The "Pure-IP" Moby Dick 4G Architecture

Jürgen Jähnert[1], *Jie Zhou*[1], *Rui L. Aguiar*[2], *Victor Marques*[1], *Michelle Wetterwald*[4], *Eric Melin*[5], *Jose I. Moreno*[6], *Antonio Cuevas*[6], *Marco Liebsch*[7], *Ralf Schmitz*[7], *Piotr Pacyna*[8], *Telemaco Melia*[7], *Pascal Kurtansky*[9], *Hasan*[9], *Davinder Singh*[10], *Sebastian Zander*[10], *Hans J. Einsiedler*[11,] *Burkhard Stiller*[13,9,12,]

[1] Rechenzentrum Universität Stuttgart, Germany, [jaehnert|zhou]@rus.uni-stuttgart.de
[2] Instituto de Telecomunicações/Universidade de Aveiro, Aveiro, Portugal, ruilaa@det.ua.pt
[3] Portugal Telecom Inovação, Aveiro, Portugal, victor-m-marques@ptinovacao.pt
[4] Institut Eurecom, Sophia Antipolis, France, michelle.wetterwald@eurecom.fr
[5] Motorola Labs, Paris, France, erik@motorola.com
[6] Universidad Carlos III de Madrid, Spain, [joseignacio.moreno|antonio.cuevas]@uc3m.es
[7] NEC Network Laboratories, NEC Europe Ltd., Heidelberg, Germany, [liebsch, schmitz, melia]@ccrle.nec.de
[8] University of Mining and Metallurgy, Cracow, Poland, pacyna@kt.agh.edu.pl
[9] Computer Engineering and Networks Laboratory TIK, ETH Zürich, Switzerland, [kurtansky|hasan|stiller]@tik.ee.ethz.ch
[10] Fraunhofer Gesellschaft, Institute FOKUS, Berlin, Germany, [singh|zander]@fokus.fhg.de
[11] Deutsche Telekom, Berlin, Germany, hans.einsiedler@telekom.de
[12] University of Federal Armed Forces Munich, Information Systems Laboratory IIS, Germany
[13] University of Zürich, Computer Science Department IFI, Switzerland

ABSTRACT

Network operators, service providers and customers are players who have different interests and raise different requirements on the functionality of future mobile communication networks. However, some new capabilities, such as mobility, security, ubiquity and quality are spelled out by all, which means that there exist some fundamental mechanisms which are in fact needed in every network. This paper concentrates on critical elements of the network infrastructure which need to be deployed in 4G networks before services can be offered. In the paper we discuss these elements, and show how they can be combined to satisfy versatile service requirements. Furthermore, the paper shows how to combine these mechanisms of three traditionally quite separate architectures - for Authentication, Authorisation, Accounting and Charging (AAAC), for Mobility (Mobile IP with Fast Handover), and Quality-of-Service (QoS). A technology-independent paging concept is also integrated in this system. The resulting integrated system architecture is general and can be deployed in heterogeneous environments. Our implementation has recently been completed, validated and verified with applications such as data transfer, voice-over-IP, video streaming and real time concurrent gaming. This prototypical implementation incorporates TD-CDMA, 802.11 WLANs and Ethernet, and treats all transmission technologies as physical and data-link layers, while higher-level functions are supported in a uniform way with an all-IPv6-based signalling.

1 INTRODUCTION

Virtually all papers with the title 'Beyond 3G Mobile Networks' consider IP as the final means of integrating access networks from any technology – wireless or fixed – and the, equally IP-based, core network. This migration from traditional circuit-switched networks toward a packet-based IP network infrastructure adds considerably to the pressure to provide commercialised services in this network. Although some of these aspects have been already addressed, there is still a long path to follow before all citizens are able to run "multimedia communications anytime, anywhere, any style".

Although transmission capabilities have increased drastically, requests of new upcoming applications have evolved in a similar way. In this sense, the fact that more resources are available does not mean that applications will be able to utilize them when needed – which is especially true when we consider the usage of the relatively scarce radio spectrum. On a different wording, resources in wireless environments must be managed in an efficient way. Efficiency within this context does not only denote technology-specific efficiency, but also economic efficiency which will lead to profit maximization – one of the reasons behind the integration pressure towards IP networks. Users paying for privileged service should be treated accordingly. This drives the need to integrate QoS support for differentiated traffic of different users, charging components and AAA architectures in an overall mobility environment. An additional fundamental aspect of these new generation networks covers security, since new communication scenarios will comprise security critical processes like e-administration. On a flexible environment, capable of supporting multiple services dynamically offered to various users, it is essential that multiple security-related aspects are assured, such as privacy of the user information and integrity of the control messages.

The work presented in this paper describes a vision of these future network architectures. We present a heterogeneous architecture supporting mobility, assuring Quality-of-Service (QoS), and with features required to deploy such a network in a provider's framework (i.e., supporting AAA functions enriched with charging). At this stage, minimal security features are further included in this network.

Our research implemented a Pure-IPv6 architecture toward 4G networks, developed under the aegis of the IST Moby Dick project [1], integrating Mobility, QoS and AAA. It supports seamless handover mechanisms across heterogeneous access networks. Mobility-enabled end-to-end QoS and AAA mechanisms are enriched with Auditing and Charging (AAAC). This network is not derived from implementation visions of industry groups, such as 3GPP2, but on evolutions from IP-based networks. Nevertheless, it seems possible to evolve the so-called 3GPP "All-IP" architectures into our own real "Pure-IP" approach.

The rest of this paper is organized as follows: In chapter 2 we will briefly summarize the overall network architecture, covering different access technologies and describing architecture elements and their functions. In chapter 3, mobility-related aspects are presented, covering the deployed FHO (Fast HandOver) mechanism and IP paging in detail. Chapter 4 presents the QoS architecture framework, covering main elements involved and their building blocks. In chapter 5 we present the AAAC system in detail. In chapter 6 we discuss the integration of those Mobility, QoS and AAAC architectures and provide an overall view of the system operation. Chapter 7 present the prototype developed and its evaluation. Finally, main conclusions are drawn in chapter 8.

2 BASIC NETWORK ARCHITECTURE ELEMENTS

The overall network was developed in order to support three different types of access technologies (Ethernet, Wireless LAN and TD-CDMA). The network architecture was developed using three key design principles:

- The network should implement as many functions as possible using standard IP-based protocols and technologies, by reusing as many commonalities in different access technologies as possible.
- The network should be able to provide real-time services with quality comparable to traditional cellular networks.
- The services should be generally accessible regardless of the access network and uninterrupted during handover (change of point-of-attachment).

The overall network architecture, depicted in Figure 1, includes the following elements:

- Mobile end-systems which can be equipped with interfaces from different technologies simultaneously. In particular, interfaces to TD-CDMA (UMTS-TDD), wireless LANs (802.11b), and fixed networks (Ethernet) are supported;
- Access Routers, providing an interface between a wireless and a wired-network domain. These nodes are associated with the traditional concept of a Base Station, the actual access point of 2G and 3G wireless technologies, but enriched with IP capabilities. One Access Router controls one IP subnet, which is directly mapped to one radio cell.
- Network management servers in the fixed network for mobility management, AAA, Charging, QoS, security and paging issues.

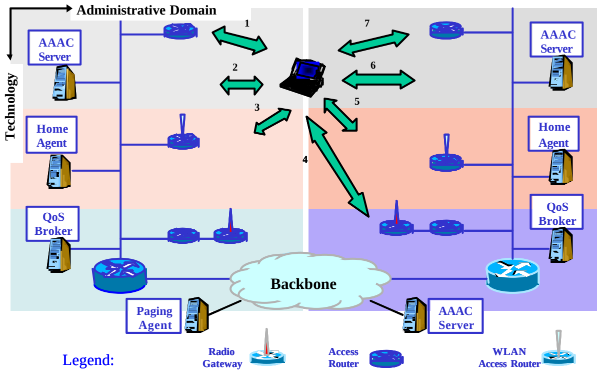


Figure 1: General Network Architecture

All signalling between these entities is exchanged at IP-layer level, in order to achieve a convergence layer independent of the technology. The mechanisms provided by the entities and their proper interworking provide an architecture which supports a cheap and efficient seamless network access on an heterogeneous environment, in an open manner. The architecture flexibility, resorting only to IP for mobility management, QoS provisioning and AAAC, presents a big step towards an IP dominated 4G network architecture, and will allow easy and simple integration of other access technologies. The architecture is completely independent of the access technology, so it can be adapted to incorporate any future spectrum-efficient wireless technologies, like UWB [2].

TD-CDMA support is particularly relevant in this type of environment: it is simply achieved by the direct connection of the base station to the IP network (eliminating several of the elements defined in the 3GPP UMTS architecture like RNC, SGSN, GGSN, and thus, simplifying the overall network) [3]. The base station was developed using a hardware platform supporting TD-CDMA mode. In the WCDMA-TDD interface, the physical layer and the layer 2 comply with the 3GPP standards. The Radio Resource Control (RRC), however, is a subset of the 3GPP protocol described in [4], designed to operate according to the project requirements and interface directly the IPv6 protocol through an adaptation component. The current implementation does not cover all 3GPP stack (unnecessary), as it supports direct connection with IPv6.

Most current radio protocols are 3GPP-compliant, and some additional work would allow the WCDMA-TDD interface in the Mobile Terminal to communicate to either a standard 3G network or a "Pure-IP" network such as described here, these networks being seen as separate Radio Access technologies. Note that the migration from 3G networks to this "Pure-IP" approach in terms of the network would basically imply pushing IP-protocols into the UTRAN. Coexistence has to be managed in the terminal, as discussed above.

In the Moby Dick architecture, network access is provided by a radio AR, which controls a single radio cell. At IP level, an IP subnet is directly mapped on such a radio cell. So, dynamic mobility management is a process managed at IP-level. To provide seamless mobility competitive to the one already achieved in existing cell networks, FHO [5] is used, with some specific enhancements and associated to context transfer techniques. This approach has the advantage to increase the scope of mobility across cells based on different access technologies. Furthermore, to increase the efficiency of both the mobile end systems and the usage of the (scarce) access medium, and simultaneously to improve a low power consumption, a novel IP paging concept was integrated into the overall architecture.

For the provision of QoS, a DiffServ-based model [6] was adopted because of its higher scalability and reduced signalling overhead in this highly dynamical network infrastructure. The association of a DiffServ framework with the use of QoS Brokers (QoSBs) supports quality of service on larger scale. The AAAC support is based on the IRTF Architecture [7], enriched with charging mechanisms to provide an overall architecture targeted to commercial use. To achieve this, additional auditing mechanisms and a metering based on [8] have been also integrated. Security is implemented based on IPSec, and adequate modifications to IPSec packages have been made to support mobility in the access link.

3 MOBILITY SUPPORT

The purpose of mobility support inside Moby Dick is twofold: implementing FHO procedures, able to operate across multiple networks, and developing paging at the network level.

3.1 FAST-HANDOVER

Mobile IPv6 [9] is considered as the basic mobility management scheme for the overall Moby Dick network architecture. While 'standard' Mobile IPv6 is deployed for global mobility management (i.e. inter-administrative domain handover) extensions for local mobility management (i.e. intra-administrative domain handover) are required in real-time environments (such as voice communications) for achieving seamless handover, i.e., for the user not to be aware that a handover is being processed.

There are several different proposals to enhance handover performance, such as hierarchical [10] and non-hierarchical approaches. The analysis and simulation results of the evaluation in [11] showed that a non-hierarchical FHO approach [5] seems to be the most suitable mechanism for this network.

Generally, a handover is composed of two different types: technology specific lower layer handover (which poses rigid technology specific constrains) and IP handover. The chosen IP FHO approach is independent of the access technology (i.e., TD-CDMA, 802.11b or Ethernet) and follows a 'make-before-break' philosophy: layer 3 handover will be prepared via the existing communication channel, before layer 2 handover is performed. In this way, the handover delay is reduced to the minimum amount of time, which is necessary for the reconfiguration of the interface or changing the interface in case of inter-technology handover. During this preparation phase, the current Access Router is informed about the intended handover and, for the duration of the handover, bi-casts all packets destined for the Mobile Terminal (MT) to the current care-off address and copies them to the new point of attachment. This mechanism reduces packet loss during the handover.

3.2 PAGING

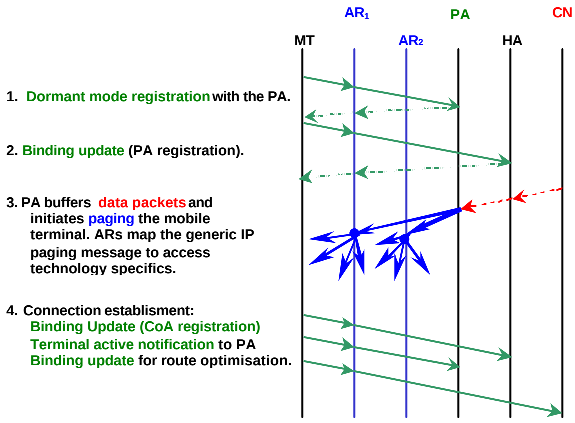
The paging architecture aims for reachability support of roaming dormant MTs and routing of data packets (destined to that MT). As a requirement of the standard Mobile IPv6 concept used in the network, an active MT acquires in each cell a different Care-off Address (CoA) [9], which identifies the IPv6 subnet in which the MT is currently located. This CoA is then to be used for registration with the Mobile IPv6 Home Agent (HA). Tracking of the MT's location, with the preciseness of an IPv6 subnet (i.e. the cell), is required if packets are to be routed to a MT. If no packets are to be routed to a MT, keeping this location information updated with the HA (Binding Update messages) is superfluous signalling and power overhead. Location update related signalling costs further increase if the mobile is moving very fast and the geographical size of an IPv6 subnet is small – which may be a realistic scenario for 4G networks. Hence, if no packets are to be routed from the network to a MT, the preciseness of the MT location in the network could be decreased. This can save frequent location updating (binding updates), decrease signalling costs, and save scarce radio bandwidth - as well as a saving the MT battery power.

The system has then many advantages in supporting “dormant” nodes, where no traffic is flowing for them. To allow MT to enter this dormant mode, reducing the sending of location update messages to the network, a mechanism has to be deployed to notify the respective dormant MT of incoming traffic. After the notification that data packets are waiting to be received, the dormant MT must re-activate itself and re-establish detailed routing information with the network. Within the Moby Dick architecture, a Mobile IPv6 managed network would know a dormant MT location with the preciseness of the registered paging area - which can be a cluster of two or more IPv6 subnets - each subnet supported by an individual AR.

According to the basic Mobile IPv6 protocol, it is required to keep paging and the specific MT state maintenance transparent to Correspondent Nodes (CNs) and to the MTs HA. The only nodes which should be aware of the MT current state (active or dormant) are a dedicated Paging Agent (PA) entity and the MT itself. The motivation for handling paging (both paging areas and related signalling) at the IP layer is to keep the paging protocol within the network, independent of and common to all access technologies. This provides an universal framework to be effectively supported in future heterogeneous access networks. Furthermore, to allow the deployment of access technology specific dormant mode and paging approaches, mapping of the common and generic IP paging protocol to access technology specific functions has to be allowed in the Access Network. This is readily achieved in Moby Dick by means of a paging control supporting function implemented in the ARs.

The PA is responsible of localising a registered dormant MT in case of incoming data destined to it. A paging attendant function is implemented with each AR, supporting paging related signalling as well as mapping between the generic IP paging protocol and access technology specifics. A MT, which decides to switch its state to dormant mode, discovers the responsible PA through the paging attendant function in its current AR, and notifies this PA of its current paging area. The paging area information is retrieved from Router Advertisement messages, since each AR advertises a specific identifier, indicating the paging area it is assigned to. After this, the MT registers the PA address with its HA by means of a standard Mobile IPv6 Binding Update message carrying the ‘Alternate Care-of Address Sub-Option’ (see Figure 2). This additional option allows a MT to register an IP address, which is different to the MTs care-of address. From this point of time the MT can enter in dormant mode and can roam within the registered paging area without the need to send location update information to the network. When roaming across paging areas, the dormant MT issues a paging area update message to the PA.

When a CN addresses data packets to a dormant MTs home address, the HA intercepts these packets and forwards them to the PA by means of IP tunnelling. The PA terminates this tunnel, and buffers the initial data packets until the paging process has resolved the MTs current location, and the MT is again able to receive traffic packets. The PA initiates the paging process sending an IP Paging Request message to (the paging attendant in) all ARs of the registered paging area (Figure 2). Individual paging attendant functions build link-level paging request messages, which are sent through the respective access technology. The current concept and implementation cover wired Ethernet, wireless LAN and TD-CDMA. On reception of one of the link-level paging messages, the MT re-activates and re-establishes detailed routing information with the network, notifying the PA and the HA of its current care-of address (Figure 2). As a result, the data packets buffered at the PA are forwarded to the MT, and further data packets intercepted at the HA are now directly forwarded to the re-activated MT. Route optimisation, through Binding Update information to the CN, is then also possible.


AR1
PA
CN
MT
AR2
HA
1. Dormant mode registration with the PA.
2. Binding update (PA registration).
3. PA buffers data packets and initiates paging the mobile terminal. ARs map the generic IP paging message to access technology specifics.
4. Connection establisment:
Binding Update (CoA registration)
Terminal active notification to PA
Binding update for route optimisation.

CN: Correspondent Node, MT: Mobile Terminal, HA: IPv6 Home Agent, PA: Paging Agent, AR1, AR2: Access Router 1,2

Figure 2: Illustration of the dormant mode registration with the PA, the paging process and re-establishment of routing information after re-activation

4 QoS Architectures for Mobile Service Provisioning

The QoS architecture has to support end-to-end QoS, easily manageable from the operator’s point of view. To achieve this objective, entities and methods had to be defined for the allocation and control of resources in the access networks, able to offer and guarantee end-to-end QoS and maintaining user connectivity and QoS while the MT is moving.

4.1 Overall Architecture

A hard constraint of the architecture is the simultaneous support of mobility and QoS. Several IETF QoS frameworks were considered before designing the final QoS architecture, taking this integrated vision in consideration. The IETF IntServ [12] approach uses per-flow resource reservation and RSVP signalling, but is mobility unaware and has well-known problems of scalability and complexity. The IETF DiffServ [6] approach uses flow aggregation per Class of Service based on priorities, and although it is scalable, it does not offer end-to-end guarantees and has no specific provision for mobility. Although more complex approaches are possible, such as the mix of IntServ (at network boundaries) with DiffServ (in the core), none seems to simultaneously support mobility and execute QoS reservations. Thus, the architecture developed resorts to the use of a DiffServ approach, combined with innovative use of QoSBs, to be able to control and manage the available resources in an efficient way [13]. This architecture is also able to assure that the SLA (Service Level Agreement) contracts are not violated (neither by the users nor by the service provider). The architecture relies on the concept that the user will be provided with a service contracted with the provider. The QoSBs are in charge of allocating resources in the access network, per user and per service (signalled by the CoS – Class of Service), according to the contractual information to the user, and further manage flow aggregation of resources in the core network. The QoSB will also control all QoS aspects of the FHO operation.

4.2 QoS Building Blocks

Figure 3 presents the building blocks of the several elements of the QoS architecture: the MT, the QoSB, and the AR. The Differentiated Services Code Point (DSCP) [14] Marking Software located in the MT marks outgoing traffic according to globally defined rules. A Layer 2 (L2) QoS entity covers specific physical-layer QoS aspects, dependent on the access technology (mostly important for the radio access). At the Access Router, a L2 QoS entity acts accordingly and is associated to a QoS Mapping function. This block includes the mapping between DiffServ QoS parameters and the appropriate parameters of the access network. (TD-CDMA is currently the only technology, which supports layer-2 QoS reservations.)
The bandwidth resources for the cell are controlled in the AR by using the respective DiffServ building blocks (classifier, meter, shaper, and policer, [6]) and it is the place where guarantees are imposed. The QoSB has the overall resource control and will guarantee the overall QoS resources.

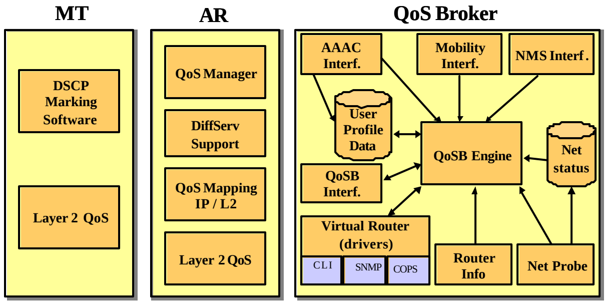


Figure 3: QoS architecture basic components

Traditional DiffServ support for shaping and policing of the traffic to be sent from and to the MT is another block of the QoS architecture. The QoS Manager performs critical control functionalities in relation with shaping, policing, and mobility. Policing functionality is performed when a new traffic flow from the MT is trying to gain access to the Core Network. The policing decision is outsourced to the QoSB following the COPS [15] outsourcing model [16]: the QoS Manager requests information for the appropriate treatment of this new traffic. The QoS Manager also is periodically sending reports of the use of resources to the QoSB. Based on these reports, the QoSB can perform decisions on the traffic in the network, selecting or modifying the appropriate SLA. When a user initiates a FHO to move to a new AR, it is also the QoS Manager that starts a procedure involving the old and new ARs, and the old and new QoSBs, as detailed in section 6.3.

The QoSB is the core entity in the QoS architecture, performing most control decisions, able to operate on a heterogeneous environment. Its core, the QoSB engine, includes all decision algorithms for the QoS management of the network. The QoSB engine operates on an abstraction of the ARs: a Virtual Router Module provides this uniform interface to the QoSB engine and maps the control decision into the specific network commands to each AR. The QoSB incorporates two interfaces: an AAAC interface, to receive authorization information from the AAAC server during the registration process of the user and a QoSB Interface for exchanging information with other QoSBs, in order to provide end-to-end QoS within and between administrative domains. For handling QoS during handover, a Mobility Interface is defined. Three different functions are dedicated to aspects of control and network monitoring: a NetProbe (monitoring network status, collected in the NetStatus database), a RouterInfo (the module that obtains router information, either manually or automatically), and a NMSInterface (that allows the Network Management System entity to define the network resources that can be controlled by the Broker).

4.3 TD-CDMA QoS Support

As already mentioned, the overall Moby Dick architecture provides modules for technology dependent layer-2 QoS support. Figure 4 presents the MT and ARs components involved in the handling of TD-CDMA radio resources. All Radio Resource Management (RRM) functions located in the Core Network of traditional UMTS infrastructures have disappeared. The IP QoS Control procedures based on DSCPs are directly mapped to the physical layer. This new approach is controlled by the QoSB. When a new service is started, it maps the IP DSCP code onto one of the radio QoS classes and sends a request to the QoS Mapping component asking to allocate the corresponding resources. This entity stores the mapping information and forwards the request to the RRC entity of the Radio Interface Protocols constituting the Access Stratum. With the help of a new RRM entity, the RRC computes the changes of the radio parameters needed in the AR and the MT to ensure a proper operation of the other Radio Interface Protocols shown in the figure: PDCP, RLC, MAC and PHY. The list of the parameters to compute is described in [17]. The computation of these radio parameters is based both on the requested QoS mapped from the DSCP code, and on the existing configuration and allocated resources. Their loading enables open of new radio bearers, usually mapped onto dedicated logical transport and physical channels, as shown in Figure 4. Once this radio bearer is available, the Non Access Stratum can use it to transfer user data through the PDCP and the lower layers, according to the previously stored mapping. When the user traffic is over, the QoSB closes the radio bearer in the exact same manner.

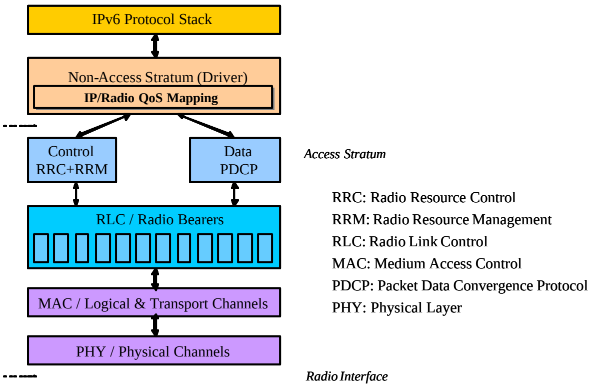

IPv6 Protocol Stack
Non-Access Stratum (Driver)
IP/Radio QoS Mapping
Control
RRC+RRM
Data
PDCP
Access Stratum
RLC / Radio Bearers
MAC / Logical & Transport Channels
PHY / Physical Channels
Radio Interface
RRC: Radio Resource Control
RRM: Radio Resource Management
RLC: Radio Link Control
MAC: Medium Access Control
PDCP: Packet Data Convergence Protocol
PHY: Physical Layer

Figure 4: Layered vision of the MT and AR TD-CDMA architecture (Radio QoS)

5 AAAC ARCHITECTURE

The migration from traditionally circuit-switched networks toward a packet-based wireless IP network infrastructure provokes a demanding pressure to provide packet-switched voice and data services which can be regarded as current key drivers for the development of new communication systems and technologies. In both,the wired Internet and in the future wireless packed-based mobile Internet, an efficient economic concept is still an open issue. Still under research are associated mechanisms required to describe, define and detect IP services and, as a consequence, determine service usage on a finer granularity and, finally charge for this usage in an efficient manner. Along with the Internet Engineering and Research Task Forces' (IETF, IRTF) AAA work, a promising base for these missing functions and mechanisms has been defined [18]. Although basic mechanisms required are available and widely understood, their efficient and scalable integration – especially in the beyond 3G mobile environment – is still a point under research. In order to close this missing gap, the Moby Dick architecture has implemented successfully basic elements and concepts of the IETF and IRTF AAA mechanisms, extended where required. An important issue covers the enhancement of the IRTF AAA architecture with charging and aditing functionality. The Moby Dick AAAC System defines the formerly known AAA Server with specific extensions, thus providing metering, charging and auditing functions. Auditing enables further functions with respect to evaluations of audit trails generated by the AAAC System and others (cf. Figure 5). In Moby Dick, different AAAC Systems communicate with each other via the Diameter AAA Protocol [19]; i.e. the AAAC Server in the home domain (AAAC.h) uses the Diameter base protocol to communicate with the AAAC Server in the foreign domain (AAAC.f).

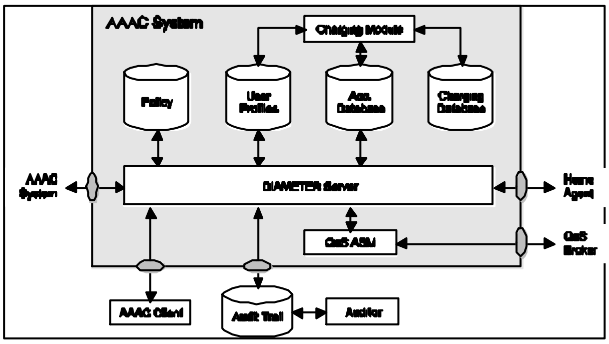


Figure 5: Enhanced AAAC Architecture

With the help of well-defined Application Specific Modules (ASM), an AAA server can communicate to various entities. In such a case, the external entity acts as a quasi AAA Client and communicates with the AAAC System using Diameter. These ASMs can either be integrated into an AAA server or in the external entity. As shown in Figure 6, the enhanced generic AAAC Architecture is applied to the QoS-enabled Mobile IPv6 environment resulting in the instantiation of the ASM and respective service equipments.

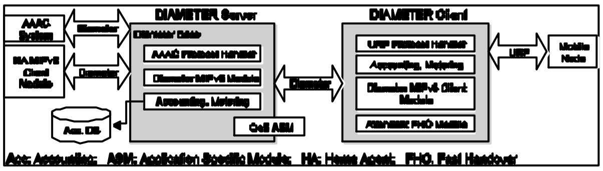


Figure 6: AAAC Architecture for the QoS-enabled Mobile IPv6 Environment

The Moby Dick architecture includes an ASM for a Mobile IPv6-centered Attendant, as well as one for a QoSB. In this particular application, the service user is the MT on behalf of the human user. It is assumed that any request with respect to the authentication and authorization may originate from alternate AAAC Systems, from the MT (therefore, the user), or from the service being supported by the AAAC System, such as the QoSB for a QoS-enabled service or from AAAC Client. The software architecture is modular, thus, separating different processes, e.g., the User Registration module handles the registration process, and other modules execute MIPv6, and Fast handover, accounting. Finding a solution to the problem of authenticating a mobile user posed a major challenge. This was solved using a novel approach to fast handover. In this approach the context of the user is transferred between associated access routers (AR) before such a fast handover is executed.

5.1 Basic Support Components for AAAC

The key support concepts for the enhanced AAAC Architecture include the session model, service definitions, the user profile, and a QoS Interaction model, each of which is discussed below.

Session Concept

The concept of sessions plays an important role in the Moby Dick AAAC environment. The notion of a session ID is introduced to bind together a set of related activities. Each session is associated with a unique session ID and is linked to one user with a unique user ID. The session ID together with the user ID allows a provider to combine accounting data for the different activities and generate a bill.

The AAAC Client creates a new session with a new session ID after several events, e.g., a user switches on his MT and requests authentication and authorisation. After these actions, the AAAC Client produces a Diameter "start" message, which is propagated to the AAAC.h Server. It will generate the first entry for this new session in the accounting database.

Within a running session, a customer can – according to his Service Level Agreement (SLA) – use services offered by the operator. It is also possible that a user has more than one open session. During a session, the AAAC Client produces Diameter "interim" messages, containing accounting data that will be written to the accounting database by the AAAC.h Server. Every authorization is bound to a certain lifetime, and before the lifetime expires, the mobile terminal must perform a re-authentication. Re-authentication must not affect running sessions and the corresponding accounting.

When a session is closed, the AAAC Client produces a Diameter "stop" message and the last entry in the accounting database is created.

User Profiles

A User Profile (UP) is defined as a data record of all user-specific data. This includes authentication data, data of the SLA and auditing information. The UP is unique and is initially created, when the user signs his SLA, which allows or prevents the customer to use services. The UP is stored in the AAAC server of that Moby Dick operator with whom the customer has an SLA established. This Moby Dick operator will offer its services in a certain geographical region, which is referred to as the home domain.

The user-specific information of the UP is needed to identify properly the customer. The next part of the UP contains tariff IDs for each service being offered. These tariff IDs are used by the charging component to select the appropriate tariff function. In addition – for auditing purposes – the UP contains the level of availability guarantee and guarantee of success registration and session setup. Those guarantees are valid for home users only. Finally, the UP contains QoS-related information, called the Network View of the User Profile (NVUP). The NVUP contains DSCPs of those services a user is allowed to use. In the case of a roaming user, the NVUP is transferred to the AAAC.f. The foreign operator will map DSCPs to actual services, according to its service table.

5.2 Metering, Accounting, Charging, and Auditing

After having seen which basic components offer support to the AAAC System's implementation, it is important to detail the metering of user data, the task of performing accounting, the charging process, and the auditing scheme.

Metering

Measuring network traffic has a long history within the IETF. A working group within the IETF, the Real Time Traffic Flow Measurement (RTFM) Working Group, was founded in 1995 and developed an architecture for traffic measurement [8]. Central to this architecture is the notion of traffic flows. An IETF RTFM flow is a bi-directional stream of packets between two endpoints, each defined by a set of attribute values, which can be determined flexibly. Via these flows, the notion of a virtual "connection" is introduced to the IP layer. Historically, this architecture was designed for accounting issues and as a base for further charging and billing processes in order to detect and economically handle the resource/service usage more accurately. Heart of the traffic flow measurement architecture is the Meter. Meters are placed at measurement points determined by the network operator. Each meter selectively records network activity as directed by its configuration settings. It can also aggregate, transform, and process the recorded activity before the data is stored. Processed and stored results are called the usage data. In Moby Dick the IETF RTFM metering framework is used. In order to fulfil Moby Dick specific requirements, the existing IETF RTFM reference architecture has been modified and integrated into the overall AAA Framework. An AAA/Meter interface was implemented in order to let the AAAC Attendant access the accounting database and to configure the meter.

Accounting

Accounting defines the process of collecting data on resource consumption. Used resources are metered by the metering infrastructure after receiving a trigger from the AAAC Attendant (also called AAAC Client) at the start of a session. The AAAC Client receives metering data and sends it to the AAAC.h via AAAC.f. The accounting data is aggregated at the AAAC.h and made available to the charging and auditing entities.

The AAAC Attendant receives answers from AAAC.f for AA requests issued. In case of a negative answer the answer is forwarded to the meter client and no accounting takes place. In case of a positive answer the AAAC Attendant configures its accounting functionality according to accounting policies contained in the AA answer. If no such policies are present it configures accounting according to a default configuration specified before the start of the client. If errors may occur after forwarding the AA answer, the client immediately generates a accounting start request and sends it to the AAAC.f. During a session runtime the attendant generates interim accounting messages and sends them to the AAAC.f.

The AAAC Attendant offers an interface to the metering functionality on the AR. It sends a message to start a meter – containing a filter spec and all other configuration parameters – when accounting is started. It indicates the meter to stop and sends messages containing a filter spec or handle previously received to stop metering for that particular session.

Charging

Charging calculates the price for a given service consumption based on accounting information and the SLA. It maps technical values into monetary units by applying a tariff. The accounting data together with the user profiles covers all information needed by the CM (Charging Module). In Moby Dick an initial scenario of the post-paid business case has been investigated. Therefore, it is sufficient that the CM will contact the accounting database periodically and extract new records that have not been yet used. Charging is session-based, i.e. only closed sessions are being used. After selecting new accounting data, the CM extracts the user (in more detail, the User ID) associated with the current session (Session ID). The User ID is the key index used by the CM to select the appropriate SLA from the user profile database. At this stage, the CM has obtained all necessary information to apply the tariff function for a given customer and a given session. The result of this charge calculation is written to and stored in the charging database located in the AAAC.h. Customers are provided with a web-based user interface allowing them to login and view their charges.

Auditing

Moby Dick's auditing objective covers the detection of a SLA violation. Moby Dick defines the following commitments: availability, success of user registration, and success of session setup. The availability guarantee defines the availability of Moby Dick entities responsible for user registration, session setup, and service delivery, in unit of time or in percentage within each period of a predefined length. Those entities encompass the AAAC Client, the AAAC Server, the QoS Manager, and the QoSB.

The success of a user registration within <n> minutes is guaranteed, if at least <r> valid retries have been made with a valid Network Access Identifier (NAI) and respective credentials. The guarantee is applied only to users in the home domain. To determine whether registration attempts are successful each AAAC Client must log all valid registration requests and responses.

The success of a session setup within <n> minutes is guaranteed, if at least <r> valid retries have been made with valid DSCP. The guarantee is applied only to authenticated users in the home domain. To determine whether session setups are successful each QoS Manager must log all valid service requests sent to the QoSB and its corresponding responses.

Figure 7 shows the respective logging and auditing architecture. Events and actions of AAAC Clients, AAAC Server, QoS Managers, and QoSBs, which need to be logged, will be stored in a local log via an integrated logger. This local log will be managed by the Local Log Management module, which has the responsibility to transfer this log to the Central Log Management module. The Central Log Management module stores this log to the Audit Trail before being fetched by the auditor to be examined with respect to the original SLA. In processing this audit trail, the auditor will make use of the user database and audit rules, which define violation conditions. The result of such an auditing process is stored in an audit report.

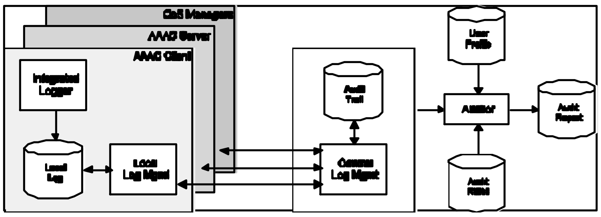


Figure 7: Logging and Auditing Architecture

6 Mobility, Quality of Service and AAAC Integration

Given the overall concepts described so far, three distinct situations arise in the overall architecture which are described in greater detail in the following section: i) registration - in this architecture, a MT/User may only start using network resources after authentication and authorisation, as in today's cellular networks; ii) authorization – the user has to be authorized to use specific services; and iii) handover - the user needs to have its existing resources reservations during a dynamic change of the AR.

6.1 Registration

The Registration (see Figure 8) process (AAA and Mobility) is initiated after a CoA is acquired by the MT via stateless auto-configuration, using unique layer-2 identifiers to avoid the need for Duplicate Address Detection. However, getting the CoA does not entitle the user to consume resources besides registration messages – but allows emergency calls. For accessing the network otherwise, the MT has to start the authentication process by sending the authentication information (message 1, Figure 8) to the AR. That request will be then forwarded to the AAAC System (message 2) responsible for that AR.

In a roaming case, which is more complex, the AAAC system of the visiting domain will issue an AA registration request to the MTs home AAAC (message w). This AAAC checks if there is a formal contractual relationship between the administrative domain this request comes from and the own administrative domain. Then, in case of a positive result, the home AAAC performs authentication by verifying the provided credentials. In the positive case, the home AAAC sends to the HA a request for that user (x) to which the HA answers (y). Then, the home AAAC finally answers the Domain A AAAC (z). One attribute of this positive acknowledgement is the user profile containing all required information to provide the services requested in the foreign domain.

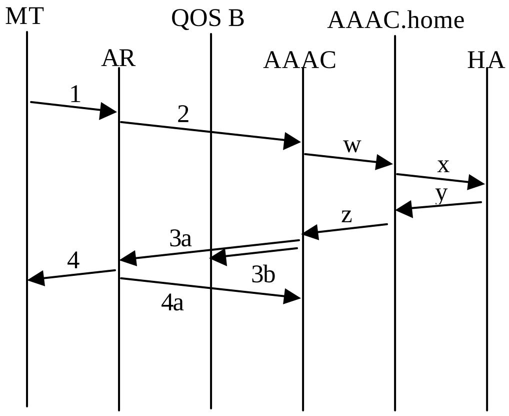


Figure 8: Registration Scenario

The NVUP will be forwarded from the AAAC server to the QoSB (3b), and the rest of the profile will be sent to the AAA Attendant located at the AR (3a). (The NVUP contains all required information relevant to the service provisioning, while the metering and security related information will be forwarded to the AAA Attendant.) The AAAC will also inform the MT of the successful registration, via the AR (3a and 4). Afterwards, the AR will initiate accounting for that user, and informs the AAAC (message 4a). The authentication process is thus completed.

6.2 AUTHORIZATION / SESSION SETUP

Figure 9 shows how each network service is authorized. First, the MT sends a packet (message 5, Figure 9) (either with real information or just a dummy packet) with the DSCP code marked to request a particular service. If the requested packet does not match any policy already set in the AR, the AR issues a request (6) to the QoSB, through the QoS manager (5a). Upon analysing the request, and based on the User NVUP and on the availability of resources, the QoSB eventually decides on an answer to the AR (7). The AR QoS manager will then configure the AR with the appropriated policy for that User/MT service (7a), or informs the User/MT of a service denial (7b). After (7a), any other packet sent from the MT that matches the configured policy rule will be able to cross the network (8). Non-conformant packets will restart the authorization process once more. Upon reaching the end-domain where the other User or CN is, the marked packet starts another QoS authorization process (8a).

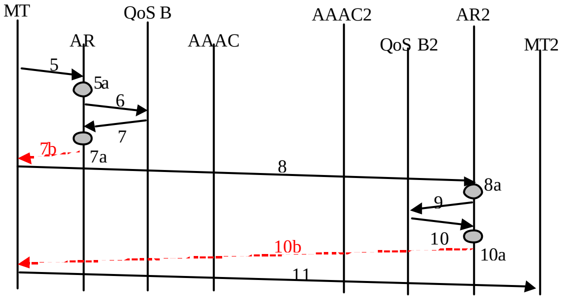


Figure 9: Session setup

The QoS manager on this AR will send a policy query to the QoSB (9), to which the QoSB answers (10). If there are resources, this answer is a positive answer and the QoS Manager will configure the AR (10a) with the right policy, otherwise, the AR will send back a service denial message (10b). After (10a), the next packets matching the installed policy will be able to reach the other terminal (11). With this approach both access networks provide contracted levels of QoS; core network is then monitored to check if the expected performance is being provided, in a DiffServ approach.

6.3 QOS ENABLED HANDOVER

One of the most difficult problems when dealing with IP mobility is assuring a constant level of QoS. User mobility is assured in our network by means of FHO techniques in combination with message exchange to and between the QoSBs during the handover, as shown in Figure 10 and Figure 11.

When a MT starts losing signal strength to the current AR (“old AR”) (message 1, depicted in Figure 10), it starts a handover procedure to a neighbouring AR (“new AR”) from which it receives a beacon signal with the network prefix advertisement (2). The MT builds its CoA and initiates the handover procedure, sending an IP-handover request to its new AR, but still through the old AR (3). The FHO module in old AR will forward this request to its QoS Manager (3a) and to the FHO module on the new AR (known by the network prefix) (4a). The QoS Manager immediately forwards this request to the QoSB (4b) (“old QoSB”). The old QoSB sends a handover request to the new QoSB (5), indicating the User's NVUP and the list of services currently being used. Basically, this acts as a context transfer from the old QoSB to the new QoSB. With this information, the new QoSB will verify the availability of resources, and sends a message to the QoS Manager on the new AR (6) indicating whether the MT may or may not perform the handover. This mechanism allows the QoSB to abort the handover due to QoS constraints (e.g., missing bandwidth resources). If the handover is possible, the QoS Manager then sends this information to the FHO module (6a) and performs the configuration of the new AR to accommodate the MT when it moves. Meanwhile, the FHO module sends the handover reply back to the FHO module on the old AR (7). Upon receiving this message, the old AR sends it to the MT (8). When the answer is positive, the MT sends a handover execute to the old AR (9), to which the old AR reacts initiating bi-casting, setting up a timer (9a) and replying to the MT (10). The MT may now perform a successful layer 2 handover (10a), because all the layer 3 resources were previously reserved, which results in sending a neighbour advertisement to the new AR (12). The new AR starts an accounting process in the AAAC system for that user (13). To complete the handover process, the MT must send the binding update to its HA (14) that replies with a binding acknowledge. Meanwhile, the bi-casting timer on the old AR expires (10b), which also makes it send all accounting information relative to that user to the AAAC system (11). The MT has now completed the handover to the new AR (15). The CN will notice the handover by receiving IP packet with the new CoA. IP packets will be sent to the new address. If the handover is done inside the domain of only one QoSB, that is, if the QoSB controlling the new AR and old AR is the same, then message 5 will not exist. Everything else is exactly the same.

Notice that AAAC attendants are also informed of the handover, and the current user AAAC parameters (for metering, e.g.) are exchanged directly via the Handover initiate messages (message 4a).

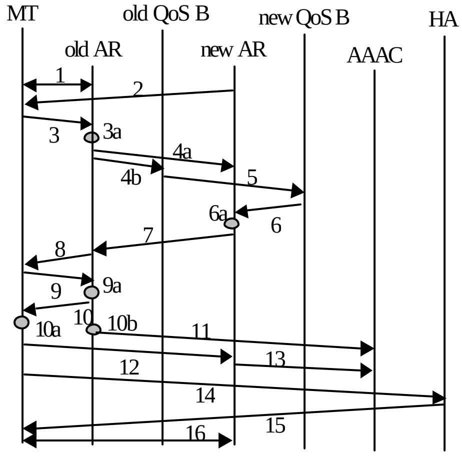


Figure 10: Integrated QoS, AAAC and Fast Handover Logical view

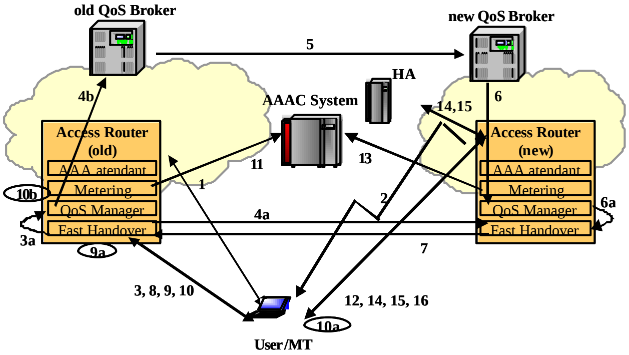


Figure 11: Integrated QoS, AAAC and Fast Handover – Entities view

7 PROTOTYPE IMPLEMENTATION AND TESTS

In order to show and to prove the Moby Dick concept, two Moby Dick test beds were built up. The test beds were located in Madrid (Spain) and in Stuttgart (Germany), and had all the elements defined in the Moby Dick architecture:

Network access is enabled via DiffServ Edge routers where the Mobile Terminals attach to: i.e. Access Routers. These Access Routers have also AAA attendants, Auditing and Paging functionalities. QoSB, Home Agent, AAAC and auditing server and Paging Agent are located in the core network. Madrid's test bed supports two access technologies, Ethernet and WLAN. Stuttgart test bed supports additionally TD-CDMA.

7.1 MOBY DICK PROTOYPE

The network technology employed in the access networks, core networks and for the interconnection of trial sites is pure IPv6. The two test beds, connect via the public IPv6 network and define two administrative domains, thus it was possible to build a roaming scenario as a basis of extending the system to a European-wide scale. All the network nodes run Red Hat 7.2 with Linux 2.4.16.

The DiffServ and QoS implementations were mostly implemented in the kernel space, while control was placed in user space. Handover modules were kernel implementations. The paging software was implemented as a user space program. The AAAC software, including auditing and metering, was based on user space modules controlled by direct access through a shell (e.g. the registration) or via a Web-based interface (e.g. access of the auditing data). TD-CDMA was a pure kernel implementation embedded in Real-Time Linux. The interconnection between the different modules was implemented via sockets or command emulation.

For demonstration purposes standard applications for voice over IP, gaming and video streaming including in the installation packet of Linux were used.

7.2 Moby Dick Concept Evaluation

The Moby Dick evaluation phase involved two kind of test: those performed internally by partners (expert evaluation), and those made by external users - students- unaware of Moby Dick details (user evaluation). The former provided valuable feedback to developers and served to perform a quantitative evaluation of network behaviour; the latter provided a qualitative view of the results achieved, via the usage of common applications over a 4G test bed. The overall usage perception of the network showed that we had successfully implemented a prototype of a 4G network.

Among the results of the expert evaluation we relate briefly two of them. The average time it took to register a user in its home domain was about 280 ms and 350 ms for a roaming user, including transfer of all the sub parts of the user profile as described in section 6.1. The major factor in this time elapsed is the Diffie Hellmann key calculation in the AAA client. The round trip time between AAAC servers in Madrid and in Stuttgart (in case of roaming) is about 70 ms. Another result is related to FHOs. For intra- and inter-technology fast handovers delays were always less than 20 ms (delay between message 3 and message 10 of Figure 10) except when TD-CDMA was involved. The largest factor in this FHO time covers the context transformation in the old QoSB.

8 Conclusions

We presented an advanced architecture being able to provide diversified services in heterogeneous access infrastructures: wired LANs, wireless LANs, and TD-CDMA cells. This architecture is a first approach to 4G systems relying on the IPv6 protocol, and replaces most technology-dependent tasks by IP-oriented approaches.

This paper provides a snapshot of the architecture, showing how proposals under development in the IETF were re-used and mastered in order to achieve this goal. The key problems solved by this architecture are concerned with efficient handover solutions across different cells which require an efficient involvement of the three traditionally separated disciplines QoS, AAAC and mobility management. The architecture puts emphasis on the use of Mobile-IP and its FHO optimisations, a DiffServ-based QoS transport infrastructure managed by QoSBs, and a coherent AAA management structure. An innovative IP paging concept was also presented and discussed. The architecture identifies similar elements across all access technologies, but maintains enough flexibility to support optimisations for physical layers. This architecture is conceptually flexible and open and provides a clear separation between technology and administrative domains.

This approach facilitates the deployment of multiple service provision models, as it decouples the notion of service (associated with the user contract) from network management tasks. It also allows for specific optimisations to be made for specific physical layers. In summary, the Moby Dick project provides a simple, flexible architecture being able to support multimedia service provision in future 4G networks, as our experts and non-expert evaluations showed.

The approach does not cover the possibility for backwards compatibility to 3G networks. However, by modifications of the Gateway GPRS Support Node (GGSN in 3G Mobile Communication) to act as an AR, the UMTS Terrestrial Radio Access Network (UTRAN) can be seen as an additional access technology.

9 Acknowledgements

The work presented was partly funded by the EU project IST-2000-25394 “Moby Dick” (Mobility and Differentiated Services in a Future IP network). We would like to thank all our colleagues, who contributed to this paper and the project.

10 Table Of Acronyms

2G	Second Generation
3G	Third Generation
4G	Fourth Generation
AAA	Authentication, Authorization, and Accounting
AAAC	Authentication, Authorization, Accounting, and Charging
AR	Access Router
ASM	Application Specific Modules
CDMA	Code Division Multiple Access
CM	Charging Module
CN	Correspondent Node
CoA	Care-of-Address
COPS	Common Open Policy Service
CoS	Class of Service
DiffServ	Differentiated Services
DSCP	DiffServ Code Point
EU	European Union
FHO	Fast HandOver
GGSN	Gateway GPRS Support Node
HA	Home Agent
ID	Identification
IETF	Internet Engineering Task Forces
IntServ	Integrated Services
IP	Internet Protocol
IPSec	IP Security
IPv6	Internet Protocol version 6
IST	Information Society Technology
L2	Layer 2
IRTF	Internet Research Task Forces
MAC	Medium Access Control
MT	Mobile Terminal
NSIS	Next Step in Signalling
NVUP	Network View of the User Profile
PA	Paging Agent

PDCP	Packet Data Convergence Protocol
PHB	Per-Hop-Behaviour
PHY	Physical Layer
QoS	Quality of Service
QoSB	QoS Broker
RLC	Radio Link Control
RRC	Radio Resource Control
RRM	Radio Resource Management
RTFM	Real Time Traffic Flow Measurement
SLA	Service Level Agreement
TD-CDMA	Time Division CDMA
UMTS	Universal Mobile Telecommunications System
UP	User Profile
UTRAN	UMTS Terrestrial Radio Access Network
UWB	Ultra Wide Band
W-CDMA	Wideband CDMA
WLAN	Wireless LAN

11 REFERENCES

[1] Moby Dick: *Mobility and Differentiated Services in a Future IP Network;* EU Project IST-2000-25394, http://www.ist-mobydick.org, January 2003.

[2] UWB: *Ultra Wideband Working Group;* http://www.uwb.org/, September 2004.

[3] *3GPP Technical Specification* TS 23.002, v5.0.0: *Network Architecture* (Release 5), October 2000.

[4] M. Wetterwald et al.: *An Original Adaptation of the UMTS Protocols for a Direct Interconnection with IPv6;* 13th IST Mobile and Wireless Communications Summit, Lyon, France, June 2004.

[5] R. Koodli: *Fast Handovers in Mobile IPv6;* IETF, work in progress, draft-ietf-mobileip-fast-mipv6-05.txt, September 2002.

[6] D. Black et al.: *An Architecture for Differentiated Services;* IETF, RFC 2475, December 1998.

[7] C. de Laat, G. Gross, L. Gommans, J. Vollbrecht, D. Spence: *Generic AAA Architecture;*, IETF, RFC 2903, August 2000.

[8] N. Brownlee, C. Mills, G. Ruth: *Traffic Flow Measurement: Architecture;* IETF, RFC 2722, October 1999.

[9] D. Johnson, C. Perkins, J Arkko: *Mobility Support in IPv6*; IETF, work in progress, draft-ietf-mobileip-ipv6-19.txt, October 2002.

[10] H. Soliman, C. Castelluccia, K. El-Malki, L. Bellier: *Hierarchical Mobile IPv6 Mobility Management (HMIPv6);* IETF, work in progress, draft-ietf-mobileip-hmipv6-07.txt>, October, 2002.

[11] H. Hartenstein, M. Liebsch, X. P. Costa, R. Schmitz: *A MIPv6, FMIPv6 and HMIPv6 Handover Latency Study: Analytical Approach;* IST Mobile & Wireless Telecommunications Summit, Thessaloniki, Greece, June 2002, pp. 100-105.

[12] Braden, R., Clark, D., S. Shenker: *Integrated Services in the Internet Architecture: An Overview;* IETF, RFC 1633, June 1994.

[13] V Marques et al.: *An IP-based QoS Architecture for 4G Operator Scenarios;* IEEE Wireless Communications Magazine. June 2003..

[14] K. Nichols, S. Blake, Fred Baker, D. L. Black: *Definition of the Differentiated Service Field (DS Field) in the IPv4 and IPv6 headers;* IETF, RFC 2474, December 1998.

[15] D. Durham (edt): *The COPS (Common Open Policy Service) Protocol;* IETF, RFC 2748, January 2000.

[16] S. Salsano: *QoS Control by Means of COPS to Support SIP-based Applications;* IEEE Network, March/April 2002, Vol. 16, No 2.

[17] *3GPP Technical Specification* 25.331 V3.12.0 (2002-09), 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Radio Resource Control (RRC); Protocol Specification (Release 1999).

[18] S. Glass, T. Hiller, S. Jacobs, C. Perkins: *Mobile IP Authentication, Authorization, and Accounting Requirements;* IETF, RFC 2977, October 2000.

[19] P. Calhoun et al.: *Diameter Base Protocol;* IETF, work in progress, draft-ietf-aaa-diameter-07.txt, IETF, July 2001.