Interworking Techniques and Architectures for Heterogeneous Wireless Networks

Karl Andersson
Pervasive and Mobile Computing Laboratory
Luleå University of Technology
SE-931 87 Skellefteå, Sweden
karl.andersson@ltu.se
Tel: +46-0910-585364

Abstract

Fourth generation (4G) wireless systems targeting 100 Mb/s for highly mobile scenarios and 1 Gb/s for low mobility communication are soon to be deployed on a broad basis with LTE-Advanced and IEEE 802.16m as the two candidate systems. Traditional applications spanning everything from voice, video, and data to new machine-to-machine (M2M) applications with billions of connected devices transmitting sensor data will in a soon future use these networks. Still, interworking solutions integrating those new 4G networks with existing legacy wireless networks are important building blocks in order to achieve cost-efficient solutions, offer smooth migration paths from legacy systems, and to provide means for load balancing among different radio access technologies.

This article categorizes and analyzes different interworking solutions for heterogeneous wireless networks and provides suggestions for further research.

Keywords:Heterogeneous Wireless Networks, Interworking Techniques and Architectures, Tight Coupling, Loose Coupling.

1 Introduction

One important trend within the area of wireless networking is heterogeneity. No single wireless radio access technology will deliver all required services to all end-users anywhere, anytime. It will rather be the case, a variety of radio access technologies together forming the wireless infrastructure in each geographical area. Overlapping coverage is a typical feature where there is a choice for the end-user to connect to more than one radio access technology, either within the same administrative domain or across administrative borders. This architecture model takes full advantage of existing investments by infrastructure owners. Furthermore, it allows for increased wireless capacity and for backward compatibility. Also, it could offer higher data rates in selected areas at a lower cost. Finally, it allows for enhanced competition and flexibility.

At the same time, a growing proportion of new mobile telephones, smartphones, and laptops are equipped with more than one radio access technology. The opportunity of automatically being able to connect to various types of wireless networks is today a reality for huge numbers of users.

Also around the corner are the fourth generation (4G) wireless systems. They are targeted toward data rates of 100 Mb/s for highly mobile scenarios and 1 Gb/s for low mobility communication. The Third Generation Partnership Project (3GPP) has proposed LTE-Advanced as their candidate system, while the IEEE has proposed the IEEE 802.16m as their candidate system. These systems will be packet-only and not be offering any circuit-switched subsystem handling voice services separately that was the case for second generation (2G) and third generation (3G) wireless networks. Also, the transition toward 4G systems will take a very long time.

As a consequence, interworking solutions integrating those new 4G networks with existing legacy

Journal of Internet Services and Information Security (JISIS), volume: 2, number: 1/2, pp. 22-48

wireless networks are important building blocks in order to achieve cost-efficient solutions, offer smooth migration paths from legacy systems, and to provide means for load balancing among different radio access technologies. This article categorizes and analyzes different interworking solutions for heterogeneous wireless networks and provides suggestions for further research.

The rest of the article is organized as follows. Section 2 presents related work, while Section 3 outlines a few scenarios. Section 4 provides background on the most common radio access technologies being used today, while Section 5 introduces a number of key technologies related to the area. Section 6 analyzes a number of standards-based interworking techniques, while Section 7 concludes the work and indicates further research.

2 Related Work

The activity level of research in the area of interworking solutions for heterogeneous wireless networks has been relatively high during the past decade. Researchers from both academia and industry collaborate in various consortiums trying to define architectural solutions to the "Always Best Connected" vision outlined by Gustafsson et al. [37] in 2003. The authors describe the concept of being best connected and discuss the user experience and business relationships in an Always Best Connected environment.

Yiping et al. [67] described a new architecture supporting Always Best Connected services covering an access discovery mechanism integrating service location protocols and location-based services. A personalized network selection scheme is proposed, as well as a Mobile IPv6-based seamless vertical handover scheme. The authors argue through analytical solutions both end-users and network operators would benefit from using their solution.

Perera et al. [46] proposed a mobility toolbox architecture for All-IP networks including support for Mobile IPv4, Mobile IPv6, NEMO, and HIP. This coexistence is effected by means of a mobility toolbox enabling mobility handling to be selected according to context. The design of the toolbox is described as a component of the Ambient Networks architecture. Feasibility and performance gains are demonstrated with a prototype implementation of network mobility.

Eastwood et al. [24] showed how IEEE 802.21 supports seamless mobility between IEEE 802.11VHT and IEEE 802.16m networks. The proposed mobility management scheme integrated these two access technologies along with IEEE 802.21 into one IMT Advanced (4G) compliant system.

Kong et al. [41] analyzed both qualitatively and quantitatively network-based approaches and host-based approaches to the mobility management problem. They stressed the key features of Proxy Mobile IPv6 and expected that it would be the mobility protocol of choice when realizing the next-generation All-IP mobile networks.

Pontes et al. [49] described integration issues between IEEE 802.11 and IEEE 802.16 networks in both infrastructure networks and ad hoc networks. They surveyed solutions from IETF, 3GPP/3GPP2 and IEEE and proposed using the IEEE 802.21 framework.

Song et al. [58] proposed a new architecture for integration of IEEE 802.16 and 3GPP networks. A new network element, the Data Forwarding Function, was introduced in order to eliminate data loss during vertical handover. By simulations, the authors show that the proposed solution is effective in minimizing data loss during vertical handovers and point out that the solution is general and can be applied to other access networks as the handover solution is IP-based.

Salkintzis [52] categorized and analyzed interworking techniques and architectures for WLAN/3G integration. The interworking scenarios covered were (1) common billing and customer care, (2) 3G-based access control and charging, (3) access to 3G packet-switched services, (4) access to 3G packet-switched-based services with service continuity, (5) access to 3G packet-switched-based services with seamless service continuity, and, finally, (6) access to 3G circuit-switched-based services with seamless mobility. Salkintzis focused on the second and the third scenarios concluding that 3G subscribers can benefit from high-throughput IP connectivity at hot spot locations and also maintain access to the same 3G packet-switched services across several radio access technologies. Session mobility would be an evolutional step beyond service roaming offered by the third scenario.

Ferrus et al. [35] developed a framework to categorize interworking solutions by defining a set of interworking levels and their related key interworking mechanisms. The proposed framework was used to analyze a few of the most relevant interworking solutions including I-WLAN, GAN, WiMAX-3GPP networks interworking, and IEEE 802.21 interworking. Ferrus et al. also proposed a number of interworking levels similar to those being propsed by Salkintzis. Ferrus et al. distinguished four interworking levels: (A) visited network service access, (B) intersystem service access, (C) intersystem service continuity, and, finally, (D) intersystem seamless service continuity. Intersystem AAA, Intersystem user data transfer, network layer handover, and network layer handover optimization were the key interworking mechanisms used at each level.

3 Scenarios for Interworking in Heterogeneous Wireless Networks

As previously mentioned, many handsets are today equipped with more than one radio access technology. Those handsets have the opportunity to connect to a wireless network in the vicinity. Coverage is typically overlapping.

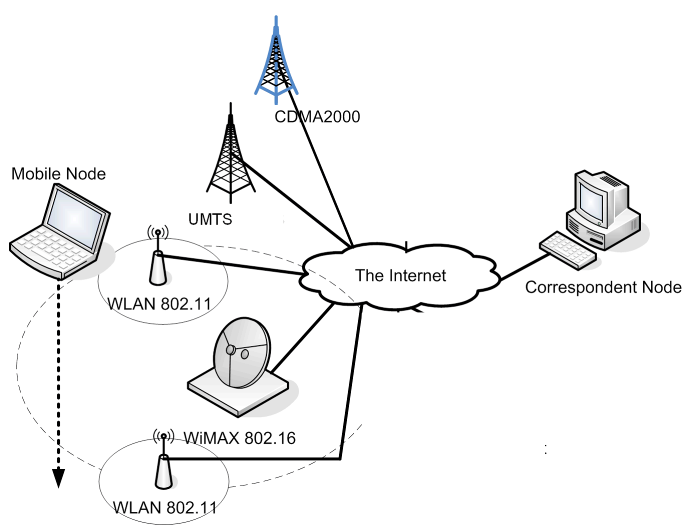

Figure 1: Interworking scenario in a heterogeneous wireless networking environment

End-user services can either be delivered in an access network specific manner or independent of the involved wireless access networks as depicted in figure 1. The scenarios discussed in this article are the following:

Scenario 1 is targeting the event of end-users visiting networks outside the domain of their home network. This scenario involves checking of credentials by the visited network to the home network and transfer of accounting information from the visited network to the home network.

Scenario 2 offers access to end-user services available only in the home network to users while roaming to visited networks.

Scenario 3 extends scenario 2 with seamlessness so that end-users do not have to re-establish their connections that were set up while them being connected to their home network.

Scenario 4 extends scenario 3 with support for realtime services where very low handover latencies can be accepted.

These scenarios offer an increased level of interworking in a heterogeneous wireless networking environment.

4 Evolution and Characteristics of Wireless Networks

Before analyzing the interworking solutions for heterogeneous wireless networking environments, this section surveys the most commonly available wireless access networks worldwide.

4.1 GSM

The most popular wireless access technology, GSM (Global System for Mobile Telecommunications), was defined in its first version in 1990 by ETSI (European Telecommunications Standards Institute). Initially designed to be used across Europe the standard is today used all over the world. Replacing first generation (1G) analogue systems like NMT (Nordic Mobile Telephony) and TACS (Total Access Communication System), GSM is often referred to as a second generation (2G) wireless access technology. GSM uses licensed spectrum, where 900 and 1800 MHz are the most common frequency bands, although 850 and 1900 MHz are used e.g. in Canada and the United States. Also, installations on the 400 and 450 MHz bands exist in some countries. GSM is used both for outdoor and indoor use.

GSM uses TDMA (Time Division Multiple Access) technology in the radio interface to share a single frequency between several users. The system assigns sequential timeslots to each user sharing one common frequency.

Users are identified via their Subscriber Identity Module (SIM) which is a detachable smart card containing the user's subscription information and his/her phone book. This feature allows users to easily switch handsets. Roaming agreements between GSM operators give the opportunity for end-users to use their handsets in other countries as well.

Communication is secured using a variety of cryptographic procedures. Initially, two codecs were used either at the data rate of 6.5 kb/s (half rate) or 13 kb/s (full rate). Later on, the Enhanced Full Rate (EFR) codec was introduced working at a data rate of 12.2 kb/s.

The GSM network is built up of the mobile station (MS), the base station subsystem (BSS), and the Network and switching subsystem (NSS) (Figure 2). In BSS the Base Station Controller (BSC) controls a number of Base Transceiver Stations (BTSs). NSS consists of two types of switches, the Mobile Services switching Center (MSC) serving subscribers in its service area, and the GMSC (Gateway Mobile Services switching Center) connecting the mobile network to the Public Switched Telephony Network (PSTN). Also, a number of databases are present in the NSS. Subscriber data is stored in the Home Location Register (HLR). In this register there is also information on the identity of the MSC that the subscribers are connected to. The Visitor Location Register (VLR) that is connected to each MSC holds finer granular location data on users in the service area. Finally, EIR (Equipment Identity Register) stores information on valid handsets, while AUC (Authentication Center) holds data on authentication and encryption parameters.

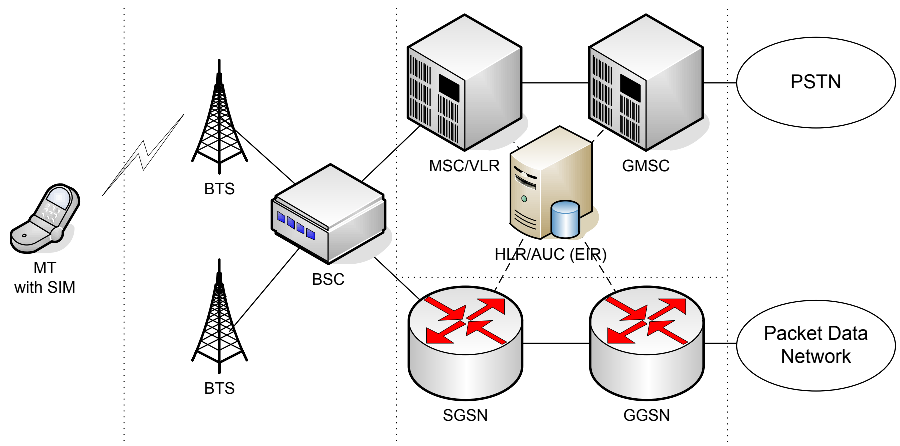

Figure 2: The structure of a GSM network

Support for packet switched data was added in Release 97 when GPRS (General Packet Radio Service) arrived. A new subsystem was added to the GPRS Core Network containing two new node types for GPRS Support: SGSN (Serving GPRS Support Node) and GGSN (Gateway GPRS Support Node). Four coding schemes (CS-1, CS-2, CS-3, and CS-4) using Gaussian Minimum Shift Keying (GMSK) were introduced allowing for different data rates at various levels of robustness. Data rates of 20 kb/s per time slot were reached using the fastest coding scheme. Using four time slots for downlink traffic and one time slot for uplink traffic gave 80 kb/s and 20 kb/s of data rates in each direction. FDD (Frequency Division Duplex) was introduced so that a pair of frequencies was allocated using one channel for downlink traffic and one channel for uplink traffic. The downlink used first-come first-served packet scheduling, while the uplink used a scheme similar to reservation ALOHA (R-ALOHA). This means that slotted ALOHA (S-ALOHA) is used for reservation inquiries during a contention phase, and then the actual data is transferred using dynamic TDMA with first-come first-served scheduling.

In 2003, EDGE (Enhanced Data rates for GSM Evolution) or EGPRS (Enhanced GPRS) was introduced. No hardware or software upgrades were needed in the core network, but EDGE-compatible transceiver units were required to be installed. Also, the BSS needed to be upgraded to support EDGE.

EDGE makes use of 8 phase shift keying (8PSK) as coding scheme allowing for data rates of 59.2 kb/s per time slot. Just like GPRS, EDGE adapts the coding scheme to the quality of the radio channel. Incremental redundancy was introduced so that the need for retransmission of disturbed packets was decreased. S-ALOHA is used for reservation inquiries just as in GPRS. Effective data rates of 236.8 kb/s and 59.2 kb/s for downlink and uplink traffic were achieved respectively if four times slots were used for downlink traffic and one time slot was used for uplink traffic. End-to-end latencies were reduced to 150 ms.

4.2 UMTS

The most important third generation (3G) mobile telephony system is UMTS (Universal Mobile Telecommunications Systems) specified in its first version by 3GPP (Third Generation Partnership Project). Although requiring a complete new infrastructure, concepts and solutions were reused from GSM. The 2100 MHz band was the original frequency band for UMTS in Europe, but operators are nowadays deploying UMTS on a wide range of frequencies in many parts of the world. Peak data rates were initially

384 kbps in both direction and delays around 100 ms.

The air interface used in UMTS is WCDMA (Wideband Code Division Multiple Access) where a pair of 5 MHz-wide channels typically is used for transmission in FDD mode. Spread-spectrum technology is employed where each transmitter is assigned a spreading code to allow multiple users to be multiplexed over the same physical channel.

A number of channel types exist divided into physical channels, transport channels (subcategorized into common transport channels and dedicated transport channels) and logical channels. Small amounts of data may be sent using a contention based uplink channel (Random Access Channel, RACH) or a common downlink channel (Forward Access Channel, FACH) using a common spreading code. Larger amounts of traffic are sent using a dedicated channel (DCH) in both uplink and downlink directions. Higher data rates can be achieved using the latter scheme at the cost of slower connection setup.

The fact that many handsets often support both GSM and UMTS with seamless dual-mode functionality and that combined core networks supporting both GSM and UMTS radio accesses are common today led many to view GSM and UMTS as one unified system, sometimes referred to as 3GSM.

The structure of UMTS networks is slightly changed from the GSM network structure (Figure 3).

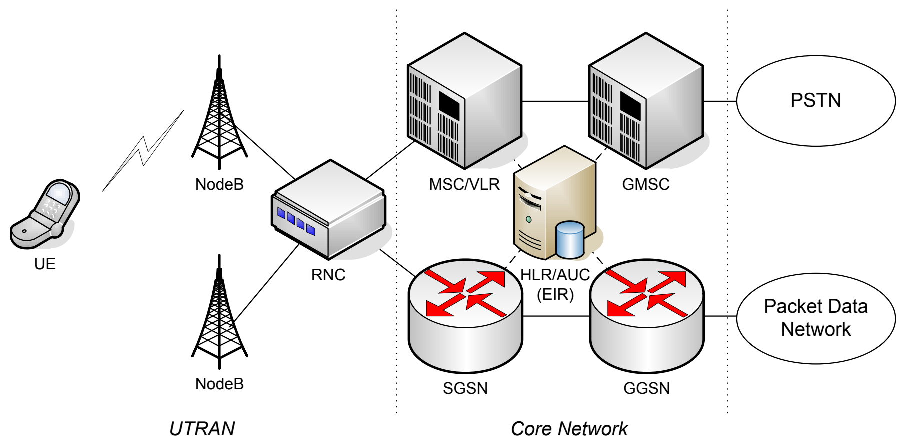

Figure 3: The structure of a UMTS network

Later as HSDPA/HSUPA (High-speed Downlink Packet Access/High-speed Uplink Packet Access) was added data rates could reach as high as 14.4 Mbps in the downlink direction and 5.76 Mbps in the uplink direction and end-to-end delays around 25 ms. The scheduling procedure was changed so that only NodeB performs this task leading to faster resource management. The Downlink Shared Channel (DSCH) was extended to a High Speed Downlink Shared Channel (HS-DSCH) so that multiple spreading codes were used and a fast feedback mechanism on channel conditions was established allowing for adaptive modulation and coding using both QPSK and 16-QAM. The minimum transmission time interval (TTI) was decreased from 10 ms to 2 ms in order to allow for reduced latencies. Retransmissions used HARQ (Hybrid Automatic Repeat Request) performed at NodeB based on feedback from the UE (ACK/NACK). HARQ combines common ARQ with Forward Error Correction (FEC). FEC was used, so that its decoding procedure was based on all unsuccessful transmissions implementing a Stop-and-Wait (SAW) protocol. Two schemes were used: chase combining meaning that same data block is sent at each retransmission or Incremental Redundancy (IR) where additional redundant information is sent at each retransmission.

Evolved HSPA (HSPA+) is expected to offer downlink data rates of 21 Mbps and uplink data rates of 11 Mbps. In HSPA+ NodeBs may connect directly to the GGSN over a standard Gigabit Ethernet connection reducing latencies to 10 ms.

4.3 cdmaOne and CDMA2000

cdmaOne and CDMA2000 form a parallel development track to GSM and UMTS using Code Division Multiple Access as channel access method and a duplex pair of 1.25 MHz radio channels. cdmaOne was first designed by Qualcomm as IS-95 (Interim Standard 95) and used a similar network structure as GSM.

Its successor CDMA2000 is nowadays standardized by Third Generation Partnership Project 2 (3GPP2) and was upgraded from the first 1X version to the Evolution-Data Optimized (EV-DO) versions Rev. 0, Rev. A, and Rev. B. Rev. 0 and Rev. A offer data rates of 3.1 Mbps and 1.8 Mbps in the downlink and uplink directions respectively. Rev. B offers data rates of 14.7 Mbps and 5.4 Mbps in the downlink and uplink directions respectively after hardware upgrade. End-to-end delays are below 35 milliseconds.

Figure 4 depicts the CDMA2000 network structure. The structure is similar to the GSM and UMTS network structures. However, AAA is handled using a Radius (Remote Authentication Dial-in User Service) server in the packet switched domain. Also, the functionality provided by SGSN and GGSN in the GSM and UMTS networks is handled by a Packet Data Service Node including a Foreign Agent (PDSN/FA) and a Home Agent (HA), respectively. Also, some Base Station Controllers are equipped with a Packet data Control Function (PDF).

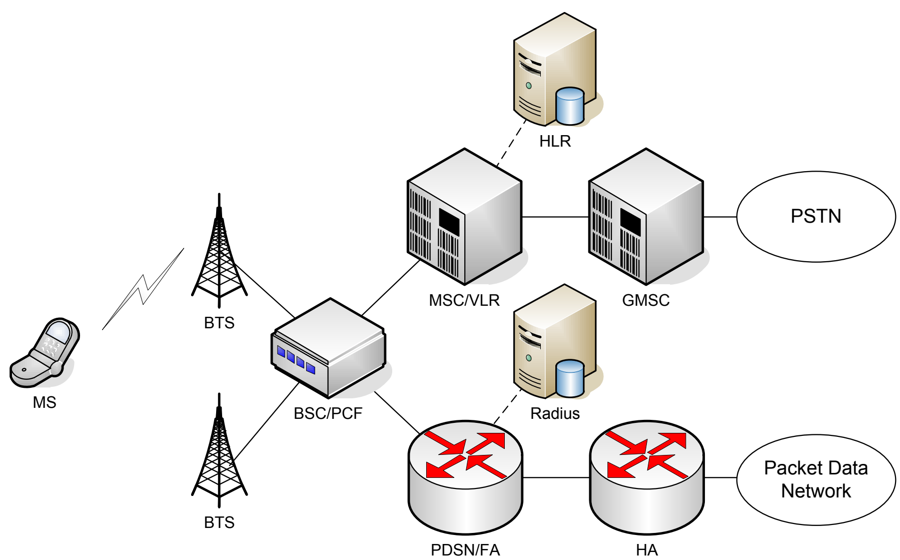

Figure 4: The structure of a CDMA2000 network

Qualcomm intended to continue the development track of cdmaOne and CDMA2000 having Ultra Mobile Broadband (UMB) as the next major step. Today, there are no such plans. However, an upgrade of EV-DO Rev. B to DO Advanced is expected to deliver downlink data rates of 32 Mbps and uplink data rates 12.4 Mbps.

4.4 LTE

3GPP Long-term Evolution (LTE) is the latest standard in the GSM/UMTS line specified in 3GPP Release 8. It replaces the WCDMA transmission scheme of UMTS so that OFDMA (Orthogonal Frequency-Division Multiple Access) is used for downlink while SC-FDMA (Single-carrier FDMA) is used for uplink traffic.

Orthogonal frequency-division multiplexing (OFDM) is an FDM type of scheme that is used as a digital multi-carrier modulation method where a number of closely spaced orthogonal sub-carriers are used to carry data. The data is divided into several parallel data streams or channels, one for each sub-carrier. A flexible resource allocation is achieved through dynamic assignment of sub-carriers to a specific node. Each sub-carrier is modulated with a conventional modulation scheme at a low symbol rate. Furthermore, MIMO (multiple-input, multiple-output) antenna technology is used in LTE. Minimum transmission time interval is 1 ms and 64QAM was added as a modulation scheme.

The Dedicated Traffic Channel (DTCH) in LTE is mapped to DL-SCH and UL-SCH (Downlink Shared Channel and Uplink Shared Channel) respectively. Just as in HSPA it uses HARQ and adapts dynamically to the link quality.

Spectrum flexibility was an important design goal for LTE and it was built to scale using bandwidths ranging from 1.4 MHz to 20 MHz in both paired and unpaired configurations. A wide range of frequency bands are expected to be used for LTE including the 700 MHz band allowing for indoor usage and wide coverage.

LTE provides data rates up to 100 Mbits/s in the downlink direction, uplink data rates up to 50 Mbps in the uplink direction and latencies in the radio access network at 10 milliseconds. The system is non-backward compatible with GSM or UMTS and hence requires a new infrastructure. The upgraded version LTE Advanced is designed to meet the requirements from the fourth generation (4G) radio access network of 1 Gbits/s in data rate for stationary applications and 100 Mbits/s for mobile applications. The first commercial LTE network was opened in Stockholm and Oslo in December 2009. A wide range of frequencies are expected to be used.

The structure of LTE networks is changed radically from the GSM and UMTS network structures (Figure 5). eNB (Evolved NodeB) is the only node type in E-UTRAN (Evolved UTRAN) responsible for all radio interface-related functions. Main node types in the EPC (Evolved Packet Core) are the MME (Mobility Management Entity) responsible for mobility, UE identity, and security management functions, the S-GW (Serving Gateway) terminating the interface towards E-UTRAN, and the P-GW (PDN Gateway) terminating the interface towards the PDN.

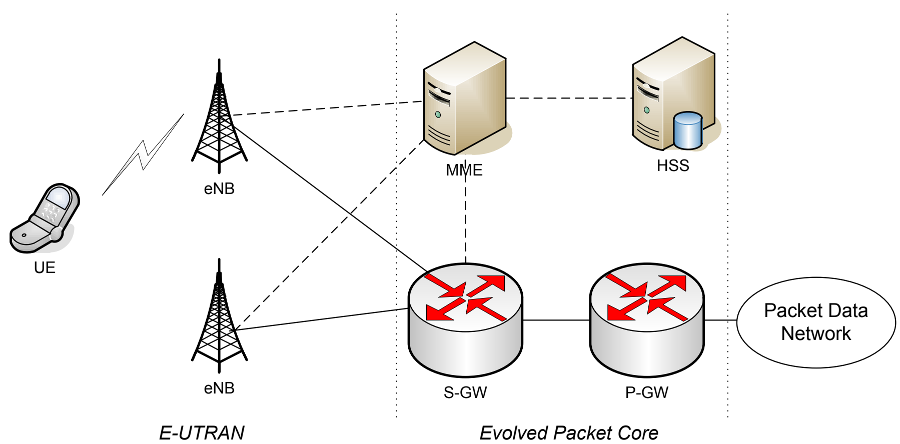

Figure 5: The structure of an LTE network

It should be noted that the circuit switched domain finally has been removed from the network architecture. In LTE voice services are not delivered through dedicated nodes in the core network, but through VoIP-based mechanisms in other subsystems like the IP Multimedia Subsystem (IMS). GSM Association launched their Voice over LTE (VoLTE) initiative in February 2010.

4.5 WLAN

The IEEE released their first version of the Wireless LAN (WLAN) standard 802.11 in 1997 enabling local area network services over the air. Using unlicensed spectrum at the 2.4 and 5 GHz bands made the standard very popular for both enterprise and consumer users. Also, Wireless Internet Service Providers (WISPs) and traditional cellular operators typically deploy 802.11-based wireless hot spots where user density is high and demands for high data rates are common.

The initial version of the standard used direct-sequence spread spectrum (DSSS) and frequency-hopping spread spectrum (FHSS) as alternate physical layer technologies. The 802.11 a, g, and n amendments then used orthogonal frequency-division multiplexing (OFDM) scheme, while the 802.11b amendment used OFDM and DSSS. Furthermore, the 802.11n amendment allows for usage of 4 multiple-input multiple-output (MIMO) streams.

New features have been added to the standard by amendments to the base standard, or as in 2007, by a new release of the entire standard. Peak data rates are 11 Mb/s for 802.11b, 54 Mb/s for 802.11a/g, and 150 Mb/s for 802.11n. Typically half those data rates are available to applications with no difference in uplink and downlink directions. Latencies are typically in the range of a few milliseconds. IEEE 802.11-based systems are used both for indoor and outdoor installations.

Security was originally week, but improved after the arrival of the 802.11i amendment. Support for both infrastructure networks (called Basic Service Set, BSS) and ad hoc networks (called Independent Basic Service Set, IBSS) is included in the standard. A typical BSS type of network is built up of one or more stations (STAs) and one access point (AP). The AP is responsible for bridging the wireless traffic to the wired local area network and to act as a base station for the STAs.

The 802.11 standard also allows stations to roam among a set of APs connected to the same wired network or distribution system (DS). That configuration is called an Extended Service Set (ESS).

Laptops are typically equipped with WLAN cards and most smartphones and PDAs today have both cellular and WLAN interfaces installed to them.

Figure 6 depicts the structure of an 802.11-based network.

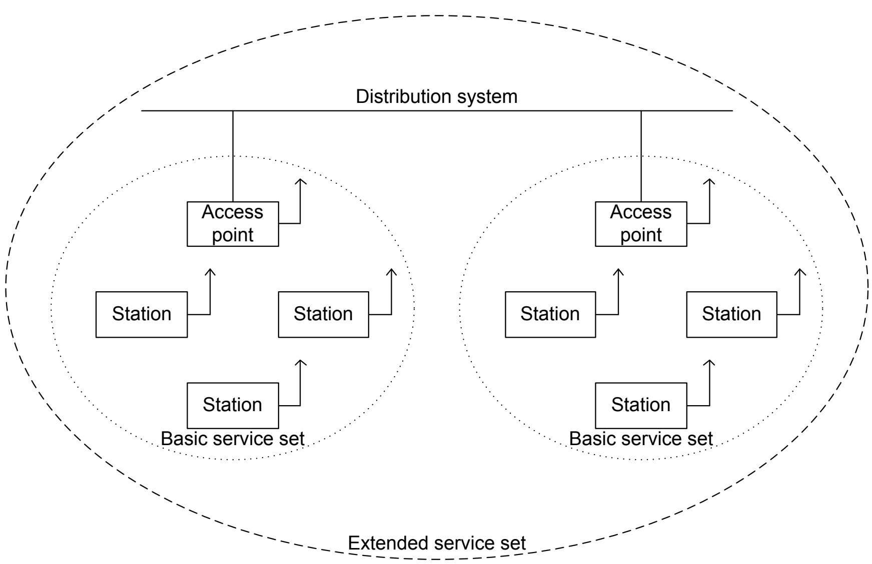

Figure 6: The structure of a WLAN network

4.6 WiMAX

WiMAX, Worldwide Interoperability for Microwave Access, is standardized under the name of 802.16 by the IEEE. WiMAX uses both licensed and unlicensed spectrum where the 2.3 MHz, 2.5 MHz, and 3.5 GHz bands are most common for licensed installations. While WLAN is a short-range technology, WiMAX is long range allowing for many kilometers of communication providing a connection-oriented MAC layer and support for quality of service operating either in a time division duplex (TDD) or frequency division duplex (FDD) mode.

The 802.16-2004 version of the standard was directed towards fixed use offering data rates up to 75 Mbps, while the 802.16e supplement was adding mobility support to the standard offering data rates up to 30 Mbps. The most recent issue of the standard is the 802.16-2009 version. The 802.16m supplement is expected to meet the 4G requirement of 1 Gbps downlink data rates for stationary usage and 100 Mbps downlink data rates for mobile usage.

The mobile station (MS)/subscriber station (SS), the access service network (ASN), and the connectivity service network (CSN) are the three main components of the WiMAX network architecture defined by WIMAX Forum.

An ASN is typically built up of a set of base stations (BSs) and one or more ASN gateways (ASN-GWs) interconnecting the ASN with the CSN. The ASN is typically delivering MAC layer services to the SS while the CSN typically delivers layer 3 services. The WiMAX business model allows an ASN provider (Network Access Provider, NAP) to sign contracts with one or more CSN providers (Network Service Providers, NSPs). Also, NSPs may have roaming agreements with other NSPs.

Figure 7 depicts the WiMAX network structure.

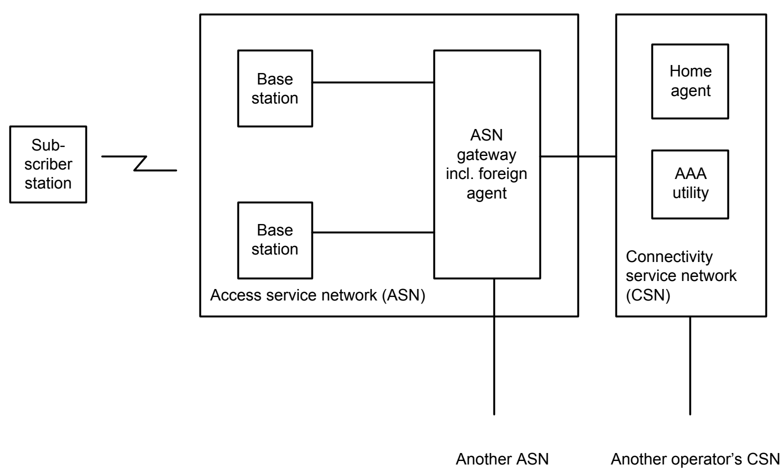

Figure 7: The structure of a WiMAX network

5 Key Technologies for Interworking Solutions in Heterogeneous Wireless Networks

5.1 Mobility Management Solutions

There are a number of mobility support architectures for heterogeneous wireless networks proposed. One important aspect is at what layer [33] in the communication stack mobility management is handled. Mobility management could be handled at the datalink layer for a single radio access technology. It might also be managed at the network layer allowing mobile nodes to have their IP address fixed while roaming across subnets. Solutions at the transport and application layers also exist.

Other design aspects of mobility support solutions are whether the solution allows for simultaneous connections to multiple radio access networks and if handover decisions are host-based or network-based. Yet other aspects are whether the mobile node is involved in mobility-related signaling and whether its IP communication stack remains unchanged or needs modification.

Mobility management consists of two fundamental operations: handoff and location management [13]. Handoff introduces a number of questions, notably how to determine the timing of the handoff, the decision on what access network to transfer the traffic to (network selection), and how to migrate existing connections smoothly. Location management is the mechanism for locating the mobile node (MN) or a user in order to initiate and establish a connection.

In addition to the ability to perform handovers within a certain radio access technology (also referred to as horizontal handovers), the ability to perform handovers across radio access technologies is needed. This important feature is referred to as vertical handover. Another way of classifying handover types is to distinguish inter domain from intra domain handovers. Inter domain mobility is called macro mobility while intra domain mobility is referred to as micro mobility.

Users of heterogeneous wireless networks with multiple access networks included need a mobility management solution at layers above the data-link layer in order to take advantage of all available technologies at a certain moment and a certain place. Today there are solutions available at the network layer, the transport layer, and the application layer.

Furthermore, cross-layer designed solutions exist as well as solutions introducing new layers in the network stack.

The following subsections describe state of the art mobility management schemes and solutions on those layers and, for completeness, also examples from the datalink layer.

5.1.1 Examples on Mobility Management at the Datalink Layer

GSM and UMTS In GSM and UMTS the MS/UE initiates communication with the PS domain through requesting a PDP (packet data protocol) context. SGSN then selects which GGSN to be used based on the Access Point Name (APN), while the Home Location Register (HLR) is responsible for authenticating the UE. After initiation, traffic is tunneled from UE via BS, RNC, and SGSN to GGSN where decapsulation occurs and standard IP routing is performed. GPRS Tunneling Protocol (GTP) is used for tunneling between SGSN and GGSN.

cdmaOne and CDMA2000 Mobility management in cdmaOne and CDMA2000 is using Mobile IP with the PDSN acting as Foreign Agent.

LTE Mobility management in LTE is using GTPv3 or Proxy Mobile IPv6 (PMIP). For other radio access technologies interconnecting with LTE both host-based (Mobile IPv4 or DSMIPv6) and network-based (PMIPv6) mobility management schemes may be chosen.

WLAN The IEEE 802.11 standard allows stations to roam among a set of APs placed so that overlapping coverage areas exist. STAs may perform seamless handoffs among APs. Mobility is handled, so that the STA first associates with the AP it wants to connect to, then re-associates with new APs, and finally disassociates from the last AP it associated with. Also, the standard allows new AP to contact old AP to get frames buffered for a STA that re-associated recently.

One important drawback of this type of configuration is that all STAs and all APs must be part of the same subnet to allow roaming.

WiMAX The mobility procedures in WiMAX are divided into two mobility levels: ASN anchored mobility for micro-mobility and CSN anchored mobility for macro mobility. The latter is based on Mobile IP where either Proxy-MIP or Client MIP is used. ASN anchored mobility is handled, so that the SS either listens for network topology advertisements or scans for neighbor BSs. Handovers are split into five steps: cell re-selection, handover decision and initiation, synchronization to a target BS downlink, ranging and network re-entry, and termination of SS context. Also, BSs can initiate handovers.

5.1.2 Mobility Management at the Network Layer

One of the basic challenges to deal with when introducing mobility management at the network layer is that network layer addresses not only are used to identifying hosts but also to finding routes between hosts on the Internet.

Handling mobility management at the network layer has several advantages since applications do not need to be aware of mobility. If the network layer handles mobility management entirely, applications can, in theory, be used as if the user was running the application in a fixed environment since the user is reachable through a fixed IP address. The network layer is extended with a suitable mobility management module taking care of the delivery of packets to the user's current point of attachment to the Internet. This mobility management solution works both for connection oriented flows (i.e. TCP connections) and connection less flows (i.e. UDP traffic).

The most well-known example of mobility management at the network layer is Mobile IP (MIP) which is defined both for IPv4 [25] and IPv6 [38].

MIP makes use of a mobility agent located in the home network, a home agent (HA), and, in MIP for IPv4, a mobility agent in the visited network, a foreign agent (FA). The HA is a specialized router responsible for forwarding packets aimed for the end-user at the MN. The MN is assigned a home address (HoA) in the same subnet as the HA. The FA is responsible for assigning a care of address (CoA) for the MN and forwarding packets for the MN. The HA holds a binding cache with mappings of HoAs to CoAs. The MN can also use a co-located address CoA. In that case, the MN acquires an IP address using regular mechanisms like DHCP and is not dependent on the existence of an FA in the visited network.

Packets are transported from the originating host, the correspondent node (CN), to the HA and then tunneled through an IP tunnel using IP in IP encapsulation to the MN (possibly via the FA). The MN continually sends binding update (BU) messages to the HA indicating its CoA. If a new CoA is indicated in the BU message, the HA updates the binding cache. The HA returns binding acknowledgments (BAck) to the MN. Packets in the direction from the MN to the CN can be sent directly to the CN. In MIPv6 route optimization techniques also exist enabling the CN to send packets directly to the MN. Thus, all packets do not need to travel through the HA.

MIP has got some drawbacks with handover latencies, introduction of tunneling overhead, and dependency of mobility agents being the most severe. Several extensions to MIP exist, including fast handovers for MIPv6 (FMIPv6) [28] and hierarchical MIP (H-MIP) [57]. Both address the problem with handover latencies where packets typically are lost and the MN is not able to send packets for a period of time.

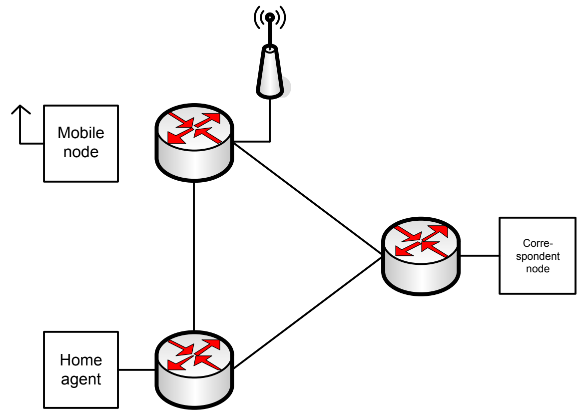

Figure 8: Mobile IP basic architecture

FMIPv6 enables an MN to provide the new access point and subnet prefix information to the current access router in a fast binding update (FBU) message.

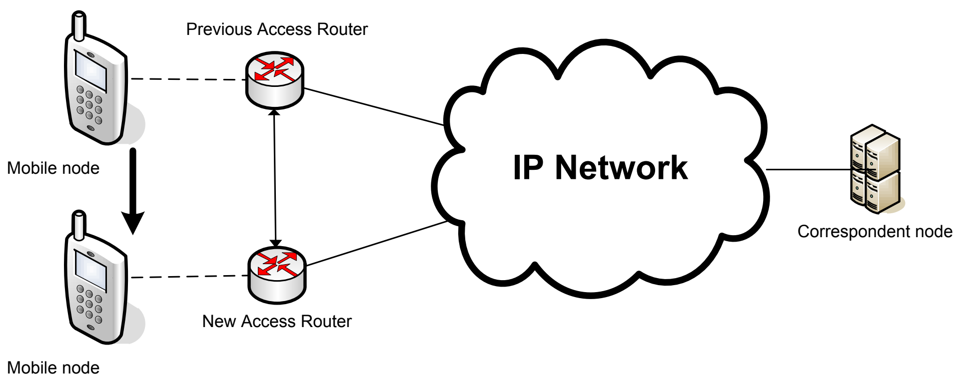

Figure 9: Reference scenario for fast hand-overs

First, the MN sends a Router Solicitation for Proxy Advertisements (RtSolPr) message to the previous access router (PAR) including the datalink layer identifiers that the MN discovered at the new access router (NAR). The PAR then sends a Proxy Router Advertisement (PrRtAdv) message including network specific information. Based on this information, the MN creates a care of address at the NAR and sends a fast binding update (FBU) message to the PAR. The PAR then sends a hand-over initiate (HI) message to the NAR which answers with a handover acknowledge (Hack) message to the PAR. A fast binding acknowledgment (FBack) message is sent both to the MN and the NAR. Packets are forwarded from the PAR to the NAR. The MN sends a fast neighbor advertisement (FNA) message to the NAR when the connection is migrated to it. This signaling scheme is referred to as predictive.

A reactive version of this hand-over scheme is also available where the MN sends an FNA message to the NAR which sends an FBU message to the PAR, which, in turn, replies with an FBack message to the NAR. Packets are forwarded from the PAR to the NAR in this version as well.

H-MIP introduces mobility anchor points (MAPs) as a new node type being basically a local HA. Information about MAPs is delivered to MNs through router advertisements. If there are multiple MAPs available it is up to the MN to decide on which MAP to connect to. It may also decide to connect to more than one MAP simultaneously.

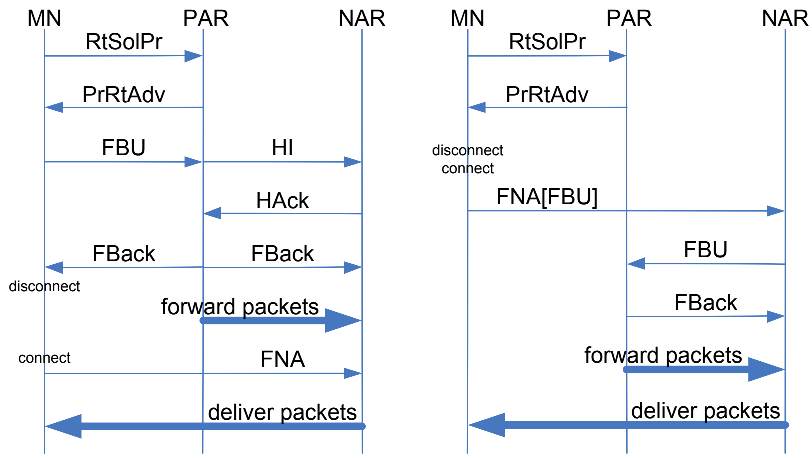

Figure 10: FMIPv6 signaling (predictive vs. reactive)

In H-MIP, the MN is assigned two addresses, namely an on-link care of address (LCoA) and a regional care of address (RCoA). The MN sends a local BU message to the MAP with separate flags set in order to inform the MAP it has formed a regional CoA (RCoA). This way a binding is created between the RCoA and the LCoA in the MAP. H-MIP thus makes use of two tunnels, one from the MN to the MAP and one from the MAP to the HA. When the MN moves within the domain of the MAP, only the tunnel from MN to the MAP needs to be altered and the tunnel between the MAP and the HA may stay unchanged.

H-MIP is also beneficial from a location privacy standpoint as only the RCoA is sent in BU messages from the MN to the HA and CNs.

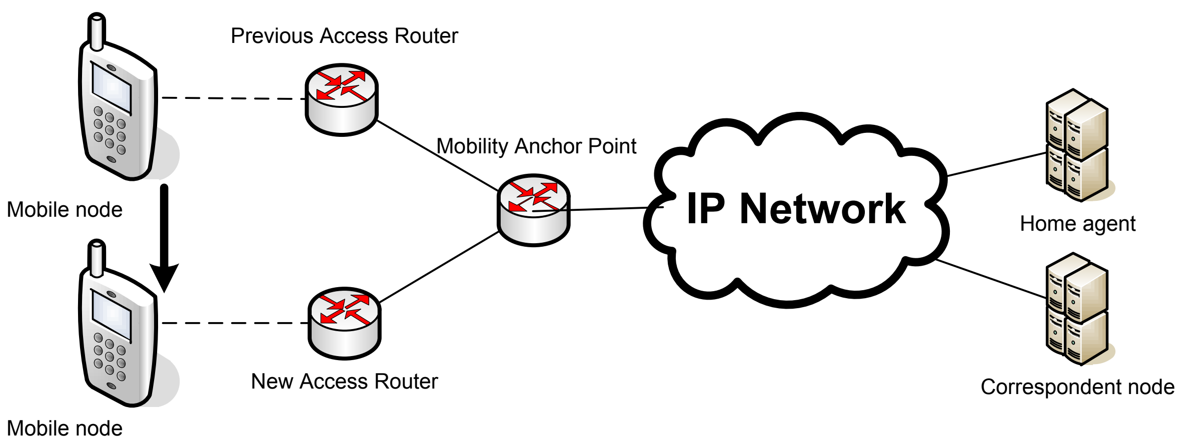

Figure 11: H-MIP architecture

Evaluations being performed combining FMIPv6 with H-MIP have shown good results when coming to reduction of handover latencies [47].

The possibility to register more than one active CoA to the HA and to CNs for a given HoA, often referred to as M-MIP (multi-homed MIP), is described in [29]. By the introduction of a binding unique identification (BID) number for each binding cache entry, multi-homing support is added to MIP.

New initiatives in the area of network-layer mobility management include development of an Internet Key Exchange (IKE) Mobility and Multi-homing Protocol (MOBIKE) [27][40] basically being a multi-homing extension to IKE. A mobile virtual private network (VPN) client could use MOBIKE to keep the connection with the VPN server active while changing IP addresses.

In addition, the Host Identity Protocol (HIP) [44], has also been proposed. HIP separates end-point identifier and locator roles of IP addresses and introduces a new layer between the network and transport layers. A new name space in addition to the IP address and DNS name spaces is also introduced. Not being deployed to a large extent, this approach is, from a theoretical view point at least, promising and interesting. However, new layers in the network stack have until now not been successfully introduced in real-world deployments.

Other ongoing IETF work includes the following working groups active in the field:

- mip4 is focusing its activities in maintaining mobility support for IPv4. The mip4 working group published an Internet Draft on flow binding support for Mobile IPv4 [31].
- mext is another IETF working group focusing on extensions to existing mobility protocol solutions. The activities of the former working groups mip6, nemo, and monami6 were transferred to mext. Support for Dual Stack Mobile IPv6 [26] has been standardized, as well as a solution for multiple care-of addresses registration [29]. An RFC describing flow bindings for Mobile IPv6 is also available [63].
- mipshop is an IETF working group that just concluded. It was targeted towards IP mobility focusing on performance, signaling and handoff optimization. The mipshop workshop published the RFCs for hierarchical MIPv6 (H-MIPv6) [27] and for fast handovers in MIPv6, FMIPv6 [28].
- netlmm is another IETF working group that just concluded. They focused on network-based mobility management and published the Proxy Mobile IPv6 protocol [30]. An IETF working group, netext, was formed to extend the Proxy Mobile IPv6 protocol even further.
- shim6 is an IETF working group focusing on site multihoming by IPv6 intermediation. The basic idea behind shim6, is to provide locator agility with failover capabilities for IPv6 nodes. Hosts use multiple IPv6 address prefixes and setup state with peer hosts. That state could be used for doing failover to a different set of locators. The shim6 protocol [45] was published as an RFC in 2009.
- lisp was an IETF working group focusing on locator/ID separation just as the Host Identity Protocol, but taking a network-based approach instead. The working group published a few Internet Drafts describing the architecture on the LISP solution [34].

One drawback of network-layer mobility management schemes is the lack of support for session, service, and personal mobility. This has made research teams to seek for solutions on higher layers.

5.1.3 Mobility Management at the Transport Layer

One part of the research community suggests handling mobility management at the transport layer [33]. The Stream Control Transmission Protocol (SCTP) [60] is an end-to-end, connection-oriented protocol that supports transport of data in independent sequenced streams. It supports multi-homing which makes it interface redundant. Furthermore, SCTP combines the datagram orientation of UDP with the sequencing and reliability of TCP.

Cellular SCTP (cSCTP) [16] is an extension to SCTP making hand-overs smoother by sending data on multiple paths during handover. Location management in cSCTP can be handled by using a SIP user agent (see Section 5.1.4) running at the application layer at both the MN and the CN.

MSOCKS [43] is yet another architecture for transport layer mobility management. MSOCKS is built on top of the SOCKS protocol for firewall traversal and uses a proxy server between the mobile client and the server. A connection identifier is used for tracking sessions between the mobile client and the proxy. The server does not need to be mobility aware.

The most notable problem with handling mobility management at the transport layer is the need for modifications of well established TCP-based applications.

5.1.4 Mobility Management at the Application Layer

Apart from handling mobility management at the network and transport layers proposals for mobility management at higher layers exist. There are descriptions of mobility management by the introduction of a separate mobility layer above the transport layer [36]. As mentioned before, adding new layers have not been a popular step previously in the Internet history.

However, the idea of handling mobility management at the application layer using the session initiation protocol (SIP) [51] as mobility management protocol is one of the most popular idea in current research.

SIP is an end-to-end signaling protocol designed for initiating, maintaining, and terminating sessions on the Internet, mainly targeted for multimedia applications, but suitable for any type of session-oriented application. In addition to the client side, where the SIP user agent (UA) resides, SIP makes use of three types of servers: SIP proxy servers, SIP redirect servers, and SIP registrars. SIP messages are carried both on top of TCP and UDP and are routed from endpoint to endpoint through a chain of servers. The session description protocol (SDP) is used for describing sessions, including IP addresses, port numbers, codecs, etc. SIP has inherited structures from both SMTP and HTTP making it easier to develop and deploy light-weight implementations when combined with email and web client software. It should also be mentioned that SIP is designed for handling both pre-session mobility management and mid-session mobility management for connection-less transport protocols, e.g. UDP. Application layer using SIP was proposed by Schulzrinne et al. [55].

SIP has become the state-of-the-art protocol for signaling in both IP telephony and other types of multimedia applications. SIP is also the core protocol of 3GPP IP Multimedia Subsystem (IMS), making its deployment to real applications even faster.

SIP has, however, some drawbacks due to its placement in the layered protocol hierarchy. SIP can not, for example, do anything to broken TCP connections due to changes of network layer addresses at handovers. Additionally, if SIP is to be used as a general mobility management solution, already existing applications need to be rewritten completely in order to be mobility-aware. Also, there exist several variants and versions of SIP making global deployment a serious problem to consider carefully.

5.1.5 Mobility Management Using Cross-layer Designed Solutions

As described in the previous sections, there are pros and cons for handling mobility management at each layer. A hot topic in current research is therefore cross-layer designed solutions for mobility management.

However, cross-layer designed solutions are seen by some researchers as violating the basic principles of the layered network stacks like the OSI reference model and the TCP/IP protocol suite. Typical violations include creation of new interfaces (layer N is not only capable of communicating with layer N+1 and layer N-1), merging of adjacent layers into a new super layer, design coupling without new interfaces, and vertical calibration (or joint tuning) across layers [59]. Furthermore, implementations typically include direct communication between layers, a shared database across the layers, or completely new abstractions.

Various examples of cross-layer designed solutions for mobility management exist. In [61] a topology-aided cross-layer fast handoff design has been proposed. A large number of proposals on combinations of MIP and SIP are present [42][50][64][15].

Since it is very hard to make a single layer responsible for mobility management some kind of cross-layer designed solution will be needed.

5.2 AAA Solutions

Key technologies in the area of AAA (Authentication, Authorization, and Accounting) include a flexible AAA framework, support for reliable and secure transportation of different authentication protocols, and a secure tunneling protocol. Today, this is achieved using the Diameter base protocol [21], the Extensible Authentication Protocol (EAP) [12], and IPSec [39]. Also, the older Radius protocol is still in use in new mobility signaling protocol as Proxy Mobile IP[66].

5.3 QoS Solutions

Quality of Service (QoS) is used to differentiate different types of flows and prioritize some flows over others in order to deliver realtime services. Key technologies in the area of QoS include access technology specific QoS mechanisms as well as policy handling and QoS handling on the IP layer. Typical parameters of interest are delay, jitter, loss rate, and throughput.
QoS on the IP layer is mostly using the Differentiated Services (DiffServ) [19] mechanism. It specifies a simple, scalable and coarse-grained mechanism for classifying and managing network traffic on the Internet using the 6-bit Differentiated Services Code Point (DSCP) field in the IP header for packet classification purposes. DSCP replaces the outdated Type of Service (TOS) field used before.

In WLAN the 802.11e amendment introduced a new coordination function: the hybrid coordination function (HCF). Traffic Categories (TC) used by 802.11 are background, best effort, video, and voice.

WiMAX, being a connection oriented technology, offers quite a strong QoS support including five traffic classes: Unsolicited Grant Service (UGS), Extended Real-time Polling Service (ertPS), Real-time Polling Service (rtPS), Non-real-time Polling Service (nrtPS), and Best Effort (BE).

UMTS networks define four QoS classes: conversational, streaming, interactive, and background.

LTE networks define two bearers: guaranteed bit-rate (GBR) and non-guaranteed bit-rate (non-GBR). The Evolved Packet System QoS concept is class-based, where each bearer is assigned one and only one QoS class identifier (QCI) by the network bearer. The QCI values are mapped to DSCP values in a mapping function.

Policies are used to dynamically allocate network resources. Policy rules control the priority, packet delay, and the acceptable loss of packets. In LTE, the Policy Control Resource Functio (PCRF) is the policy server.

6 Interworking Solutions for Heterogeneous Wireless Networks

This section analyzes a number of standards-based interworking solutions for heterogeneous wireless networks. A few interworking solutions are characterized in terms of being tightly or loosely coupled to the 3GPP architecture. These principles can be described according to the following:

A. Loose coupling The architectural model behind loose coupling is an independent interconnection of those wireless access networks participating in the heterogeneous wireless network. Different mechanisms for mobility management, authentication, and billing can be used in the individual wireless access networks. Minimal changes are needed in the existing wireless access networks and the model is quite straight forward. Mobile IP is often used as the mobility management solution basically forming a mobility overlay network. However, other mobility management solutions working at the network or higher layers can also be used. The most important drawback of this architectural model is longer handover latencies compared to the architectural model behind tight coupling.

B. Tight coupling In a tightly coupled architecture interconnection between the wireless access networks takes place in one of the participating wireless access networks' core network or radio network. The most common example is interconnection at the GGSN, SGSN, or RNC level of a 3G network. This model is much more complex compared to the model behind loose coupling and requires installation of gateways for the connected wireless access networks.

6.1 Multimode Terminals

All current architecture proposals for wireless heterogeneous networks are built on the assumption that IP is the common network layer protocol. Applications and a variety of transport protocols are run on top of IP, which in turn are run over a number of access technologies. This is sometimes referred to the hourglass model (Figure 12).

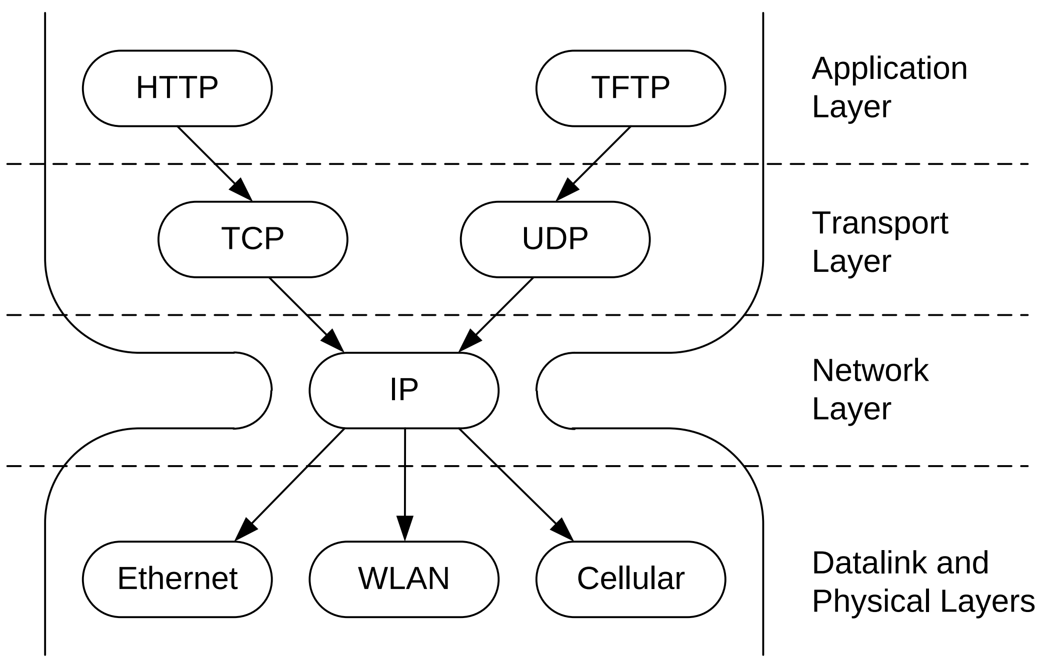

Figure 12: Hourglass model

In order to allow a multi-RAT (multi Radio Access Technology) enabled mobile node to work in a heterogeneous wireless networking environment the following components are needed:

- Interface monitors for each interface installed in the mobile node
- A Network selector taking decisions on what interface to activate

- A Handover manager actually executing the handover decisions
- A Policy repository storing the user policies

The implementation of this functionality is sometimes referred to as a connection manager. IETF recently started a working group focusing on this, the multiple interface group (mif). Mif is focusing its activities in the area of multiple interfaces with the aim to describe the issues of attaching hosts to multiple networks. RFCs are published in terms of current practices [65] and problem statement [20]. Non-working group Internet Drafts on connection managers [56] and DNS server selection [53] are also available.

6.2 Wireless LAN Interworking (I-WLAN)

3GPP defined Wireless LAN Interworking (I-WLAN) [5] in its Release 7. This is a loose coupling technique covering AAA and intersystem user data transfer. AAA is handled using the Diameter protocol between a proxy server in the IEEE 802.11 access network communicating with a 3GPP AAA server. This was enabled when EAP extensions for the 802.1X access control were added allowing authentication based on UMTS credentials.

Payload traffic is transported over a secured channel (IPsec) to a new node in the core network, the Packet Data Gateway (PDG). This node acts like a GGSN to WLAN users.

Mobile nodes using I-WLAN services need 802.11 network cards supporting EAP/802.1X authentication (like Wireless Protected Access, WPA) and IPsec support in the IP protocol stack.

6.3 Unlicensed Mobile Access (UMA), Generic Access Network (GAN)

Another initiative supporting heterogeneous wireless networking is Unlicensed Mobile Access (UMA) [3] providing roaming and hand-over services for users between GSM/UMTS, WLAN, and Bluetooth networks. By the introduction of a UMA Network Controller (UNC) users can connect to and be reachable via a GSM/UMTS network through e.g. a residential WLAN access point and a broadband IP network connection (Figure 13).

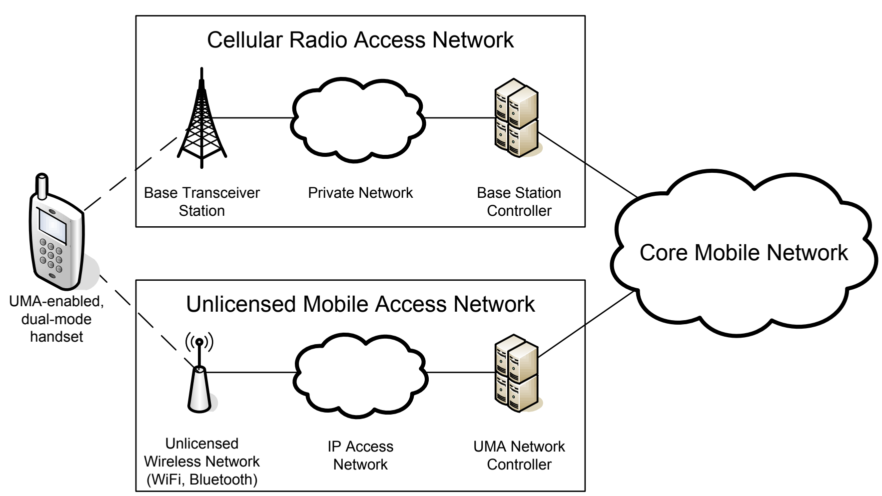

Figure 13: UMA architecture

The UNC appears to the core network as a base station subsystem (BSS). It includes a security gateway (SEGW) providing mutual authentication, encryption and data integrity for signaling, voice and data traffic. UMA uses a very simple tunneling technique so that GSM/UMTS packets are encapsulated in 802.11 packets. Support for video sessions etc. is not included.

The UMA specifications were transferred to 3GPP in 2005 and were part of 3GPP release 6 being referred to as Generic Access Network (GAN) [9]. The UNC is therefore today, not surprisingly, called Generic Access Network Controller (GANC). This node acts like a RNC to WLAN users.

The UMA/GAN model is sometimes referred to as tight coupling.

6.4 Voice Call Continuity (VCC)

To support handsets supporting both WiFi and cellular technologies, the 3GPP defined the Voice Call Continuity (VCC) specifications [4] in Release 7. It described persistence of voice calls when roaming between the circuit switched (CS) and packet switched (PS) domains. A client application in the handset may send information to the VCC application in the network enabling triggering and controlling of handovers. This allows for transfer of voice calls between the two domains, transparently to the end user.

Since VCC was only targeted for voice services, other features (like the short message service, SMS) needed to be replicated in the packet switched domain as well.

6.5 IP Multimedia Core Network Subsystem (IMS)

In the field of multimedia distribution in heterogeneous networking environments, the Third Generation Partnership Project (3GPP)-led standardization of the IP Multimedia Subsystem (IMS) [6] and the 3GPP2-led standardization of the Multimedia domain (MMD) [11] are promising efforts in terms of defining a separation of service logic and service infrastructure from the physical infrastructure and different access networks [22][48] (see figure 14). Working together with the IETF the basic architectural idea has been to re-use as much as possible from existing Internet protocols and solutions and to make IMS-specific amendments when needed.

By introducing an overlay network of SIP servers, named Call Session Control Functions (CSCFs), and standardizing AAA functions, implementing the Diameter protocol, 3GPP and 3GPP2 are contributing to the vision of creating seamless mobile multimedia applications. Furthermore, the support for policies and Quality of Service provisioning, as well as standardized codecs and interworking technologies for communication with legacy circuit switched networks (like the PSTN) are promising.

The straight-forward approach for media distribution with real-time transmission protocol (RTP) [54] over UDP was also a natural step. Primarily being developed as an extension of the emerging 2G/3G networks, IMS is today operating with various types of access networks, both wireless like WLAN and wired like DSL.

6.6 Interworking in LTE

In 3GPP Release 8 architecture enhancements for non-3GPP accesses were defined [8]. IP mobility is handled using Mobile IP and Proxy Mobile IPv6 with home agent functionality located in the Packet Data Network-Gateway (PDN-GW) node. Also, functionality for access network discovery and selection (ANDSF) was added. The functionality provided by the 3GPP ANDSF entity covers data management and control functionality thus providing access network discovery and selection assistance to the mobile node. As for the case with the IEEE 802.21 Information Service, the ANDSF entity may both respond to requests from mobile nodes and push data to them. Information exchange is handled using the OMA Device Management Protocol [14] with the ANDSF Management Object being specified in [7]. Mobile nodes may connect to ANDSF entities both in the visited network and in their home networks. IETF has published an RFC on mechanisms for discovery of 3GPP ANDSF services using DHCP options [23].

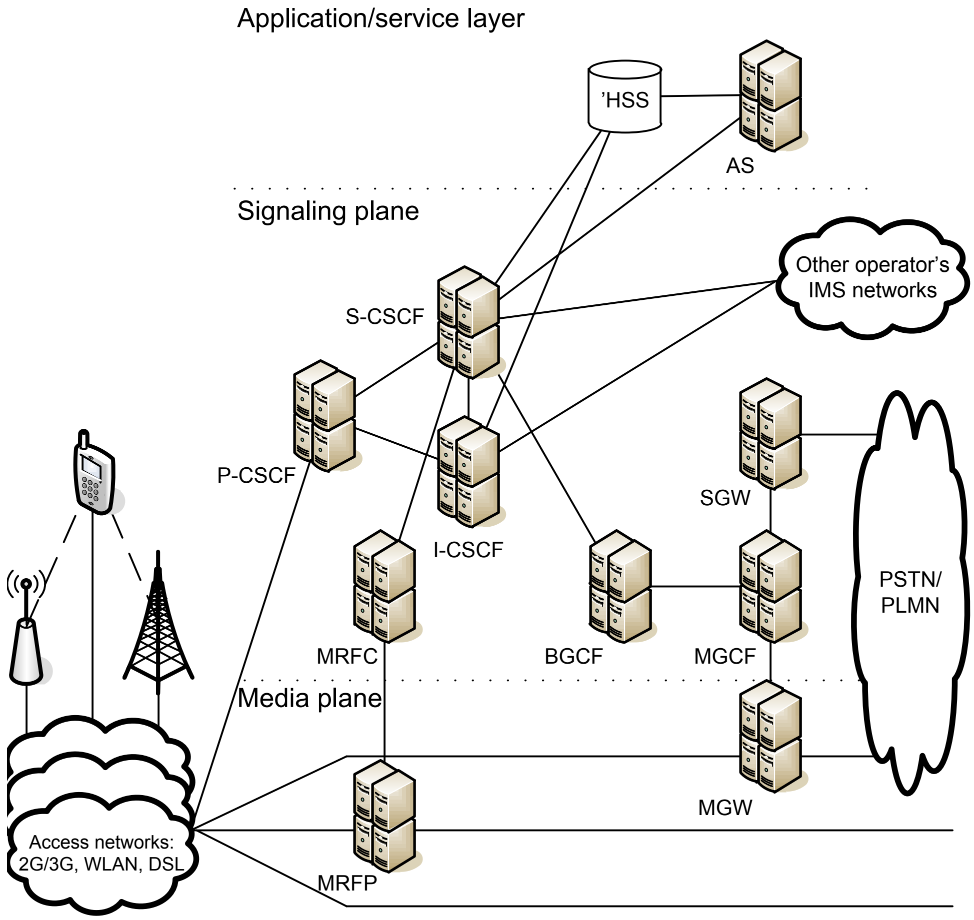

Figure 14: IMS architecture

Currently, 3GPP is also standardizing support for IP flow mobility and seamless WLAN offload [10] building the solution on Dual-stack Mobile IPv6 [26], multihoming [29], and flow-based mobility handling [63] using traffic selectors [62]. The goal is to provide mobile data offload support for LTE networks allowing multimode terminals to move certain flows to WLAN hotspots when overlapping coverage is available in the proximity of the terminal.

6.7 Media Independent Handover Services

To improve handover performance in heterogeneous environments, the IEEE decided to standardize a media-independent handover (MIH) framework under the name of IEEE 802.21 [1]. It defines mechanisms for exchanging handover-related events, commands, and information. Handover initiation and handover preparation are covered but not the actual handover execution. The actual mobility management mechanism can be of any type, working at either the network, transport or application layer. Furthermore, the IEEE 802.21 standard allows for both network-controlled handovers and host-controlled handovers and it defines three main services: Media-independent Event Services (MIES), Media-independent Command Services (MICS) and Media-independent Information Services (MIIS).

Media-independent Event Services (MIES) MIES define events representing changes in the link characteristics either originated from the link layer or from the MIH function. Such characteristics could be information on link status or link quality, for example. Events can be subscribed to and be either local or remote. They may indicate changes in the state and transmission behavior of the physical, data-link and logical-link layers. Events can also predict state changes of these layers. Remote events are transported over the network in MIH protocol messages and typically contain information on link events originated from the point of attachment.

Media-independent Command Service (MICS) The MICS defines commands for controlling the link state and can be invoked either locally or remotely. By using the MICS, the user may control the configuration and selection of a specific link. Remote commands are, like remote events, transported over the network by MIH protocol messages and may result in a link command or an MIH indication in the peer Media-Independent Handover Function (MIHF) entity.

Media-independent Information Service (MIIS) The MIIS defines a set of information elements (IEs), their structure and their representation. Furthermore, it defines a query-response-based mechanism for information retrieval. Such information can be used to take more accurate handover decisions. The idea is that using information on available access networks in the proximity of the user may help to radically improve the decision-making process for handovers. Information is exchanged through binary type-length-value (TLV) coded messages. Also, complex queries are supported.

IEs can be of general type indicating either the network type, operator identifier, or a service-provider identifier. They can also be access-network specific providing specific information on Quality of Service (QoS), security characteristics, revisions of current technology standards in use, cost, and roaming partners. Also, some IE types deliver Point-of-Attachment (PoA)-specific information such as the MAC address of the PoA, its geographical location, data rates offered, and channel information. IEs may also be vendor-specific.

Figure 15 shows the MIH framework and communication between local and remote MIHF entities. The IEEE 802.21 standard also defines a set of interfaces defined by a number of Service Access Points (SAPs). The interface between MIH users and the MIH function is referred to as the MIH_SAP while the interface to the lower layers is referred to as the MIH_LINK_SAP which is generic to all access technologies. The primitives in the MIH_LINK_SAP are mapped to technology-specific primitives included in the IEEE 802.21 standard. MIH_NET_SAP defines the exchange of messages between MIH entities.

MIH protocol messages are either sent at layer 2 or by using IP.

Currently, security extensions are being standardized (IEEE 802.21a), as well as handling of handovers for downlink only technologies (IEEE 802.21b).

In order to support the new IEEE 802.21 standard, the IETF defined a framework for mobility support [32] and mechanisms for discovery of mobility services using DHCP options [18] and the DNS [17].

6.8 IEEE 802.11u

The IEEE recently published an amendment to the 802.11 standard, the 802.11u amendment [2]. It covers interworking with external networks allowing IEEE 802.11 devices to interwork with external networks. IEEE 802.11u aids network discovery and selection, enabling information transfer from external networks, and enabling emergency services. It provides information about the networks prior to association. The aim is to deliver MAC layer enhancements allowing higher layer functions to provide the overall end-to-end interworking solution.

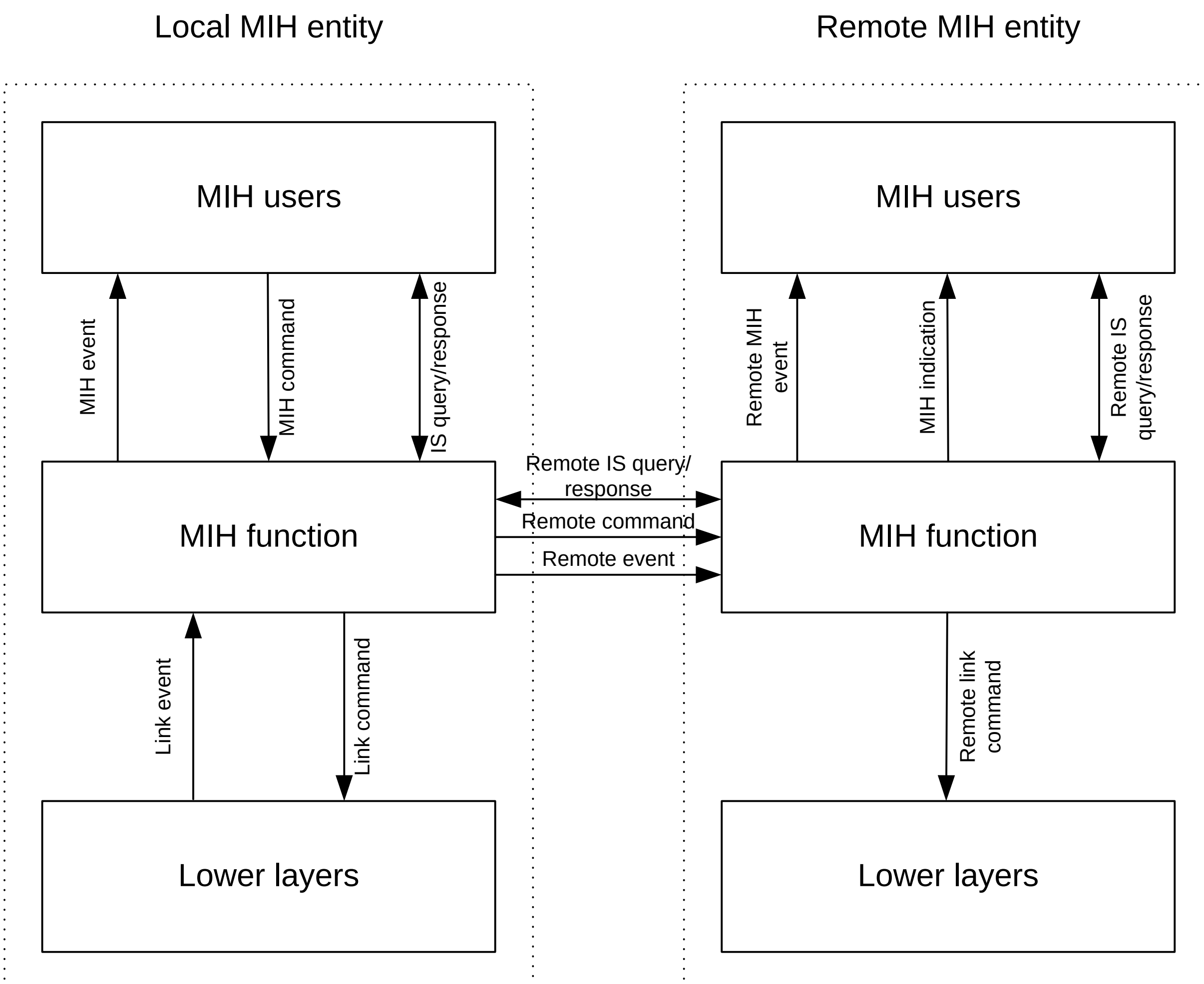

Figure 15: Media-independent handover framework

7 Conclusion and Further Research

This article presented some important interworking solutions for heterogeneous wireless networks. I-WLAN, UMA/GAN, and VCC are today quite mature standards that paved the way for the newer standards like IEEE 802.21 being a generic solution and 3GPP IP Flow Mobility and Seamless WLAN Offload being a tailored 3GPP-centric solution. The IMS will also play a very important role in delivering multimedia services to users that connect to wireless networks where the circuit switched subsystem is no longer available.

Moreover, the mext IETF working group closed and re-chartered under the name of dmm, distributed mobility management [2]. The purpose of that working group is to develop new IP mobility management protocols that do not rely on the centralized role of any anchor node as the home agent in Mobile IP or the Local Mobility Anchor in Proxy Mobile IP. This effort comes to address yet a current trend in research about designing flatter mobile architectures.

Once the 4G systems have arrived to the market even more advanced interworking solutions will be needed optimizing handovers even further and taking full advantage of the potential in those new systems. More research in this area is therefore needed.

Acknowledgment

The work presented in this article was partially supported by the NIMO project (Nordic Interaction and Mobility Research Platform) funded by the European Union Interreg IVA North programme.

References

[1] Ieee standard for local and metropolitan area networks-part 21: Media independent handover services. IEEE 802.21, January 2009.

[2] Ieee standard for local and metropolitan area networks-part 11: Wireless lan medium access control (mac) and physical layer (phy) specifications. amendment 9: Interworking with external network. IEEE 802.11u, February 2011.

[3] Smart wi-fi powered by uma. `http://www.smart-wi-fi.com`, last viewed February 2012, 2012.

[4] 3GPP. Voice Call Continuity (VCC), Stage 2. TS 23.507, 3rd Generation Partnership Project (3GPP), December 2007.

[5] 3GPP. 3GPP system to Wireless Local Area Network (WLAN) interworking, System description. TS 23.234, 3rd Generation Partnership Project (3GPP), December 2009.

[6] 3GPP. IP Multimedia Subsystem (IMS), Stage 2. TS 23.228, 3rd Generation Partnership Project (3GPP), September 2009.

[7] 3GPP. Access Network Discovery and Selection Function (ANDSF) Management Object (MO), V9.1.0. TS 24.312, 3rd Generation Partnership Project (3GPP), March 2010.

[8] 3GPP. Architecture enhancements for non-3GPP accesses. TS 23.402, 3rd Generation Partnership Project (3GPP), September 2010.

[9] 3GPP. Generic Access Network (GAN), Mobile GAN interface layer 3 specification. TS 44.318, 3rd Generation Partnership Project (3GPP), March 2010.

[10] 3GPP. IP Flow Mobility and Seamless WLAN Offload. TS 23.261, 3rd Generation Partnership Project (3GPP), June 2010.

[11] 3GPP2. All-IP Core Network Multimedia Domain. X S0013-000-0, 3rd Generation Partnership Project 2 (3GPP2), July 2005.

[12] B. Aboba, L. Blunk, J. Vollbrecht, J. Carlson, and H. L. (ed.). Extensible authentication protocol (eap). IETF RFC 3748, June 2004. `http://www.ietf.org/rfc/rfc3748.txt`.

[13] I. F. Akyildiz, J. McNair, J. S. M. Ho, H. Uzunalioglu, and W. Wang. Mobility management in next-generation wireless systems. *Proceedings of the IEEE*, 87(8):1347–1384, August 1999.

[14] O. M. Alliance. Oma device management protocol. `http://www.openmobilealliance.org/`, February 2007.

[15] K. Andersson, M. Elkotob, and C. A. hlund. A new mip-sip interworking scheme. In *Proc. of the 7th International Conference on Mobile and Ubiquitous Multimedia (MUM'08), Umeå, Sweden*. ACM Press, December 2008.

[16] I. Aydin, W. Seok, and C. Shen. Cellular sctp: A transport-layer approach to internet mobility. In *Proc. of the 12th International Conference on Computer Communications and Networks (ICCCN'03), Dallas, TX, USA*. IEEE, October 2003.

[17] G. Bajko. Locating ieee 802.21 mobility services using dns. IETF RFC 5679, December 2009. `http://www.ietf.org/rfc/rfc5679.txt`.

[18] G. Bajko and S. Das. Dynamic host configuration protocol (dhcpv4 and dhcpv6) options for ieee 802.21 mobility services (mos) discovery. IETF RFC 5678, December 2009. http://www.ietf.org/rfc/rfc5678.txt.

[19] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss. An architecture for differentiated services. IETF RFC 2475, December 1998. http://www.ietf.org/rfc/rfc2475.txt.

[20] M. Blanchet and P. Seite. Multiple interfaces and provisioning domains problem statement. IETF RFC 6418, November 2011. http://www.ietf.org/rfc/rfc6418.txt.

[21] P. Calhoun, J. Loughney, E. Guttman, G. Zorn, and J. Arkko. Diameter base protocol. IETF RFC 3588, September 2003. http://www.ietf.org/rfc/rfc3588.txt.

[22] G. Camarillo and M. Garcia-Martín. *The 3G IP Multimedia Subsystem (IMS): Merging the Internet with the Cellular World, 2nd edition*. Wiley, 2006.

[23] S. Das and G. Bajko. Dynamic host configuration protocol (dhcpv4 and dhcpv6) options for access network discovery and selection function (andsf) discovery. IETF RFC 6153, February 2011. http://www.ietf.org/rfc/rfc6153.txt.

[24] L. Eastwood, S. Migaldi, Q. Xie, and V. Gupta. Mobility using ieee 802.21 in a heterogeneous ieee 802.16/802.11-based, imt-advanced (4g) network. *IEEE Wireless Communications*, 15(2):26–34, April 2008.

[25] C. P. (ed.). Ip mobility support for ipv4. IETF RFC 3344, August 2002. http://www.ietf.org/rfc/rfc3344.txt.

[26] H. S. (ed.). Mobile ipv6 support for dual stack hosts and routers. IETF RFC 5201, June 2009. http://www.ietf.org/rfc/rfc5201.txt.

[27] P. E. (ed.). IKEv2 Mobility and Multihoming Protocol (MOBIKE). IETF RFC 4555, June 2006. http://www.ietf.org/rfc/rfc4555.txt.

[28] R. K. (ed.). Fast handovers for mobile ipv6. IETF RFC 5268, June 2008. http://www.ietf.org/rfc/rfc5268.txt.

[29] R. W. (ed.), V. Devarapalli, G. Tsirtsis, T. Ernst, and K. Nagami. Multiple Care-of Addresses Registration. IETF RFC 5648, October 2009. http://www.ietf.org/rfc/rfc5648.txt.

[30] S. G. (ed.), K. Leung, V. Devarapalli, K. Chowdhury, and B. Patil. Proxy mobile ipv6. IETF RFC 5201, August 2008. http://www.ietf.org/rfc/rfc5201.txt.

[31] S. G. (ed.), K. Leung, G. Tsirtsis, H. Soliman, and A. Petrescu. Flow binding support for mobile ipv4. IETF Internet-draft (work in progress), July 2010. http://tools.ietf.org/html/draft-gundavelli-mip4-multiple-tunnel-support-03.txt.

[32] T. M. (ed.), G. Bajko, S. Das, N. Golmie, and J. Zuniga. Ieee 802.21 mobility services framework design (msfd). IETF RFC 5677, December 2009. http://www.ietf.org/rfc/rfc5677.txt.

[33] W. M. Eddy. At what layer does mobility belong? *IEEE Communications Magazine*, 42(10):155–159, October 2004.

[34] D. Farinacci, V. Fuller, D. Meyer, and D. Lewis. Locator/id separation protocol (lisp). IETF Internet-draft (work in progress), February 2012. http://tools.ietf.org/html/draft-ietf-lisp-21.txt.

[35] R. Ferrus, O. Sallent, and R. Agusti. Interworking in heterogeneous wireless networks: Comprehensive framework and future trends. *IEEE Wireless Communications*, 17(2):22–31, April 2010.

[36] D. Funato, K. Yasuda, and H. Tokuda. Tcp-r: Tcp mobility support for continuous operations. In *Proc. of the 1997 International Conference on Network Protocols (ICNP'97), Atlanta, GA, USA*. IEEE, October 1997.

[37] E. Gustafsson and A. Jonsson. Always best connected. *IEEE Wireless Communications*, 10(1):49–55, February 2003.

[38] D. Johnson, C. Perkins, and J. Arkko. Ip mobility support in ipv6. IETF RFC 3775, June 2004. http://www.ietf.org/rfc/rfc3775.txt.

[39] S. Kent and R. Atkinson. Security architecture for the internet protocol. IETF RFC 2401, November 1998. http://www.ietf.org/rfc/rfc2401.txt.

[40] T. Kivinen and H. Tschofenig. Design of the IKEv2 Mobility and Multihoming (MOBIKE) Protocol. IETF RFC 4621, August 2006. http://www.ietf.org/rfc/rfc4621.txt.

[41] K. Kong, W. Lee, Y. Han, M. Shin, and H. You. Mobility management for all-ip mobile networks: Mobile

ipv6 vs. proxy mobile ipv6. *IEEE Wireless Communications*, 15(2):36–45, April 2008.

[42] T. T. Kwon, M. Gerla, and S. Das. Mobility management for voip service: Mobile ip vs. sip. *IEEE Wireless Communications*, 9(5):66–75, October 2002.

[43] D. Maltz and P. Bhagwat. Msocks: An architecture for transport layer mobility. In *Proc. of the 7th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '98), San Francisco, CA, USA*. IEEE, March-April 1998.

[44] R. Moskowitz, P. Nikander, P. J. (ed.), and T. Henderson. Host identity protocol. IETF RFC 5201, April 2008. `http://www.ietf.org/rfc/rfc5201.txt`.

[45] E. Nordmark and M. Bagnulo. Shim6: Level 3 multihoming shim protocol for ipv6. IETF RFC 5201, June 2009. `http://www.ietf.org/rfc/rfc5201.txt`.

[46] E. Perera, R. Boreli, S. Herborn, M. Georgiades, J. Eisl, and E. Hepworth. A mobility toolbox architecture for all-ip networks: an ambient networks approach. *IEEE Wireless Communications*, 15(2):8–16, April 2008.

[47] X. Pérez-Costa, M. Torrent-Moreno, and H. Hartenstein. A performance comparison of mobile ipv6, hierarchical mobile ipv6, fast handovers for mobile ipv6 and their combination. *ACM SIGMOBILE Mobile Computing and Communications Review*, 7(4):5–19, October 2003.

[48] M. Poikselkä, G. Mayer, H. Khartabil, and A. Niemi. *The IMS IP Multimedia Concepts and Services in the Mobile Domain*. Wiley, 2004.

[49] A. Pontes, D. dos Passos Silva, J. Jailton, O. Rodrigues, and K. L. Dias. Handover management in integrated wlan and mobile wimax networks. *IEEE Wireless Communications*, 15(5):86–95, October 2008.

[50] W. A. Romijn, D. Plas, D. Bijwaard, E. Meeuwissen, and G. van Ooijen. Mobility management for sip sessions in a heterogeneous network environment. *Bell Labs Technical Journal*, 9(3):237–253, July 2004.

[51] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler. Sip: Session initiation protocol. IETF RFC 3261, June 2002. `http://www.ietf.org/rfc/rfc3261.txt`.

[52] A. K. Salkintzis. Interworking techniques and architecture for wlan/3g integration toward 4g mobile data networks. *IEEE Wireless Communications*, 11(3):50–61, June 2004.

[53] T. Savolainen and J. Kato. Improved dns server selection for multi-homed nodes. IETF Internet-draft (work in progress), October 2011. `http://tools.ietf.org/html/draft-ietf-mif-dns-server-selection-07.txt`.

[54] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson. Rtp: A transport protocol for real-time applications. IETF RFC 3550, July 2003. `http://www.ietf.org/rfc/rfc3550.txt`.

[55] H. Schulzrinne and E. Wedlund. Application-layer mobility using sip. *ACM SIGMOBILE Mobile Computing and Communications Review*, 4(3):47–57, July 2000.

[56] P. Seite, G. Feige, T. Melia, and J. Zuniga. Connection manager requirements. IETF Internet-draft (work in progress), October 2010. `http://tools.ietf.org/html/draft-seite-mif-connection-manager-02.txt`.

[57] H. Soliman, C. Castelluccia, K. E. Malki, and L. Bellier. Hierarchical mobile ipv6 mobility management (hmipv6). IETF RFC 4140, August 2005. `http://www.ietf.org/rfc/rfc4140.txt`.

[58] W. Song, J. Chung, D. Lee, C. Lim, S. Choi, and T. Yeoum. Improvements to seamless vertical handover between mobile wimax and 3gpp utran through the evolved packet core. *IEEE Communications Magazine*, 47(4):66–73, April 2009.

[59] V. Srivastava and M. Motani. Cross-layer design: A survey and the road ahead. *IEEE Communications Magazine*, 43(12):112–119, December 2005.

[60] R. Stewart, Q. Xie, K. Morneault, C. Sharp, H. Schwarzbauer, T. Taylor, I. Rytina, M. Kalla, L. Zhang, and V. Paxson. Stream control transmission protocol. IETF RFC 5201, October 2000. `http://www.ietf.org/rfc/rfc5201.txt`.

[61] C. Tseng, L. Yen, H. Chang, and K. Hsu. Topology-aided cross-layer fast handoff designs for ieee 802.11/mobile ip environments. *IEEE Wireless Communications*, 43(12):156–163, December 2005.

[62] G. Tsirtsis, G. Giaretta, H. Soliman, and N. Montavont. Traffic selectors for flow bindings. IETF RFC 6088, January 2011. `http://www.ietf.org/rfc/rfc6088.txt`.

[63] G. Tsirtsis, H. Soliman, N. Montavont, G. Giaretta, and K. Kuladinithi. Flow bindings in mobile ipv6 and

nemo basic support. IETF RFC 5201, January 2011. http://www.ietf.org/rfc/rfc5201.txt.

[64] Q. Wang and M. A. Abu-Rgheff. Interacting mobile ip and sip for efficient mobility support in all ip wireless networks. In *Proc. of the 5th IEE International Conference on 3G Mobile Communication Technologies (3G 2004), London, UK*, October 2004.

[65] M. Wasserman and P. Seite. Current practices for multiple interface hosts. IETF RFC 6419, April 2008. http://www.ietf.org/rfc/rfc6419.txt.

[66] F. Xia, B. Sarikaya, J. K. (ed.), S. Gundavelli, and D. Damic. Radius support for proxy mobile ipv6. IETF Internet-Draft (Work in Progress), January 2012. http://tools.ietf.org/html/draft-ietf-netext-radius-pmip6-08.

[67] C. Yiping and Y. Yuhang. A new 4g architecture providing multimode terminals always best connected services. *IEEE Wireless Communications*, 14(2):36–41, April 2007.

Karl Andersson received his M.Sc. degree in Computer Science and Technology from Royal Institute of Technology, Stockholm, Sweden, in 1993. After spending more than 10 years as an IT consultant working mainly with telecom clients he returned to academia and earned his Ph.D. degree from Luleå University of Technology (LTU) in 2010 in Mobile Systems. Following his Ph.D. degree Karl was appointed Assistant Professor of Pervasive and Mobile Computing at LTU in 2011. Currently Karl is also a postdoctoral research fellow at Internet Real-Time Laboratory, Columbia University, New York, USA. His research interests are centered around mobility management in heterogeneous networking environments, mobile e-services, and location-based services. Karl is a senior member of the IEEE.