve the overhead, packet loss, and path reestablishment latency experienced by

macromobility protocols during handoff. In general, the solutions adopted confine control message exchanges to a reduced area and set up mobility agents representing that area and allowing interoperability with macromobility schemes. The final goal of both solutions is to offer the user a reliable network capable of keeping alive the connections all the time, independently of the actual position of the node, inside a single domain (micromobility) or even inside the whole Internet (macromobility). The following subsections give a brief overview of some of the solutions found in the literature.

2.1 Macromobility (Mobile IP)

Mobile IP is a network layer protocol conceived to provide macromobility to mobile terminals. Mobile IP is being designed by the Internet Engineering Task Force (IETF) in two versions. After various improvements, the latest Mobile IPv4 proposed standard is described in RFC 3344 [PeAl02]. Mobile IPv6, though, is still an IETF draft [JoPe03]. The objective of both protocols is to allow users moving in large areas to maintain their network connections while changing their point of attachment to the network.

Mobile IPv4 Overview

Mobile IPv4 introduces four functional entities:

- Mobile Node (MN): A mobile device.
- Home Agent (HA): A router of the home network that manages localization of the MN.
- Foreign Agent (FA): A router of the foreign network that cooperates with the HA to provide mobility.
- Correspondent Node (CN): A fixed or mobile node, with which the MN communicates.

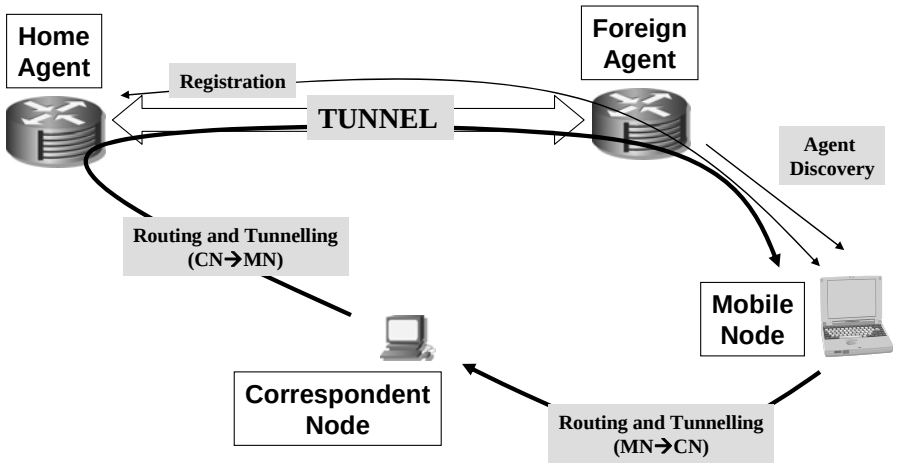

Figure 1. Mobile IPv4 overview

The protocol establishes four phases (figure 1). In the first one (*Agent Discovery*), the MN has to be able to detect if it is attached to the home network or to a foreign network. For this purpose, HA and FA send periodically Agent Advertisements (an ICMP Router Discovery extension). When a MN receives this message, it determines in which

network it is attached, and if it is on a foreign network, it obtains a Care-of-Address (CoA). The CoA is the IP address temporarily assigned to the MN while in the foreign network. The MN can also request an Agent Advertisement sending an Agent Solicitation to accelerate the process.

During the *Registration* phase, the MN registers its CoA in the HA. The MN sends a Registration Request to the FA, which forwards it to the HA. The HA replies with a Registration Reply to accept the request. At this point, the HA knows the localization of the MN and the communication with CN can be initiated, or continued in case of handoff.

In the third phase, called *Routing and Tunnelling*, the CN communicates with the MN (and vice versa). When a CN sends an IP packet to a MN, the destination address is the home address of the MN, i.e. the address assigned to this node when it was in the home network. When this packet arrives at the home network, it is intercepted by the HA. The packet is encapsulated and forwarded to the FA, which decapsulates and delivers it to the MN. On the other hand, when the MN sends a packet to a CN, it is directly sent using the home address as source. This asymmetric routing, which often is not the optimal, is known as triangle routing (see figure 1). This generates a series of inefficiencies such as longer packet delivery delays or increased load in the network. Though there are optimizations to solve these problems (route optimization), they require the modification of the CN, which may be any host in the Internet, and thus, their wide deployment is difficult.

In the fourth phase, known as *Handoff Management*, the MN moves from a subnet to another one by changing its point of attachment. The MN must obtain a new CoA, register it in the HA, and, once accepted, the MN is able again to communicate with CN. During the *Handoff Management* process the HA is not able to localize the MN, thus, some packets may be lost between the CN and the MN.

Mobile IPv6 Overview

Mobile IPv6 is very similar to Mobile IPv4. However, unlike in IPv4, in which mobility issues were not considered in its initial design, when IPv6 was developed, mobility was taken into account, and it is perfectly integrated into the protocol. Mobile IPv6 is more efficient and avoids some problems suffered by Mobile IPv4. Among others, Mobile IPv6 (figure 2) does not need FAs because IPv6 address autoconfiguration provides the required functions for the *Agent Advertisement* phase. During *Registration* and *Routing and Tunnelling*, packets are directly sent from the HA to the CoA of the MN.

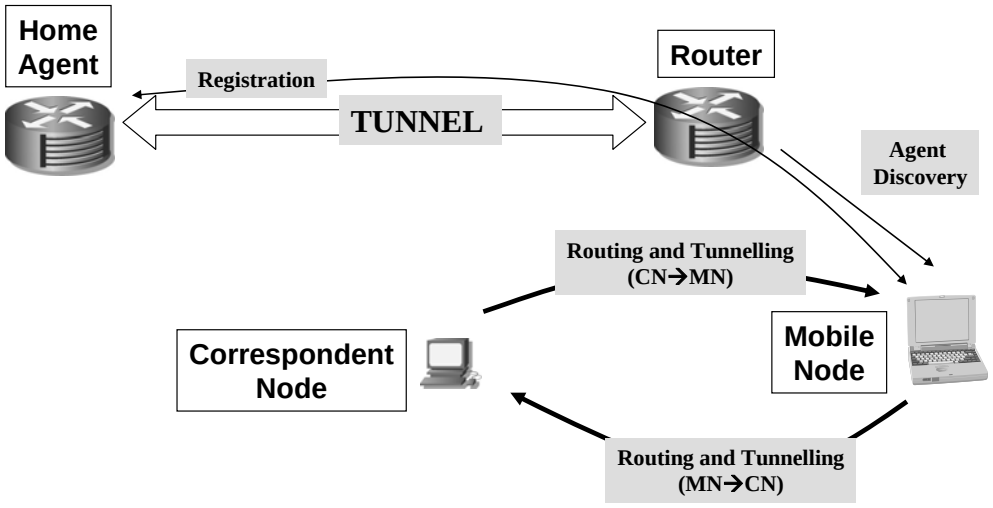

Figure 2. Mobile IPv6 overview

Mobile IPv6 also avoids triangle routing because when a CN sends a packet to the home address of a MN, the HA intercepts, encapsulates, and forwards the packet to the MN. However, the MN can also directly send a Binding Update (BU) to the CN. This message includes the CoA of the MN, and it is cached on the CN Binding Cache. At this point, any CN sending a packet first checks its Binding Cache for the IP destination address of the packet. If there is an entry, it will directly send the packet to the MN using the MN's registered CoA. This feature is inherent to IPv6, and no additional modification needs to be done to CNs to make them mobile-aware.

2.2 Micromobility

There are many environments where applications running in mobile nodes may become unusable if they change frequently their point of attachment to the network. For example, many real-time applications, like voice-over-IP, experience noticeable degradation of service if handoff is frequent. This problem is especially relevant when very large volumes of wireless subscribers should be supported.

The basic mobile IP protocol based on tunnelling mechanism introduces network overhead in terms of increasing delay, packet loss and signalling. The establishment of new tunnels can introduce additional delays in the handoff process, causing packet loss and delayed delivery of data to applications. This delay is inherent in the round-trip introduced by Mobile IP as the registration request is sent to the home agent and the response sent back to the mobile node (or sometimes to the foreign agent).

Micromobility protocols [CaGo02] aim to handle movement within a domain of MNs, with minimum or zero packet loss, minimum signalling, reduced power consumption and by just interacting with Mobile IP in the access network gateway (ANG), i.e. the node through which the domain connects to the Internet. This has the benefit of reducing delay

and packet loss during handoff and eliminating registration between MNs and possibly distant home agents when MNs remain inside their local coverage areas. All IP micromobility protocols share the same operational principles related to fast handoff, e.g. reduced location updates, fast security or even the quality of service.

Support for fast handoff is an important characteristic of micromobility protocols. Handoff is influenced by handoff management, buffering and forwarding techniques, radio behaviour, movement detection and prediction, and coupling and synchronization between the IP and radio layers.

Typically, fixed hosts connected to the Internet remain on-line for extended periods of time, even though most of the time they do not communicate. Mobile subscribers expect a similar service. MNs maintaining location information for being continuously reachable require frequent location updates, which would consume precious bandwidth and battery power resources. This signalling overhead and MN power consumption can be reduced by means of *paging*. Idle MNs do not have to register if they move within the same paging area, which is composed by all access points that share the same ANG. Rather, they only register if they change of paging area.

Networking functions like security or billing invoked during handoff, should be designed to help fast operation. While authenticating location update messages seems necessary in most cases, data encryption over the air interface or in the fixed network may not always be needed. User authentication for authorization or accounting may be required in some cases, while anonymous free access is sufficient in others.

Micromobility protocols try to guarantee the arrival of packets and to reduce signalling by hiding local migrations from home agents. Hierarchical mobility protocols do it by registering in the HA the address of the ANG instead of the CoA assigned to the MN in the visited domain. In this way, when a MN moves from one access point to another one (which is reachable through the same gateway) the HA needs not be informed. The role of micromobility protocols is to ensure that packets arriving at the ANG are forwarded to the appropriate access point. In order to route packets to the MN's actual point of attachment, protocols maintain a location database that maps host identifiers to location information. There are two styles of micromobility: hierarchical tunnelling and mobile-specific routing.

In hierarchical tunnelling, the location database is maintained in a distributed way by a set of mobility agents. Each agent reads the incoming packet's original destination address and searches its list of visitors for a corresponding entry. The entry contains the address of the next lower level agent. Entries are created and maintained by registration messages transmitted by MNs. Some proposals rely on a tree-like structure of mobility agents. However, in the most recent version of HMIP (Hierarchical Mobile IP), one of the main hierarchical tunnelling proposals, mobility agents directly interact with MNs without the need for establishing a tree-like structure [SoCa03].

Mobile-specific routing approaches avoid the overhead introduced by decapsulation and reencapsulation schemes of tunnelling approaches. These schemes typically introduce implicit or explicit signalling to update host-specific routes. In the case of Cellular IP [CaGo00] MNs attached to an access network use the IP address of the gateway as their Mobile IP care-of address. The gateway decapsulates packets and forwards them towards the access point. Inside the access network, MNs are identified by their home address and data packets are directly routed without tunneling or address conversion. The routing protocol ensures that packets are delivered to the MN's actual location.

3. Quality of service

Mobility is often associated to wireless links. These links present characteristics (e.g. fading or interferences) that may substantially vary depending on the surrounding environment, thus affecting the communication. Furthermore, mobility of nodes is implicitly associated with lightweight nodes that may have diverse processing, user interface, or power consumption characteristics. It also implies the potential for handoffs as nodes move. In this environment, the goal of handoff management schemes with QoS is to solve both, the routing issues for the correct delivery of the packet through the new path and the transport issues related with the reestablishment of the QoS state along this path. In such a varying environment, there might be some minor effects hidden to the user by means of application layer adaptation. However, there might be other severe variations that require the intervention of the network, particularly in the presence of handoffs. This section is mainly concerned with the problems that appear and possible solutions to handoff management with QoS.

The goal of mobility architectures that take into account QoS is to try not just to keep the communication alive, but to maintain the requested QoS for the MN, even in the case a handoff occurs. RFC 3583 [ChAl03] states the main requirements imposed over a solution to provide QoS for mobile IP, namely minimization of interruption of QoS during handoff, reestablishment of the affected parts of the QoS path, releasing the QoS state along the old path, interoperability with mobility protocols, support for heterogeneous QoS paths (due to different QoS provisioning philosophies), QoS support along multiple packet paths, and interaction with wireless link-layer support for QoS.

As explained above, the depth of the handoff, i.e. the magnitude of the associated changes (e.g. change of access point and/or technology and/or domain) involved determines the complexity in providing QoS. That is, if only the access point is changed while remaining in the same subnet, the handoff is simpler than in case that also the subnet and/or domain are changed (greater depth). Therefore, the QoS state reestablishment latency is likely to increase with handoff depth. Macromobility protocols have also been improved since their initial conception to provide better packet handling, particularly for Mobile IPv6. However, these solutions, even when used in conjunction with QoS signaling

(e.g. RSVP), do not scale for large mobility environments due to the signaling overhead and the latency in state reestablishment. Micromobility solutions confine mobility management to localized areas, thus providing shorter latencies and less overall overhead in the network. But these solutions only apply within an administrative domain. There is still the lack of a global integrated solution [MaLo 02].

At a coarse level of QoS provisioning, appropriate for diffserv-like operation, some kind of statistical admission control might be carried out at the network edge. This could be the task for ANGs and Access Routers (ARs), the former for the traffic destined to MNs of the domain and the latter for that generated by the MNs. ANGs and ARs upon request of a communication might forward the request to a global QoS broker in charge of managing the resources of a given domain. In turn, for finer-grained QoS offerings inside a domain, a close coupling between micromobility and flow-based QoS solutions (Intserv) might be in order. This coupling might vary in intensity, and could range from using handoff events for triggering QoS reservation messages to jointly designing and integrating micromobility protocols and QoS reservation solutions. In this case, QoS objects could be carried inside registration messages, thus establishing QoS state in the network at the same time that the new path is being established after a handoff. Modifications in network nodes might also allow to confine messages to track changes in the QoS reservation of the path to a small area, thus avoiding the need for end-to-end reestablishment of the reservation. This would minimize QoS state reestablishment latency and signaling overhead at the expense of added complexity in the network and dependency of the QoS solution on the micromobility protocol in use.

Aside from enhancing the latency and overhead of the QoS architecture, the provision of QoS guarantees to a given session running in a moving node requires mechanisms for both admission control priority and advanced reservations in all the cells that might be visited during the session lifetime. Admission control would be in charge of giving a higher priority to connections entering a new cell after handoff over new connection requests, the basic idea being the reservation of resources in neighboring cells in anticipation of potential handoffs. Advanced reservation mechanisms would be in charge of explicitly signaling the QoS needs to the cells that might be visited by the MN during the session. Examples of such mechanisms are Mobile Resource Reservation Protocol (MRSVP), which follows an Intserv approach, and ITSUMO, which follows a diffserv approach [MaLo 02].

Alternatively, mechanisms for carrying out pre-handoff negotiations, like the context transfer protocol being developed by the IETF Seamoby working group, might help in determining which neighboring cell is capable of offering the needed QoS and transferring the QoS state information so that when handoff eventually occurs, everything is in place to offer the requested QoS to the MN without having to start a new reservation request [KeAl02]. In this way, the potential waste of resources due to advanced reservations might be minimized at the risk of higher connection rejection.

4.Security

In order to preserve Internet's security, mobility support protocols must provide the same level of security available in the fixed Internet. However, the complexity is increased because of the implications that the mobility of nodes carries. Some issues to consider are: implications of the visiting node over the foreign network, implications over the home network when the node is abroad, and security implications to the mobile node itself when visiting a foreign network [MiPä00]. These issues are dealt with by means of the mechanisms explained in this section.

Security in Mobile IP

To accomplish the security goal, when a node receives a message binding a Home Address with a Care-of Address, it must verify that both addresses belong to the same node. In MIP4 [PeAl02] only the Home Agent processes such messages. Since it is reasonable to assume that a trust relationship exists between the HA and the MN, binding messages are protected using a pre-established security association between them.

In MIP6 [JoPe03], both the HA and the CN have to process binding messages, called Binding Update (BU) messages. BU messages sent from the MN to the HA are protected using a pre-established security association and IPSec, similar to the MIP4 case.

BU messages sent from the MN to the CN cannot be protected with such mechanism, since it does not seem reasonable to assume a trust relationship between the MN and all the potential CN of the Internet. An alternative method, called Return Routability (RR), is then used to acquire authorization information for the BU messages. The RR procedure verifies that the same node is reachable through the home address and the CoA. During the RR procedure, the MN requests two keys from the CN: one key is sent to the home address and the other key is sent to the CoA. Then, the MN generates the BU authorization information by hashing both keys and some additional information. Since both keys are used to generate the authorization information, the node generating the BU message has to be able to receive packets sent both to the home address and to the CoA.

When the CN receives the BU message, it first verifies the authorization information and if the verification succeeds, it processes the BU message.

Through the security mechanisms detailed above, mobility support protocols provide mobile communications with the same level of security available for fixed Internet communications. For further information about mobile IP security, the reader is referred to [NiAr03].

Integrating Mobile IPv6 and AAA Infrastructure

Equally important in the security framework to support mobility is the integration with an Authentication, Authorization, and Accounting (AAA) infrastructure. Mobile IPv6, like MIPv4, does not consider multi-domain network environments, understanding domain as a logical entity which has its own rules and policies and which could have business agreements with other domains. Besides, to allow a node to move doing roaming between different domains, service level and business agreements are needed between operators. Some of these issues need to be addressed by a complementary infrastructure for AAA [DIAM]. This infrastructure allows to authenticate end-users, processes and devices (the act of verifying the identity of an entity), to authorize them (the act of determining whether a requesting entity will be allowed to access to a resource, i.e. the own network is considered as a resource), and finally to monitor end-users operations over the network (e.g. for charging purposes). Therefore, integrating both elements, that is, making Mobile IPv6 a AAA services-aware protocol, will enable the roaming of mobile users in multi-domain scenarios.

An important issue in this context is the protocol to be used to carry AAA information between the MN and the equipment, which is named attendant, and that also takes part in the deployed AAA infrastructure (in fact, it is in charge of receiving the access request of a Mobile IPv6 user and forwarding it to the back-end AAA deployed infrastructure using the Diameter protocol). One of the most interesting proposals is the protocol defined by the IETF PANA Working Group [PANA]. PANA stands for Protocol for carrying Authentication for Network Access and its goal is to allow clients to authenticate themselves to the access network using IP protocols. Such a protocol allows a client to interact with a site's back-end AAA infrastructure to gain access without needing to understand the particular AAA infrastructure protocols that are in use at the site.

On the other hand, one of the objectives of authentication and authorization process is the establishment of a security association (SA) between the mobile node (MN) and service equipment (SE or attendant). It is supposed that this entity has a pre-established trust relationship with the AAA infrastructure. In order to get this MN-SE security association, a key distribution scheme(i.e. [LeFa02]) between the mobile node and the service equipment is needed. The idea more widespread is that end-user’s home AAA server is in charge of distributing these keys [FaLe02].

Finally, Mobile IPv6 (MIPv6) protocol itself can benefit from integration with AAA. MIPv6 needs to authenticate some of its management packets (binding updates, binding acknowledgments) [JoPe03] in order to avoid security problems. So, the issue is to make use of a trustworthy infrastructure as AAA infrastructure, to make more reliable the authentication of these MIPv6 packets.

5. References

[CaGo00] A. T. Campbell, J. Gomez, S. Kim, Z. Turanyi, C.-Y. Wan, and A. Valkó. "Design, Implementation and Evaluation of Cellular IP." *IEEE Personal Communications* 7 (4): 42-49, August 2000.

[CaGo02] A. T. Campbell, J. Gomez, S. Kim, A. Valkó, C.-Y Wan, and Z. Turanyi, "Comparison of IP Micromobility Protocols." IEEE Wireless Communications Magazine, 9(1): 72-82, February 2002

[ChAl03] Chaskar H., ed. “Requirements of a Quality of Service (QoS) Solution for Mobile IP,” IETF RFC 3583, September 2003.

[DIAM] Open Source Diameter Server http://sourceforge.net/projects/diameter

[FaLe02] Stefano M. Faccin, Franck Le “Mobile IPv6 Authentication, Authorization, and Accounting Requirements”, November 2002. http://www.ietf.org/internet-drafts/draft-le-aaa-mipv6-requirements-01.txt

[GuJo03] E. Gustafsson, A. Johnson. “Always Best Connected”. IEEE Wireless Communications 10(1): 49-55, February 2003.

[JoPe03] D. Johnson, C. Perkins, J. Arkko, “Mobility Support in IPv6.” Internet Draft, draft-ietf-mobileip-ipv6-24, June 2003.

[KeAl02] Kempf J. ed. ”Problem description: reasons for performing context transfers between nodes in an Ip access network.” IETF RFC 3374, September 2002.

[LeFa02] F.Le, S.M. Faccin “Dynamic Diffie Hellman based Key Distribution for Mobile IPv6”, Internet Draft, April 2002.

[MaLo 02] Manner J., López A., Mihailovic A. et al. "Evaluation of mobility and quality of service interaction." Computer Networks 30: 137-163, 2002.

[MiPä00] Mink S., Pählke F., Schäfer G., and Schiller J. “Towards secure mobility support for IP networks.” IFIP International Conference on Communication Technologies (ICCT): 555-562, August 2000.

[NiAr03] P. Nikander, J. Arkko, T. Aura, G. Montenegro, E. Nordmark, “Mobile IP version 6 Route Optimization Security Design Background”, Internet-Draft, draft-nikander-mobileip-v6-ro-sec-02, December 2003.

[PANA] Protocol for carrying Authentication for Network Access (PANA) http://www.ietf.org/html.charters/pana-charter.html

[PeAl02] C. Perkins, ed. “IP Mobility Support for IPv4.” IETF RFC 3344, August 2002

[SoCa03] H. Soliman, C. Castelluccia, K. Malki, L. Bellier, "Hierarchical Mobile IPv6", Internet Draft, draft-ietf-mipshop-hmipv6-00, October 2003.