

Towards Flat and Distributed Mobility Management: a 3GPP Evolved Network Design

Carlos J. Bernardos^{*}, Juan Carlos Zúñiga[†], Alex Reznik[†]

^{*} Universidad Carlos III de Madrid

E-mail: cjbc@it.uc3m.es

[†] InterDigital

E-mail: {JuanCarlos.Zuniga, Alex.Reznik}@InterDigital.com

Abstract—The last few years have seen an exponential increase in mobile data use, coming primarily as a result of the “iPhone revolution”. Mobile operators are currently facing traffic demands higher than what their networks can support, which has triggered the need of implementing offloading mechanisms at different levels of the architecture. One of the problems is that existing mobility procedures are highly centralized, requiring all data traffic to traverse the mobile operator’s network core. This paper presents an evolved 3GPP architecture that supports full distributed mobility management, while allowing incremental deployment and co-existence with existing solutions.

I. INTRODUCTION

We are witnessing an exponential increase in the use of data services by mobile subscribers. This is motivated by a variety of different reasons: 3G and WLAN accesses are widely available (combined, coverage reaches almost 100% of dense populated areas in developed countries) and affordable by users (most mobile handsets are 3G and WLAN capable, all laptops and netbooks are equipped with WLAN interfaces, 3G USB modems are quite cheap and operators offer flat rates to their customers). Additionally, the number and popularity of smart phone applications that make use of Internet connectivity is increasing every day.

This huge amount of mobile data traffic seriously impacts the dimensioning and planning of current mobile networks, as *a*) spectrum is limited and expensive, so available bandwidth in the access cannot be easily increased; and *b*) deployed mobile networks are highly hierarchical and centralized, introducing serious scalability and reliability concerns. In order to tackle the first issue by increasing system capacity of the wireless access, operators are looking at deploying femto and pico cells [1], as well as to selectively offload traffic from 3G to WiFi [2]–[4]. Different approaches are currently being developed to address the second aspect, such as the Local IP Access and Selected IP Traffic Offload (LIPA-SIPTO) within the 3GPP¹, and the Distributed Mobility Management (DMM) effort at the IETF².

This paper focuses on the scalability and reliability problems caused by the use of current hierarchical and centralized anchoring and mobility approaches, and proposes a 3GPP evolved architecture following a DMM approach. The rest of

the paper is organized as follows: Section II briefly introduces the background and motivation for this work. In Section III, we present and provide details of our designed architecture. Since there are other approaches that can be followed to tackle the same problems, we analyze the most relevant ones in Section IV, comparing them with our proposed solution. Finally, Section V is devoted to concluding the paper and outlining some future work.

II. BACKGROUND AND MOTIVATION

Recent mobile architectures such as WiMAX and the Evolved Packet System (EPS) are intended to be IP-based both for data and voice communications, triggering a real need to optimize IP protocols for mobile networks.

IP mobility management plays a key-role in providing the *always-on* and ubiquitous service envisioned by future technologies. Unfortunately, IP mobility protocols standardized so far have not met the deployment expectations, most of them being customized with proprietary solutions instead.

The mobility management schemes standardized by IETF for IPv6 networks are extensions or modifications of the well known Mobile IPv6 protocol (MIPv6) [5], such as Proxy Mobile IPv6 (PMIPv6) [6], Dual Stack Mobile IPv6 (DSMIPv6) [7] and Hierarchical Mobile IPv6 (HMIPv6) [8]. However, they come at the cost of handling operations at a central point – the mobility anchor – and burdening it with data forwarding and control mechanisms for a great amount of users. Table I shows, for the main mobility protocols and architectures, the equivalence between the principal mobility roles and the logical entities playing them. The mobility anchor is usually far away from the edge and deep into the core network, and although HMIPv6 proposed to split the management hierarchically, this only shifts the problem close to the edge without really addressing the flat IP architecture demand.

In order to address such issues, a new paradigm of solution, the so-called Distributed Mobility Management (DMM), is currently being analyzed by both academic and standards communities. DMM basically develops the concept of a flatter system, in which the mobility anchors are placed closer to the user, distributing the control and data infrastructures among the entities located at the edge of the access network.

¹3rd Generation Partnership Project: <http://www.3gpp.org/>

²The Internet Engineering Task Force: <http://www.ietf.org/>

	MIP	Proxy MIP	GPRS & UMTS	EPS
Mobility anchor	HA	LMA	GGSN	PGW
Signaling agent	FA	MAG	SGSN	SGW/ePDG/eNB

Table I

EQUIVALENCE BETWEEN MAIN MOBILITY ROLES AND LOGICAL ENTITIES.

Centralized mobility solutions, such as Mobile IPv6 or the different macro-level mobility management solutions of 3GPP (for GPRS & UMTS and EPS), base their operation on the existence of a central entity, e.g., Home Agent (HA), Local Mobility Anchor (LMA), Packet Data Gateway (PGW) or Gateway GPRS Support Node (GGSN), which anchors the IP address used by the mobile node and that is in charge of coordinating the mobility management (MM), sometimes helped by a third entity like the Mobility Management Entity (MME) or the Home Subscriber Server (HSS). This central anchor point is in charge of tracking the location of the mobile and redirecting traffic towards its current topological location. While this way of addressing mobility management has been fully developed by the Mobile IP protocol family and its many extensions, it brings several limitations that have been identified [9]:

- Sub-optimal routing. Since the (home) address used by a mobile node is anchored at the home link, traffic always traverses the central anchor, which leads to paths that are, in general, longer than the direct one between the mobile node and its communication peer. This is exacerbated with the current trend in which content providers push their data to the edge of the network, closer to the users. With centralized mobility management approaches, user traffic will always need to go first to the home network and then to the actual content location, adding unnecessary delay and wasting operator's resources.
- Scalability problems. Existing mobile networks have to be dimensioned to support all the traffic traversing the central anchors. This poses several scalability and network design problems, as the central mobility anchors need to have enough processing and routing capabilities to be able to deal with all the users' traffic simultaneously. Additionally, the entire operator's network needs to be dimensioned to be able to cope with all the users' traffic.
- Reliability. Centralized solutions share the problem of being more prone to reliability problems, as the central entity is a potential single point of failure.
- Signaling overhead. By allowing mobility management to be dynamically enabled and disabled on a per application basis, some signaling can be saved, as well as the associated handover latency.

III. 3GPP EVOLUTION TOWARDS DISTRIBUTED MOBILITY MANAGEMENT

This section describes our proposed 3GPP architecture evolution supporting distributed mobility management, by first introducing the architecture, and then providing more details about the designed solution.

A. DMM-based architecture

In this section we present a DMM-based solution for the network-based and client-based mobility architectures currently supported in the latest 3GPP specifications. For the former, we consider GPRS Tunnelling Protocol (GTP) and Proxy Mobile IPv6 (PMIPv6) solutions, while for the later we assume Dual Stack Mobile IPv6 (DSMIPv6). For simplification, we only focus on the non-roaming scenarios. Figure 1 shows the interface diagram, highlighting in light blue those parts that are new compared with latest 3GPP specifications. The key novelty introduced by our proposal is the Distributed Gateway (D-GW), which is a new logical network entity located at the edge of the network, and can be considered as the result of splitting and distributing the PGW functionality, so that it sits closer to the users.

The distributed gateway (D-GW) implements the functionality of the PGW, in addition to some additional operations required for DMM operation. In terms of capacity, a distributed gateway is not expected to manage a large number of subscribers, as multiple D-GWs should be deployed. The actual number of distributed gateways (or the ratio number of D-GWs / number of PGWs) is up to the mobile operator and the specific scenario needs. We next describe the additional functionalities that the distributed gateway implements.

For the case of the network-based DMM variant (Figure 1(a)), the distributed gateway behaves as access router and mobility signaling agent³. This basically matches with the Mobile Access Gateway (MAG) functionality, according to the Proxy Mobile IPv6 specification [6], if the PMIPv6 network-based DMM solution is used. This functionality is performed on a per User Equipment (UE) and per-IPv6 prefix granularity. This means that a single D-GW instance may behave as a MAG when handling traffic of a given UE's IPv6 prefix, while operating differently when handling traffic of a different prefix that might or not belong to the same UE. The MAG functionality terminates the S5* interface (which is basically a DMM-enabled version of the S5 interface) with another distributed gateway implementing the LMA counterpart functionality. If the GTP network-based variant solution is used, the D-GW also behaves logically as a mobility signaling agent, but using in this case GTP for both control and data planes.

The distributed gateway also implements the functionality of mobility anchor. This corresponds with the Local Mobility Anchor (LMA) functionality according to the Proxy Mobile IPv6 specification, if the PMIPv6 network-based DMM solution is used. This is already a functionality of the packet data gateway, but the D-GW implements it differently, as it should be performed also on a per-UE and per-IP prefix granularity. The LMA functionality terminates the S5* interface with another distributed gateway implementing the MAG counterpart functionality. If the GTP network-based variant solution is used, the D-GW also behaves logically as a mobility anchor, but using in this case GTP for both control and data planes.

³Table I shows the equivalence between mobility roles and logical entities of the principal mobility protocols and architectures

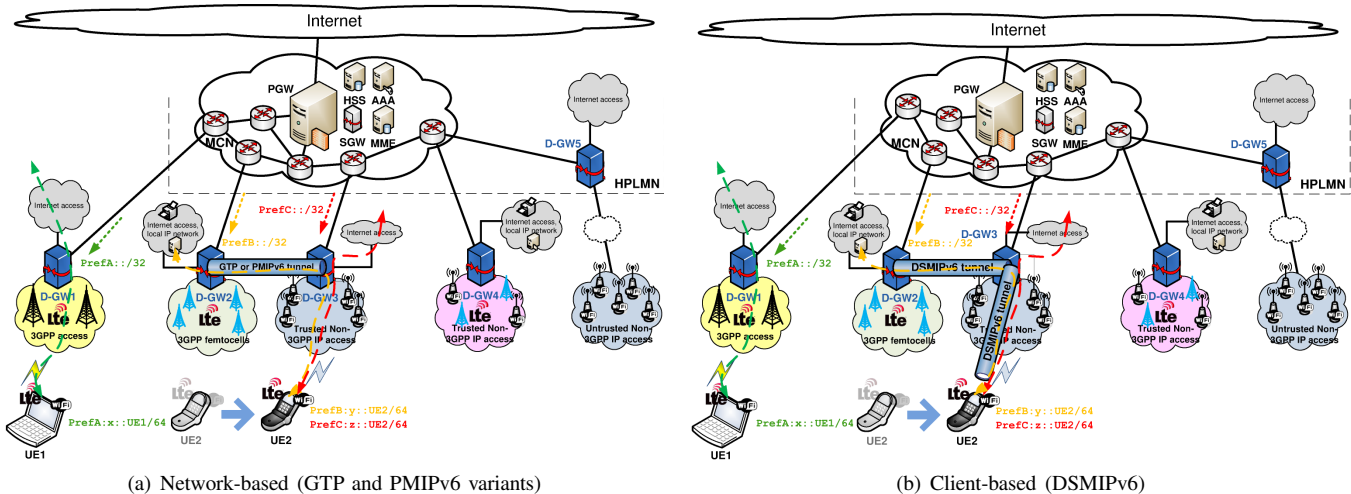


Figure 2. DMM-Based Mobile Network System Design: solution overview.

B. Solution operation

This section provides an overview of the operation of the solution, for both network and client-based mobility management approaches (see Figure 2).

D-GWs are distributed at the edge of the network. For 3GPP and Trusted Non-3GPP IP accesses, these D-GWs are placed close to the UE, at the access network level (i.e., close to the SGW or even to the Home evolved Node B/Local Gateway for 3GPP access, and close to the Access Points for WiFi Trusted Non-3GPP IP access). For Untrusted Non-3GPP IP access, the D-GW is located at the edge of the HPLMN of the operator, instead of the ePDG, which is the operator-managed entity closest to the user equipment.

The distributed gateway behaves almost as a packet data gateway from the viewpoint of the user equipment and the rest of the network, with some DMM modifications. The distributed gateway may provide Internet access (local breakout à la SIPTO) and connectivity to local resources (LIPA scenario). Each D-GW has a pool of IPv6 prefixes anchored at the D-GW (i.e., IP routing delivers packets addressed to those prefixes to the D-GW) available for delegation to UE.

When a user equipment initially attaches to the network, the APN requested by the UE (or the default one, if none provided) is checked together with its profile at the home subscriber server. Assuming a single packet data network connection, if it is decided to be handled locally, then the following steps described on subsequent subsections apply. Otherwise, already specified 3GPP procedures would be followed, being the D-GW transparent (i.e., a mere relay in most of the procedures).

We now use a simple example to explain how the solution works, using the scenario shown in Figure 2 (both network and client-based cases are included). Every time a PDN connection is requested by a UE, it is handled by the distributed gateway, which assigns an IPv6 prefix from its pool to the user equipment. This prefix is conveyed to the UE so it can auto-configure an IPv6 address. We assume stateless auto-configuration (i.e., the distributed gateway sends Router

Advertisements carrying the assigned prefix), but other options are possible (e.g., use of DHCPv6). The D-GW updates on the HSS (via the MME for 3GPP access and the AAA Server for non-3GPP access) the IPv6 prefix assigned to the user equipment, including the D-GW Identifier (and IPv6 address if the D-GW Identifier is not enough to derive the address). The UE can then start sending and receiving IPv6 packets, which are routed via the distributed gateway, without traversing the Mobile Core Network (MCN) (for the case of Untrusted Non-3GPP access, packets need to traverse the HPLMN, but not the MCN). If we take UE1 in Figure 2, it is attached to D-GW1 and configures $\text{PrefA}::\text{UE1}/64$ address out of the prefix $\text{PrefA}::/64$ assigned by D-GW1.

For the sake of brevity in the explanation of the handover scenario, we first focus on the network-based solution (Figure 2(a)). If the user equipment moves and attaches to another access network, there are two different kind of procedures that take place. First, the packet data network connections that the UE has established needs to be maintained (i.e., address preservation). This requires, for each of the PDN connections of the user equipment, that the distributed gateway anchoring the IP address used by the UE plays the role of packet data gateway (i.e., mobility anchor) for that PDN connection, meaning that the distributed gateway performs the local mobility anchor functions for that UE and that PDN connection. The serving distributed gateway (i.e., the D-GW the UE is currently attached to) has to play the role of signaling agent (e.g., MAG) for each of the UE's packet data network connections that are anchored at other distributed gateways. The serving D-GW obtains the required information about the on-going PDN connections of the UE, the IPv6 prefixes used and the D-GWs anchoring them, by interacting with the HSS/AAA. Second, the user equipment requests a new packet data network connection (or several) to the serving distributed gateway. This provides the user equipment with an IPv6 address anchored at the serving D-GW, which can be used by the UE to enjoy optimal routing while making the

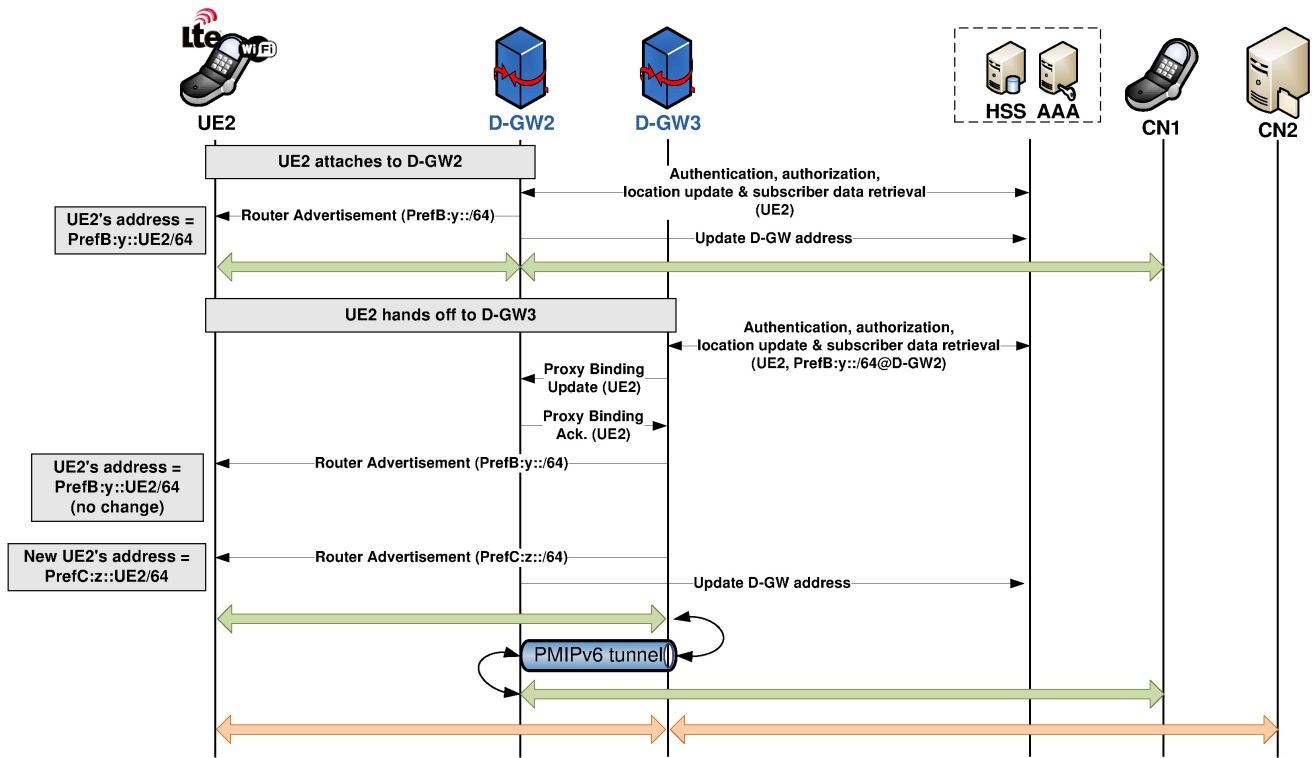


Figure 3. DMM network based (PMIPv6 variant) example MSC.

best use of the operator's network resources.

Referring to the scenario illustrated in Figure 2(a), we show in Figure 3 an example Message Sequence Chart (MSC) of the DMM network-based (PMIPv6 variant) solution. UE2 initially attaches to D-GW2, where it establishes a PDN connection and configures $\text{PrefB:y}::\text{UE2}/64$ as IP address (anchored at D-GW2). Using this locally anchored IPv6 address, UE2 can directly communicate (i.e., no tunneling, no sub-optimal path) with any Correspondent Node (CN) of the Internet, such as CN1. Later on, UE2 moves and attaches to D-GW3. The original PDN connection is handed over, by D-GW3 playing the role of MAG and D-GW2 playing the role of LMA, establishing a PMIPv6 tunnel (in case of the GTP variant, a GTP tunnel would be setup instead) between them to forward traffic addressed to $\text{PrefB:y}::\text{UE2}$ to the current location of UE2. This allows UE2 to keep using $\text{PrefB:y}::\text{UE2}$ and therefore to seamlessly maintain any running services/applications/connections using that address (e.g., the communication with CN1). Besides, UE2 establishes a new packet data network connection at D-GW3, configuring a new IPv6 address ($\text{PrefC:z}::\text{UE2}/64$), anchored at D-GW3, that can be used by UE2 for new connections (for example with CN2, avoiding any tunneling and sub-optimal routing). Note that Proxy Mobile IPv6 signaling is shown in Figure 3, and that GTP signaling would be used in case the GTP variant was deployed.

Note that for the case of client-based mobility (Figure 2(b)), the basics of the solution are very similar. A user equipment requesting a PDN connection configures an IPv6 out of a

prefix assigned by the distributed gateway. As before, if the UE moves and attaches to another access network, the UE requests a new packet data network connection (or several) to the D-GW the UE is currently attached to. For each of the PDN connections that the user equipment had previously established and needs to be maintained, the D-GW anchoring the IP address used by the UE plays the role of PGW (i.e., HA), meaning that the distributed gateway performs the home agent functions for that UE and that PDN connection. The user equipment has to signal to each of the anchoring D-GWs its current location and establish an IPv6-in-IPv6 tunnel, so data packets can be redirected to the UE. In order to do that, the user equipment uses the address obtained from the currently attached D-GW (i.e., serving D-GW) as care-of address, and sends a Binding Update message per Home Address (i.e., each of the addresses assigned by the previously visited serving distributed gateways, which now play the role of anchoring D-GWs). Note that this is only done for those addresses that are still being used by the UE.

It is important to highlight the role of the user equipment. Most of the advantages brought by a DMM approach are enabled by UE smart IP address management. This basically means that the IP address selection mechanisms used by the UE should be enhanced to allow the UE to always prefer an IPv6 address anchored at the distributed gateway the UE is currently attached to. In this way, new communications make use of the locally anchored IPv6 addresses, while old communications are seamlessly maintained by ensuring IPv6 address continuity. It is also important that as soon as

communications using old IPv6 addresses finish, the UE is aware and signals to the network that reachability for those addresses is no longer required, so no further signaling is generated and used tunnels are removed. This UE enhanced intelligence to manage IPv6 addresses can be implemented as part of the connection manager, and should also support the use of policies from the network.

IV. COMPARISON WITH OTHER APPROACHES

In this section, we overview other existing approaches that try to address the challenges caused by the increase of the mobile data traffic demand from users and the convergence and integration of different types of networks (i.e., different access technologies, enterprise and home networks, etc.), but restricting the scope to solutions that are designed for the 3GPP architecture. Although there are several mechanisms that follow a DMM approach [13]–[16], to the best of authors knowledge, none of them goes into all the details of how to implement them in existing 3GPP architectures.

If we first look at the ongoing efforts inside the 3GPP itself, there are two main solutions that try to alleviate the load of the mobile network core: Selected IP Traffic Offload (SIPTO) and Local IP Access (LIPA) [17]. SIPTO enables an operator to offload certain types of traffic at a network node close to the UE's point of attachment to the access network, by selecting a set of GWs (SGW and PGW) that is geographically/topologically close to the UE's point of attachment. LIPA, on the other hand, enables an IP capable UE connected via a Home eNB (HeNB) to access other IP capable entities in the same residential/enterprise IP network without the user plane traversing the mobile operator's network core. In order to achieve this, a Local GW (L-GW) collocated with the HeNB is used. LIPA is established by the UE requesting a new PDN connection to an access point name for which LIPA is permitted, and the network selecting the Local GW associated with the HeNB and enabling a direct user plane path between the Local GW and the HeNB.

Both SIPTO and LIPA have a very limited mobility support, specially in 3GPP specifications up to Rel-10. In Rel-11, there is currently a work item on LIPA Mobility and SIPTO at the Local Network (LIMONET) [18] that is studying how to provide SIPTO and LIPA mechanisms with some additional, but still limited, mobility support. Without going into much details, LIPA mobility support is limited to handovers between HeNBs that are managed by the same L-GW (i.e., mobility within the local domain), while seamless SIPTO mobility is still limited to the case where the SGW/PGW is at or above Radio Access Network (RAN) level. Seamless mobility at the local network is still not considered in SIPTO. Therefore, although SIPTO and LIPA allow offloading traffic from the network core similarly to the DMM approaches, even with LIMONET they just provide localized mobility support, requiring packet data network connections to be deactivated and re-activated when the UE is not moving locally.

On the research side, there are also some proposals to extend current 3GPP mechanisms towards a flatter network

architecture, like [19] and [20], but they do only consider 3GPP accesses (E-UTRAN), not addressing the distribution of anchors in non-3GPP accesses.

V. CONCLUSIONS AND FUTURE WORK

In this paper we have presented an evolution of current 3GPP architecture towards a flat and fully distributed mobility network design. This architecture allows pushing the data anchors towards the edge, alleviating hence the overloaded network core infrastructures of mobile operators.

The proposed solution follows the distributed mobility management paradigm, which has been so far mainly discussed at the IETF, but takes into consideration the 3GPP architecture specifics. A new logical entity, called distributed gateway, is located close to the users, anchoring the data communications and supporting mobility when they move to a different D-GW. Both network-based (GTP and PMIPv6) and client-based (DSMIPv6) procedures are provided. Last but not least, the solution can be deployed incrementally, as both DMM and non-DMM operation modes are supported, and the D-GW can also be co-located with existing 3GPP network nodes.

Future work includes the definition of the mechanisms required on the UE to fully benefit from this DMM approach, conducting an analytic and experimental evaluation of the solution, and considering different network deployments, user traffic patterns and user mobility models.

REFERENCES

- [1] V. Chandrasekhar, J. Andrews, and A. Gatherer, "Femtocell networks: a survey," *Communications Magazine, IEEE*, vol. 46, no. 9, pp. 59–67, 2008.
- [2] A. de la Oliva, C. Bernardos, M. Calderon, T. Melia, and J. Zuniga, "IP flow mobility: smart traffic offload for future wireless networks," *Communications Magazine, IEEE*, vol. 49, no. 10, pp. 124–132, 2011.
- [3] K. Lee, I. Rhee, J. Lee, S. Chong, and Y. Yi, "Mobile data offloading: how much can WiFi deliver?" in *Proceedings of the 6th International Conference, ser. Co-NEXT '10*. New York, NY, USA: ACM, 2010, pp. 26:1–26:12.
- [4] T. Melia, C. Bernardos, A. de la Oliva, F. Giust, and M. Calderon, "IP Flow Mobility in PMIPv6 Based Networks: Solution Design and Experimental Evaluation," *Wireless Personal Communications*, vol. 61, pp. 603–627, 2011.
- [5] C. Perkins, D. Johnson, and J. Arkko, "Mobility Support in IPv6," RFC 6275 (Proposed Standard), Internet Engineering Task Force, Jul. 2011.
- [6] S. Gundavelli, K. Leung, V. Devarapalli, K. Chowdhury, and B. Patil, "Proxy Mobile IPv6," RFC 5213 (Proposed Standard), Internet Engineering Task Force, Aug. 2008.
- [7] H. Soliman, "Mobile IPv6 Support for Dual Stack Hosts and Routers," RFC 5555 (Proposed Standard), Internet Engineering Task Force, Jun. 2009.
- [8] H. Soliman, C. Castelluccia, K. ElMalki, and L. Bellier, "Hierarchical Mobile IPv6 (HMIPv6) Mobility Management," RFC 5380 (Proposed Standard), Internet Engineering Task Force, Oct. 2008.
- [9] H. Chan, "Problem statement for distributed and dynamic mobility management," Internet-Draft (work in progress), Mar. 2011.
- [10] R. Draves and D. Thaler, "Default Router Preferences and More-Specific Routes," RFC 4191 (Proposed Standard), Internet Engineering Task Force, Nov. 2005.
- [11] R. Draves, "Default Address Selection for Internet Protocol version 6 (IPv6)," RFC 3484 (Proposed Standard), Internet Engineering Task Force, Feb. 2003.
- [12] A. Matsumoto, T. Fujisaki, R. Hiromi, and K. Kanayama, "Problem Statement for Default Address Selection in Multi-Prefix Environments: Operational Issues of RFC 3484 Default Rules," RFC 5220 (Informational), Internet Engineering Task Force, Jul. 2008.

- [13] H. A. Chan, H. Yokota, J. Xie, P. Seite, and D. Liu, "Distributed and dynamic mobility management in mobile internet: Current approaches and issues," *Journal of Communications*, vol. 6, no. 1, 2011.
- [14] F. Giust, A. de la Oliva, and C. J. Bernardos, "Flat Access and Mobility Architecture: an IPv6 Distributed Client Mobility Management Solution," in *3rd IEEE International Workshop on Mobility Management in the Networks of the Future World (Mobiworld 2011)*, in conjunction with *IEEE INFOCOM 2011*, Apr. 2011.
- [15] F. Giust, C. J. Bernardos, S. Figueiredo, P. Neves, and T. Melia, "A Hybrid MIPv6 and PMIPv6 Distributed Mobility Management: the MEDIEVAL approach," in *Sixth Workshop on multiMedia Applications over Wireless Networks*, Jun. 2011.
- [16] F. Giust, A. de la Oliva, C. J. Bernardos, and R. P. F. D. Costa, "A Network-based Localized Mobility Solution for Distributed Mobility Management," in *International Workshop on Mobility Management for Flat Networks (MMFN 2011)*, 2011.
- [17] 3GPP, "General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access," 3rd Generation Partnership Project (3GPP), TS 23.401, Sep. 2011.
- [18] —, "LIPA Mobility and SIPTO at the Local Network," 3rd Generation Partnership Project (3GPP), TR 23.859, Jul. 2011.
- [19] W. Hahn, "Flat 3GPP Evolved Packet Core-Improvement for multiple network connections," in *WPMC*, 2011.
- [20] —, "3GPP Evolved Packet Core support for distributed mobility anchors," in *ITST*, 2011.