Optimized Low Mobility Support in Massive Mobile Broadband Evolved Packet Core Architecture

Marius Corici, Jens Fiedler, Dragos Vingarzan, Thomas Magedanz
Fraunhofer FOKUS Institute
Kaiserin Augusta Allee 31, 10589
Berlin, Germany
{marius-iulian.corici, jens.fiedler, dragos.vingarzan, thomas.magedanz}@fokus.fraunhofer.de

***Abstract*—With the deployment of new wireless connected devices like sensors and actuators for metering, automotive, healthcare, tracing etc. the wireless core network has to support a new level of scalability for all its functionality, including the mobility management. Being designed for human communication only, the current network architecture ensures seamless mobility in all network locations for all devices being highly redundant for low mobile devices which are rarely and in reduced areas or never moving. This paper introduces a new low mobility management concept based on subscriber device differentiation and differentiated forwarding mechanisms, along with its integration with the current functionality through dynamic management parameters updates. Furthermore, the concept is exemplified on the 3GPP Evolved Packet Core architecture and evaluated as a testbed realization based on the Fraunhofer OpenEPC testbed.**

Mobility Management, Machine Type Communication, Evolved Packet Core, Low Mobility

I. INTRODUCTION

With the development and the gradual deployment of new radio technologies such as HSPA, high throughput WiFi, LTE, femto and small cells etc. the future wireless environment is able to offer packet based broadband communication for a massive number of device simultaneously.

The radio technology deployment is followed closely by the gradual adoption of the new communication means by the users. This translates for the network operator into a linear increase of devices using data services such as smartphones, tablets and laptops along with an exponential increase in their aggregated traffic.

The development of novel self-managed wireless connected devices such as sensors and actuators enables the wireless environment with new means of remote communication. It provides the means for the gathering, analyzing, processing and control for other business branches such as energy, auto, health, logistics, device maintenance etc. Through these new services the telecommunication environment is transforming into a total convergence of the different industries.

It is foreseen that the rapid deployment of these new machine devices augmenting the rapid acceptance of data services by the mobile users will further increase the requirements for scalability on the network. This requires a better customization of the core network functionality including among others the resource reservation of data path resources, the mobility management, the security and trust as well as the charging control.

The current network architecture is designed with uniform handling of the subscribers. For the mobility management case, a set of tunnels which enable the forwarding of data packets between the current location of the mobile device and a fixed location in the network representative for the device are considered. Through this ubiquitous mechanism seamless mobility is ensured for all the devices in all the network locations while maintaining it transparent to the applications and services.

When defining the requirements for the future mobile networks, 3rd Generation Partnership Project (3GPP) ([7]) presumes that a large number of the wireless connected devices will not require ubiquitous mobility support. These devices, commonly known as low mobile devices are classified into three categories, depicted also in Figure 1:

1. Fixed devices - Devices which do not move and are always connected through the same radio access base station e.g. power metering devices;
2. Small mobile devices - Devices which move only within small area and thus are connected to a limited number of radio access base stations e.g. health monitoring at home;
3. Infrequent mobile devices – Devices which do not move often, but may move within wide area e.g. mobile sales terminals.

For these categories of devices a high overhead is considered from the part of the current mobility management solutions which are uniform to all the devices acting the same for all the wireless connected devices and supporting uniformly for all of them high mobility rates in wide network areas.

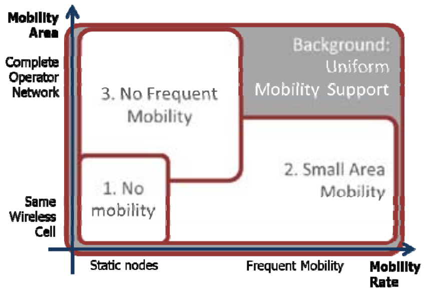

Figure 1. Mobile Devices Classification

In order to reduce the redundant network procedures related to the devices from the previous categories, this paper proposes a low mobility management solution and its dynamic integration with the current uniform mobility management based on the device mobility characteristics. The solution includes the separation of the devices into different classes based on their allocated IP addresses, the specific mobility management procedures of each class as well as the procedure of transferring one of the devices between mobility management classes.

The procedures related to the new concept are then exemplified using the 3GPP Evolved Packet Core (EPC) ([1], [2]) architecture, the reference all-IP multi access core network integrating both 3GPP e.g. GSM, UMTS, LTE etc. and non-3GPP e.g. CDMA, WiFi, WiMAX access technologies. The concept here presented integrates in the EPC as an extension of the mobility management functionality which supports seamless mobility, transparent to applications between the heterogeneous accesses.

The remainder of this paper is organized as follows: Section II provides the background of the proposed method while Section III describes the concept and an example evaluation of the solution. Section IV describes the exemplification of the concept on the 3GPP architecture followed by the description of the Fraunhofer OpenEPC testbed and in Section V conclusions are provided.

II. BACKGROUND

As depicted in Figure 2, the current mobile network architectures for packet switched communication have as main goal to offer an efficient and access network convergent mobility management functionality. It is presumed that any wireless connected devices can move to any location while being able to communicate with the correspondent nodes in a transparent manner ([6], [4]).

Through this convergent processing of the mobility related procedures independent of the mobility characteristics of wireless connected device, a certain level of scalability is attained as no multiple network entities have the same functionality for groups of mobile subscribers. However, with the high increase in mobile devices foreseen by the evolution towards machine communication, the wireless environment has to face a new level of scalability in which it is foreseen that same processing of all the subscribed devices may not be possible.

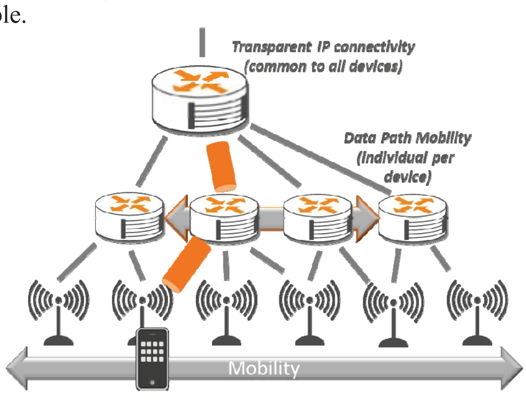

Figure 2. Current Mobility Management

The current network level mobility management solutions such as Mobile IP, Proxy Mobile IP ([4]) and GPRS Tunneling Protocol (GTP) ([5], [6]) offer such as ubiquitous mobility solution, transparent to the correspondent devices. All the solutions maintain a binding between the current location and the public IP address of the device. The public IP address is allocated at the device attachment to the network by an anchoring function. In the same time, one or multiple tunnels are established between the current network location of the wireless connected device and the anchoring function including the access technology base station and aggregation router.

Whenever the wireless device changes the location in the network, the set of tunnels from the source location is replaced by another set of tunnels to the target location. Through this mechanism, the communication is enabled transparent to the applications, regardless of the location of the device into the network.

As the current solutions assume that the mobile devices may move to any network location at any given time, no aggregation of the binding information is considered. This implies that in order to be able to send a data packet to any wireless connected device, the anchoring point has to execute a lookup on a list of information in which for each public IP allocated to a wireless connected device an entry is available. Also, this device specific information has to be maintained in all intermediary entities of the multiple tunnels which establish the data path for a subscriber.

These procedures are redundant for devices which move in small areas or do not move at all. For example, for each of the fire sensors of a building a mobility binding and a set of tunnels have to be maintained although they can share the complete communication path.

Also, these procedures are redundant for devices which rarely move, but may move in wide areas. These devices when static can share a location with other devices. Otherwise when dynamic they have to be independently processed. For example a tracking device on a truck when static can be grouped with the other devices from a parking lot while when dynamic act as a common mobile smartphone.

However, the core network currently maintains a subscription profile for each connected device including the access networks to which it is allowed to connect to and other authentication and authorization information. The subscription profile also includes the current network location of the device from which the mobility state can be determined i.e. if the device is mobile or not. Thus the core network contains the device classification mechanisms for determining when the mobility is redundant. Based on this feature a new concept of mobility management is developed.

As example network architecture, the 3GPP Evolved Packet Core (EPC) was chosen ([1], [2], [3]), due to its capability of providing all-IP network convergence between different types of accesses by offering a single unified mechanism of mobility management implemented through a multitude of network level protocols.

As depicted in Figure 3, the EPC architecture includes a set of gateways which are transparently unifying the parameters of the different access technologies like LTE, UMTS, WiMAX, cdma2000, WiFi etc. The Serving GW (S-GW) and the evolved Packet Data Gateway (ePDG) are responsible for the 3GPP and untrusted non-3GPP technologies while other technology specific gateways are used for trusted non-3GPP accesses e.g. ASN GW for WiMAX.

These gateways act either as intermediary nodes between two legs of the tunnels or as termination node. From the mobility management perspective, they maintain the unique binding between the wireless device and the anchor point.

The Packet Data Network Gateway (PDN GW) is acting as the anchor point for all the subscribers. It decides which IP address is allocated to the mobile devices at the initial attachment. It also establishes the tunnels on the data path according to the mobility management parameters. First a mobility protocol is selected. Then using this protocol, the access network specific gateways are signaled. The PDN GW terminated the core network towards the Internet core, thus having the role of advertising the IP addresses reachability.

In case of mobility, the PDN GW ensures that the source tunnels are replaced with the target tunnels. This is executed whenever a request is received from a target access network gateway. However, when mobility does not require the change of the access network gateway, the operation is transparent to the PDN GW.

The EPC uses a Home Subscriber Server (HSS) for maintaining the subscription information for each UE. This information is used for authentication and authorization in the access networks and as subscriber information in the Policy and Charging Rules Function (PCRF) for the data path resource reservation according to the requirements of the service providers, generically named Application Functions (AFs).

For the 3GPP accesses, the intra-technology mobility management is handled by the Mobility Management Entity for LTE access and by the SGSN for the 2G/3G accesses. They also have the role of selecting the appropriate PDN GW and S-GW for the communication. The SGSN together with the eNodeB for the LTE access terminate the core network tunnel toward the radio access network.

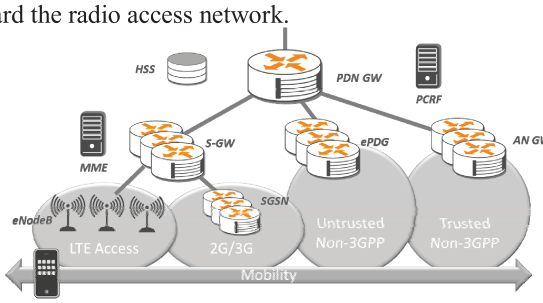

Figure 3. Simplified 3GPP Evolved Packet Core Architecture

For the EPC example, this article addresses the problem on how to execute the appropriate mobility management procedures for low mobility devices. The functionality described here extends the currently deployed mobility management solutions.

III. CONCEPT

This article proposes a novel mobility management solution for wireless connected devices which requires low mobility support. In parallel with the current dynamic forwarding scheme based on tunnels some static forwarding schemes have to be deployed. The selection of the appropriate mobility management scheme for a device is done based on the subscription profile. When the device is connected to the network it receives an IP address. The address is allocated in different ranges for each mobility management solution. The general concept is depicted in Figure 4.

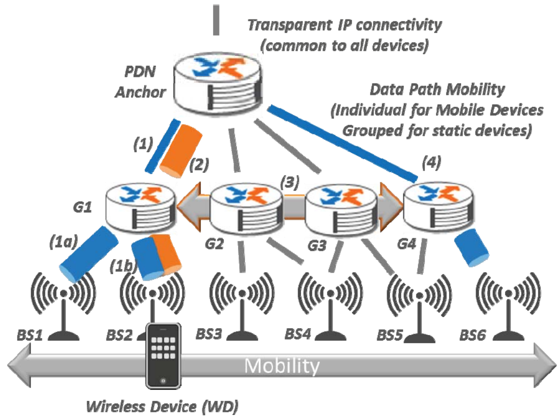

Figure 4. Proposed Concept Description

The low mobility management solution considers that for a specific device which is moving only in a given area, the closest to the device entity which covers that specific area will act as an anchor terminating the tunnels from the mobile device. On the rest of the data path, between this anchoring node and the end of the operator domain, static routing is executed. This implies that all the other data path nodes through which the data traffic passes do not have to maintain mobility related functionality, thus requiring less functionality and signaling per wireless device.

However, the system has to also sustain the current uniform mobility, thus implying that all the entities in the data path have to sustain the current tunneling solution.

The selection of the mobility management level is done during initial attachment. The network classifies the device based on its identity and the information from the subscription profile. If the device is recognized as a fixed or small mobility device then the low mobility management procedures are executed. Otherwise, the uniform mobility procedures are executed.

A device may also change its mobility pattern depending on exterior factors, which require also a dynamic change in the mobility management level. In order to cover also this mobility cases, this article proposes a mechanism to re-evaluate the mobility management solution used when specific events occur i.e. when the wireless connected device moved at the border of the area covered by the specific anchor point.

First, during the attachment procedure, the wireless connected device will require an IP address over which to communicate. This requirement is also bound to the identity of the device itself e.g. International Mobile Subscriber Identifier (IMSI), MAC address etc. The network checks the user profile and selects one mobility management level.

If the mobile device does not move more than the area covered by the gateway G1, then a set of tunnels are established only between the G1 and the wireless connected device (1a and 1b) while for the part between G1 and the PDN anchor node static forwarding is used. This allows the device to move between the (1a) and the (1b) base stations without requiring any procedure from the side of the PDN anchor node.

If uniform mobility management level is selected then the tunnels are established between the mobile device and the packet data network (PDN) anchor node as considered in case (2), allowing the mobile device to move to any base station (3).

For case (1) the PDN anchor node is not maintaining any mobility state for the wireless device while for case (2) it maintains the current levels.

So, through a simple selection based on the user profile at the initial attachment of the device to the network, the PDN anchor is using less functionality for the same number of device.

This concept can be further extrapolated to maintaining mobility management to a single base station i.e. static routing on the complete data path. However, this case is not considered feasible in future network deployments because the current evolution of the radio technologies concentrate into base stations which change their coverage area depending on the current network situation and based on some self-organizing algorithms.

As this concept is based on modifying the initial attachment procedure and as it refers to the mobility management of the wireless connected devices in IP networks, the extended attachment procedure is described having these initial hypothesis.

The differentiation between the different wireless devices (WD) is done based on the IP address allocated at the initial attachment as depicted in Figure 5. For this, the PDN anchor has a specific IP range for devices which require uniform mobility e.g. "PDN Anchor Range" and a set of IP addresses for the devices which connect from specific gateways and which require low mobility at the level of those gateways e.g. G1 to G4 anchor range.

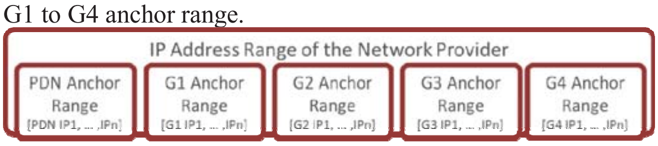

Figure 5. Example on IP address range classification

Step 1: The Wireless Device (WD) connects to the network. It requests an IP address for the communication over the network.

Step 2: The data path entities query the user profile on the mobility characteristics for the specific WD. This can be done at the same stage with the authentication to the network.

Step 3: If the WD requires uniform mobility, then an IP address from the PDN Anchor range is allocated, together with the establishment of the data tunnels between the PDN anchor, passing through the gateway and the base station for the mobile device.

Step 3': If the WD requires low mobility sustained at the Gateway level, then an IP address is allocated from the specific IP range of the gateway. Additionally, a tunnel is established between the gateway, the base station and the mobile device.

When downlink data traffic is received by the PDN anchor on one IP address, if the address is from the PDN anchor range then the specific tunnel is looked up and the data packet is forwarded through it, as in case of the current solution. If a data packet is received for an address pertaining to one of the gateways, then the packet is forwarded using static routing mechanisms in which one entry into the routing table is introduced for each of the gateways. Through this mechanism, the number of search steps is reduced from a maximum of all IP addresses available in the PDN GW to the maximum number of gateways which should be at least two orders of magnitude less. Also, the size of the table with the downlink tunnels is reduced to only the devices which require such tunnels. The routing table in the PDN anchor node will look like:

[PDN IP1, … IPn] → to local tunneling functionality
[G1 IP1, … IPn] → to G1
[G2 IP1, … IPn] → to G2
….
[Gp IP1, … IPn] → to Gp

Figure 6. Example of Forwarding Table

where the ranges allocated to the specific gateways can be ordered by the longest prefix first as in the IP routing mechanisms for a faster matching and each gateway has a variable size of the allocated range.

However, in some scenarios, the device changes its mobility management pattern during the functioning for example a car which was static for two days thus receiving low mobility, becomes very dynamic for a short period of time and then becomes again static, for example passing from (1) to (2) to (4) in Figure 4.

In order to be able to cope with these scenarios, the IP address allocated in the first stage, which pertained to a specific gateway passes into the uniform mobility management level, and then again into a state in which pertains to a second gateway.

For example for the WD, an IP address was allocated using the previous described procedure from [G1 IP1 … IPn] range – G1 IPx address. If the device becomes mobile, then the G1 receives this event and announces the PDN anchor that for the specific WD a higher level of mobility is required. A tunnel is established between G1 and PDN anchor for the subscriber following the steps which were previously taken out of the attachment procedure due to the low mobility. Also the forwarding table is modified as following:

[PDN IP1, … IPn] → to local tunneling functionality
[G1 IP1, … IPn] → [1, ….1, 0,1….1] to G1, [0,….0,1,0….0] to tunneling
[G2 IP1, … IPn] → to G2
….
[Gp IP1, … IPn] → to Gp

Figure 7. Example of Extended Forwarding Table

where a second link list of binary values is maintained, with one entry for each destination. These entries are ordered by the number of IP addresses which have the specific node as destination. One entry is a set of bits equal in size with the size of the IP address range which has the respective second link list. A bit is 1 if the destination is the one associated with the entry and 0 if the destination should be searched in the next elements.

[PDN IP1, … IPn] → to local tunneling functionality
[G1 IP1, … IPn] → [1, ….1, 0,1….1] to G1, [0,….0,1,0….0] to G4
[G2 IP1, … IPn] → to G2
….
[Gp IP1, … IPn] → to Gp

Figure 8. 2nd Example of Extended Forwarding Table

This allows for a linear search which is maximum equal to the number of IP addresses in the IP range, which is the case of the current state of the art mechanism, however, being less efficient than the static routing mechanisms. When the WD stops and it requires mobility management only at the level of the new gateway, then the range is modified as to have as destination of the packets the respective gateway as depicted in Figure 8.
This procedure does not introduce any modifications in the duration of the processing of the decision for forwarding the data packet. Also, from a logical perspective, although the WD has as anchor another gateway, it still pertain to the previous one, so in case the WD moves back to the original area, the forwarding table will be further reduced to the initial stage.

The mechanism here presented can be applied also for the current IP routing protocols. However, as the main goal of this paper is the support of low mobility in operator core network, the concept is applied only in the 3GPP Evolved Packet Core.

IV. 3GPP Exemplification and OpenEPC Testbed

For the 3GPP Evolved Packet Core the concept is implementable directly on the data path entities namely on the access network gateways and on the PDN GW.

As the PDN GW contains the functionality for allocating IP addresses it is considered that the mobility management class selection will be also executed by the PDN GW. Prior to any operation, the IP address range allocated to the PDN GW is separated into smaller ranges and separated based on the assessment on the number of devices into different classes based on whether the PDN GW has to sustain tunneling for the mobile devices or whether the downlink data packets have to be routed to a specific access network gateway.

Also the access network specific gateways are able to terminate the mobility managed path and to forward directly the data traffic to the PDN GW in the uplink for the devices requiring only low mobility. Two procedures are considered to be updated compared to the current standards for which the data flows are presented in a high level form next: the network attachment and the mobility management layer modification.

For the access network attachment, the data flow for the LTE was chosen as example. A simplified form is depicted in Figure 9.
First the User Endpoint (UE) sends an attachment request to the eNodeB – the LTE base station which is forwarded to the MME (1), the identity is assessed (2) and the device is authenticated and authorized (3). During this procedure, the user profile is fetched from the HSS (4) which includes also the low/high mobility requirement for the device. The MME will send a modified request for establishing a data session including this indication (5) which will be forwarded to the PDN GW (6). The PDN GW makes a decision based on the indication and on the IP addresses free at the attachment time which mobility management and which IP address to allocate to the mobile device (7). If the uniform mobility management is chosen, then the PDN GW side of the tunnel is established. If the low mobility management is chosen, no mobility related procedure is executed. The decision is communicated together with the selected IP address to the Serving GW (8) which establishes its side of the tunnels and forwards the message to the MME (9) which in its turn notifies the eNodeB on acceptance of attachment to the MME (10). At this moment, a tunnel or static routing is established for the IP address allocated to the mobile device between the S-GW and the PDN GW. Then the link is configured (11) and the attachment is completed (12). Data can be uploaded by the UE through the tunnel established between eNodeB and S-GW and through the mobility management level chosen between S-GW and PDN GW (13). For downlink data (14) there may be an additional procedure of resource reservation modification not depicted.

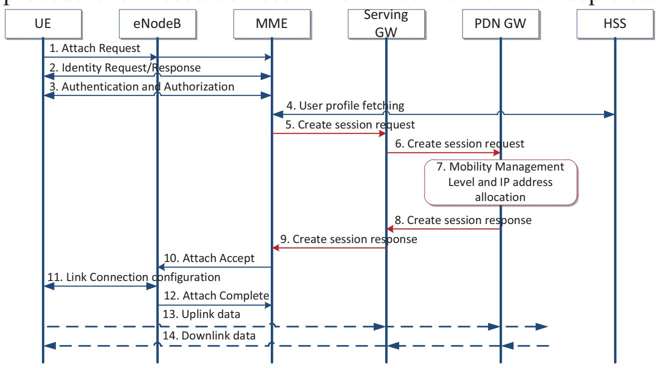

Figure 9. Attachment Procedure

Through this procedure a specific mobility management level is allocated to the UE and the appropriate IP address is allocated.

The modification of the mobility management level of a UE can be triggered by various changing parameters such as a handover to a new eNodeB, a long camping time period in a specific area, the modification of the subscription profile etc. A simplified procedure initiated by the MME is depicted in Figure 10.

When the trigger is received by the MME (1), the UE is already exchanging data through the eNodeB, the Serving GW and the PDN GW with one of the multiple mobility management levels i.e. with a forwarding tunnel between PDN GW and S-GW or through static forwarding. For changing this the MME sends a request for the modification of the session over the specific IP address to the S-GW (2) which forwards it to the PDN GW (3). The PDN GW makes the decision for the modification of the mobility management level and transmits the response to the S-GW while adapting the mobility support accordingly – modifying the forwarding table and the establishment or the termination of the data tunnel (4). A response is issued to the S-GW (5) which in its turn modifies the mobility support accordingly and forwards the request to the MME (6). The data will be further send bi-directionally on the same data path, however, with another mobility management level between S-GW and PDN GW.

The modification of the mobility management level procedure is transparent to the UE. It also is independent of any handover procedures which may occur simultaneously. In this case the modification of the bearer part of the procedure can be integrated with the modification of the bearer in case of the handover between S-GWs. However, this procedure should occur rarely, as the network should be able to detect from the subscription profile which type of wireless device is connected to the network.

Also an interface may be introduced between the PDN GW and the HSS for the PDN GW mobility management level selection functionality which in this case can make the decisions independently of the information received from the MME.

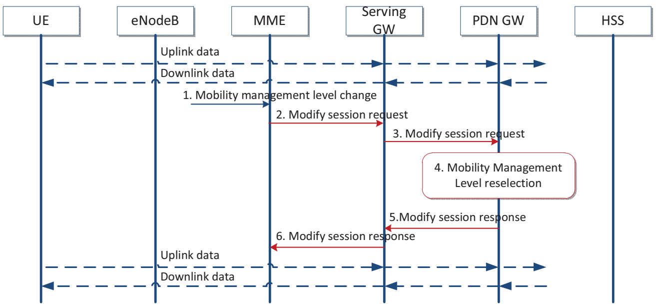

Figure 10. Mobility Management Level Change Procedure

As the data path between the S-GW and the PDN GW can pass through multiple intermediary forwarding nodes, even when static forwarding is used, the data packets should be encapsulated into the backhauling transport header which can be another IP level. However, this can be achieved in a bulk mode for all the devices with low mobility connected to a specific S-GW and not independently for each subscribed node, furthermore reducing the necessary forwarding procedures.

To be able to test the applicability of the presented solution in a real testbed environment the Fraunhofer FOKUS OpenEPC toolkit was used ([8]), depicted in Figure 11. OpenEPC provides the basis functionality for mobility management and resource reservation in both the IPv4 and IPv6 environments, including a standard mechanism of allocating IP addresses to the mobile devices from the PDN GW from a specific address pool. It also provides mobility management signaling using Proxy Mobile IP and GPRS Tunneling Protocol (GTP).

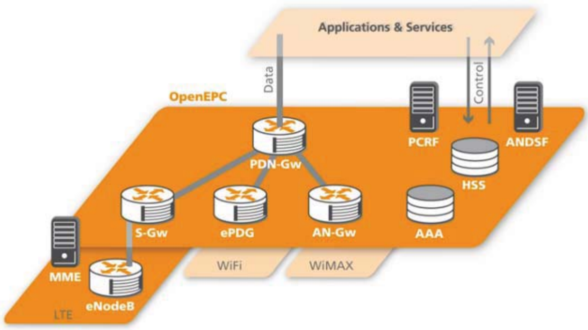

Figure 11. OpenEPC testbed

A UE is able to connect to the OpenEPC platform using an LTE type of access in which the radio link is simulated using WiFi technology and the core network includes the GTP forwarding between an eNodeB stub and a S-GW and a control path through an MME. It may also connect to a public 3G network for which the data traffic is tunneled to a UMTS S-GW or a direct WiFi connection to an ePDG gateway. These ePDG and S-GWs are using GTP or PMIP protocol for the mobility management procedures with the PDN GW. In this testbed the functionality of the gateways was extended as well as the functionality of the PDN GW to support multiple mobility management levels. Also the IP address allocation mechanism separates between static forwarded IP addresses which are allocated to devices which require low mobility management and the IP addresses which require mobility management support at the PDN GW level. Proof of concepts usage scenarios were developed for the here described functionality, demonstrating that it is feasible to deploy such a solution in the current networks. However an in-depth testing of the scalability of the solution has to be further developed through a stress use case, as well as the implementation of the major exception cases such as UE moving out of the mobility area allocated and remaining without mobility management support.

V. CONCLUSIONS AND FURTHER WORK

This article presented a new concept for low mobility support as well as for classifying the wireless connected devices and allocated them to the appropriate mobility management level. This novel approach is required due to the new levels of scalability for the core network which have to be reached in the evolution towards massive wireless broadband environment. From another perspective, this concept allows for a one time signaling and the complete removal of the mobility support for static nodes or for nodes which move into a reduced area, as well as supporting the dynamic selection of the appropriate mobility management level for devices which do not move often.

The concept was exemplified on the 3GPP Evolved Packet Core and evaluated and validated on a prototype implementation using the Fraunhofer OpenEPC toolkit. However a thorough full analysis of the scalability of the solution should be further done as well as the simulation of a very dense environment as well as the extension towards the access networks of the concept enabling a complete static routing for the devices which remain connected to single base station.

Also, the implications of using multiple PDN GWs in a system will be further considered as well as the optimization of the forwarding tables here presented.

REFERENCES

[1] 3GPP TS 23.401, “General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access”, www.3gpp.org;

[2] 3GPP TS 23.402, “Architecture enhancements for non-3GPP accesses”, www.3gpp.org;

[3] 3GPP TS 23.203, “Policy and Charging Control Architecture”, version www.3gpp.org

[4] IETF RFC 5213, “Proxy Mobile IPv6”, www.ietf.org, August 2008;

[5] 3GPP TS 29.274, “General Packet Radio Service (GPRS) Tunnelling Protocol for Control plane (GTPv2-C)”, www.3gpp.org;

[6] 3GPP TS 29.281, “General Packet Radio System (GPRS) Tunnelling Protocol User Plane (GTPv1-U)”, www.3gpp.org;

[7] 3GPP TR 23.888, “System improvements for Machine-Type Communications (MTC)” www.3gpp.org;

[8] Fraunhofer FOKUS OpenEPC toolkit, http://www.openepc.net/