An Overview of Macro-Mobility Management in Next Generation All-IP Based Wireless Network Infrastructures

Mate Boban
University of Zagreb,
Faculty of Organization and Informatics, Varaždin, Croatia
mboban@foi.hr

Abstract: *Next generation wireless systems are most likely to have an IP based core network. Mobile IP is a proposed standard of the IETF (Internet Engineering Task Force) designed to assist different mobile users. Mobile IP provides an elegant solution for node mobility when the mobile node infrequently moves. However, it implies a relatively large overhead and delay, due to indirect communications. In this paper, three macro-mobility management architectures based on different layers are reviewed, first of which uses Mobile IP to manage mobility for non-real-time communications, the second is based solely on SIP, while the third is based on a new solution introducing third party to manage macro-mobility. Finally, the most suitable architecture for macro-mobility management is proposed*
Keywords: *All-IP networks, Wireless networks, macro-mobility, IP mobility*

1. INTRODUCTION

Over the past several years a number of IP mobility management protocols have been proposed. Mobility Management Framework proposed in [11], distinguishes 2 categories of device (terminal) mobility management solution, namely localized mobility (micro-mobility, intra-domain mobility) and macro-mobility (inter-domain mobility). The movement of devices between two subnets within one domain is referred to as localized (intra-domain, micro) mobility; for example, the movement of MD (Mobile Device) from subnet A to subnet B shown in Fig 1. Localized mobility in UTRAN networks is device movement from one BS (base station) to another, both BSs belonging to the same RAN (Radio Access Network), while in WLAN it is device movement between two APs (Access Points) of the same network. The movement of devices between two network domains is referred to as macro-mobility (inter-domain mobility); for example, the movement of MD from domain 1 to domain 2 shown in Fig. 1. One domain represents an administrative body, which may include different access networks, such as WLAN, second-generation (2G), and third-generation (3G) networks [2]. However, each of these networks can alone represent one domain, which basically is the case with most of the existing networks today. NG (Next Generation) all-IP based wireless network will include various heterogeneous networks [13], each of them using possibly different access technologies. Therefore, satisfactory macro-mobility solution supporting all of the access technologies is needed.

In this paper, three most promising macro-mobility management solutions are reviewed: SIP-based (Session Initiation Protocol, [8]) approach introduced in [10], hybrid SIP/MIP (Mobile IP) approach introduced in [5] and [10], and finally, macro-mobility management solution based on AMC (Architecture for Ubiquitous Mobile Communications) introduced in [1]. The first two solutions are based on existing protocols (SIP and MIP) and network entities, while the third solution is yet to be implemented in novel network entities running some of the existing protocols and some yet to be realized protocols. However, the localized mobility management in all of the presented solutions is performed using some of the existing protocols.

The rest of the paper is structured as follows: Section 2 explains why there is a need for macro-mobility solution other than Mobile IP, the proposed standard of the Internet Engineering Task Force (IETF). Section 3 presents a brief overview of two localized mobility management solutions compatible with reviewed macro-mobility solutions, one of which is routing-based and the other is tunnel-based. Namely, CIP (Cellular IP) [12] and Mobile IP Regional Registration [4] are presented. Section 4 explains SIP-based, hybrid SIP/MIP and AMC macro-mobility management solutions, respectively, and provides their qualitative comparison. Finally, in section 5 the most suitable solution for supporting macro-mobility in heterogeneous IP-based infrastructures is proposed.

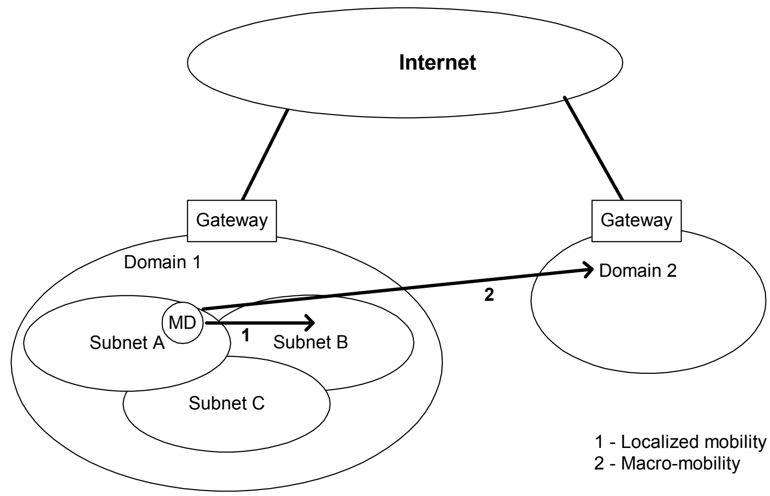

Figure 1. Localized mobility and macro-mobility

2. MOBILE IP (MIP)

Mobile IP (MIP) [2] is a standard proposed by a working group within the Internet Engineering Task Force (IETF). It is designed to solve the problem of MD mobility outside the home network (domain). While roaming, MD maintains two IP addresses, a home IP address assigned to it by its home network, the home agent (HA) specifically, which is used in all transport layer connections, and care-of address (CoA) when visiting a foreign network (domain). The CoA reflects the current point of attachment of the MD, and is obtained through a foreign agent (FA) of the foreign network or auto-configuration process. HA is in charge of maintaining a connection between MD's home IP address and CoA. MD sends registration requests to inform the HA of any change in CoA or to renew a mobility binding. As shown in Fig. 2, all packets sent by correspondent device (CD) to MD are received by HA (step 1 Fig. 2), which employs tunneling to re-direct packets to the MDs current location characterized by the CoA (step 2 Fig. 2). The FA of the foreign network receives the packets sent to MD, decapsulates the packets and forwards them to the MD (step 3 Fig 2.). With a collocated CoA, the encapsulated packets reach the MD, which then decapsulates them. While all the packets from CD to MD had to be sent indirectly, MD sends the packets directly to the CDs home address (step 3 Fig 2.). This is also referred to as triangular routing and is considered, as well as tunneling, one of the biggest problems of macro-mobility based on MIP. It is intuitively visible that MIP provides an elegant solution for node mobility when the mobile node infrequently moves. However, MIP suffers from indirect communications, which increase the delay and cause an overhead due to tunneling, which decreases the bandwidth utilization. Therefore, new solutions have been designed.

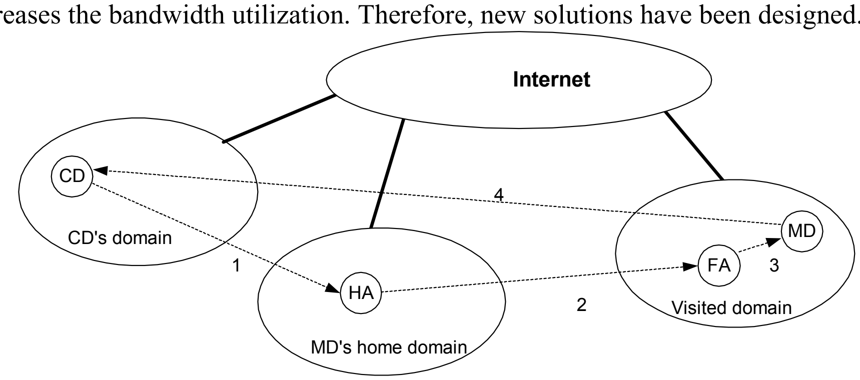

Figure 2. Mobile IP

3. LOCALIZED MOBILITY SOLUTIONS

As mentioned above, localized mobility implies device movement between two subnets within one domain. Note that Reinbold et al. [7], as well as many others, refer to localized mobility as micro-mobility, but the term micro-mobility can be misleading in this context. A localized mobility solution may actually provide mobility for large geographical areas [11].
To reduce signaling load and delay to the home network during movements within one domain, many localized mobility solutions have been proposed. They can be broadly classified into two groups: tunnel-based and routing-based localized mobility schemes [2]:

- Tunnel-based schemes use local or hierarchical registration and encapsulation concepts to limit the scope of mobility-related signaling messages, thus reducing the global signaling load and handoff latency. Mobile IP regional registration (MIP-RR) [4], hierarchical Mobile IP (HMIP) [9] and TeleMIP [3] are tunnel-based micro-mobility protocols.
- Routing-based schemes maintain host-specific routes in the routers to forward packets. The host-specific routes are updated based on host mobility. Cellular IP (CIP) [12] and Handoff-Aware Wireless Access Internet Infrastructure (HAWAII) [6] are routing-based localized mobility protocols.

3.1 MOBILE IP REGIONAL REGISTRATION

MIP-RR [4] introduces the concept of gateway foreign agent (GFA), whose purpose is to reduce signaling delay by performing registrations locally inside one network (domain): when an MD first arrives at foreign network, it sends registration request to it's HA. During the registration, the HA registers the CoA of the MD with the GFA of the foreign network (in MIP, the HA would register the CoA of the MD with the FA). When an MN changes FAs within the same foreign network, it performs only a regional registration to the GFA to update its CoA, without contacting it's HA (this is possible because GFA manages MDs movement between the FAs of the same foreign network), thus reducing the overall signaling. However, when an MD moves from one foreign network to another, it has to perform regional registration with it's HA. The packets for the MD are first intercepted by it's HA, which tunnels them to the registered GFA. The GFA checks its visitor list and forwards the packets to the corresponding FA of the MD. The FA further forwards the packets to the MD. In this way, the GFA introduces a layer of hierarchy between the HA and the FA of the MD. The use of the GFA avoids any signaling traffic to the HA as long as the MD is within the same foreign network.

3.2 CELLULAR IP

The design of CIP is based on four fundamental principles [2]:

- Distributed caches are used to store location information of MDs.
- Location information of an active MD is updated by regular IP datagrams originated by itself. For an idle MD, this is achieved by the use of dummy packets that are sent by the idle host at regular intervals.
- Location information is stored as soft states.
- Location management for idle MDs is separated from location management of MDs that are actively transmitting/receiving data.

Architecture of CIP is shown in Fig. 3. Each access network (domain) is connected to the Internet through a gateway, which handles the mobility within one access network. Packets sent to MD are received by gateway first. Gateway forwards the packets to MD using host-specific routing path. Important concepts of CIP are *distributed paging cache* and *distributed routing cache* for location management and routing, respectively. CIP supports power saving mode for the MDs using distributed paging cache, which maintains the position of the passive MDs for efficient paging, whereas the routing cache maintains the position of an active MD up to subnet level accuracy. This means that CIP manages passive and active MDs separately. When active MD changes access networks (domains), the routing states in the routing cache are dynamically updated. The handoff process of CIP is automatic and transparent to the upper layers. When the strength of the beacon signal from the serving BS (base station) is lower than that of a neighboring BS, the MD initiates a handoff. The first packet that travels to the GW through the new BS configures a new path through the new BS, while keeping the old path, thus using bicasting of the traffic through old and new path to reduce the packet loss. In order to reduce a potential synchronization problem between the packet flows coming through the two paths, the CIP devices must implement an additional delay device. This device delays the packets transferred through the new path during the handoff. If the MD is capable to listen to both flows at the same time, the handoff is soft. Otherwise, hard handoff is performed.

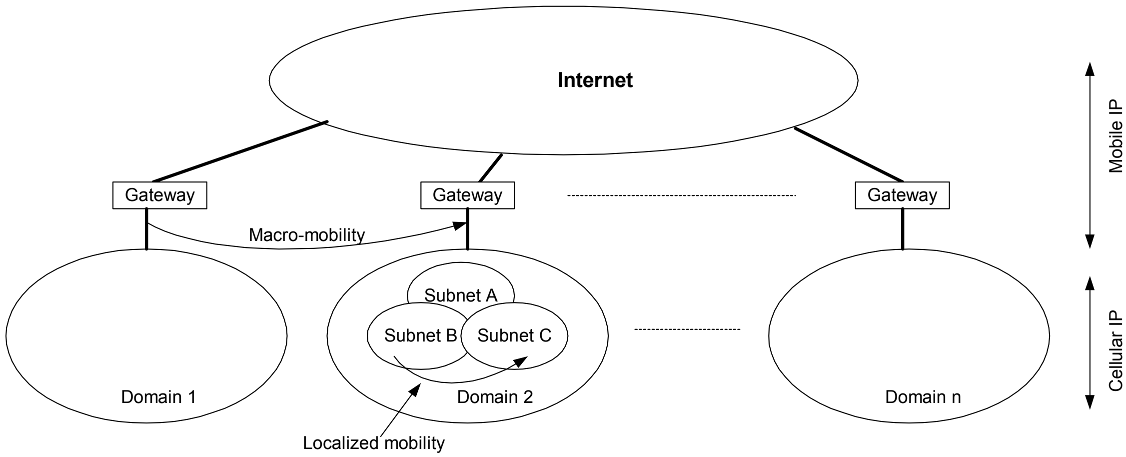

Figure 3. Cellular IP

4. MACRO MOBILITY SOLUTIONS FOR THE NEXT GENERATION ALL IP NETWORKS

This paper considers three macro-mobility solutions for the support of the next generation all IP networks. All three solutions consider the handling of the localized mobility through the use of an IP localized mobility protocol (e.g., Cellular IP, HAWAII and MIP-RR). The first solution is SIP-based, proposed in [10]. The second is Hybrid SIP/Mobile IP solution (Fig. 5) explained in [5] and [10], and the third solution is AMC (Architecture for Ubiquitous Mobile Communications) introduced in [1].

4.1 SIP-BASED SOLUTION

This solution (Fig. 4) is based on the use of SIP protocol to handle macro-mobility (inter-domain mobility). The mobile hosts are identified with their private home addresses within the localized mobility areas. This provides location privacy and application transparency, while the mobile host roams inside a visited (network) domain. In order to provide efficient interworking between SIP and localized mobility management solution, a new type of control information in localized mobility management route/paging update packets is used: the SIP user identifier, an email-like address of the form “user@host”, where “user” is a user name and “host” is a domain name or numerical address. User identifier is inserted in the payload of the first route-update packet after inter domain handoff and may be repeated in a few subsequent route/paging-update packets for reliability. Upon receiving the first route-update packet, the gateway of the visited domain performs local registration to register the corresponding host in its caches.

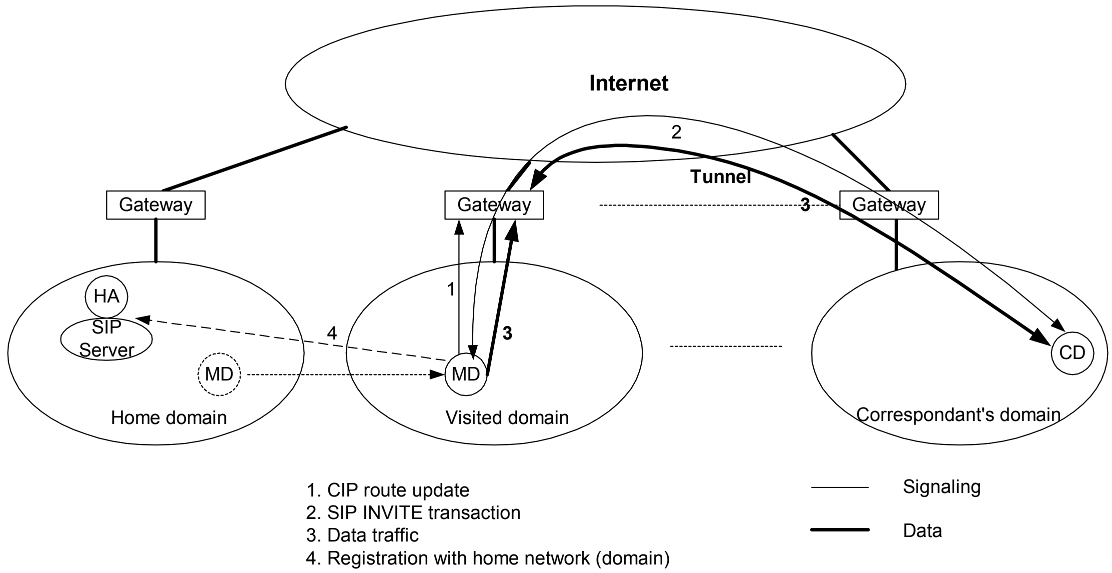

Figure 4. SIP-based macro-mobility

While MD moves during session, SIP user agent (UA) sends a SIP re-INVITE request message to each of its corresponding devices (CDs). In this message original SIP user identifier into the form field of the SIP header is included. It also includes the corresponding gateway's address into the contact field of the SIP header, in order to inform the CD where it wants to receive future SIP messages, and it also includes it into c (connection) field of the header that contains a description of the session, in order to redirect the data traffic flow towards its new location.
SIP responses destined to MD are first received by the visited network's gateway. Gateway is equipped with a SIP message tracking-agent in order to forward the SIP responses to their original destination. This agent checks the gateway's binding caches to determine whether a SIP response message must be forwarded towards a registered MD. This forwarding is accomplished after the destination IP address has been modified. This assures that the SIP re-INVITE transaction is correctly completed upon handoff. A CD after receiving the SIP INVITE message uses IP encapsulation to forward data to the MD. The encapsulated data packets are captured by the visited domain's gateway, which decapsulates and forwards them to the MD using localized mobility routing. It is important to mention that data traffic from the MD is routed without the use of tunneling. The MD completes the handoff by sending a SIP REGISTER message towards a SIP server on its home network (Home Registration).

4.2 HYBRID SIP/MOBILE IP SOLUTION

Macro-mobility (intra-domain mobility) in this solution is based on SIP signaling to support macro-mobility for real-time traffic, while Mobile IP is used to support non-real-time traffic, as shown in Fig. 5. Traffic to/from MD is separated on the domain edge router of the visited network. As mentioned, real-time traffic is supported by SIP through network address translation (NAT). NAT is required on domain edge routers, since MDs are identified by their private home addresses within one domain. By using NAT, IP encapsulation on the devices is avoided. As mentioned before, non-real-time traffic is supported by Mobile IP in a way described in Section 2 and shown in Fig. 2. Triangular routing and tunneling are employed. Data packets from CD are received by MD's HA, which encapsulates them and sends them to the gateway of the visited network in which MD is located. The FA in the visited network, located on the gateway, decapsulates data packets and forwards them to the MD. Data packets sent from MD to CD are sent directly to CD, first crossing the gateway of the visited network, of course. Tunneling and IP encapsulation are not so critical to non-real-time traffic. However, real-time traffic is supported by SIP-based approach, as described in the previous subsection.

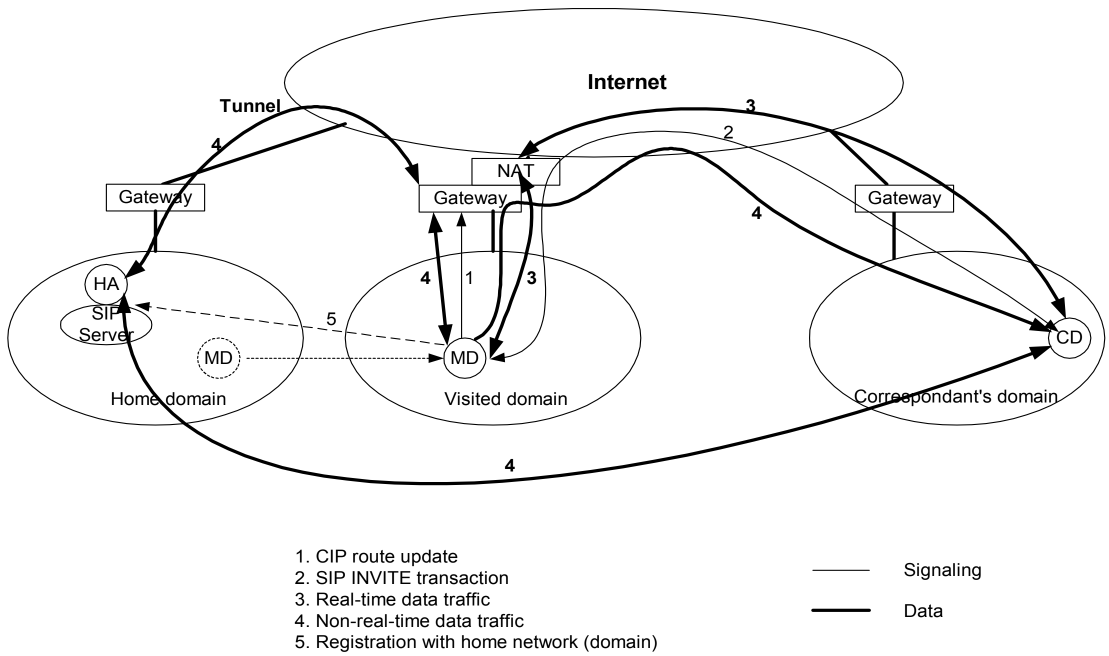

Figure 5. Hybrid SIP/Mobile IP macro-mobility

4.3 AMC (ARCHITECTURE FOR UBIQUITOUS MOBILE COMMUNICATIONS)

As proposed in [1], AMC is a next-generation macro-mobility management solution which intelligently integrates the existing wireless systems so that users may receive their services via the best available wireless system. This implies that each operator domain shown in Fig. 6. can be one of the existing wireless systems (e.g., Bluetooth, IEEE 802.11, UMTS, or satellite network). AMC is a solution that does not interfere with either one of the localized mobility management solutions. On contrary, AMC is designed to inter-operate with any one of them, and is destined to work with future localized mobility management solutions as well. AMC integrates heterogeneous wireless systems using a third party, Network Inter-operating Agent (NIA). Thus, it eliminates the need for direct (bilateral) SLAs (service level agreements) among different network operators, a policy problem which causes a lot of difficulties in roaming between operators. If a third party doesn't exist, an operator has to negotiate SLA with every other operator he wants to enable roaming with. AMC achieves transparency to the heterogeneities of individual systems by using Internet Protocol (IP) as the inter-connection protocol. In addition, AMC proposes algorithms for the best network selection. Besides NIA which resides in the Internet, another entity is defined by AMC, namely, the Interworking Gateway (IG). The IG, which resides in each operator domain and acts as a gateway (as shown in Fig. 6.) is connected to only one entity, the NIA, instead of being connected to every other domain (it would have to be connected to every other domain if only bilateral SLAs would have been negotiated). This means that each operator has to create SLA only with NIA, which is in charge of interworking (authentication, authorization, billing and macro-mobility management issues) between the operators. IG can be implemented as a separate entity or integrated with the gateways through which individual systems are connected to the Internet; e.g., PDSN (packet data serving node), GGSN (gateway GPRS support node), GS (gateway station), and AR (access router), for cdma2000, GPRS, satellite, and WLAN, respectively. The architecture of NIA and IG is thoroughly explained in [1]. As mentioned above, AMC implements algorithms to determine the best available network. The criteria for the selection of the best network can be monetary cost, network conditions, power consumption, user activity history or the required QoS from applications. Moreover, the best network selection also affects the distribution of the overall system load. After the best available network is selected, the next challenge is to determine the right time to start handoff procedures. A dynamic RSS (received signal strength) threshold is proposed to eliminate the effect of signaling delay variation. AMC also predicts handoff signaling delay and the speed of MDs in advance using physical and MAC layer sensing to determine the appropriate time for handoff initiation.

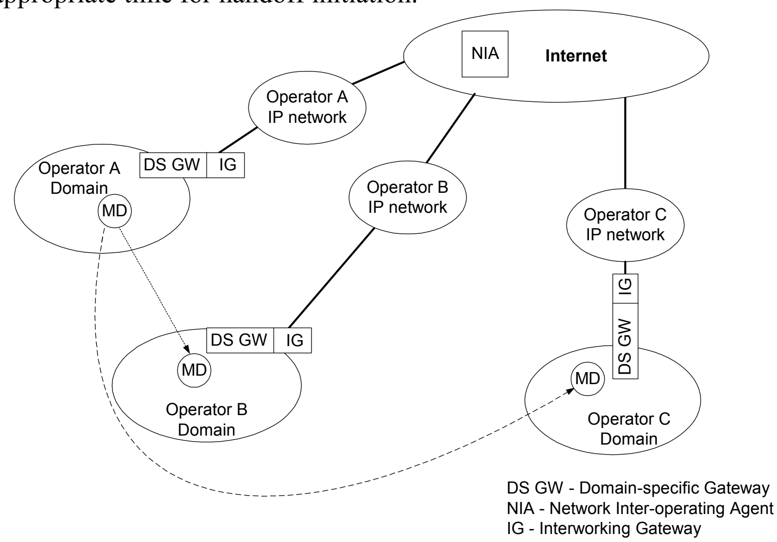

Figure 6. AMC (Architecture for Ubiquitous Mobile Communications)

4.3 COMPARISON

Table 1 illustrates the most important characteristics of the reviewed solutions against Mobile IP. Same as Mobile IP, all of the reviewed solutions support transparent device mobility. Unlike Mobile IP, all of them support seamless handoff and paging. As mentioned in [10], SIP-based macro-mobility management solution requires modifications in the IP stack of end hosts to support IP-in-IP encapsulation, and thus it will experience problems concerning wide acceptance. Concerning single point of failure, all of the novel macro-mobility management solutions don't have single point of failure in terms of *device failure.* However, AMC has single point of failure in terms of *third party* (if third party owning NIA experiences any kind of policy problems, such as security breach, failure could happen). This is why there will be a lot of work in finding appropriate NIA managing organization. Also, besides its advance handoff support and fully overlapping or partially overlapping network selection, AMC proposes Mobile IP to handle MD's roaming, which implies triangular routing problem. This could be quite a drawback because of the fact that MD could be far from its home network (and therefore far from it's HA, e.g. if MD moves to another continent from its home WLAN network, using satellite overlay network), thus significantly increasing propagation and overlay delay. On the other hand, AMC supports sophisticated handoff prediction and best available network selection.

	Mobile IP	**SIP-based**	**Hybrid SIP/Mobile IP**	**AMC**
Transparent device mobility	+	+	+	+
No single point of failure	-	+	+	-
Seamless handoff support	-	+	+	+
No modifications to IP stack	+	-	+	+
Paging support	-	+	+	+
Handoff prediction	-	-	-	+
Best available network selection	-	-	-	+
No triangular routing for real-time-data	-	+	+	-

Table 1. Comparison of macro-mobility management solutions

5. CONCLUSION

This paper considers three macro-mobility management solutions for next generation all-IP network infrastructure. All of the schemes present a significant improvement of the Mobile IP, proposed standard of IETF. While SIP-based and Hybrid SIP/Mobile IP solutions rely on SIP to handle real-time data traffic, AMC proposes improvement in means of algorithms for handoff prediction and selection of best available network. Also, AMC supports network selection between fully overlapped networks, unlike two other solutions. Hybrid SIP/Mobile IP solution differentiates between real-time and non-real-time traffic, thus allowing mobility awareness on a higher above IP layer, where certain knowledge about the type of application can be exploited [10]. SIP-based solution has a serious drawback in the fact that, to be implemented, it requires modifications in the IP stack of end hosts to support IP-in-IP encapsulation. Provided that a reliable third party for managing NIA could be found, because its scalability and ability to support multiple operators interconnection without bilateral SLAs and because it requires no modifications to the existing heterogeneous radio technologies, AMC would most likely be the best solution for next generation all-IP network infrastructure.

REFERENCES

[1] Akyildiz, I.F.; Mohanty, S.; Xie, J. (2005) *A Ubiquitous Mobile Communication Architecture for Next-Generation Heterogeneous Wireless Systems*, IEEE Radio Communications Magazine, pp. 29-36, June 2005.

[2] Akyildiz, I.F.; Xie, J.; Mohanty, S. (2004) *A Survey on Mobility Management in Next Generation All-IP Based Wireless Systems,* IEEE Wireless Communications Magazine, vol. 11, no. 4, pp. 16-28, Aug. 2004.

[3] Das, S.; Misra, A.; Agrawal, P.; Das, S. K. (2000) *TeleMIP: Telecommunications-enhanced mobile IP architecture for fast intradomain mobility*, IEEE Personal Communications Magazine, Vol. 7, pp. 50–58, Aug. 2000.

[4] Gustafsson, E.; Jonsson, A.; Perkins, C. E. (2004) *Mobile IPv4 Regional Registration*, Internet draft, IETF, draftietf-mobileip-reg-tunnel-09.txt, June 2004, work in progress.

[5] Politis, C., Chew, K., Tafazolli, R. (2003) *Multilayer Mobility Management for All-IP Networks: Pure SIP vs. Hybrid SIP/Mobile IP*, IEEE VTC Spring Conference, 21-24th April 2003, Jeju, Korea

[6] Ramjee, R.; Porta, T. L.; Thuel, S.; Varadhan, K.; Wang, S. Y. (1999) *HAWAII: A domain-based approach for supporting mobility in wide-area wireless networks,* International Conference on Network Protocols (ICNP), (Toronto, Canada), Nov. 1999.

[7] Reinbold, P.; Bonaventure, O (2003) *IP micro-mobility protocols*, IEEE Communications Surveys & Tutorials, vol. 5, no. 1

[8] Schulzrinne, H; Wedlund, E. (2000) *Application-Layer Mobility using SIP*, ACM Mobile Computing and Communications Review, Vol. 4, No. 3

[9] Soliman H. et al. (2004) *Hierarchical Mobile IPv6 Mobility Management (HMIPv6)*, Internet draft, IETF, draft-ietfmipshop-hmipv6-02.txt, June 2004, work in progress.

[10] Theofilatos, D; Dagiuklas, T; Gatzounas, D. (2003) *Mobility Management Schemes for Supporting Multimedia Services in All-IP Network Infrastructures*, Proceedings of the 5th European Personal Mobile Communications Conference (EPMCC), Glasgow, Scotland

[11] Uskela, S. (2002) *Mobility Management in Mobile Internet*, Nokia Finland, http://www.nokia.com/downloads/aboutnokia/research/library/communication_systems/CS2.pdf

[12] Valk, A. G. (1999) *Cellular IP: A new approach to internet host mobility,* ACM Computer Communication Review, Vol. 29, pp. 50–65, Jan. 1999.

[13] Xie, J (2004) *Mobility Management in Next Generation All-IP Based Wireless Systems,* Georgia Institute of Technology, Atlanta Apr. 2004