



IP-based mobility management for heterogeneous wireless access

H. J. Einsiedler¹ and D. von Hugo²

¹Telekom Innovation Laboratories, Seamless Network Control, Berlin, Germany

²Telekom Innovation Laboratories, Seamless Network Control, Darmstadt, Germany

Correspondence to: D. von Hugo (dirk.von-hugo@telekom.de)

Abstract. Future high quality communication services will be offered in an integrated or converged network infrastructure maintaining both fixed wireless and mobile access via multi-mode user terminals. A support of various scenarios of user and/or terminal mobility within a common IP-based infrastructure requires intelligently designed control protocols. A major challenge is to provide seamless (i.e. lossless and low delay) handover between different radio cells and operator domains to enable continuation of unicast and multicast sessions while using network resources most efficiently. IETF (Internet Engineering Task Force) is specifying related IP mobility management protocols to be applicable also to a flat architecture as envisaged by Next Generation (Mobile) Networks (NGNs/NGMNs). The contribution will describe operator requirements towards such an approach. Both single-domain and multi-domain scenarios will be discussed based on federation ideas. Already existing solutions are taken into consideration and application of solution proposals towards a Distributed Mobility Management (DMM) currently under evaluation within IETF will be outlined.

1 Introduction

Evolution of mobile wireless technology including cellular radio access (e.g. LTE-A, Long Term Evolution Advanced system, by 3GPP) and local access (e.g. WLAN, Wireless Local Area Network, by IEEE) is heading towards higher data rates, lower transmission delay, and better efficiency to meet the observed mobile data traffic increase due to new customer services and applications as exemplary shown in Fig. 1.

Thus a multi-dimensional more “natural” communication service including real-time multimedia applications used in different environments via various user equipment terminals is possible. To intelligently cope with the resulting increasing complexity and serve the customer always in the best and

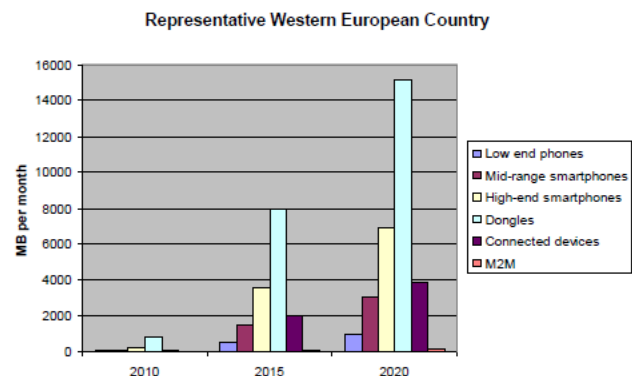


Fig. 1. Expected monthly mobile traffic per device for a representative Western European country according to UMTS-Forum (2011).

most efficient way a converged heterogeneous network infrastructure has to be established. Thus the customers will experience high quality integrated fixed and mobile services using multi-mode terminals (i.e. able to access as well 3GPP and IEEE networks). Further operator requirements are to provide high reliability and availability (in terms of user data throughput and large coverage) at affordable cost and (resource-) efficient operation.

2 Architecture of heterogeneous access infrastructure

A heterogeneous access network infrastructure to allow such a flexible service provision will rely on a common IP-based infrastructure. This concept allows a future integrated mobile as well as fixed network access as outlined by 3GPP (2011).

Challenge of such a design is to provide seamless session continuity to a user moving between mobile and fixed access assuming an IMS-based control (3GPP, 2010). A prospective

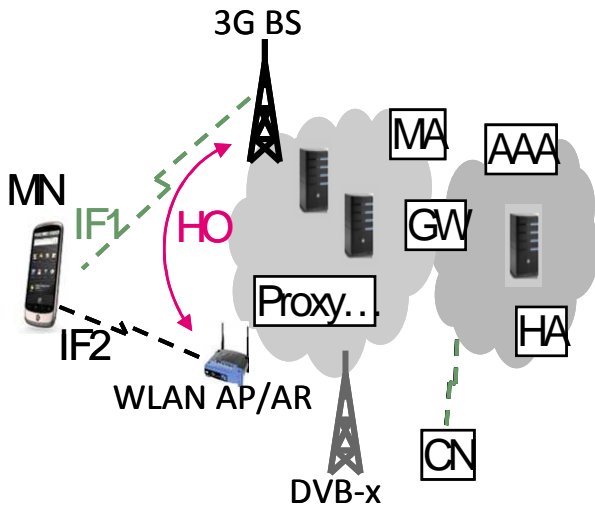


Fig. 2. Components of mobile IP transmission for illustration.

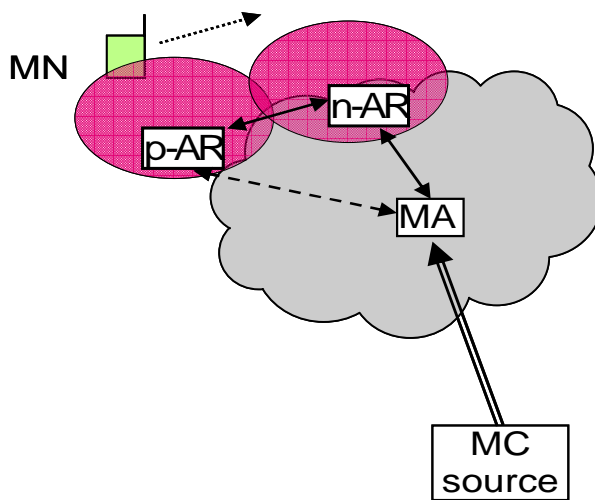


Fig. 3. Schematic view of relations and signalling paths between infrastructure entities in protocols for enhanced handover.

future-proof solution would have an optimum trade-off for both users and operators.

In any case to achieve service continuity a switching of the mobile terminal between different access nodes (points of attachment) of potentially different technologies shall be enabled by a properly designed mobility control protocol.

3 Mobility support overview

3.1 (IP) Mobility support general overview

Regarding the changing service characteristics and usage scenarios of mobile applications (e.g. with mobile subscriptions replacing residential fixed network access) a differentiation between true terminal mobility (i.e. usage “on the move”) and portability (using the service at different locations) has to be considered. In addition, a change of access point in terms of Handover (HO) may take place also without actual user movement, e.g. due to network policies, resource constraints, or changing propagation conditions. Thus an efficient handling of terminal connectivity regarding the aspect of mobility already is and in future will even more be an important network capability.

The entities or parties which are involved and have to be considered in mobility protocols are those participating in a mobile conversation. These cover first-of-all the Mobile Node (MN) as session originator and the node at the other session endpoint (which may also be a MN or a fixed terminal or server etc.) generally denoted as Correspondent Node (CN). Of course in-between are several network entities such as Access Point and/or Access Router (AP/AR), Gateways (GWs), Mobility Anchors or Mobility Agents (MAs). Further logical or physical components which may be involved in mobility provision are those serving AAA (Authentication, Authorization, Accounting) needs and any “Proxy” entities which resemble the functionality of mostly centrally located functionalities near the edge, i.e. in the access network.

Finally – since the MN will change its location during movement there is need for a stable entity which maps the MN’s current position or address to its stable identity. This is traditionally done within the in Home Agent (HA).

Examples of a scenario with heterogeneous access technologies such as WLAN, cellular 3G (i.e. UMTS) and different types of Digital Video Broadcasting (DVB) and the corresponding Base Stations (BSs) are shown exemplarily in Fig. 2.

3.2 Existing mobility protocols designed for different purposes

Since the scenarios where a mobility protocol is required may differ much from each other and an evolution of solutions towards complexity and sophistication has been occurring since at least two decades (see Johnson et al., 2004) there is a bunch of approaches described in detail e.g. by Zhu et al. (2011). In general one main differentiation is whether location or identity will be used for reaching a node, i.e. a routing- or a mapping-based approach is chosen. The classification would be to apply either dynamic routing with fixed IP address regardless of location change – or to continuously map an MN’s stable identifier to its dynamically changing IP address.

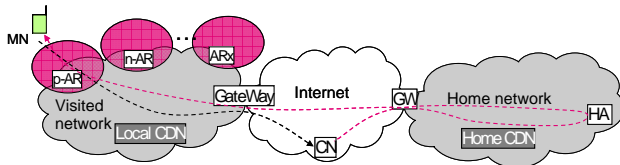


Fig. 4. Traditional support protocol for IP mobility and corresponding typical scenario.

Further classification is the amount of entities which have to participate in the mobility process, i.e. whether both session end points (MN and CN) have to be mobility-aware or the MN only – or even none of them. In the latter case only the network (operator) controls the mobility-related activities whereas in the other approaches also the user terminals may be involved. Thus it is called a client-based solution (requiring e.g. software on the terminal).

Approaches may further differ by the range of mobility covered: Local and global-scale mobility protocols serve different sizes of geographical and/or topological service areas, a single or multiple administrative domains etc. The protocol performance in terms of supported speed of user/terminal and real-time requirements of an application is another characteristic – however the velocity in terms of route optimization has trade-offs such as the required administrative relationship between MN and CN to secure the corresponding signalling messages against fraud and privacy of user location towards session partners.

Based on the considerations above from an operator point of view several requirements have to be kept in mind which will be listed in the next section.

3.3 Criteria for decision on “best” approach

Design goals and principles of mobility management approaches have to respect several aspects. Some of the most important ones are listed below.

3.3.1 Scalability

The amount of nodes which can be handled by a system concurrently puts a limitation which may be caused by e.g. signalling effort per handover and network entity. In addition centralized components often constitute a single Point of Failure, to be counteracted by backup entities if possible. An extensively created core traffic volume is another issue which needs careful inspection of connection bandwidth to mobility agents etc.

3.3.2 Application friendliness

Aim of new more efficient protocols is to increase customer satisfaction – thus all approaches which require update of installed code on user terminals but also on corresponding

application servers within CDNs or other 3rd party service providers should be excluded as far as possible.

3.3.3 Local routing capability

To reduce the amount of data traffic in the overall network often local cache servers are installed for content distribution. To make best use of such features a dynamic support of “local breakout” (i.e. enabling direct access to local available content) is favourable.

3.3.4 Low signalling

In order to spend scarce resources economically each new feature related to new options and messages should introduce only a limited amount of additional signalling overhead.

3.3.5 Security and ease of AAA/ID management support

With global ubiquitous use of mobile services and introduction of additional service and content providers in the mobile communication chain persistently secured interests of each partner has to be granted. Thus the control of all administrative issues across a handover has to be achieved in an efficient way, i.e. without introducing additional delay or unduly high user interaction.

3.3.6 Universal usage

Finally an ideal protocol for mobility handling (such as holds also for other service features) should operate in a technology unaware manner, support as well uni- and multicast traffic, and also a mobility control between different domains (i.e. operators, network clusters, etc.).

4 Comparison of mobility approaches

4.1 Traditional centralized approach and advantage of a distributed approach

Current centralized approaches such as Mobile IPv6 (MIPv6) by Johnson et al. (2004) and described in section 3.1 can be deployed in an efficient manner and offer inherent security features. The drawbacks, however, are a weak scalability in case of large amount of mobile nodes, the vulnerability due to single points of failure, and a high congestion probability since in multi-domain scenarios any gateway towards the HA resembles a bottleneck for data and signalling traffic. In addition in a roaming scenario the latency and routing path inefficiency introduced by the distant HA affects the overall performance. Furthermore a high signalling overhead is established independent of whether the mobile node actually moves between access nodes or remains within the coverage area of a single cell.

Table 1. Amount and Average Length of Messages Exchanged on Corresponding Links for Different Protocol Proposals.

	MN-MAG		Inter-MAG		MAG-LMA		Total msg ovh in B
	Amount	Avg. length	Amount	Avg. length	Amount	Avg. length	
RFC 6226: Base Multimob	4	84	0	0	3	64	528
CXTP enhanced HO	2	92	5	88.8	4	72	916
RAMS predictive	2	92	0	0	5	65.2	510
RAMS reactive	2	92	0	0	5	76	564
PMIPv6 ext. (user profile)	2	92	0	0	3	64	376
PMIPv6 ext. (PBU/A-M)	2	92	0	0	5	69.6	532
PFMIPv6 predictive	3	68	4	72	1	76	636
PFMIPv6 reactive	4	84	4	72	1	76	700

The Distributed Mobility Management (DMM) approach which is currently in focus at IETF (Internet Engineering Task Force) could improve

- Scalability and robustness by set of multiple distributed MAs (e.g. proxy HAs)
- Routing path length and delay by placing mobility entities (MAs) closer to edge of network (e.g. at AR level)
- Core network load via off-loading/local break-out near edge of network and flexible and dynamic split of data flows along different paths
- Efficiency of (radio) resource management
- Mechanism tuned for different MN status: Non-moving, paging, idle mode
- Energy saving thanks to less signalling, scanning etc. (mainly at the MNs)
- Separation of control and data planes by splitting location and routing anchors
- Central control and distributed data paths for minimum inter-MA signalling
- Support for next generation service-centric content- and cloud-based networks

4.2 Exemplary enhancements in IP mobility protocols

When designing enhanced protocols a reuse of existing approaches is essential since minimum changes allow for faster adoption of technology. In that sense the adaptation of complex MIPv6 protocol (Johnson et al., 2004) suite towards serving a mobility-unaware MN has resulted in Proxy MIP (Gundavelli et al., 2008). With a Mobility Access Gateway (MAG) caring for mobility on behalf of an MN and the Local Mobility Anchor (LMA) replacing the HA in the neighbourhood of the MNs current position MN a step towards decentralization has been done.

Multimob WG has been initiated to work on protocols supporting multicast service provisioning to mobile nodes. Base Multimob protocol has been approved as RFC and the next steps head towards optimization of Handover speed and reliability.

Beside others the application of Context Transfer Protocol (CXTP) (Loughney et al., 2005) has been proposed to apply as Extension for Mobile Multicast to achieve lossless and low delay session continuity after a MN's handover. Features are:

- No new subscription by MN via new Access Router (n-MAG) required
- Exchange between n-MAG and (previous) p-MAG via CXTP context data
- Enabling buffering and forwarding of content at p-MAG
- Enabling buffering and packet re-ordering at n-MAG
- Reduced load and delay on wireless link MN-MAG

A quantitative comparison with other approaches is currently in progress described in the next section.

5 First analytical results for handover improvement approaches

A performance analysis has been performed for four protocol proposals beside the standard solution ("Base Multimob", RFC 6224) by Schmidt et al. (2011a): These cover the draft proposal RAMS ("Rapid Acquisition of Multicast Subscription") by Contreras et al. (2011b) in both proactive and reactive version, which does not require any inter-AR communication but solely relies on AR-MA message exchange and the CXTP-extension method described by von Hugo and Asaeda (2011). Further also the PMIP extension proposed by Asaeda et al. (2011) (where subscription information is either transferred via user profile or PBU/PBA extensions) and an approach based on Proxy Fast Handover (PFMIPv6) protocols (both predictive and reactive mode) such as proposed by Schmidt et al. (2011b) are included in the comparison. Handover performance is characterized by additional

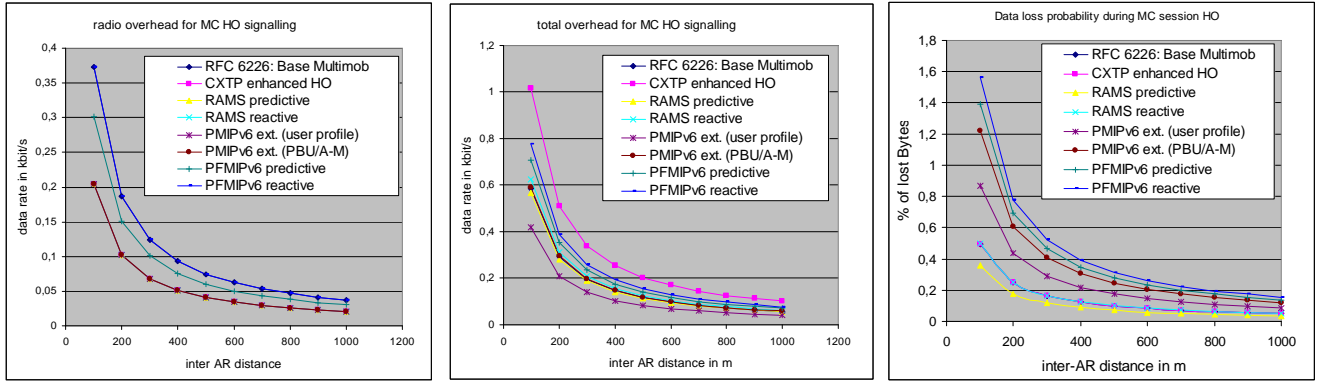


Fig. 5. Results of analytical calculation for loss and overhead for different approaches to improve multicast HO (MN-MAG: 5 ms, inter-MAG: 5 ms).

overhead due to signalling message exchange and by delay during handover process assuming typical latencies on different radio and non-radio paths within the scenarios.

From the protocol definitions given in (Contreras et al., 2011b), (Schmidt et al., 2011a), (von Hugo and Asaeda, 2011), (Asaeda et al., 2011), and (Schmidt et al., 2011b) the amount and size of messages on the connection paths are derived and summarized in Table 1. Assuming the message structures as proposed in the different descriptions and adding typical IPv6 address header lengths the sum of total message overhead in Bytes or octets follows as given in the last column of Table 1.

Though the overhead due to different approaches may vary up to more than 100 % a more important aspect is which delay is introduced due to the message exchange, i.e. where within the topology or between which entities the messages have to be exchanged. Similar to the approach described by Contreras et al. (2011a) we add up the delay components due to radio links (MN-MAG), within the access network (inter-MAG, i.e. between previous and new MAG), and in the back-haul network (between MAG and LMA). Here we have neglected for a first attempt the additional processing delay introduced by an expected average MLD Query response time (Vida and Costa, 2004). Furthermore whereas the tunnel between MAG and LMA is expected to be operational before handover occurrence an additional delay would have to be accounted for set-up of a tunnel between neighbouring MAGs as well as for e.g. look-up of MN user profile information in LMA. These latency figures have also been neglected until reliable data will be available. The expected data loss for the different approaches has been calculated as the time between “Movement of MN to n-MAG” and resume of multicast transmission to MN or start of buffering multicast data at p-MAG where applicable.

Assuming different radio cell sizes or vehicular velocities v_{MN} and typical delays on radio d_{RF} , inter-MAG d_{RAN} , and

core network links d_{CN} a calculation of the expected total handover delay d_{HO} and the corresponding relative data loss l_{tot} (assumed to be proportional to delay and bandwidth) and of additionally introduced signalling overhead has been performed as a function of distance between neighbouring access points Δ_{RF} :

$$l_{tot} = d_{HO} \cdot b_{RF} \quad (1)$$

with

$$d_{HO} = \Delta_{RF}/v_{MN} \cdot (d_{totRF} + d_{totAN} + d_{totCN}) \quad (2)$$

where

$$d_{totRF} = \Delta_{RF}/c_0 + d_{RF} \cdot n_{RF} + b_{RF} \cdot N_{RF} \quad (3)$$

with c_0 as the speed of light, n_{RF} denoting the amount of messages exchanged on radio path (i.e. between MN and MAG), b_{RF} describing the transmission bandwidth on the radio path in bits/s, and N_{RF} representing the amount of bits required for the additional messages on the radio path as shown in Table 1. Correspondingly the suffix AN denotes the situation in the access network, i.e. distance, delay, bandwidth between MAGs, and the expressions with suffix CN represent the parameter of the core network, i.e. which apply between MAG and LMA.

The results of additional data rate required for messages are separated into portion occurring on the radio link (MN-MAG) and within the access network (AN), i.e. on inter-MAG links, and core network (CN) where the MAG-LMA tunnel is assumed to be located is simply calculated as product of message amount n and size N for each area.

According to requirements of NGMN the latency of radio access shall be “less than 5 ms in unload condition” (3GPP, 2009) whereas “core latency shall be less than 10 ms, however, it is desirable that the core latency be reduced to below 5 ms” (3GPP, 2006).

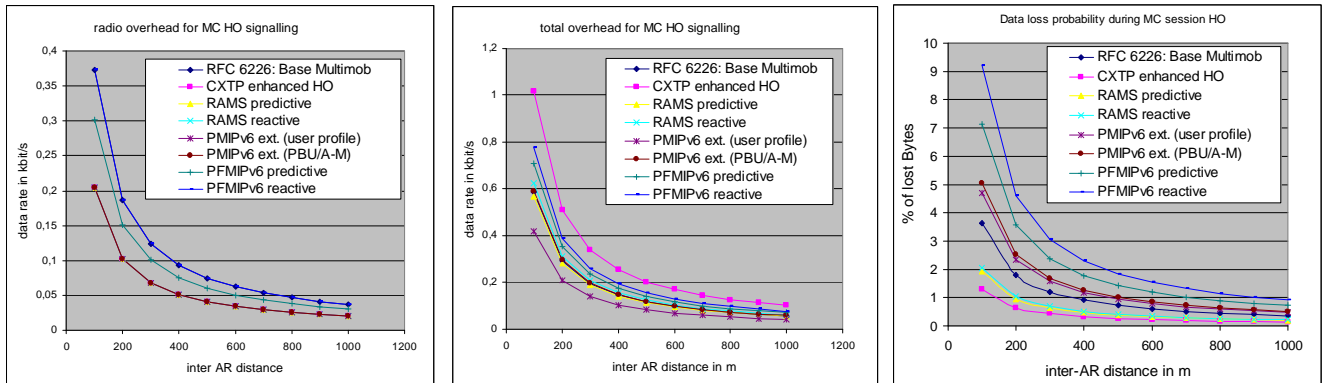


Fig. 6. Results of analytical calculation for loss and overhead for different approaches to improve multicast HO (MN-MAG: 63.5 ms, inter-MAG: 5 ms).

Exemplary results for an LTE-like topology with MN-MAG delay, inter-MAG and MAG-LMA delay of same order of magnitude (i.e. 5 ms) are shown for a terminal velocity of 50 km h^{-1} in Fig. 5.

As can be seen the radio overhead is reduced for all modifications as compared to the base protocol. CXTTP extension has a slightly higher message overhead on non-radio links and loss figures in the same order as RAMS reactive approach.

Considering on the other hand that delay components for 3G (Rel. 99/Rel. 4) as described in TR 25.853 (3GPP, 2001) may add to between 50 ms and 70/80 ms while on RNC level up to additional 20 ms are accounted for. Calculations for figures as mentioned above have been reported in (Contreras et al., 2011b) and for sake of comparability we present results of our formula assuming an average MN-MAG delay of 61.3 ms (UTRAN) and the backhaul delay of 5 ms in Fig. 6.

As can be seen the overhead is not affected but a higher loss due to longer radio delay can be clearly depicted on the right graph. Here CXTTP enhanced HO provides improved figures for small inter-MAG distance (i.e. in a very dense urban environment).

The comparisons show that to improve the performance of seamless session handover for multicast services in terms of data loss versus network signalling load several aspects have to be considered: In addition to an optimum choice of access technology and network topology also the impact of the protocol structure has to be thought about.

Additional factors such as e.g. additional delay for IP-tunnel set-up should be accounted for in future work which is currently under way by members of the IETF WG Multimob.

6 Federation issues

As mentioned in the introduction network heterogeneity not only covers different technologies and topologies but – espe-

cially when trying to achieve better i.e. more efficient use of resources – a stronger cooperation between several operators in terms of sharing network components and offering roaming (see e.g. Atalla, 2010). Thus a handover may also include change of administrative control which introduces additional requirements to AAA of user and network and grant of security and QoS maintenance across operator boundaries. To enable a fast and reliable interoperation for the customer a federation between operators offers a solution described in the following sections.

6.1 Impact of handover between operator domains

New mechanisms are needed to cluster roaming when in future service provision scenarios roaming becomes more and more important. The model of federation has to respect a number of additional requirements such that each operator's interests are properly considered.

Requirements cover

- Administrative domains
- User profile retrieval
- Mobility operations
- QoS negotiation
- Charging/Billing

Typical federation classes range between lowest level without any federation in place, via “basic roaming” (exchange of accounting and charging information only), to “personalized roaming” (sharing also personalization information, e.g. a users service preferences). Highest classes are a “premium roaming” (allowing for session handover across domains) and finally the full federation between the parties. A model has been developed within the framework of EU projects SWIFT (2010) and Daidalos (see Aguiar et al., 2007,

and Jähnert et al., 2008) which allows for different levels of trust and could enhance the business model for all participating parties.

Federation offers the possibilities of contract definition between operators and can also enhance the mobility experience of the user when identity functions and federation are extended to the networks while addressing usability and privacy concerns (SWIFT, 2010).

7 Discussion and conclusion

In view of the expected still increasing mobile Internet traffic predicted for the next years an adaptation of current mobile network architectures seems necessary. Technically and economically efficient solutions have to support evolving core and access network technology to meet demands in terms of data and signalling volume and transmission quality (carrier grade operator services).

This paper has discussed ongoing research and engineering effort and presented exemplary issues of existing mobility management protocols and improvements.

Especially heterogeneous – both fixed and mobile - networks and flat architecture designs may benefit from a distributed approach where functional intelligence is moved towards the edge of the infrastructure.

Federation models will contribute to grant both the user and the operator the needed level of privacy and security while being economically beneficial within an increasing diversification of applications and scenarios.

Acknowledgements. Support and contribution of members in the IETF WG MultiMob as well as members of our project field team are gratefully acknowledged.

References

- 3GPP: Delay Budget within the Access Stratum (Release 1999&Release 4), TR 25.853, 2001.
- 3GPP: Requirements for Evolved UTRA (E-UTRA) and Evolved UTRAN (E-UTRAN) (Release 9), TR 25.913, 2009.
- 3GPP: IP multimedia subsystem (IMS), Stage 2 (Rel. 10), TS 23.228, 2010.
- 3GPP: 3GPP System-Fixed Broadband Access Network Interworking; Stage 2 (Rel.11), draft TS 23.139, 2011
- Aguiar, R.L., Sarma, A., Bijwaard, D., Marchetti, L., Pacyna, P., and Pascotto, R.: Pervasiveness in a Competitive Multi-Operator Environment: the Daidalos Project, IEEE Commun. Mag., 45, 24–26, 2007.
- Asaeda, H., Seite, P., and Xia, J.: PMIPv6 Extensions for PIM-SM, work in progress, IETF draft, 2011.
- Atalla, S.: Mobile Operator Network Sharing the Trend to Watch in 2010, <http://www.zawya.com/story.cfm/sidZAWYA20100116073017/?relcontent=ZAWYA20111206040256>, 2010.
- Contreras, L. M., Bernardos, C. J., and Soto, I.: RAMS: A Protocol Extension to PMIPv6 for Improving Handover Performance of Multicast Traffic, Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, 2, 67–82, 2011a.
- Contreras L. M., Bernardos, C. J., and Soto, I.: Rapid acquisition of the MN multicast subscription after handover, work in progress, IETF draft, 2011b.
- Gundavelli, S., Leung, K., Devarapalli, V., Chowdhury, K., and Patil, B.: Proxy Mobile IPv6, RFC 5213, 2008.
- Jähnert, J., Roque, R., and Kaldanis, V.: Daidalos II, White Papers on Key Concepts: Federation, <http://www.ist-daidalos.org/daten/publications/publications.htm>, 2008.
- Johnson, D., Perkins, C., and Arkko, J.: Mobility Support in IPv6, RFC 3775, 2004.
- Loughney, J., Nakhjiri, M., Perkins, C., and Koodli, R.: Context Transfer Protocol (CXTP), RFC 4067, 2005.
- NGMN Alliance: Next Generation Mobile Networks beyond HSPA & EVDO, White Paper, 2006.
- Schmidt, T.C., Waehlich, M., and Krishnan, S.: Base Deployment for Multicast Listener Support in Proxy Mobile IPv6 (PMIPv6) Domains, RFC 6224, 2011a.
- Schmidt, T.C., Waehlich, M., Koodli, R., and Fairhurst, G.: Multicast Listener Extensions for MIPv6 and PMIPv6 Fast Handovers, work in progress, IETF draft, 2011b.
- SWIFT, EU project: Secure Widespread Identities for Federated Telecommunications, <http://www.ist-swift.org>, 2010.
- UMTS Forum, Mobile traffic forecasts 2010–2020, UMTS Forum Report 44, 2011.
- Vida, R. and Costa, L.: Multicast Listener Discovery Version 2 (MLDv2) for IPv6, RFC 3810, 2004.
- Vogt, C. and Arkko, J.: A Taxonomy and Analysis of Enhancements to Mobile IPv6 Route Optimization, RFC 4651, 2007.
- von Hugo, D. and Asaeda, H.: Context Transfer Protocol Extension for Multicast, work in progress, IETF draft, 2011.
- Zhu, Z., Wakikawa, R., and Zhang, L.: A Survey of Mobility Support in the Internet, RFC 6301, 2011.