Telecommun Syst (2016) 61:257–275
DOI 10.1007/s11235-015-9999-3

Mobility management solutions for current IP and future networks

Jozef Wozniak

Published online: 17 March 2015

Abstract Enormous progress in the design of portable electronic devices allowed them to reach a utility level comparable to desktop computers, while still retaining their mobility advantage. At the same time new multimedia services and applications are available for IP users. Unfortunately, the performance of base IP protocol is not satisfactory in mobile environments, due to lack of handover support and higher layer mobility management mechanisms. In this paper, we outline the most important current methods of handling mobility in IP networks that are expected to play an important role in the future (covering the following ISO–OSI layers: 2+, 3, 3+, 4 and 7 of mobility solutions), as well as describe the newly proposed methods.

Keywords IP networks · Mobility protocols · Analysis · Comparison

1 Introduction

There is no doubt that wireless technologies and new generations of mobile devices were one of the main drivers of telecommunications in the last two decades. Over the past years we have been witnessing a very rapid growth in the popularity of various mobile devices processing and presenting digital data. A large and fast growing number of such multifunctional terminals include such devices as laptops, palmtops, PDAs, smart phones, as well as multi-functional GPS navigation devices, and even portable MP3/MP4 media players. Analyzing utilization of these devices one can observe a progressive convergence causing more and more functions to be integrated in a single device, thus increasing the range of their applications. Current and estimated trends in popularity and usefulness of different portable devices reported by Cisco [1] are presented Fig. 1.

The vast majority of multi-function terminals can use the IPv4 protocol, however more and more of them implement and utilize the new version—IPv6. The prevalence of the use of IP protocols provides opportunities to create new, useful services, as well as new uses of known solutions. It is estimated that within the next few year, IP traffic generated by a variety of mobile devices will increase significantly (see Table 1)

The above-mentioned trends, including the growing number of mobile terminals, and a huge increase in traffic volume generated by these devices make it necessary to implement the protocols that allow for dynamic management of mobile stations as they move between network attachment points.

However, implementation and deployment of solutions allowing mobile users to maintain high quality of network communication in IP-based network environment proves to be a complex and difficult task.

Mobile users need to communicate without interruption while moving across different access networks, which requires not only an ability to seamlessly change points of physical network access (handover), but also to dynamically manage IP-level configuration to deal with, for example, involuntary changes of user's IP address. It is evident, that efficient support of mobile users requires us to address both seamless handover between network attachment points and network-layer mechanisms for uninterrupted IP connectivity.

There are several mobility scenarios. *Service mobility* provides continuous access to the service, even when the user moves, irrespective of the network type or technology. *Network mobility* (see e.g., [2]) refers to the case when all net-

J. Wozniak (✉)
Department of Computer Communications, Faculty of Electronics, Telecommunications and Informatics, Gdansk University of Technology, Narutowicza 11/12, 80-233 Gdańsk, Poland
e-mail: jozef.wozniak@eti.pg.gda.pl

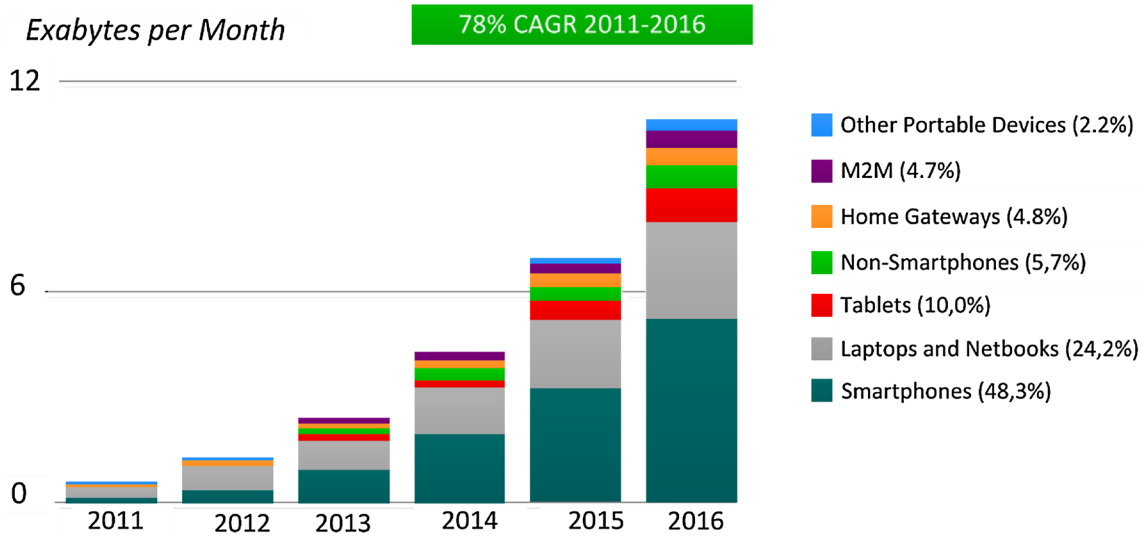

Fig. 1 Traffic shares for different portable devices

Table 1 Parts of IP traffic generated by different ICT segments (*Source* Cisco VNI Mobile, 2012 [1])

IP Traffic, 2011–2016	2011	2012	2013	2014	2015	2016
By type (PB per month)						
Fixed internet	23,288	32,990	40,857	50,888	64,349	81,347
Managed IP	6,849	9,199	11,846	13,925	16,085	18,131
Mobile data	597	1,252	2,379	4,215	6,896	10,804

work elements (i.e., nodes, access points, routers, etc.) are mobile, i.e., changing their positions in time. If a personal identity can be used in multiple terminals (e.g., GSM SIM card), it is referred to as *personal* (*or user*) *mobility*. This includes the ability to access the network from different terminals under the same identity. *Terminal mobility* is the ability of a user terminal to access the network when the terminal moves. *This type of mobility has an influence on multiple protocols at communication layers and is the main subject of this paper*.

Many mobility management solutions have been proposed to overcome serious limitations of layered (TCP/IP and ISO–OSI) architectures designed with only stationary users in mind. In majority of them to address the changing location of moving hosts, an additional IP address, named care-of address (CoA), is allocated to the hosts. For moving hosts, the CoA is utilized for routing (to the current location), while the original IP home address (HoA), is treated as identifier (ID) for the connection management. In addition, the tunneling or IPv6 extended headers are used to keep sessions with moving hosts [3–7].

Performance analysis of mobility management solutions utilizing centralized rendezvous points, like Home Agents in the classical Mobile IP, shows their ineffectiveness in large and highly mobile environments due to long-lasting registration processes and generation of unnecessary signaling traffic. To partly overcome such problems the distributed mobility management (DMM) concept can be considered [8]. In this approach, mobility agents are allocated in the distributed manner and exchange information about the current locations of Mobile Hosts (about their CoAs). The DMM concept (see e.g. Global Home Agent to Home Agent protocol [9]) can be applied to various mobility support protocols, including Mobile IP (MIP) and PMIP (Proxy Mobile IP) [10].

One of key issues discussed in the last years, in the context of new Internet services, especially ones requiring mobility, is how to overcome limits and mismatches of Internet's namespaces and addresses, that seem to be not sufficiently adjusted to new challenges. One of postulated, however controversial, in the Internet community, solution is to divide an address into a separate Identifier and a separate Locator. This proposal has been thoroughly discussed within both IETF and IRTF bodies, but was always rejected as unworkable, due to required global changes in the Internet protocols. In this paper we are trying to show, based on proposals known from literature as well as solutions standardized by IETF, what are the current developments in the mobility protocols patterned on Mobile IP, pointing also out, that evolving the naming in

the Internet by splitting the address into separate Identifier and Locator names can provide an elegant integrated solution to many key issues, at least, in separate domains without losing compatibility with the unmodified reminder of the IP environment.

One of very interesting and promising ideas behind the mobility management is to introduce a new abstract identifier, named host identity tag (HIT) (see e.g., host identity protocol [11]). Since each host is identified by its own ID that is also used for connection identification purposes, the change of IP address (as a result of host movement) does not affect the continuity of its session. To support host mobility, additional mechanisms are required.

Proper answer to the question how to address the needs of a whole mobile environment is a big challenge for current and Future Internet [12]. Among Future Internet architectures projects, one well-known is MobilityFirst (MF). This scheme was designed with the aim to solve the issue of effective support of mobile hosts (MHs) [13,14]. In this architecture, network endpoints are represented by IDs that are unique in the global sense. They can be not only hosts, but also other abstract objects, e.g., users, contents, and even contexts.

Comprehensive surveys of mobility management protocols together with related taxonomy can be found e.g., in [15–18].

In the paper, we provide an overview of selected mobility management concepts and protocols together with their classification. In Sect. 2, a basic taxonomy is presented. In the following sections essential requirements for smooth and seamless mobility handovers, together with example mobility protocols are presented.

2 Mobility management patterns, definitions and general requirements

In order to provide proper mobility management, a number of fundamental issues must be solved, and a number of requirements for efficient Internet mobility support must be satisfied. They include [4,5]:

Handover Management This requirement concerns maintaining the ongoing (active) communication session while a mobile node/host (MN/MH) moves and changes its point of attachment. The major objective is to provide the minimum value of service disruption time during handover. New services and applications have a strong impact on precise requirements of communication quality during handovers.

Location Management This includes identification of the mobile node current location, as it moves.

Multihoming To enhance throughput and at the same time to make implementation of the Always Best Connected (ABC) concept possible, it is desirable for an MN equipped with multiple interfaces to have multiple communication paths across either the same, or different access networks in mobility scenarios.

Applications Internet mobility should support current services or applications without requiring them to be modified.

Security When providing Internet mobility support, a very important issue is protection provisioning against various misuses of the mobility features and the respective mechanisms.

More in-depth requirements for mobile-oriented Internet environments are presented in [19,20]. The first set of requirements concerns the node identification (ID) and location (LOC), commonly represented by a pair of IP addresses, used not only in the network layer, but also in the upper layers. The present mobility solutions often assume the need for changing the mobile host location, leaving the identification used by the upper layers intact. What is more, network nodes use separate addressing for each of their interfaces, which may be cumbersome in multihoming environments. It also poses a problem of general node identification.

The mentioned issues can be formalized into the following three requirements concerning the nodes' static ID structure:

1. ID for identification should be separated from LOC used for routing.
2. Mobile hosts should not possess a static LOC.
3. An addressable entity (for example: network node) should possess one ID not related to any of its interfaces nor its location within network structure.

The next set of requirements, formulated in the paper, concerns the architecture of mobility support in the network. The assumption is *that a single anchor responsible* for controlling traffic delivery to a mobile host's changing network location leads to non-optimal routing, additional traffic overhead, and single point of failure case. Hence, two more requirements are needed for scalable, mobile host dominant environment:

4. Mobility support should be provided natively rather than as an additional feature.
5. Traffic forwarding for mobile hosts needs to be realized in a distributed manner.

A general requirement related to quality of service states that the control and data planes should be separated. This requirement comes from the assumption that the actual mobility provision needs more control messages in comparison to traffic between static host forwarding. Hence, the sixth requirement is formulated in the following way:

6. The control plane should be separated from the data plane.

The next three requirements are related to the Future Internet concept and concern the issues of common delivery mechanisms for heterogeneous and diverse networks, the way of mobility provisioning, and the routing scalability. They are formulated as follows:

7. There should be a possibility of different protocols usage in mobile environments.
8. Both the host-based (end-to-end) and network-based solutions should be considered.
9. Both mobility and scalability issues should be considered in Future Internet addressing architecture.

In addition to the above requirements for current and mobile-oriented Internet mobility support, there are also performance requirements for mobile environments. While developing any Internet mobility solution, the following performance measures are the most relevant:

Handover Latency—time elapsed from the moment of receiving the last packet via the old network to the moment of receiving the first packet via the new network after the handover.
Packet Loss—defined as the number of lost packets measured during the handover process.
Signaling Overhead—defined as the number of messages exchanged between networking components for the handover and location procedures.
Throughput—the amount of data successfully transmitted via a mobile Internet in a given time period.

The mobility support protocol has to fulfill multiple functions that are not present in networks supporting only stationary clients. *Registration* is the process in which the network is informed about the device and user that connects to the network and is ready to receive requests. The procedure typically includes authentication, authorization and accounting (AAA). *Paging* is the procedure used to determine the location of a mobile device within the network. The procedure used by the mobile device to inform network about its new position is called *location update*. *Handover* is the procedure that controls the transition of the mobile device between the points of attachment to the network. Its performance has a direct and profound impact on user satisfaction. Finally, *rerouting* is the modification of the routing information that is typically required after handover.

A change of node's network point of attachment can lead to various results, as far as network mechanisms are concerned. Example scenarios include:

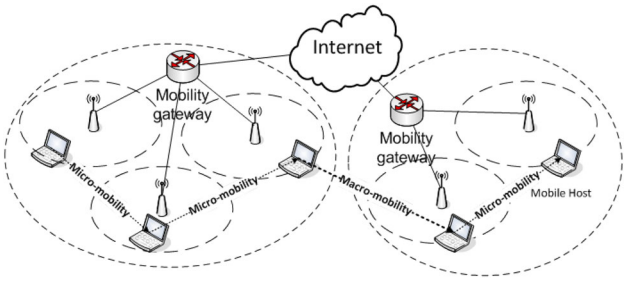

Fig. 2 Micro and macro-mobility

- a change of access point in a homogeneous network (including a horizontal or intra-technology handover),
- a change of access technology (both a vertical as well as inter-technology handover),
- a more advanced case of change of access router requiring network layer information like IP addressing (inter-Access Network handover).

It should be noted that, if a mobile terminal is equipped with more than one network interface, it can use one of them to obtain connectivity through a new point of network access during handover, while still continuing using the old one. That way connectivity disruption is vastly minimized, as connectivity through new point of network access is already functional, when the old one is disconnected. Such an approach is called soft-handover, in contrast with usual, single-interface procedure, requiring terminal to terminate network access before attempting to connect to a new point of access (hard-handover).

Terminal mobility can also be classified as inter- and intra-domain (Fig. 2), where the term "domain" refers to a network under a single management and authority. Such a distinction opens up the possibility to apply methods designed for specific functionality.

The more complex of these two scenarios, inter-domain mobility (sometime called macro-mobility or global-scale mobility) occurs, when mobile user changes his point of network attachment between two separate domains. The fact, that he moves between two independently managed network systems not only requires a full low-layer handover, authentication process, network layer configuration (including IP address assignment and verification), registration by mobility management mechanisms, but also significant data path changes.

The second scenario, intra-domain mobility (also called micro-mobility, or local-scale mobility), takes place when mobile user moves within domain boundaries. Due to the fact, that he remains within the same network system, many of the abovementioned steps can be simplified, for example: IP address can be retained or fast re-association can be used in place of full association/authentication procedure.

Table 2 Main sources of handover latency

Layer	Item	IPv4 best case (ms)	IPv4 worst case (ms)	IPv6 best case (ms)	IPv6 worst case (ms)
L2	**802.11 scan (passive)**	0 (cached)	**1000 (wait for Beacon)**	0 (cached)	**1000 (wait for Beacon)**
L2	**802.11 scan (active)**	20	**300**	20	**300**
L2	802.11 assoc/reassoc (no IAPP)	4	20	4	20
L2	802.11 assoc/reassoc (w/ IAPP)	20	80	20	80
L2	**802.1x authentication (full)**	**750**	**1200**	**750**	**1200**
L2	802.1x Fast resume	**150**	**300**	**150**	**300**
L2	Fast handoff (4-way handshake only)	10	80	10	80
L3	**DHCPv4 (6 to 4 scenario only)**	**200**	**500**	0	0
L3	**IPv4 DAD**	0 (DNA)	**3000**	0	0
L3	Initial RS/RA	0	0	5	10
L3	**Wait for more RAs**	0	0	0	**1500**
L3	**IPv6 DAD**	0	0	0 (UDAD)	**1000**
L3	MN-HA-BU	0	**200**	0	**200**
L3	MN-CN BU	**100**	**200**	**100**	**200**
L4	**TCP adjustment**	0	**Varies**	0	**Varies**

Bold values represent the most time consuming procedures

The elementary mechanisms for mobility support tend to be relatively simple in their base architecture. However, the necessity of taking into account the above mentioned requirements while providing adequate performance creates the need for sometimes complex optimizations of these solutions.

In the following sections, we present the justification for the need of improvements and describe the major mechanisms, optimization methods, and procedures supporting terminal mobility.

3 Functional requirements for internet mobility support

Apart from low-layer handover procedures regarding fast and seamless change of physical point of network attachment, a number of issues concerning IP protocol operation must also be addressed. Required exact mechanisms depend on a particular mobility type that brings specific challenges and requires specialized solutions.

In general, the traditional TCP/IP protocol stack and a significant number of wireless access technologies have been designed for stationary computer networks, which causes many problems when serving mobile hosts.

At the same time, modern mobile smart devices, such as phones or tablets, offer functionally comparable to desktop computer systems, even when multimedia capabilities are concerned. In this situation, it is only natural for their users to expect comparable quality of experience during both stationary and mobile use. However, with standard IPv4/IPv6 network stack designed sorely for stationary use scenario, multiple challenges appear, of which ability to provide necessary bandwidth is often the easiest to fulfil.

3.1 Limitations of data link layer

Most of wireless access technologies currently available on the market provide some kind of mobility support only as an optional, and often not standardized, solution. Moreover, such mobility mechanisms are at best applicable only in the case of homogeneous networks and address only the link layer [21] mobility issues. As such, they are not sufficient to address Internet mobility across heterogeneous networks. In general, due to network heterogeneity characteristics, it can be advantageous to locate mobility support functions at higher layers. Today, wireless local area networks (WLANs) are mainly based on IEEE 802.11 technology. Mobility in 802.11 means change of physical access point when the MN moves between access points in the extended service set (ESS) infrastructure. The whole handover (handoff) procedure is composed of a number of phases [21], among which detection, scanning and IEEE 802.1x authentication are operations that consume the most time (see Table 2). Minimization of scanning time (Tscan) can be achieved by employing one of the algorithms proposed in literature (see e.g., [22]). IEEE 802.1x authentication also introduces a significant overhead due to multiple security procedures involving the wireless client and AAA server. Reduction of IEEE 802.1x delay is addressed in the IEEE 802.11r standard [23]. Moreover, solutions of proprietary type proposed by hardware vendors, most frequently based on dedicated wireless network controllers, could be used as well.

Table 3 Mobility impact on protocol layers and mobility management issues

Protocol layer	Impact and proposed solutions
Physical layer	The radio link quality changes as the device moves.
Data link layer	The access link quality and availability changes, frame loss can occur, interface queue may encounter overflow. Moreover, functionalities of link-layer solutions differ significantly in respect to medium access efficiency, predictability, QoS support etc. Due to these differences, Data Link layer solutions are tightly coupled with specific wireless technologies.
Network layer	Mobile host address can change and/or data path through the network needs to be modified due to change of its location (and, as a result, its point of network access). Example mobility-related features (e.g., addressing management and routing control) necessary due to modifications of network topology being result of MH's location changes, are provided at the network (IP) layer. In order to perform their tasks, the network—as well as host-based mobile solutions use techniques such as proxy, tunneling, separating etc. to address mobility issues. Additional centralized or decentralized rendezvous points like gateways are usually proposed to handle the interworking and interoperating issues when roaming among heterogeneous access networks.
Transport layer	A session can be broken or its quality can deteriorate as the device moves. The main idea of providing the transport layer mobility is based on removal of network layer dependences using e.g., migration, indirection, tunneling or multihoming approaches.
Application layer	Connection-aware applications need to be adapted as the host/network configuration changes. A common example is the necessity of tracking mobile client's current IP address. Many new elements supporting mobility can be applied. In the case of Session Initiation Protocol (SIP) based Internet telephony, new logical entities, like: user agent, redirect server, proxy server and registrar are defined for this purpose. They hide user location and relay messages, process registration requests and provide a location service that connects SIP URI (Uniform Resource Identifier) with one or more IP addresses. Standard Internet services, such as DNS or Dynamic DNS can be used to map domain names into IP addresses.

Apart from the necessity to provide seamless communication at low layer, there is a need for operations in the ISO–OSI Network Layer if high-quality communication is to be provided to mobile network elements. The exact range of necessary mechanisms depends on the mobility type and scope. In the case of host-based mobility (HBM), all functionalities are included and implemented at the MH in its IP protocols suite. In such a scenario, a radio access link of limited bandwidth is used to transport both mobility-related signaling and the MH's data packets. Moreover, MH data delivery is often conducted with the use of tunneling, which causes additional bandwidth reduction at this critical network location. In contrast, in the case of network-based mobility (NBM), a network-side proxy mobility agent is used in place of a client-side agent that performs signaling and management on behalf of the MH. It is evident that the NBM approach can bring significant efficiency advantages in the MH wireless access link utilization, but also imposes strict requirements for consolidated management of all network infrastructure devices.

3.2 Limitations of IP address (network Layer)

[17] In case of IP network-layer protocols, network address both identifies the node and describes its location within the network. When a mobile node moves, it eventually becomes necessary to change its point of network attachment. In such case, node's network address must also be changed, to reflect its new location within the network structure. Without additional mobility support mechanisms, such change makes it impossible to maintain uninterrupted communication with mobile node. Even if MN is able to obtain a new network address quickly, existing transport layer connections will not be transferred.

Most of IP mobility solutions address this issue by assigning two IP addresses to a mobile host, i.e., one persistent (maintained by mobility mechanisms) and one changing (assigned by local network management). The traffic addressed to persistent IP address is then tunneled to the changing one.

3.3 Lack of cross-layer cooperation and signaling mechanisms between layers

Traditional design of transport-layer protocols follows strictly layered approach, in which its mechanisms relay on services provided by network-layer. Such approach, while making the protocols media-independent, does not allow them take into account the wireless link properties. As a result, the congestion control procedures of transport-layer mechanisms cannot distinguish.

3.4 Limitation of applications

[17] Applications developed for static TCP/IP network environment tend to encounter multiple difficulties in mobile environments. The most common problems are caused by highly unpredictable degradation of QoS parameters caused by host mobility, handovers and network layer configuration changes —real-time services are particularly suscepti-

ble to the above effects. Furthermore, some application-layer services include IP addressing information in their working parameter sets, without mechanisms to modify it, short of creating completely new communication session.

Table 3 summarizes the impact of the terminal mobility on each layer of the ISO–OSI protocol stack in case of IEEE 802.11-based network access. It is evident that there are two most prominent sources of handover delay—the Link Layer handover itself and basic IP stack configuration.

Link Layer handover delay is caused by the necessity to discover a new point of network attachment (up to 1 s) and authenticate client (up to 1.2 s). Since, due to IEEE 802.11 procedures, client is required to disconnect from its current point of network attachment before it attempts to find a new one, the network connectivity is interrupted for the whole duration of handover. Various optimizations have been proposed to minimize that interval, allowing for a significant reduction of both discovery and authentication time.

Handover-related IP stack configuration procedures causing the most significant delay refer to IP address configuration and verification (mainly Duplicate Address Detection). With default IP configuration settings, the delay can be as long 3.5 s, so alternative methods of address assignment have been proposed, which aim to minimize the address acquisition time and remove the need to verify it.

4 Internet mobility support based on IP layer

The basic classification of different mobility management strategies refers to the mobility scale. Therefore, we have IP macro- (or global-) mobility and micro- (or local-) mobility management protocols. A comprehensive survey of IP mobility management solutions, representing both groups can be found in [15–18].

4.1 Macro-mobility concept

IP macro-mobility mobility concept is aimed to support movement (roaming) of mobile nodes between subnets in two different administrative domains without disconnection of established communication sessions. Such scenario most probably requires support for very large number of geographically dispersed mobile nodes, which makes scalability of employed solution one of the most important factors. As a result macro-mobility protocols tend to closely cooperate with IP routing mechanisms to integrate fixed and mobile networks.

4.1.1 Mobile IP

Mobile IP (MIP) is one of the most popular, well known and widely adopted of macro-mobility solutions for IP systems. It is proposed by the Internet Engineering Task Force (IETF) to enable the mobile node to access the Internet and roam freely between different subnets without losing the connection, offering mobility support in the network layer, isolating higher layers from mobility issues. Mobile IP has two versions, namely Mobile IPv4 (MIPv4) [3,4], and Mobile IPv6 (MIPv6) [6,7].

The Mobile IPv4 network architecture includes three functional entities:

- *Mobile host/node (MH/MN)* A host or router changing its access point (i.e., moving to another subnet) without updating its IP home address.
- *Home agent (HA)* A router located on the MH/MN home network.
- *Foreign agent (FA)* A router located in each foreign network visited by MH/MN, which enables the MH/MN to maintain its network connectivity.

Utilization of several addresses for user movement management is the key idea introduced in Mobile IP. The MH owns its own IP address which can be referred as traditional (or persistent). Mobile IP introduces the term "home address" for such an address. Each time the MH connects to the network, a temporary IP address for the current network is obtained. The host remains reachable by means of both home and temporary addresses. For Mobile IP, the temporary IP address is termed CoA. A correspondent node/host (CN/CH) sends packets destined to the MH using its IP home address, and the packets are intercepted by HA at MN home network and tunneled via mobile IP infrastructure to the mobile node current location (its current CoA).

Mobile IPv4 also facilitates generation of CoA and utilizes the concept of foreign agents (FA) located in any network. Such Foreign Agents can be visited by the MH. In the Mobile IPv6, the mobile host is able to create its own CoA based on its link-local address together with automatic address configuration functionality (i.e., merge a subnet prefix with its own hardware address).

The key element of both MIPv4 and MIPv6 is the HA. It is located in the home network, which is defined as the network that mobile IP address belongs to. It must maintain updated information about current CoA of the MH. It is up to the MH to notify the appropriate Mobile IP entity of its location. In the case of MIPv4, the MH informs the HA of its current CoA with the assistance of an FA. In MIPv6, the process is simplified and the MH informs the HA directly.

To enable the MH to detect changes of its location and to discover addresses of appropriate mobile IP agents, they periodically send Agent Advertisement messages. In the case of MIPv6, this advertisement is an extension of ICMP Router Advertisement message with fields devoted to mobility sup-

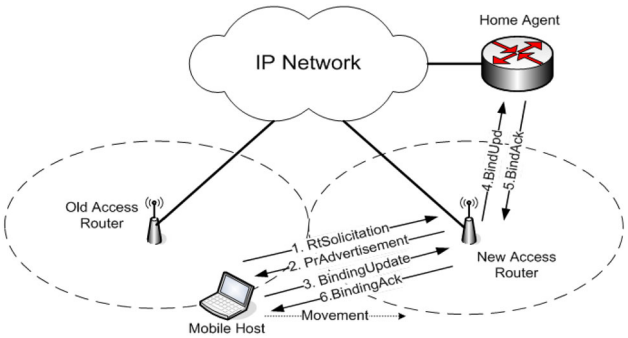

Fig. 3 Mobile IPv6 handover

port. The MH can solicit Agent Advertisement by sending the Agent Solicitation message.

Traffic destined for the MH home address is intercepted by the HA. In case the MH is outside the home network, and therefore is not able to directly receive such traffic, it can use tunneling techniques to send the traffic to an external network being the current location of the MH. In the case of MIPv4, the tunnel is created to the FA in the MH's current network, which terminates the tunnel and delivers the traffic to the MH. In the case of MIPv6, the procedure is simplified, as the tunnel is created directly to the MH.

An example of MIPv6 handover is presented in Fig. 3. When the MH leaves its home IP network, it detects foreign networks based on Agent Advertisement messages that can be solicited. To begin reception of data sent by other hosts to its home address, the MH updates bindings with its HA.

Summing up, from the perspective of the CN/CH, the MH is identified by its unchanging home address. When a packet is sent to the MH by another CN/CH, the HA (which stores the registration mapping between the MH's home address and its current CoA) intercepts it based on the home address of the MH. The packet is then tunneled from the HA to the MH (directly or through an FA).

It should be noted that the described MIPv4/MIPv6 mechanisms require that traffic addressed to MH should always be routed to its home network and only then tunneled by HA to its final destination. Such approach results in unnecessary network resource consumption and negatively impacts quality of service, due to long transmission path and tunneling overhead. At the same time, traffic sent by MH is delivered to its destination directly, causing severe asymmetry of transmission path within a single, bidirectional communication session. Fortunately, this problem (called "triangle routing problem") has been long recognized, and many respective solutions have been introduced—for example Global Home Agent to Home Agent MIP extension presented in this paper.

4.2 Micro-mobility concept

Mobile IP doesn't scale well with the number of mobile nodes because every handover between FAs triggers a Binding Update message, irrespectively of a node activity. In environments (such as access networks) where MHs change their network points of attachment frequently, mobile IP introduces significant network overhead. Due to this fact, various micro-mobility protocols have been proposed. Local mobility management solutions handle local movement of MHs without interaction with the mobile IP and HAs in the global Internet. In the case of micro-mobility protocols the current IP address of mobile host known by a macro-mobility protocol's HA no longer reflects an MH's exact point of attachment. Usually, it denotes the address of a gateway being jointly utilized by a potentially large number of access points. Such an ability brings the benefit of reducing delay and packet losses during handoff, eliminates time-consuming registration, and significantly reduces the signalling load. Thus, the local mobility management contributes to the scalability of global mobility management.

It is the overall role of micro-mobility solutions to ensure a delivery of packets arriving at the gateway to an appropriate access point, within its domain of operation. For that task, it they maintain a "location database" mapping host identifies to location information.

Micro-mobility protocols widely use a new mechanism, called IP paging, to determine MH location within a specific domain. Based on this, they create and constantly maintain paths from MHs to a border gateway. When packets destined to an MH are received by the gateway, they are sent directly to the MH's current location. High MH mobility does not cause high control overhead in the Internet. As long as an MH moves in the area of a wireless access network, its CoA from the HA does not change. Among a variety of protocols belonging to the micro-mobility group, Cellular IP [24] and HAWAII (Handoff Aware Wireless Internet Infrastructure) [25] are the most representative There are also many other solutions, including Columbia, HMIP, FMIP, TIMIP (see [15]) and a promising relatively new proxy mobile IP—PMIP [26]. TIMIP and PMIP offer network-based mobility management.

4.2.1 Mobile IPv4 extensions

The Mobile IPv4 handoff latency can exceed the threshold required to support multimedia and real-time applications. This is the reason for development of several techniques that aim to achieve lower latency handoffs in Mobile IPv4 [3, 27]. One of the solutions is Low Latency Handoff (LLH) for Mobile IPv4 described in [28] and utilizing Link Layer triggers.

An L2 trigger is a representation of an L2 event related to the L2 handoff process. It is an abstraction mechanism separating a upper layers from low layer technology-specific trigger.

In case of a wireless link providing a trigger in advance of the actual handoff, the event is early noticed of an upcoming change in the L2 point of attachment of the mobile node to the access network. Although the mobility protocols make use of specific L2 information, they should be kept independent of any specific L2.

LLH utilizes so-called cross-layer mechanisms combining low-layer (L2 and L3) handovers. The fact that Mobile IPv4 was designed without any assumptions about data link layer makes this solution highly compatible, but also results in negative consequences for handoff delay. A strict separation between the IP Layer and the Link Layer can be a major cause of handover delay. The first reason for this effect is the fact that an MH can only correspond with a directly connected FA. As a consequence, the MH cannot communicate with a new FA until L2 handoff has completed, which creates two sources of delay: L2 handoff and event propagation to the IP layer. The latter is Mobile IPv4 Registration process latency. During this time, it is not possible to send/receive to/from the MH any IPv4 packets.

There are three handoff techniques proposed by LLH, the most advanced of which is the Pre-Registration handoff method, allowing the MH to prepare its registration state on a new FA before the L2 handoff commences. It requires cooperation between FAs, and assumes that every FA stores mapping between neighbor FAs' IP addresses and L2 identifications (for example Basic Service Set IDs—BSSIDs). More specifically, the FA stores addresses of neighbors that the MH may move to. The mapping table can be provided by the administrator, or can be acquired by a neighbor discovery protocol.

4.2.2 Mobile IPv4 fast handovers

In Mobile IPv4 Fast Handovers (FMIPv4) [29], handover performance is improved by preconfiguring the new access router (NAR), while the mobile host is still connected to its previous access router (PAR). When handover necessity is detected by MH (for example by means of Layer 2 triggers similar to the LLH approach), it informs its PAR of the intention to change its point of network attachment from PAR to NAR. In response PAR informs NAR about the upcoming handover and creates tunnel allowing it to forward traffic intended for MH to NAR. This tunnel allows MH to receive traffic immediately after Layer 2 handover, by using its old CoA (obtained from PAR), despite now being connected to NAR. When MH completes Layer 2 handover between PAR and NAR, obtains new CoA and registers its presence at NAR, the new access router can modify MIPv4 registration to point incoming IP transmissions for MH's home address to a new CoA. This procedure complete, the tunnel between PAR and NAR can be disconnected.

By allowing MH to use its old CoA at new network location, delay introduced by IP Layer configuration and MIPv4 mechanisms does not result in corresponding lack of connectivity with MH. However, additional mechanisms are required to obtain mapping between Layer 2 address of new point of network access (avaliable from Layer 2 trigger mechanism) and IP address of NAR (necessary for tunnel establishement).

Both LLH and FMIPv4 are strongly dependent on unspecified L2 trigger when handover begins. This trigger cannot be relied upon in IEEE 802.11 networks as handover detection is the protocol bottleneck and can take more than one second. This delay can lead to a situation where the MH loses its connection with PAR before described procedure is completed.

There are many Mobile IPv6 extensions proposed in the literature. Example solutions include Fast Handover for Mobile IPv6 [30] and Hierarchical Mobile IPv6 [31].

4.2.3 Fast handovers for MIPv6

Fast Handover for Mobile IPv6 [30] proposes improvements aimed to minimize Mobile IPv6 handover delay. It defines two possible ways to initiate the handover: network-initiated handover and mobile-initiated handover [26]. MIPv6 Fast Handover, similarly to LLH, and FMIPv4, assumes obtaining a new CoA before proceeding with the handover, and starts using this address just after completing the L2 handover.

The protocol is initiated with the L2 trigger informing that MH is about to proceed with the handover to the particular Access Router. In the station-initiated handover MH sends Router Solicitation for Proxy (RtSolPr) to PAR. The message provides link layer address of the next attachment point, e.g., BSSID for IEEE 802.11 network. It should be noted, that the specification does not determine the way MH can obtain link layer address of the prospective access point. In response to RtSolPr, the PAR sends a Proxy Router Advertisement (PrRtAdv), which provides a link layer address, the network prefix information, and next care-of address (NCoA) for the NAR.

When the RtSolPr is received by PAR, it sends the Handover Initiation (HI) message to NAR. The Previous Access Router is able to map link layer address provided by MH into IP address of Next Access Router. Handover Initiation message initiates tunnel establishment between oAR and nAR. Such a solution makes it possible for the station to use previous care-of address (PCoA) in a new location. When the nAR receives HI, it verifies if NCoA can be used on the link and sets up a route entry for PCoA to configure tunnel end-

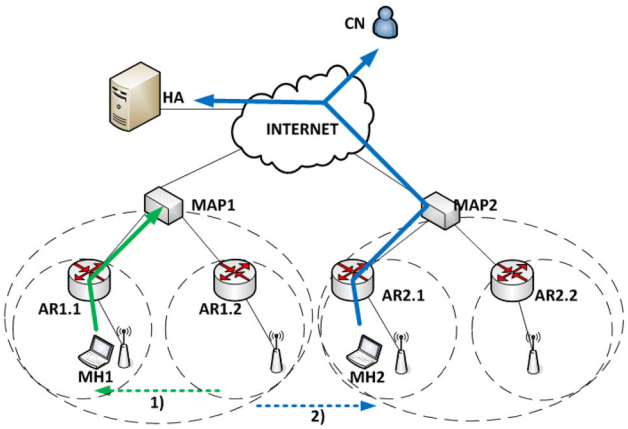

Fig. 4 Hierarchical Mobile IPv6 scenarios *1* Micro-mobility handover (in the scope of the same MAP), Local Binding Update proceed to MAP, *2* Macro-mobility handover, Home Binding Update proceed to HA and CN

point. nAR responds with Handover Acknowledge (HAck) message.

4.2.4 Hierarchical mobile IPv6 (HMIPv6) mobility management

HMIPv6 is an extension to the Mobile IPv6 protocol that enhances the efficiency of micro-mobility support [31]. This solution intends to reduce signaling load (mainly Binding Updates), limit the handover delay, optimize routing and improve scalability. HMIPv6 adds Mobility Anchor Points (MAPs), network entities that serve local handovers. MAP has a functionality similar to an HA, but it can be located not only at the domain border but anywhere within the routers hierarchy. Routers hierarchy is an effect of subdomain implementation with hierarchical IP addressing. Handling mobility in smaller scope (inside one subnetwork) improve handover efficiency.

Apart from its home address, the mobile node is assigned two addresses: *Local CoA (LCoA)* and *Regional CoA (RCoA)*. When using RCoA, MAP is a local HA for a mobile node. The specification defines a method that the mobile node can use to discover MAPs, thus creating a comprehensive solution. HMIPv6 scenarios are illustrated in Fig. 4.

Another similar solution, namely Seamless Handover for Mobile IPv6 (S-MIPv6) algorithm is proposed in [32]. This method evolved from Fast Handover for Mobile IPv6 (F-MIPv6). The new algorithm solves problems of triangle route and sequence disorder, present in F-MIPv6. A general characteristic of the S-MIPv6 is that it does not build tunnels as well as reduces delay need for registration. It is able to cooperate with a Mobility Anchor Point (MAP) to gain advantage from hierarchical networks.

4.2.5 Global home agent to home agent

This interesting solution [9] aims to improve the efficiency of MIPv6 and similar mobility protocols based on the use of HAs. It utilizes a distributed set of HAs and relies on anycast routing capability of the underlying network, which significantly limits possible deployment environments for this method.

A number of HAs distributed through the network join an anycast group and announce the same home prefix. This allows a MN to easily contact an HA closest to its own network location, by using its anycast address. Having contacted an HA, the MN registers its presence as per MIP protocol, making the chosen HA the so-called Primary HA for the MN. Information about the registration is then distributed through all HAs in the system, keeping them synchronized. A moving MN will continue to use HA anycast address to update its registration; this will possibly result in changing its Primary HA, as anycast routing mechanisms will always deliver its update message to the closest HA in the synchronized set.

When a CN wants to contact an MN, it will send the data to the MN's anycast address from within anycast prefix advertised by the synchronized HA set. Anycast routing will cause the data to be delivered to HA closest to CN. Since such an HA contains information about current MN's Primary HA, it will forward the data in encapsulated form to the Primary HA, from where they will be delivered to the final destination.

Due to the use of anycast routing, it is assured that data always enters and exits the HA set in points closest to intended endpoints, thereby allowing to avoid direct communication between corresponding nodes (and the associated triangle routing problem) without significant loss of efficiency.

4.2.6 Proxy mobile IP

As opposed to Mobile IPv4 and Mobile IPv6 which are host-based mobility standards, Proxy Mobile IPv6 (PMIPv6 [26]) presents a network-based approach that does not require any kind of client-side mobility agent. Such a paradigm shift brings numerous advantages, such as simplified management, the ability to support legacy clients and better efficiency of radio-link utilization.

In general, a proven MIPv6 idea is reused and extended by PMIPv6. However, in case of PMIPv6, no modification of a standard mobile node IPv6 stack is needed. A network-side proxy mobility agent replaces an MIP client-side agent, and provides management and signaling functionality on behalf of the MH. As a result, thanks to PMIPv6, the efficient solution is achieved without incorporating tunneling and signaling overhead on a radio access link. However, due to the lack of standardized macro-mobility mechanisms and proce-

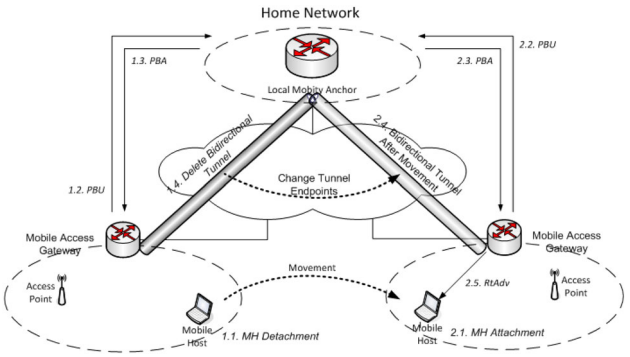

Fig. 5 Proxy mobile IPv6 domain

dures, Proxy Mobile IPv6 cannot be implemented as a global mobility system in a standalone form.

Proxy Mobile IPv6, as defined in RFC 5213 (Fig. 5), uses two specialized network elements, namely Mobile Access Gateways (MAGs) and Local Mobility Anchors (LMAs). MAG is responsible for several tasks: it tracks the MH's movement between APs and creates a bidirectional tunnel to a Local Mobility Anchor, thus managing the connectivity between the MH and the LMA. LMA plays a role similar to HA in typical client-side Mobile IPv6. It is also responsible for maintaining routes to all Mobile Hosts in the domain and forwarding traffic to and from them.

PMIP protocol is a relatively new concept, but an optimization proposal has already been published in RFC 5949. It is called Fast Proxy Mobile IPv6 and addresses the issue of handover performance. It proposes mechanisms for reducing packet loss and handover latency. In this approach, a direct tunnel between oMAG (old MAG) and oMAG (new MAG) is used to forward traffic during the handover. When the MH is disconnected during handover, its downlink data (traffic addressed to the MH) may be buffered in the nMAG.

4.2.7 Protocols utilizing NAT

Another group of protocols utilizes the network address translation (NAT) concept that is widely used in IPv4 networks [33,34]. The authors argue that NAT-based solutions are easier to deploy, but at the cost of reduced functionality. For example, reverse address translation (RAT) is the macro-mobility approach based on NAT that supports only UDP traffic [33]. On the other hand, Mobile NAT provides both micro- and macro-mobility support, and can be deployed as a mobile IP replacement [34].

5 Mobility support offered by higher layers

The next group of mobility management protocols consists of solutions offering mobility support in higher layers. This group includes transport and application layer solutions that are typically employed in heterogeneous networks to allow specific applications to function over different network technologies with different features. By locating mobility mechanisms high in ISO–OSI protocol stack we are able to provide mobility support largely independent of underlying network features (for example, there is no need for additional mechanisms at network devices), but the support we obtain is limited to specific transport protocols, applications or even particular implementations of a given user application.

5.1 Mobility support in the transport layer

A transport layer performance is strongly influenced by the mobility of network elements. In order to enable transport layer mobility, it is necessary to remove network-layer dependences by using indirection, migration, tunneling or multi-homing techniques. Example solutions dealing with the improvements of TCP efficiency in mobile environments include E2E (End-to-End) [35] communication and Mobile Stream Control Transmission Protocol (M-SCTP) [36]. In both cases, Dynamic DNS concept is used to track the mobile nodes and update their current location.

Other example protocols of this type are Host Multiple Address Service for Transport (MAST) [37], IKEv2 Mobility, and Multihoming (MOBIKE) [38].

5.2 Mobility support in the application layer

In the application layer there are several attempts to support Internet mobility. SIP [39] and Extended SIP Mobility [40] are good examples.

The SIP is an application-layer protocol used to maintain multimedia sessions [39]. It is a standard proposed by the IETF and is mainly used for Internet telephony. SIP allows two or more participants to manage a session consisting of different media stream types. For example, video and voice streams can be directed to the appliances specialized to receive particular workload. SIP end-points are addressed with an email-like address "user@host" named SIP Uniform Resource Identifier (URI).

Extended SIP Mobility [40] is a macro-mobility approach that utilizes SIP. SIP already supports user mobility (Fig. 6a). Users after handover, should perform registration procedure, subsequently they are able to initiate or response invitation of a new call (Fig. 6a–1). At the same time existing connections are broken due to IP address change (Fig. 6a–2). However, the protocol has to be extended to support an active session while the user is moving. As the mobile node is identified by SIP URI, no home IP address is required. The problem of mobility in SIP is considered as increased roaming frequency and IP address change during the session. If the mobile node moves during the session, it has to send a re-invitation (Fig.

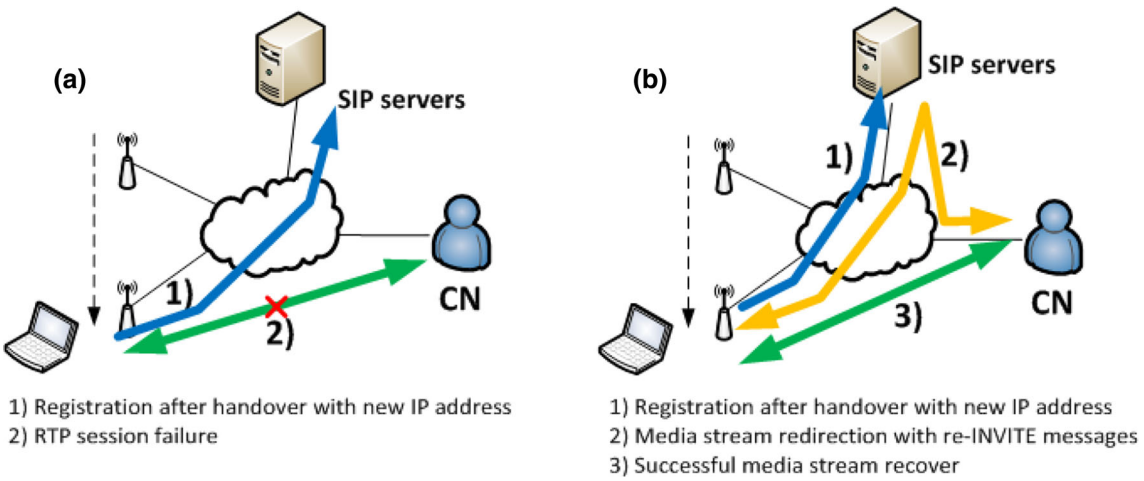

Fig. 6 Illustration of the extended SIP concept

6b–2) to the CN with a new IP address for media stream (Fig. 6b–3). Considering signaling overhead, SIP Mobility is a costly solution that has to be implemented for each application separately.

6 Mobility proposals for future internet

One of key issues discussed in the last years, in the context of new Internet services, especially ones requiring mobility, is how to overcome limits and mismatches of Internet's namespaces and addresses, that seem to be not sufficiently adjusted to new challenges. One of postulated, however controversial, in the Internet community, solution is to divide an address into a separate Identifier and a separate Locator. This proposal has been thoroughly discussed within both IETF and IRTF bodies, but was always rejected as unworkable, due to required global changes in the Internet protocols. In this paper we are trying to show, based on proposals known from literature as well as solutions standardized by IETF, what are the current developments in the mobility protocols patterned on Mobile IP, pointing also out, that evolving the naming in the Internet by splitting the address into separate Identifier and Locator names can provide an elegant integrated solution to many key issues, at least, in separate domains without losing compatibility with the unmodified reminder of the IP environment.

The TCP/IP protocol stack, used in the current Internet, was designed with the assumption of establishing connections between stationary or slowly changing their location terminals. This protocol stack is not able to meet all the needs of the Future Internet. As shown in many analyses (also here), its use does not satisfactorily solve many of the problems standing on the way of providing effective service to different user groups existing in the current and still evolving Internet. Therefore many efforts are undertaken to design new network mechanisms there are not simple modifications of existing protocols, but are designed, from the very beginning, taking into account the requirements of mobile users. The prevailing trends take into account differentiated solutions based on a variety of virtualization techniques. The very characteristic element of many of such proposals is the introduction of mechanisms that allow the separation of the so-called upper layers from the transport network (e.g., IP layer). The transport network may be, in practice, any communication system that enables data transmission between two devices—employed solutions include both the data link (such as Ethernet) and network layer (eg, IPv4, IPv6) technologies.

Thanks to virtualization, the techniques used in the transport network, as well as its structure and configuration do not have a direct impact on the logical structure of the system as perceived by higher layers. Additionally, all aspects of the transport network operation can be changed in a way practically invisible to them. This makes, from higher layers point of view, a highly flexible environment to implement their functionality.

Another advantage is the possibility to use any, abstract identifier of the target object, which is not determined by its location in the physical structure of the network. This is a huge advantage for handling mobile devices, as the identifier in a natural way may remain unchanged. What's more, one can assign identifiers not only to network devices, but also to other resources such as specific data, people, or whole groups of such objects. This allows a relatively easy implementation of systems of content aware network elements (content aware networks—CANs).

6.1 Host identity protocol (HIP)

Host identity protocol (HIP) [11] introduces an additional layer between the transport and network layers and assigns to the node a cryptographically generated public key. The

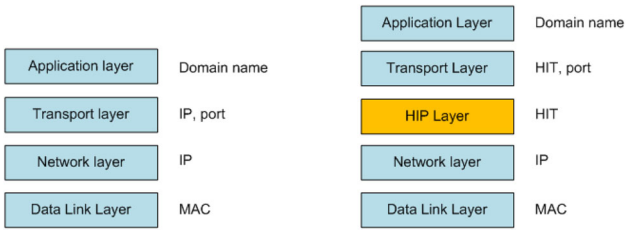

Fig. 7 Comparison of network architecture stack TCP / IP and HIP

proposed approach introduces a distinction between an identifier (the public key) and a node locator (IP address). In practice, instead of using the public key for addressing, the nodes use its hash values, called Host Identity Tags (HITs).

The binding (matching) between the identifier and locator is stored in dedicated network infrastructure components called Rendezvous Servers (RVSs). Each node is assigned to one of RVSs that monitors its current location. To establish a connection between a correspondent node (CN) and the MH, CN first queries the DNS server for information about MH. In response, the IP address of the proper RVS is returned.

Next, the CN directs the first packet of intended transmission to this specified RVS server, which retransmits it to the destination MH. As a result of this transfer, the correspondent node and mobile station can communicate directly (as exchanged datagrams contain actual IP addresses of both corresponding parties).

After changing the point of network attachment, the mobile terminal shall inform RVS about the new IP address. Fig. 7 presents a comparison of simplified network stacks—the traditional TCP / IP and HIP. In color it was marked an additional layer between the network and transport layers.

6.2 Localized mobility management

LISP-Mobility [41] is a relatively new approach introducing an additional abstraction layer between the network and transport ISO–OSI layers. Mechanism of the LISP-Mobility protocol responsible for handling mobile nodes is based on the use of Locator / ID Separation Protocol (LISP) [42]. This protocol has been developed to improve the flexibility and scalability of IP routing and introduces, in place of a single IP address, a pair of independent parameters: the node identifier (Endpoint Identifier—EID) and localization information (Route Locator—RLOC). This concept allows for the elimination of one of the most serious limitations of IP addressing, namely separates the node connection identifier (ID) from its current location in the structure of an IP network. In the case of classical IP network layer mechanisms, the nodes had to always obtain an IP address from the pool available at a given location, determined by the routing structure. In practice, any change in the location of the node, resulting in a change of its network access point, resulted in turn in the need to obtain a new IP address, and changing its identity.

An essential postulate of the LISP protocol, in the design process, was to preserve compatibility with legacy systems, so both EID and RLOC take the form of IP addresses. In addition, in its basic version, the solution does not require any modification of the network stack of end-nodes, since all mechanisms are located on access routers.

When the device wants to send data to another terminal, it creates an IP packet using EID tags in the fields of source and destination address. Access router that supports the source node (called Ingress Tunnel Router—ITR), having received such a packet, communicates with a Map-Server. The purpose of this unit is to create and store a mapping identifier EID—RLOC localization information. For the purposes of scalability, information stored by RLOC Map-Server is not unique for each EID, but points the access router (supporting LISP mechanisms) responsible for managing the target node group (so-called LISP Site)—see Fig. 8.

With this information, ITR sends the received packet to the access router that supports destination node (Egress Tunnel Router—ETR) using encapsulation mechanisms. ETR unpacks the package and delivers it to the destination node.

As a result, we have a scalable solution enabling for movement of nodes from one location to another one transparent

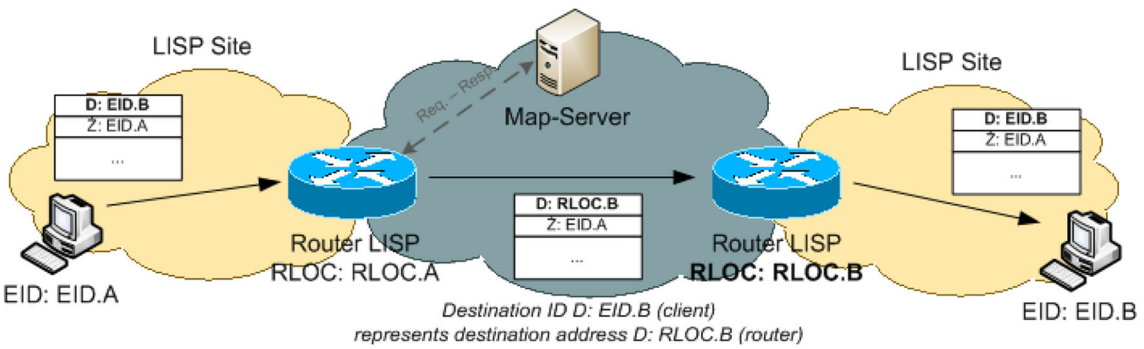

Fig. 8 Diagram of the LISP protocol

for the upper layers, without changing their address information; this in turn offers a wide range of applications, such as improved flexibility in the use of available address pools, easier change of the location of service infrastructure elements, faster response on failures, etc. In addition, it should be noted that this solution is also a convenient tool for migration, as the network layer (responsible for transferring the data) can be changed without necessity to change the mechanisms of the upper layers.

Note, however, that the LISP solution is not directly mentioned as a method to support the mobility of nodes, because the area of its operation would be limited to network access routers equipped with powerful mechanisms required by this solution. For the operation of mobile nodes modification of the protocol LISP, namely LISP-Mobility is recommended.

LISP-Mobility resigns in part from the scalability offered by the above solution, because it requires that each of the nodes is a stand-alone mobile LISP Site. So, each mobile node has its own EID-RLOC mapping, maintained in a Map-Server device, continuously updated. The LISP-Mobility solution preserves unchanged LISP node ID (EID), used by the higher layer for communication with it, while the network layer uses localization information that is subject to change with progressive changes of node location, contained in the parameter RLOC.

This approach enables the macromobility, but on the other hand, requires implementation of the LISP-Mobility components in a MN (it is not required for the CN), and also causes much larger number of mappings maintained by a Map-Server device.

Undeniable advantage of the LISP-Mobility solution is, however, compatibility with the base LISP solution, and therefore also with existing IP solutions and the related easiness of implementation. The LISP architecture is also flexible and easily extensible, which could provide a platform for a greater number of additional network services.

In its base concept, LISP-Mobility may be seen as similar to Mobile IP. ITR and ETR functions need to be implemented on each mobile node. and when a mobile node moves into a network and has to change its RLOC, it updates EID-RLOC mapping in its preconfigured Map-Server. However, while MIP has been designed as a dedicated mobility support solution, LISP-Mobility is an adaptation of LISP architecture, resulting in higher flexibility but also in higher MN resource requirements and lack of MIP many extensions.

We should also note, that another solution based on similar principles, Indentifier–Locator Network Protocol, has been proposed in [43]. It accepts an abstract network protocol, based on IPv6 and splitting the IP address into separate Identifier (representing a virtual or physical node) and Locator (being an IPv6 address prefix and describing a single IP subnet). Usage of these two separate names can provide an elegant integrated solution to the key issues referring to routing and mulihoming, without changing the core routing architecture, while offering incremental deployability through backwards compatibility with IPv6.

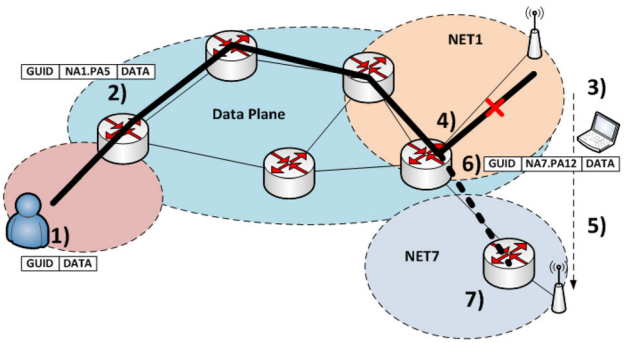

Fig. 9 Handover scenario in mobilityfirst architecture (*NA*—network address, *PA*—port address)

6.3 MobilityFirst

Current Internet protocols of the TCP/IP stack were designed taking into account a connection oriented communication model and a slow-changing network topology. As can be seen from the previous analysis, such an approach causes many difficulties in mobility support, which need to be added as an additional set of mechanisms in existing network infrastructure and/or clients.

The MobilityFirst project proposes a new approach, according to which all nodes, both mobile and fixed, support a uniform set of mechanisms. MobilityFirst introduces new principles for networking such as: separation of naming and addressing (using long GUID— Globally Unique ID, Network Address—NA), fast global resolution service (GNRS), hybrid GUID/NA storage-aware routing, in-network storage and computing options at routers [13,14]. These ideas allow to prevent from implicit (or explicit binding of sources and destinations to the current network topology. An example handover with handling disconnection is presented in Fig. 9.

This solution can be thought of as a layered approach. To communicate with another entity or access a given resource, a user presents the network stack with its "friendly name". This identifier is highly abstract, designed to be of easiest possible use for a human user and can be highly service-oriented (for example: "all voice-communication capable devices of person X").

A dedicated service is responsible for interpreting the friendly name and returning a GUID (or a set of GUIDs in case when friendly name translates into a set of resources/entities) (1). The GUID is an internal identifier, uniquely identifying resource/destination within a network, but still not directly connected with its precise network location.

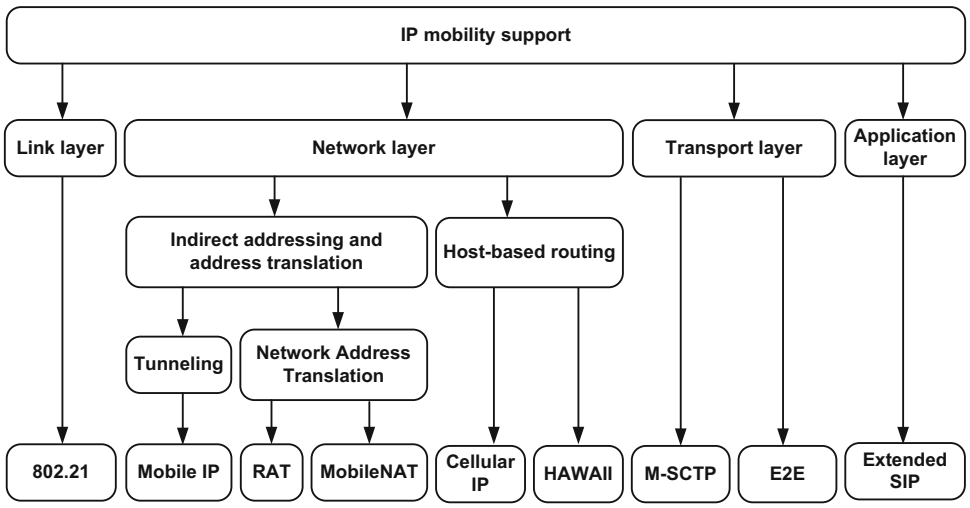

Fig. 10 Classification of unicast-based IP mobility approaches

However, GUIDs are constructed in such a way that they fulfill a number of additional functions—being, for example, a public key of a given entity and being able to indicate a network region (for example, an Autonomous System) where more precise information about this entity can be found.

GUIDs are then mapped to routable network addresses of data transport network, which are used to deliver data to its intended network destination. To facilitate the process in large network, the precise information about the entity current location is available only in selected network areas indicated by a GUID (for example, Autonomous Systems) most probable for a node to be present in. It should also be noted that Content Aware Network mechanisms are proposed at this layer, which allow accessing resources available in multiple points of the network in an efficient manner (by mapping GUID to the best of possible network addresses).

Transport network addressing information (for example IP address) can then be added to the packet header (2). Changing network address during handover between attachment points (3, 5) causes data delivery failure (4) at a network router, which then initiates *a late binding procedure* to dynamically resolve the destination GUID to a new network address (6), while concurrently buffering the incoming data in order to prevent its loss. Having obtained the updated GUID to IP binding, the router resumes transmission along the modified path (7).

7 IP mobility protocols—comparison

The IP mobility protocols, described in the previous section, address different aspects of mobility, which can be used for classification of mobility management concepts and solutions. One such classification, based on the mobility execution layer, i.e., the layer offering the essential mobility support, is presented in Fig. 10.

Mobility protocols that operate in the data link layer are typically designed only for a particular underlying protocol, but can provide better performance over the generic solutions. The IEEE 802.21 protocol is an exception from this rule, being located between the link and network layers and providing some support for IP mobility [44]. The network layer protocols are divided into addressing (or mapping)-based and host-based (with host-based routing). The protocols from the first group incorporate different techniques to obtain mobility via IP address modification. The host-based protocols manage mobility by managing route table entry for a specific host in the traditional routing infrastructure. Host-based protocols introduce smaller changes to the current network architecture at a cost of limited functionality. The addressing-based protocols can in turn be split into two main categories, utilizing tunneling or address translation.

The last group consists of application-layer solutions that are typically employed in heterogeneous networks to allow specific applications to function over different network technologies. IP mobility protocols may also be classified by their specific features. The basic characteristics of mobility support protocols are presented in Table 4. Different protocols should be used in different scenarios, depending on the mobility and handover type. Link detection, registration type and address translation properties are strictly dependent on the protocol design. Several alternative IP mobility protocols were introduced that address deployment issues related to the Mobile IP. Reverse Address Translation (RAT) is the macro-mobility approach competitive to the Mobile IP, based on the NAT procedure. It advertises easier deployment over MIP at the cost of limited functionality (e.g., no TCP session support). Mobile NAT provides both micro- and macro-mobility support and can be deployed as a mobile IP replacement. In contrast to Mobile IP, that solution is based on NAT instead of tunneling. Proxy mobile IP also falls into that category, as it addresses the problem of MIP implementation avail-

Table 4 Comparison of IP Mobility Protocols

Protocol	Mobility type	Handover type	Link detection	Registration	Address translation
Mobile IPv6 (basic)	Macro	Hard	Router advertisement	At home agent	Encapsulation
Hierarchical mobile IPv6	Universal	Hard	Router advertisement	At mobility anchor point	Encapsulation
Proxy mobile IPv6	Universal (network-based)	Hard	Events or DNAv6	At mobile access gateway	Encapsulation
Fast handovers	Universal	Hard	Proxy router adv.	At home agent	Encapsulation
Cellular IP	Micro	Semi-soft or hard	Network specific	Route updates	No
HAWAII	Micro	Forwarding and non-forw. schemes	Network specific	Path updates	No
RAT	Macro	Hard (no TCP sup.)	Network specific	At registration server	NAT
MobileNAT	Universal	Hard	Using DHCP ext.	At home NAT	NA(P)T
Extended SIP	Macro	Hard (no TCP sup)	Network specific	At SIP router	via SIP server

Fig. 11 Example classification of micro-mobility solutions

ability for the different types of mobile hosts by locating all necessary mechanisms at the network side.

The protocols can be also categorized by the optimization they introduce. The protocols optimized for routing or topology strive to limit the complexity of the architecture. The handover optimized protocols are designed to limit the delays introduced when registration point changes. A protocol may also be optimized for deployment in an existing network.

Another, probably the most popular division of IP mobility management protocols refers to two main streams: local-scale (or micro-) management versus global-scale (macro-) management mobility proposals. The first group of protocols can be further split into a number of categories, presented in Fig. 11.

Most of the well-developed, ready-to-deploy standards for mobility support in IP networks are network layer-based solutions. Included in this group are both client- and network-side mechanisms, such as MIP or PMIP, along with their multiple extensions and optimizations. They can be utilized by all protocols and applications residing above the network layer, which makes them fairly universal, as far as their usage is concerned. On the other hand, they require network layer mechanisms to be modified and/or extended, resulting in severe deployment problems.

Depending on a particular deployment scenario, client- or network-side solution may be preferable. Network-side solutions allow client device to remain unmodified, but require extensive and widespread modification of the network infrastructure.

On the other hand, client-side approach requires mobility support to be included in end-user devices, without the need for extensive network-side support. Recent introduction of general-purpose operating systems into mobile devices and their resulting unification make such approach practical. It should be noted, however, that strictly client-side solutions

are rare, as most proposals require at least one element (for example, a registration server, home agent etc.) within the network.

The network-side approach allows a network operator to efficiently provide mobility support to all of its users (due to transparent support for all client devices), but only within its administrative domain. Moreover the costs of modifying the network infrastructure can be high.

A mobile device implementing client-side mobility support such as MIP, will retain it regardless of the network administrative domain in which it currently resides. Unfortunately, the client-side approach tends to be somewhat less efficient in terms of network resource utilization than network-side solutions.

Due to the described difficulties in implementing and deploying network-layer IP mobility support, a number of higher-layer solutions have been proposed. They by definition do not match network-layer solutions' universality, but are much easier to deploy. All of them require client-side modifications to function, as layers in which they operate can be absent within the communication network itself (for example, a specific application-layer solution or transport-layer TCP mechanisms). Examples of such solutions include a limited number of ISO–OSI layer 4-6 proposals, and a number of application-layer products. In general, higher-layer solutions promote ease of deployment while limiting the scope of the solution from transport-layer mechanisms (able to provide mobility support for TCP connections between mobility-aware hosts), to applications specifically designed and implemented to function in a mobile environment, supported by a control protocol such as SIP.

It is worthy of note that except strictly network- or client-side proposals, all of the mentioned mobility support solutions, at some point require a service discovery mechanism to help them locate other elements of the employed mobility solution or obtain configuration parameters. It is a common practice to employ the Dynamic DNS service for this role, taking advantage of its high popularity and compatibility.

In recent years, some new proposals have included an additional layer in the protocol stack specifically to deal with mobility. Most prominent examples comprise HIP and LISP Mobility, which place these new elements directly above the network layer ("layer 3.5"). Their common approach is to employ a new identifier, independent of a node's network location or routing structure, in place of a mobile node's home address. This identifier is then mapped by mobility mechanisms to the appropriate network-layer address called locator address. Such an approach allows the identifiers to be assigned according to various needs, without limitations caused by network-layer mechanisms. Proposed applications of this ability include: security (public keys as identifiers), content-aware networks (routing to resources rather than network nodes), high-level service integration (addressing services instead of elements of infrastructure), etc. Another advantage of such an abstraction layer is the ability to resolve addresses across various network layer technologies (IPv4/IPv6/NonIP) or substitute another communication network without modifying higher-layer mechanisms, which can be a tremendous advantage in the case of technology migration.

Most recent proposals, mainly related to various Future Internet initiatives, propose to include mobility support as an inherent element of a network protocol stack. One such example is MobilityFirst, similar to the "addressing abstraction layer" solutions mentioned above, but taking advantage of the fact that all network nodes and devices are compelled to include mobility support mechanisms.

8 Conclusions

From the above analysis, is seems evident that despite of extensive research and development activities concerning mobility protocols, there is still a need for an universal solution able to meet the demands of users and applications.

A number of ideas that have been proposed so far all have some inevitable limitations as they are still based on the fixed-host assumption inherent in the original Internet. Therefore, Future Internet initiatives, which are likely to incorporate revolutionary changes concerning interworking solutions, require a more efficient architecture for mobility-oriented environment.

The initial development of mobility management solutions proceeded relatively slowly (see Fig. 12). More than 10 years have passed since specifying the first MIP solutions, to the point where one could see the first implementation suitable for use on large-scale production systems. The vast majority of that time was devoted to theoretical research and development of effective optimizations of its operation, but still difficult for practical implementation. It was not until the relatively recent development of mobility management protocols at the network layer runs in conjunction with the development of practical network systems. Mainly due to serious implementation difficulties, the need has forced the introduction of application layer mobility management solutions. Their development is characterized by the fast design and immediate implementation of a relatively large number of solutions dedicated to particular services (or even specific implementations of services).

Currently observed convergence of systems and networks leads to standardization of used protocols and mechanisms. Mobility management based on IP protocols allows for the integration of different, often heterogeneous systems and networks, so closely fits to the above philosophy. In conjunction with possibility of co-operation of network layer protocols with multiple lower layer transmission techniques and the

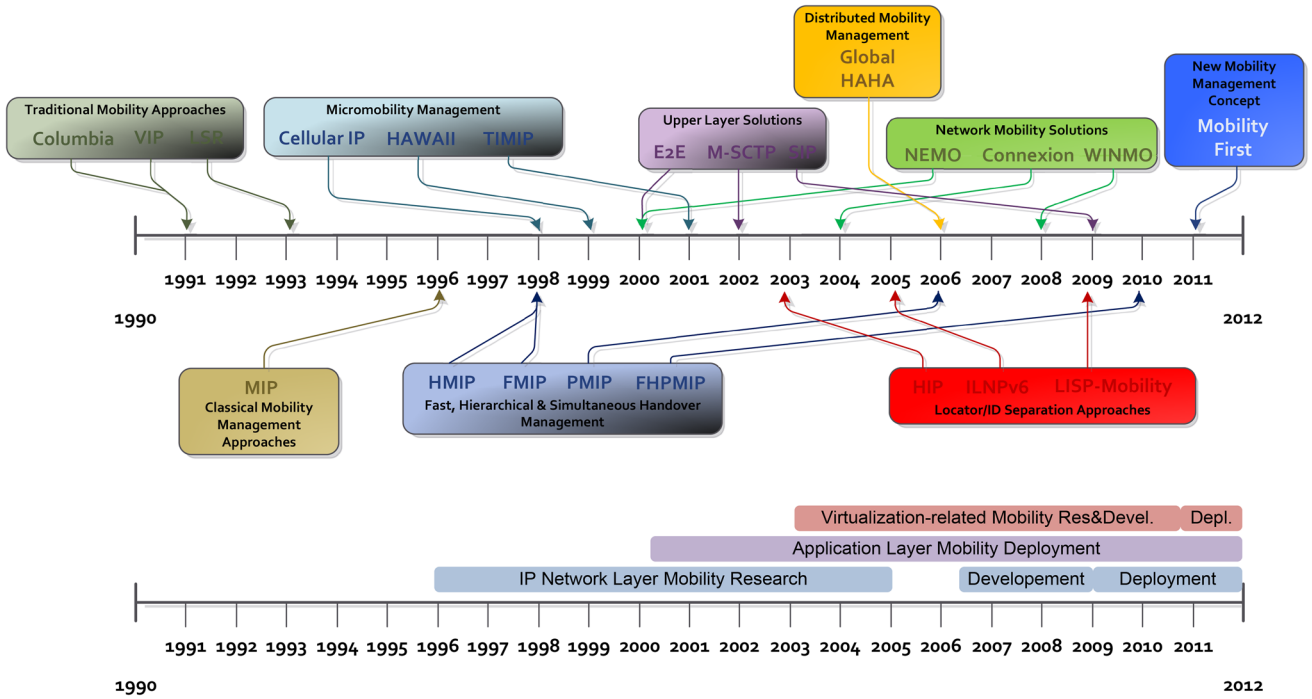

Fig. 12 Development (research, implementation, and deployment) diverse groups of mobility management protocols

lack of restrictions on the ways of realization of services in higher layers, the network layer is a natural place for locating the mobility management mechanisms.

The direction of development of various techniques used in computer networks shows that in the near future the IP protocol will still remain a homogeneous network layer protocol used in different networking environments. In the paper we presented, among others, two main groups of IP layer mobility protocols—mobile IP and Proxy Mobile IP. However, they show significant limitations due to the lack of distinction between the identifier and mobile terminal locator offered via IP protocol, which is reflected in the week efficiency in terms of movement and switching. We also presented ways to optimize these protocols and improve mechanisms handling mobile terminals.

The latest trends comprise a variety of solutions based on virtualization techniques, especially aiming at the separation of the recipient's identification from its location in the network and separation of higher layers operation from the structure of the so-called transport network.

This trend is characterized by a parallel R&D projects and implementations in small network environments. The current results allow to consider possible deployments of Internet-scale global systems.

Sample proposals for such architectures—HIP, LISP-Mobility and MobilityFirst are described in the final part of the paper, as solutions able to manage mobile users on the Future Internet.

However, despite the advances in research, development and deployment of network-layer mobility management solutions, it is still necessary to remember the requirement of providing effective support for low-layer handover procedures, which will guarantee shorter transmission breaks and smaller distortions while switching between points of network attachement.

Acknowledgments The author expresses sincere thanks to colleagues: Krzysztof Gierłowski, Michael Hoeft, Jerzy Konorski and Jacek Rak for valuable comments and assistance in preparation of this article.

References

1. Cisco Visual Networking Index (VNI). (2011). Global mobile data traffic dorecast, 2011–2016, February 2012.
2. Lin, T. M., Lee, B.-S., Yeo, C. K., Tantra, J. W., & Xia, Y. (2009). A terminal-assisted route optimized NEMO management. *Telecommunication Systems*, *42*, 263–272.
3. Perkins, C. (1996). IP mobility support, RFC 2002 work in progress, October.
4. Perkins, C. (2002). IP mobility support for IPv4, RFC 3344, August.
5. Perkins, C. (2010). IP mobility support for IPv4, revised, RFC 5944, November.

6. Johnson, D., Perkins, C., & Arkko, J. (2004). Mobility support in IPv6, RFC 3775, June.
7. Li, Q., Jinmei, T., & Shima, K. (2009). *Mobile IPv6: Protocols and implementation*. San Francisco: Morgan Kaufman.
8. Chan, H., et al. (2011). Problem statement for distributed and dynamic mobility management. IETF Internet-Draft, draft-chan-distributed-mobility-ps-04, October.
9. Wakikawa, R., Valadon, G., & Murai, J. (2006). Migrating home agents towards internet-scale mobility deployment. ACM CoNEXT.
10. Jung, H., & Koh, S. (2011). Distributed mobility control in proxy mobile IPv6 networks. *IEICE Transactions on Communications*, *E94–B*(8), 2216–2224.
11. Moskowitz, R., Nikander, P., Jokela, P., & Henderson, T. (2008). Host identity protocol. IETF RFC 5201, April.
12. European Future Internet Portal, http://www.future-internet.eu/home/future-internet-assembly.html. Accessed 10 Nov 2012
13. Seskar, I., Nagaraja, K., Nelson, S., & Raychaudhuri, D. (2011). MobilityFirst future internet architecture project. In *MobilityFirst project proc. ACM AINTec* (pp. 1–3).
14. MobilityFirst FIA Project Overview, http://mobilityfirst.winlab.rutgers.edu/. Accessed 10 Nov 2012.
15. Zhu, Z., Wakikawa, R., & Zhang, L. (2011). A survey of mobility support in the internet. RFC6301—Informational, July.
16. Akyildiz, I. F., Xie, J., & Mohanty, S. (2004). A survey of mobility management in next-generation all-IP-based wireless systems. *IEEE Wireless Communications*, *11*, 16–28.
17. Le, D., Fu, X., & Hogrefe, D. (2005). A review of mobility support paradigms for the internet. Tech. Rep. Georg-August-Universität Göttingen, Institut für Informatik.
18. Campbell, A. T., & Gomez-Castellanos, J. (2000). IP micro-mobility protocols. *Mobile Computing and Communications Review*, *4*(4), 45–53.
19. Homepage of Mobile Oriented Future Internet (MOFI). http://www.mofi.re.kr. Accessed 10 Nov 2012.
20. Jung, H., & Koh, S. (2012). New inter-networking architecture for mobile oriented internet environment. In *Proc.future network & mobile summit'12*.
21. Wozniak, J., et al. (2011). Comparative analysis of IP-based mobility protocols and fast handover algorithms in IEEE 802.11 based WLANS. *CN2011, CCIS 160*. Berlin: Springer.
22. Ramani, I., & Savage, S. (2005). SyncScan: Practical fast handoff for 802.11 infrastructure networks. In *Proceedings of IEEE Infocom* (pp. 675–684).
23. IEEE Std. (2008). IEEE 802.11r-2008—IEEE Standard for Information technology—Local and metropolitan area networks—Specific requirements—Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 2: Fast Basic Service Set (BSS) Transition. http://standards.ieee.org/findstds/standard/802.11r-2008.html. Accessed 10 Nov 2012.
24. Campbell, A., et al. (1999). Cellular IP. *Internet Draft, draft-ietf-mobileip-cellularip-00*, Work in progress, December.
25. Ramjee, R., Varadhan, K., & Salgarelli, L. (2002). HAWAII: A domain-based approach for supporting mobility in wide-area wireless networks. *IEEE/ACM Transcations on Networking*, *10*, 396–410.
26. Gundavelli, S., Leung, K., Devarapalli, V., Chowdhury, K., & Patil, B. (2008). *Proxy mobile IPv6*. RFC 5213, August.
27. Machan, P. (2012). Effectiveness of IP traffic handover in IEEE 802.11 wireless local area networks. *Ph.D. Dissertation*, GUT, Gdansk.
28. El Malki, K. (Ed.). (2007). *Low latency handoffs in mobile IPv4*. RFC 4881, June.
29. Koodli, R., & Perkins, C. (2007). Mobile IPv4 fast handovers. RFC 4988, October.
30. Koodli, R. (Ed.). (2009). Mobile IPv6 fast handovers. RFC 5568, IETF Proposed Standard, July.
31. Soliman, H., Castelluccia, C., El Malki, K., & Bellier, L. (2008). Hierarchical mobile IPv6 (HMIPv6) mobility management. RFC 5380, October.
32. Chen, W.-M., Chen, W., & Chao, H.-C. (2009). An efficient mobile IPv6 handover scheme. *Telecommunication Systems*, *42*, 293–304.
33. Singh, R., Teo, Y.C., & Yeow, S.W. (1999). RAT: A quick (and dirty) push for mobility support. In *2nd IEEE workshop on mobile computer systems and applications*, February.
34. Buddhikot, M., Hari, A., Singh, K., & Miller, S. (2005). MobileNAT: A new technique for mobility across heterogeneous address spaces. *Mobile Networks and Applications (MONET)*, *10*(3), 289–302.
35. Snoeren, A., & Balakrishnan, H. (2000). End-to-end approach to host mobility,. In *ACM Mobicom* (pp. 155–166).
36. Wolisz, & Muller, H. (2002). M-SCTP: Design and prototypical implementation of an end-to-end mobility concept. In *Proc. 5th Int. workshop on the internet challenge* (pp. 1–8).
37. Crocer, D. (2003). Multiple address service for transport (MAST). An extended proposal, September.
38. Eronen, P. (2006). IKEv2 mobility and multihoming protocol (MOBIKE). RFC 4555, June.
39. Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., Handley, M., & Schooler, E. (2002). SIP: Session initiation protocol. RFC 3261, IETF Proposed Standard, June.
40. Schulzrinne, H., & Wedlund, E. (2010). Application-layer mobility using SIP. *Mobile Computing and Communications Review*, *4*(3), 47–57.
41. Kempf, J. (2007). Problem statement for network-based localized mobility management (NETLMM). IETF RFC 4830, April.
42. Farinacci, D., Fuller, V., Meyer, D., & Lewis, D. (2009). Locator/ID separation protocol (LISP), work in progress, March.
43. Atkinson, R., Bhatti, S., & Hailes, S. (2009). ILNP: Mobility, multi-homing, localised addressing and security through naming. *Telecommunication Systems*, *42*, 273–291.
44. Machan, P., & Wozniak, J. (2010). Simultaneous handover scheme for IEEE 802.11 WLANs with IEEE 802.21 triggers. *Telecommunication Systems*, *43*, 83–93.

Jozef Wozniak holding PhD and D.Sc. degrees in Telecommunications from Gdansk University of Technology (GUT) is a Full Professor at the Faculty of Electronics, Telecommunications and Computer Science of GUT. His research activity includes a number of research projects, over 250 journal/conference papers, and four books on computer networks, communication protocols, and data communications. He participated in various research and teaching activities, including visits at Vrije Universiteit Brussel, Politecnico di Milano, and Aalborg University, Denmark. In 2006 he was invited to Canterbury University in Christchurch, New Zealand as a Visiting Erskine Fellow. He has been a TPC member of a number of national/international conferences, also chairing or co-chairing several of them. He is a senior member of IEEE, and the chair of Working Group 6.8 (Wireless and Mobile Communications) IFIP TC6. His current research is focused on modeling and performance evaluation of wireless and mobile communication networks.