Handover Management Architectures in Integrated WLAN/Cellular Networks

George Lampropoulos*, Alexandros Kaloxylos^, Nikos Passas*, Lazaros Merakos*

*Department of Informatics and Telecommunications
University of Athens
Athens, Greece
{glambr, passas, merakos}@di.uoa.gr

^Department of Telecommunications Science and Technology
University of Peloponnese
Tripoli, Greece
kaloxyl@uop.gr

Abstract- **The integration of existing cellular systems with new wireless access technologies, such as wireless LANs, has attracted considerable attention during the past few years. The challenges to be addressed include authentication, security, QoS support and mobility management. Efficient mobility management, and especially handover management, is considered as one of the major factors towards seamless provision of multimedia applications across networks of different technologies. A large number of solutions have been proposed in an attempt to tackle all relevant technical issues concerning handover management. In order to evaluate these solutions, a more systematic categorization is needed. This survey gives an overview of the most recent handover management architectures for integrated WLAN/Cellular networks, focusing mainly on 802.11-based WLANs and GPRS/UMTS cellular networks. The various proposals are categorized based on the point of system integration, their main characteristics are presented and their advantages and shortcomings are discussed in an overall evaluation section.**

Keywords: ***handover management, 802.11, GPRS/UMTS, loose coupling, tight coupling, very tight coupling***

1. Introduction

During the past few years, a race takes place among cellular operators to upgrade their infrastructure towards Third Generation (3G) systems (mainly, the Universal Mobile Telecommunication System – UMTS). At the same time, a recent phenomenon is the rising popularity of wireless LANs (WLANs), especially with the wide use of 802.11-based networks. WLANs offer low deployment cost and high communication rates (up to several hundreds of Mbps) in the unlicensed frequency bands of 2.4 GHz and 5 GHz. Based on proper planning to handle interference problems, these systems are considered as perfect candidates especially for wireless hot-spots, where users can enjoy increased bandwidth in limited geographical areas.

The high penetration and data rates of WLANs triggered operators and manufacturers to investigate the possibility of integrating them into 3G systems, in order to provide better quality and a wider range of services to their users. These new heterogeneous infrastructures are often referred to as Beyond Third Generation (B3G) or Fourth Generation (4G) systems.

Design and implementation of these systems face many technical challenges, such as unified Authentication, Authorization and Accounting (AAA), global security provision, common Quality-of-Service (QoS) support, integrated location management and inter-system handover support (known as "vertical handover") [1]. Despite the efforts in WLAN/UMTS interworking standardization [2, 3, 4], work is still required towards more efficient handover management schemes.

Intensive efforts from the research community during the past few years have tried to identify the unresolved issues and propose specific solutions for handover management. Their main difference depends on the way coupling of cellular networks with WLANs is performed. Since different categorization schemes have been proposed [2, 4-9], the adoption of a simple way of grouping the different solutions is considered important in order to highlight major and minor similarities and differences between them. Moreover, the systematic study of the interworking architectures can result in the revelation of more in-depth characteristics regarding vertical handover management.

This survey paper analyzes the most recent research efforts in the area of handover management in integrated WLAN/Cellular networks, attempting to categorize and comment on the proposed solutions. The focus is placed mainly on the ways to integrate two different architectures (i.e., 802.11-based WLANs and GPRS/UMTS cellular networks) and on the supported functionality of the integrated system.

The rest of the paper is organized as follows: section 2 presents an overview of the fundamental characteristics of handover management. Its implementation aspects in UMTS networks and WLANs are discussed in section 3. Section 4, describes various categorization attempts and defines the categorization scheme that is followed in this paper. In section 5, major handover management solutions in integrated WLAN/Cellular networks are presented, categorized as loose, tight and very tight coupling. Section 6 summarizes the most important characteristics of the presented solutions and discusses advantages and shortcomings. Standardization activities related to 3G/WLAN interworking are described in section 7. Finally, section 8 concludes the survey.

2. Handover management principles

Handover management is a fundamental operation for any mobile network. Although its functionality and implementation differs among the various technologies, some basic characteristics are common.

Handover management enables the network to keep active connections during the Mobile Terminal (MT) movement or even balance the network load evenly among different areas. The handover process can be divided into three stages: *initiation*, *decision* and *execution* [10]. Handover initiation is responsible for triggering the handover according to specific conditions such as radio bearer deterioration or network congestion. During the handover decision stage, the decision for the most appropriate new Access Point (AP)[1] is taken. At this stage several parameters (e.g., signal strength of neighboring APs, available radio resources, etc) are considered before a final decision is reached. Finally, at the last stage, the required signaling exchange for communication re-establishment and data re-routing through the new path is made.

Three are the main alternatives for handover decision depending on the way the network and the MT contribute to it: *network-controlled* handover, *mobile-assisted* handover and *mobile-controlled* handover [11, 12]. In a network-controlled handover, the network decides based on the measurements of the received radio signal from the MTs at a number of APs. This is a completely centralized solution that provides the network with the ability to apply its policy. The main drawbacks of this approach are: i) the requirement for considerable computational power at a central point of the system, and ii) the lack of knowledge about the current conditions at each MT. In a mobile-assisted handover, the MT performs several measurements but the network takes the final decision. In this way, real-time conditions at the MT can be taken into account, although the network still faces major signaling and computational load. In a mobile-controlled handover, the MT has the authority and intelligence to select the target AP based on its own measurements. Obviously, this is a distributed solution where the MTs share the handover decision-making load. This, however, may have impacts on aspects such as network stability, fairness and security, since a global policy cannot be applied.

One of the major aims of handover management is to preserve the communication quality during the handover of a MT. This is strongly related to the way the old links are released and the new ones are established. In this respect, two basic types of handover exist: the *soft* and the *hard* handover. A handover is identified as soft if at least one active link exists between the MT and any AP (old or new) during the entire handover period. This type of handover is mostly used in CDMA systems where a MT can communicate using two different codes in the same frequency and at the same time. If the MT has an active link with only one AP at a time, the handover is referred to as hard [11]. This type of handover is usually found in

[1] The term “Access Point” is used to describe any point of attachment between the mobile terminal and the network.

TDMA and FDMA systems where a period of time is needed for the MT to get synchronized in the new time slot or frequency. The challenge is to always perform soft handovers if these are supported by a system and if this is possible. In some occasions however (i.e., abrupt signal loss due to physical obstacles), the execution of hard handovers cannot be avoided.

3. Handover management in UMTS and WLANs

Before proceeding with the description of handover management for integrated UMTS/WLAN networks, it is necessary to briefly recall its main characteristics in each infrastructure.

According to 3GPP's Release 6 specifications [13], UMTS infrastructure is logically divided into the Core Network (CN) and the Access Network (AN). The UMTS CN is further logically divided into i) the packet-switched domain (PS-domain), where packets are routed independently, ii) the circuit-switched domain (CS-domain), where dedicated resources are granted for voice calls, and iii) the IP Multimedia Subsystem (IMS) that provides IP multimedia services over the PS-domain. In the CN, the routing of data between the UMTS network and the external network is performed at the Serving GPRS Support Node (SGSN) via the Gateway GPRS Support Node (GGSN) for the PS-domain. Similar functionalities are performed at the Mobile Switching Centre (MSC) and the Gateway Mobile Switching Centre (GMSC) for the CS-domain. The Home Subscriber Server (HSS), which maintains the users' profiles, is common in both domains. The CN can connect to different types of ANs concurrently. An AN can be either a Base Station System (BSS), offering GSM/GPRS services to Mobile Stations (MSs), or a Radio Network System (RNS), accustomed for UMTS services to User Equipments (UEs). A BSS consists of Base Tranceiver Stations (BTSs) and one Base Station Controller (BSC) that are responsible for radio communications and radio resource control respectively. Similar functionalities are provided by the respective RNS entities, Node-Bs and the Radio Network Controller (RNC). The part of the network that consists of RNCs and Node-Bs is the UMTS Terrestrial Radio Access Network (UTRAN). The UMTS architecture with the interfaces between the respective network components is depicted in Figure 1.

Besides the two basic handover types (i.e., hard and soft), UMTS also defines *softer* handover, utilizing the ability of the terminal to have two active communication paths (macro diversity) with the same access point [14]. Apart from the hard handover, when communication is abruptly lost, softer and soft handovers are performed in most of the cases. Softer handovers occur when the terminal moves within the area of the same Node-B (intra Node-B), while soft handover applies to the case of movement between different Node-Bs (inter Node-B/intra RNS), between different RNSs (inter Node-B/inter RNS/intra SGSN) and between different SGSNs (inter Node-B/inter RNS/inter SGSN). The UMTS handover procedure involves quite complicated protocols such as the Radio Resource Control protocol (RRC) [15]. RRC is responsible for cell selection, paging, UE measurements, RNS changes and control of radio bearers, physical and transport channels. Most of the RRC functionality is implemented in the RNC. In RRC, different functional entities handle the signaling to UEs, the paging and the broadcasting. Furthermore, some other protocols are also involved in the mobility procedure [16]. GPRS Mobility Management protocol (GMM) is used to support the mobility of the terminals [17]. Attach, routing area updates, detach, paging and authorization are only part of its functionality. GMM is placed on top of another signaling protocol, the Radio Access Network Application Part (RANAP), which is responsible for the establishment of different signaling channel to each UE.

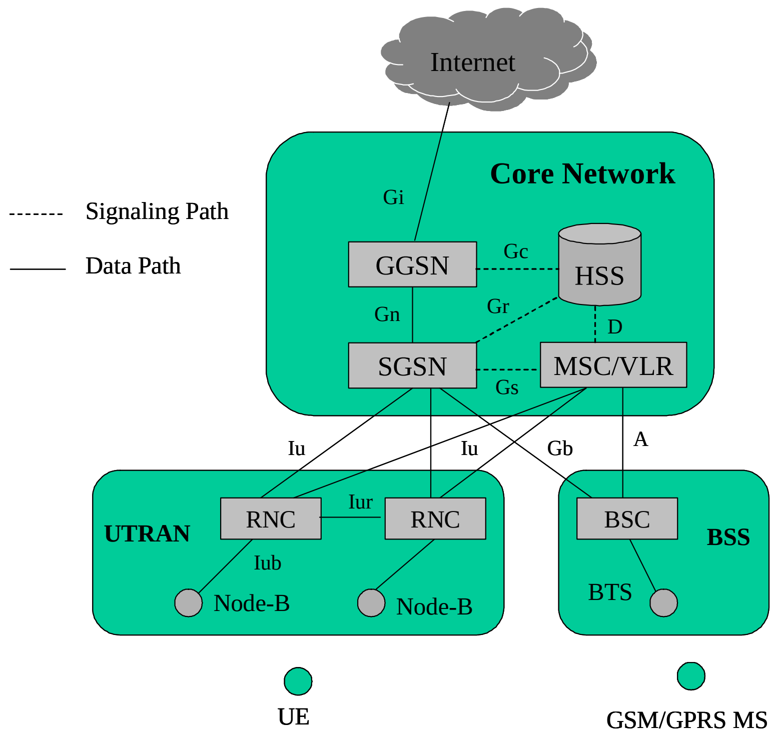

Figure 1. UMTS architecture

On the other side, handover management in WLANs is performed in a much simpler way. Since WLANs are IP-based networks, the Mobile IPv4 protocol (MIPv4) is widely used for handover management [18]. Its purpose is to maintain IP connectivity for the terminal after a movement to a different subnet. In MIPv4, the main functional entities are the Mobile Node (MN), the Home Agent (HA) and the Foreign Agent (FA). These are illustrated in Figure 2.

As the MN moves between subnets, it obtains different temporary IP addresses (referred to as "care-of addresses") and reports them to its HA, located at its home network. The role of the HA is to capture packets destined to the constant IP address of the MN and forward them to its current care-of address. Forwarding is performed through "tunneling", a well-known IP technique where the original packet is encapsulated into a new packet with a new destination address (the care-of address). The FA, located at the visited network de-tunnels the original packet and delivers it to the MN. In the opposite direction, packets are routed directly from the MN towards its Correspondent Node. MIPv4 has further evolved into Mobile IPv6 (MIPv6) [19], where extended addressing, elimination of foreign agent necessity and route optimization capabilities have been included. The MIPv4 and MIPv6 protocols are mainly suitable for inter-domain mobility, due to intrinsic latencies in move detection and registration. For intra-domain mobility, a plethora of protocols that can perform fast handover have been proposed. HAWAII, Cellular IP and Hierarchical IP are only some of the representatives in this area [20].

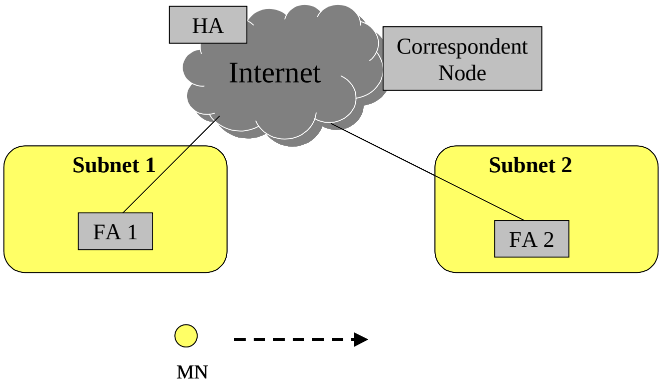

Figure 2. Mobile IP architecture

In order to assist both inter-domain and intra-domain handover and further reduce handover latency, some solutions suggest establishing communication between neighboring APs. For example, a lot of effort has been placed on the development of the Inter Access Point Protocol (IAPP) [21], standardized by IEEE 802.11 Task Group f as a recommended practice. The IAPP provides the means for transferring context from one AP to another, through a fixed distribution system. An example of the use of IAPP can be found in [22], where the protocol is used for transferring the QoS context that pertains to a particular MN. Using this context, the new AP can initiate the signaling required for re-establishing the reservation paths of active MN flows, in order to reduce the handover latency. In different case, the MN would have to initiate this process itself, after re-association with the new AP, resulting in additional delays and signaling overhead in the radio interface.

4. Categorization of integrated WLAN/Cellular architectures

Several ways have been proposed in the literature in order to classify the different interworking architectures [2, 4-9]. Although the same terms are usually used in these classifications, there are important deviations in their meaning. The first standardization effort towards the categorization of solutions has been made by ETSI [2]. There, "*loose coupling*" indicates a way of interconnecting independently the two networks, utilizing only a common subscription. Moreover, "*tight coupling*" suggests that WLAN appears as another access network to the cellular core network, thus, both data and signaling traffic is transferred through the cellular network.

Additional proposals have followed after the work of ETSI. In [5], "*mobile IP approach*" is the minimum way to interwork two different networks by using Mobile IP mechanisms [18]. In the same categorization scheme, the "*gateway approach*" suggests an interconnection way where data traffic can bypass part of the cellular network, while the "*emulator approach*" is the same with the "*tight coupling*" of the ETSI's categorization. In [6, 7], "*no coupling*" suggests the minimum way of interconnecting two networks. In "*no coupling*", the networks are considered as peers and two options are provided: (i) two separate subscriptions (independent networks with separate subscriptions) and (ii) one common subscription (same as ETSI's "*loose coupling*"). Moreover, "*loose coupling*" and "*tight coupling*" indicate that the two networks appear as one network from the network layer and above. This means that if one of the networks is considered as the *master* network, then its characteristics from that layer and above are adopted by both networks. The network that is absorbed by the *master* network is referred to as *slave network*. The "*loose coupling*" (referred to also as "*medium coupling*" in [7]) differs from ETSI's "*loose coupling*" as it provides a direct interface to the UMTS CN. "*Tight coupling*" is similar with ETSI's "*tight coupling*", but incorporates also a possibility to connect the WLAN directly to a GGSN. Another categorization proposes "*open coupling*", "*loose coupling*", "*tight coupling*" and "*integration*" [8]. The first two ways of interworking are the same with the two variations of "*no coupling*" (separate or common subscriptions respectively), while "*tight coupling*" with the definition of ETSI's "*tight coupling*". "*Integration*" suggests interworking at the access network level and not at the cellular core network level as in "*tight coupling*". The same definitions are used in [9]. However, instead of "*integration*", the term "*very tight coupling*" is used.

From the aforementioned definitions, 4 major interworking possibilities can be derived with respect to the interworking point in the system architecture (Figure 1):

- *After the GGSN (Gi interface) with multiple subscriptions*
- *After the GGSN (Gi interface) with common subscription*

- *At the UMTS CN (Gn or Iu interfaces).*
- *At the UMTS AN (Iur or Iub interfaces).*

More recent standardization efforts by 3GPP [4] have tried to propose an interworking categorization based on the actual service that users experience. In that respect, six scenarios have been specified based on the service experience by the user. This categorization suggests implicitly the level of interworking between the two networks. The six scenarios are:

i) *scenario 1*: indicates only common billing and customer care,

ii) *scenario 2*: provides 3GPP-based access control,

iii) *scenario 3*: enables access to 3GPP PS services from WLANs,

iv) *scenario 4*: allows services to continue after inter-system handover,

v) *scenario 5*: promises seamless functionality in the previous scenario and

vi) *scenario 6*: provides access to 3GPP CS services from WLANs.

It is clear from the previous analysis that definitions are still evolving and that a common base for the categorization of different solutions is needed. According to our approach, a clear base could be the point of integration concerning the system architecture. For proper differentiation of the interworking solutions, the standardization efforts from ETSI [2] and 3GPP [3, 4] should also be taken into account. Following that philosophy, the categorization scheme followed in this paper assumes three major types of interworking: *loose*, *tight* and *very tight coupling*.

Loose coupling indicates that the interworking point is after the GGSN (either common or different subscription schemes) and suggests interconnection of networks using Mobile IP [18] mechanisms. It has the advantage of simple implementations, with minimal enhancements on existing components, but at the expense of considerably larger handover execution time. Since minor architectural adjustments are required, these proposals focus mainly on the elaboration of the handover initiation/decision process and the introduction of novel mechanisms towards performance improvement (in terms of handover execution time and effective management of resources on the wireless link). Although they may not deal with the provision of 3GPP PS services (e.g. WAP and MMS services) from WLANs (scenario 3), they promise service continuity (scenario 4), but without seamless functionality at all times (scenario 5). These solutions enable the deployment of scenarios where the UMTS and the WLAN infrastructures may belong to different providers.

On the other hand, tight and very tight coupling assume interworking at the UMTS CN level (i.e., SGSN or GGSN) and the UMTS AN level (e.g., RNC) respectively. The focus of such proposals is on the technical challenges that arise from the extension of current protocol standards in order to interoperate with each other. Fast mobility functions and seamless functionality (scenario 4 and scenario 5 respectively) are the main targets of these proposals, at the expense of considerable complexity introduced by the enhancements required on existing components. These solutions are usually preferred by the cellular operators that control both the cellular and the WLAN infrastructures.

5. Handover management solutions in integrated WLAN/Cellular networks

According to the categorization scheme followed in this paper, the different interworking architectures between 802.11-based WLANs and GPRS/UMTS networks fall into three major categories: loose, tight and very tight coupling solutions. These are described in the following three sub-sections, while their respective characteristics are summarized in the end of each sub-section.

5.1. Loose coupling solutions

The common factor for all the proposed architectures in this category is the use of Mobile IP as the basic instrument for inter-system mobility and the high level perspective of the integration process. In these solutions, interconnected networks are considered as independent networks concerning the handling of data traffic. Therefore, FAs are placed at the borders of each access network to supply with roaming capabilities. In addition to that, HAs are situated at the home networks along with authentication entities (AAA components for WLAN and HSS for UMTS authentication). If users have different subscription to each network (e.g., SIM-based authentication for UMTS and login/password in WLAN), then they cannot experience service continuity while roaming. On the other hand, a single subscription to one network with roaming privileges to another access network can help to avoid service disruption, as long as the different authentication entities are closely cooperating. Such co-operation has already been standardized by 3GPP ([3]).

Another major characteristic of loose coupling frameworks is the proposal of architectures that try to improve the offered quality by focusing on various aspects of the handover procedure. These aspects include:

- acceleration of Mobile IP procedures,
- advanced handover initiation/decision algorithms, and
- policy-based architectures.

Acceleration of Mobile IP procedures suggests ways of minimizing the handover latency experienced by the execution of the Mobile IP procedures. Proposed mechanisms try to decrease the handover duration by improving the signaling exchange for both intra and inter-domain cases or by taking advantage of triggering events from the link layer [23-24]. The former can be achieved by restricting signaling in the visiting area of the terminal or maintaining mobility information in different parts of the network, while the latter aims at the reduction of the network layer handover delay by receiving a signal from the link layer concerning an upcoming handover. Handover initiation/decision algorithms address issues such as the proper time to trigger a handover or which is the best new point of attachment. Research work focuses on intelligent algorithms that often take into account various parameters apart from the received signal strength such as service type, cost, user preferences, etc [25-27]. Finally, policy-based solutions investigate the effect of a plethora of parameters (as in handover decision) on the handover procedure, but elaborate on these by proposing new functional entities in order to enhance the handover process.

Since some of the loose coupling frameworks regarding fast Mobile IP solutions and handover algorithms have already been surveyed in the literature [23, 25-27], the main focus of this sub-section is on policy-based architectures. The interest on this research field stems from the necessity to deal with the numerous challenges introduced in 4G environments [25]. In *policy-based* schemes, handover is based not only on signal measurements, but also on policies derived from criteria such as the users' profiles, the bandwidth requirements, the terminal capabilities, etc. They differ from fast Mobile IP solutions and handover decision schemes since they propose sophisticated entities for network context management, handover decisions and inter-operability between networks of different providers. Some of these solutions are based on special purpose entities that use proprietary protocols to communicate with each other. There are, however, quite a few efforts that are based on standardized frameworks, such as the IETF Policy Framework [28]. This framework is mainly concerned with means to provide policy-based control over admission control decisions (i.e., decisions for accepting or not a new session). This is done with the application of policy rules in a form of policy models [29], while the communication between the architectural elements is feasible with the COPS protocol [30].

In addition to policy-based architectures, a subset of the implementation activities for loose coupling solutions in the context of academic or industrial-driven projects will also be presented in order to highlight the performance limits set by this type of interworking solutions.

5.1.1. Proprietary policy-based solutions

A framework with advanced handover management perspective in future networks is presented in [31]. It is characterized as a mobile-assisted solution, since useful measurements are gathered from different parts of the system and the terminal as well. The proposed architecture is depicted in Figure 3.

It consists of two main components, the Context Repository and the Adaptability Manager. The Context Repository gathers, manages and evaluates context information from different parts of the network. This information is categorized as either "dynamic" or "static", depending on how often it should be refreshed to reflect up-to-date conditions. Typical dynamic parameters are the current QoS network parameters and terminal's location, while static parameters are the capabilities of the terminal and the user perceived QoS requirements. The Adaptability Manager decides about adaptation to context changes and handover execution. It is divided into two main processes, one for the vertical handover decision and one for the QoS mapping. The former process is responsible for network selection based on the evaluation of terminal's location changes and the QoS of the current and alternative networks. The latter process filters any stream towards the target network, based on context information about the available resources, the terminal status, etc. Heterogeneous networks within a domain form a Domain Network Cluster, which consists of three different types of entities: an Adaptability Manager, a Context Repository and a Proxy. Only one pair of Adaptability Manager and Context Repository exists in a Domain Network Cluster, but there is one Proxy for each different network. The Proxy is responsible for the communication between different domains, such as a Correspondent Host Domain and a Network domain.

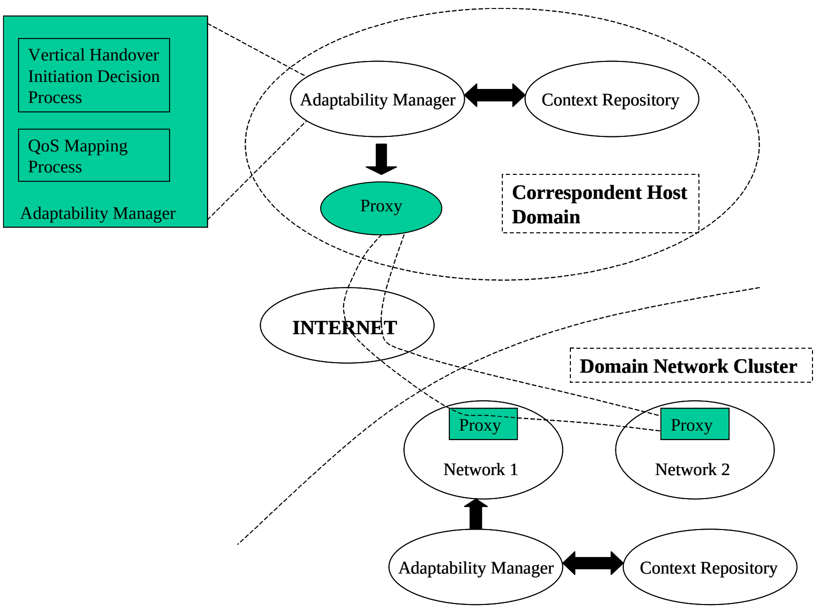

Figure 3. A vertical handover architecture with context information consideration

The proposed architecture incorporates an analytical context categorization and a detailed handover decision algorithm. The introduced components combine context gathering and handover decision processing. Prototype experiments have shown that smart decision mechanisms are necessary for smooth adaptation of the communication streams to different conditions. A drawback of the proposed architecture is that context information gathering is performed at a single point, i.e. the Context Repository. This requires frequent communication between the terminal and the network, resulting in increased overhead on the radio link. Moreover, only one decision entity per cluster exists forming a single point of failure, while handover decision time can be increased considerably in large network infrastructures.

Similarly to the previous solution, the idea to improve the handover performance with advanced context management is also explored in [32]. However, in the last approach, a more flexible software-like deployment scheme is followed, which minimizes the handover decision time. This is achieved by software agents that are used for the preparation of the collected context data and the algorithm needed for the handover at the context collection point. The software module that includes this information is downloaded at the decision point (e.g., at the terminal) in advance and invoked at handover time. Experiments in a real prototype have indicated that, using a gathering phase of 1.8 ms, the system can have a fast response to load changes. The signaling overhead, however, may reach 9.9% of the UMTS bandwidth (2 Mbps) in worst-case conditions.

An information-gathering phase is also proposed in [10], but it consists of more sophisticated entities for the measurement and evaluation functions than the previous approaches. The description has been extended with flow charts analyzing the communication between these entities in a mobile-controlled handover framework. Furthermore, policy export from the network towards the terminal is proposed in [33]. As in the previous solutions, frequent information exchange between the mobile terminal and the network takes place, while proprietary protocols are used for the policy exchange. Testbed evaluation demonstrated that handover losses are minimized if policies are applied before the handover initiation.

5.1.2. Policy-based solutions according to IETF Policy Framework

In contrast to the solutions of the previous group, these solutions use the IETF Policy Framework. This strengthens their deployment prospects as they follow a standardised way to introduce advanced handover capabilities.

A representative solution that uses the IETF Policy Framework has been proposed in the context of the project M-Zones [34-36]. It is a mobile-assisted handover scheme, where system functionality is separated into the network functions (network policy engine) and the mobile functions (mobile policy engine). A couple of Policy Enforcement Point (PEP) and Policy Decision Point (PDP) exist in both sides, along with policy repositories. A PEP is responsible for the execution of a policy that is decided inside a PDP. The policy repositories define the policies that have to be followed for proper handover decision. The proposed system architecture is depicted in Figure 4.

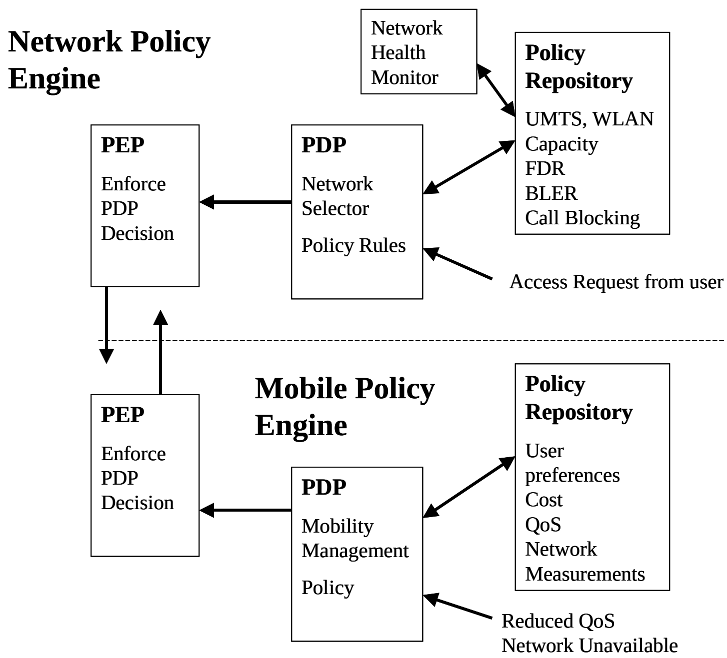

Figure 4. A system architecture using IETF Policy Framework

Two major inter-related procedures are considered in the system for call admission control and handover management. In call admission control procedure, PEPs located at the terminals consult a PDP residing at the network for available resources. A network PDP can be located at the RNCs in UMTS networks, in BSCs in GSM/GPRS and in gateway routers in WLANs. The decision made at the network PDP is mainly based on link level parameters for active connections and current load information for the cells. Performance parameters, such as Frame Drop Rate (FDR), Block Error Rate (BLER) and Call Blocking, are stored in a policy repository and fetched each time the network's load (network's "health") needs to be estimated. If under-loaded cells are available, the call is admitted and a response is sent from network PDP to terminal PEP to inform it about the granted resources. Handover management is more sophisticated. A PDP placed at the terminal is able to determine the proper time to initiate a handover based on information stored locally, such as user preferences, QoS parameters, cost policy and signal measurements. The mobile PDP may also suggest possible new points of attachment to the network PDP. If the handover initiation is approved, a request is sent to the network PDP and the request is handled as in the case of call admission control.

In this way, the proposed architecture introduces a two-level decision mechanism during handover, which gives flexibility to the terminal and the network to make the best possible handover decision. Another advantage is that load balancing is also possible by the network, and this has been demonstrated through preliminary simulations. In a scenario with 200 users, FDR for video sessions was preserved under 1% during most of simulation time with policy network selection, while peaks over 2% were frequently experienced for FDR with random network selection in the same scenario. The scenario assumed UMTS and EDGE access technologies, but no WLAN option was included. Additionally, one disadvantage of the proposed solution is that the call admission control procedure does not consider the various parameters at the terminal. The network solely takes the decision to accept a session without considering the PDP in the terminal. In this way, neither the current conditions of the terminal nor user's preferences are taken into account.

Similarly to this effort, in [37] the IETF Policy Framework is used to extend the management framework and incorporate service, user, terminal, network access and Always Best Connected managers for better data collection and handover decision. The proposed framework correctly adopts a Local PDP (LPDP) scheme located in the mobile terminal, similar to the two-level decision scheme in the previous solution. LPDP, however, has the authority to decide about the best point of attachment for the user only when the terminal connects with the network for the first time (in this case the network PDP is out of reach). This makes the LPDP less effective in contrast to the mobile PDP of the previous solution. Another disadvantage of the proposed scheme is that it does not specify the inter-domain communication.

A different policy-based solution is proposed in [38]. Although the advantage of a PDP in the mobile terminal is not exploited as in the two previous solutions, reduction of the handover latency is accomplished by minimizing the time required to discover the new access point. A drawback of this proposal is that methods and latencies to fetch context information during the candidate access point classification procedure are not considered. Another important approach has been proposed in [39]. Despite the fact that the proposal focuses on QoS management, handover management could be provided in a similar way. This framework introduces a hierarchical structure of PDPs that can be used for policy exchange between

WLANs and UMTS networks. According to that scheme, PEPs are incorporated in WLAN routers, while the highest-level PDPs relay and translate the different networks' policies. The major advantage of this proposal compared with all the previous ones of this sub-section is that it can interconnect efficiently and evenly WLAN and UMTS domains controlled by different operators.

5.1.3. Implementation activities

Apart from design proposals on policy-based handover, various efforts, mainly in the framework of research projects and industrial initiatives, have tried to implement concrete and detailed architectures. A major framework in this area was defined in the context of the CREDO project [40]. The architecture consists of different radio technology segments (i.e., IEEE 802.11b, GSM/GPRS and DVB-T), which are interconnected in an IP backbone. As in [39], interconnection between different operators is possible. However, in contrast to [39], the composite network is owned by the cellular operator, which borrows/rents resources from other operators.

The main components of the architecture are the Network and Service Management System (NSMS) and the Terminal Station Management System (TSMS). A NSMS is responsible for the resource management of one access network segment. TSMS is placed at the terminal and informs its supervising NSMS about radio conditions and QoS characteristics of active connections at the terminal. Their communication is performed through the CREDO TSMS-NSMS Protocol (CTNP), while various NSMSs interact with each other for inter-system handover. A NSMS consists of two major entities, as shown in Figure 5: i) the Terminal Context Manager, which is responsible for the communication between NSMS and TSMS, and ii) the Network Manager, which performs the selection of access network, decides on the QoS provided to the services and monitors the network.

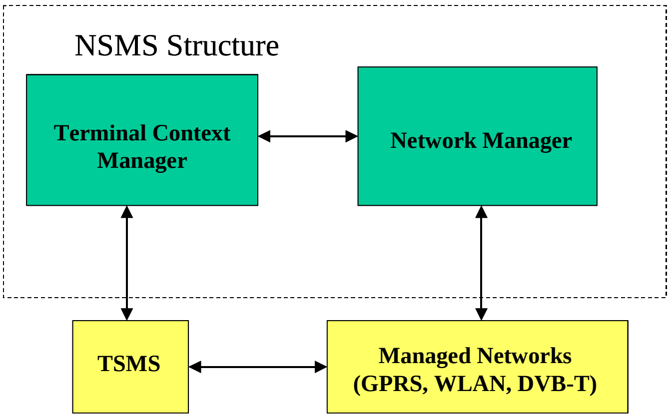

Figure 5. NSMS structure

CREDO provides a mobile-assisted handover architecture including a detailed network element definition, policy characteristics and simple decision algorithms. Although it offers sophisticated resource management in real time, it requires frequent transmission of terminal status information and available network list from the terminals towards the NSMS, resulting in increased overhead. Moreover, the results obtained through simulation and a real test-bed have shown that the required time to change to a different network and initiate a new service is nearly 0.345 seconds in the GPRS to WLAN case and about 1.35 seconds in the WLAN to GPRS case. It has also been shown that CREDO avoids congestion by distributing the terminals to alternative networks, maintaining at the same time the QoS levels for the users.

Two important implementations in the context of industrial activities have been presented in [41] and [42]. Although they do not follow a policy-based philosophy as in [40], QoS issues are taken into account during the handover process. In the first solution, a WLAN gateway (Integration Of Two Access technologies - IOTA gateway) is introduced to support loose integration. This entity supports mobility management using Mobile IP and additionally, web caching, QoS and accounting. It offers delays up to 610 milliseconds for Mobile IP FA discovery on the new interface and Mobile IP registration, while QoS guarantees can be maintained for different classes of users. Despite the performance enhancement, this solution is tailored to independent Wireless Internet Service Providers (WISPs) that interwork with CDMA 2000 networks [43]. Support of UMTS is more complicated and requires significant changes to the accounting, mobility and billing mechanisms.

In [42], a WLAN segment is loosely coupled with a public GPRS network using MIPv6. This work's scope is to examine the impacts of vertical handovers on the TCP protocol [44] performance and identify the network parts that contribute to the overall handover latency in order to improve it. Therefore, accounting issues are not considered as in [41]. Measurements from a test-bed have shown that WLAN to GPRS handover requires about 4 seconds, while the same time for GPRS to WLAN handover is nearly 7 seconds due to increased buffering in GPRS GGSN. These delays can be reduced if techniques, such as Fast Router Advertisement, client-based Router Advertisement Caching, client-Assisted Simulcast of Binding updates and soft handovers with Router Advertisement caching, are used. With the last technique, performance in handover time can be improved more than 10 times, but with a high probability of out-of-order data packets delivery.

A WLAN/GPRS system has also been tested in [45]. The main innovation of this approach is the integration of two novel entities for detection of network changes and maintenance of end-to-end connectivity. This approach differs from [40] as the network changes are based only on link layer parameters and not on any sophisticated policy (e.g., service quality degradation or congestion). Experiments have verified that an accurate handover decision leads to a reduced number of handovers, resulting in over-doubling the TCP throughput.

5.1.4. Loose coupling solutions' characteristics

Loose coupling schemes concentrate on mechanisms required for advanced handover initiation/decision and execution. The use of Mobile IP and the independency between cellular networks and WLANs have solved the basic architectural problems. In addition, a plethora of different parameters influence the handover procedure. Therefore, a major trend is to introduce various management entities for the collection and processing of context information in the network and the mobile terminal, and to use this information in order to minimise handover latencies and allocate the resources fairly. The use of proprietary protocols gives the advantage of more flexible architectures, but their application in industrial implementations is generally considered a risk. On the other hand, using standard policy frameworks (e.g., IETF Policy Framework) provides a common architecture with changes made mainly in the policy design. The adoption of context management techniques results in performance improvement that is favourable for both the users and the network operators, but at the same time it provides significant implementation complexity and signaling overheads.

Another feature of loose coupling schemes is the almost seamless handover, as handover latencies vary from some hundreds of milliseconds to some seconds. The type of handover supported is mainly the hard handover, although cross-layer techniques can result in smaller handover latencies [23]. Furthermore, in all proposed schemes collaboration between the network and the terminal is necessary for a proper handover decision and better QoS provision. Mobile-assisted handover schemes are mostly preferred. Despite that, some solutions consider the terminal as the best place for the final handover decision, but problems in applying a global network policy and bandwidth restrictions prevent from implementation.

Furthermore, it could be stated that an ideal policy-based scheme should incorporate:

- a standardised policy framework as the one proposed by IETF,
- context gathering entities for better parameter collection,
- decision entities both at the terminal and the network for enhanced handover decisions, and,
- capability of interconnecting multiple networks of different providers.

Concerning the implementation of fast Mobile IP handover schemes, it is worth mentioning that the effect on higher layer protocols (e.g., TCP) should be also considered apart from network layer measurements, as it is often contributing to the handover delay seen by the application user.

5.2. Tight coupling solutions

Tight coupling is considered to apply in cases where a WLAN is directly attached to a UMTS CN component (either GGSN or SGSN) affecting the functionality of this component. Such solutions may use part of the UMTS functionality at the integration point or apply Mobile IP procedures, offering soft or hard handovers respectively. Another characteristic is that mobile-assisted and network-controlled handovers are exclusively used as these comply with the respective handover types existing in UMTS [14].

5.2.1. Coupling at the GGSN level

A way to tightly integrate UMTS and WLAN is presented in [46]. Despite the interworking degree, the two networks remain peer and handle their own subscribers independently. In this proposal, a new logical node called Virtual GPRS Support Node (VGSN) is introduced. VSGN is used to interconnect the UMTS and WLAN backbones. Its main functionality is to exchange subscriber and mobility information, and to route packets between the two networks. VGSN acts in this way as a gateway in WLAN and as a GPRS Support Node in UMTS. VGSN can be implemented as an independent node or integrated in the WLAN gateway or SGSN or GGSN. The proposed architecture is depicted in Figure 6.

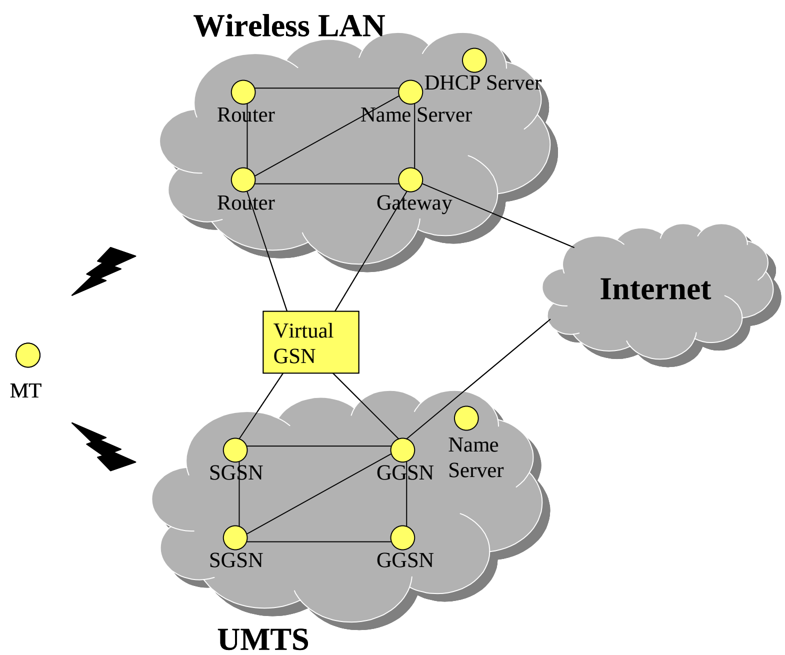

Figure 6. UMTS/WLAN interworking with VGSN node

In case of UMTS to WLAN roaming VGSN acts as the new SGSN to which the terminal handovers. Packets towards an Internet host can bypass the UMTS network, while in the reverse direction tunnelling between GGSN and VGSN manages to route the packets to the terminal in the WLAN. When a WLAN subscriber enters the UMTS network, VGSN acts as the GGSN in the UMTS network and incoming packets from the Internet traverse WLAN gateway, VGSN and SGSN sequentially before reaching the terminal.

With the VGSN approach, UMTS and WLAN handle their subscribers independently and no Mobile IP functionality is needed. Moreover, simulations have shown that VGSN approach manages to provide an average bandwidth per user similar to that in a loose coupling case where Mobile IP is used. Handover latency is also considerably lower when compared with that case. The main drawback of this proposal is that unnecessary routing delays for both roaming scenarios could be caused if the location of VGSN is not properly selected. A replication of VGSN functionality would ease the network design and increase its availability in case of failure. Another disadvantage is that the usage of VGSN asks for stronger co-operation than a simple roaming agreement between the two network providers, since sensitive UMTS network information could be revealed to the WISP, without UMTS provider's allowance.

A combined SGSN/GGSN emulator (GSN') is used to interconnect UMTS and WLAN in [6]. In contrast to the previous proposal, the terminal obtains a new IP address while roaming between different networks. This requires higher layer mobility solutions to be used for enhanced handover support (e.g., SIP [47]), although some delays to the handover procedure and extra complexity are added. The main advantage is that only UMTS signaling traffic passes through the UMTS infrastructure, leaving unaffected the UMTS CN from the high amount of WLAN data traffic. This means that the GSN' approach manages to make better use of the UMTS CN resources than in [46], where only uplink packets can bypass the cellular infrastructure. Moreover, GSN' is considered to provide flexible deployment also in cases where different operators share the WLAN and the UMTS networks. This, however, asks for proper co-operation between the two operators as in [46], since UMTS signaling has to reach indirectly the UMTS CN. Another drawback of GSN' is that it is implemented as a separate node and may result in a single-point of failure, while VGSN in [46] can be incorporated in different network elements and survive after hardware failures.

5.2.2. Coupling at the SGSN level

Another solution proposed in [48] describes an architecture where the coupling is done at the SGSN. For that purpose, a Gateway Interworking Function (GIF) connects the WLAN network with SGSN. GIF is responsible for hiding WLAN particularities at SGSN. The key to the proper functionality of GIF is the WLAN Adaptation Function (WAF). WAF is placed on top of WLAN radio subsystem and makes the use of WLAN radio transparent to upper GPRS protocols. Moreover, WAF is found both in the terminal and the GIF protocol stacks. Communication between peer WAF entities provides for the exchange of GPRS signals and WLAN data. WAF functions include the activation of WLAN interface, discovery and paging procedures, QoS support at the terminal and the GIF, and transport services. The system's architecture is shown in Figure 7.

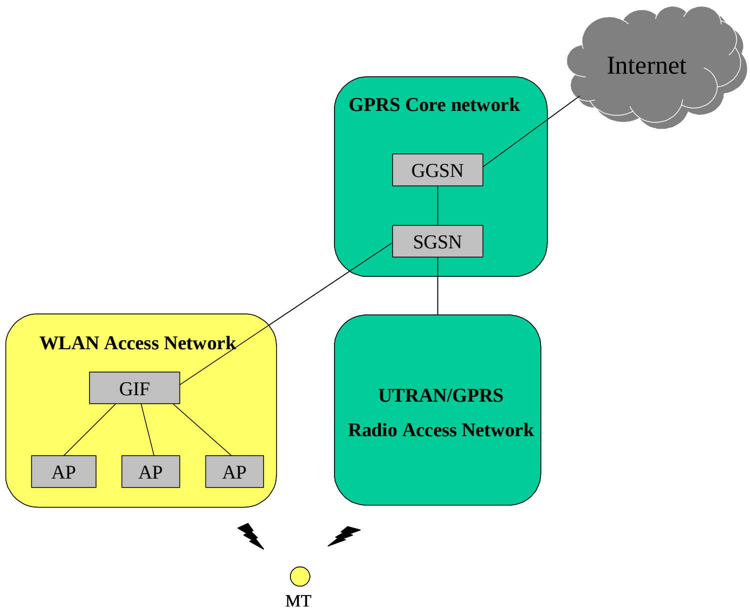

Figure 7. Tight coupling at the SGSN with GIF

The main advantage of this solution is the enhanced mobility management, based on existing GPRS functionality that ensures at least service continuity (scenario 4, and scenario 5 if QoS support in WLAN is assumed), including authentication, authorization, accounting and billing. A large part of the GPRS infrastructure is reused, minimizing the cost of deployment. Such solutions, however, are tailored to WLANs deployed by cellular operators and few opportunities for profits apply for independent WISPs. Additionally, several extensions are needed in SGSN and GGSN nodes to support the large amount of data from WLAN users through the UMTS network, while terminals should also include integrated GPRS/WLAN functionality. Another attempt to re-use AAA procedures and offer access to 3GPP PS services from WLAN is proposed in [49]. The main advancement is the use of UMTS instead of GPRS, but no solution for session continuity is given.

Similarly to [48], an enhanced WLAN access point is connected directly to SGSN in [50]. This node supports authentication and access capabilities (scenario 2), but does not provide service continuity as in [48]. The proposed scheme focuses on load balancing issues, such as the importance of shifting data users from UMTS to WLAN in order to increase the availability of the network. Simulation results have shown that load balancing and latencies for standard UMTS procedures are improved when the WLAN interface is used. Despite that, data session set-up delay and application response time in UMTS/WLAN terminals are increased, compared to the case where only WLAN functionality is present. This is the result of additional overhead added by the UMTS session management protocol in the dual-mode terminals. This proposal is expected to be more easily implemented than in [48] as seamless functionality is not supported. However, [48] is superior in terms of QoS support, because WAF also includes a scheduling mechanism for the transmission of packets both in the terminal and GIF.

Another coupling solution at the SGSN is presented in [51]. WLAN is connected to the UMTS network through border routers. The key difference from [48] and [50] is that Mobile IP functionality is used. In turn, this means that it performs less efficiently than these approaches. The main contribution of this approach, compared to all the previous tight coupling solutions, is that a mobile node can maintain one data connection through WLAN and one voice connection through UMTS at the same time. This can happen in a hot-spot scenario where WLAN micro-cells are included in larger UMTS macro-cells. However, in the hot-spot scenario, UMTS is restricted only for voice calls and WLAN only for data communications. A similar load balancing policy has also been followed in [50], where data users were shifted to WLAN in order to release UMTS resources for voice calls. However, both policies result into inefficient use of the available resources, since the transfer of voice calls through WLAN or the overloading of WLAN are not considered. One advantage of this proposed approach is that the use of IP protocols at the terminal makes the dual mode terminal implementation easier.

In contrast to all the previous tight coupling architectures, two policy-based approaches are considered in [9] and [52] respectively. A sophisticated architecture has been proposed in the context of the EVEREST project [9]. The main architectural characteristics of this solution are: (i) the adoption of IP transport capability at the RNC, the introduction of an Access Gateway (AG) for interconnection of WLAN to SGSN and the inclusion of policy entities according to IETF Policy Framework (adopted by 3GPP [53]). RNC’s IP transport capability enables the network operator to install PEPs at the different radio access networks (e.g., AG, RNC), while PEPs are also installed at the core network routers (e.g., at the GGSN). These

PEPs communicate with PDPs for QoS and bandwidth management (Wireless QoS Broker and Bandwidth Broker) as shown in Figure 8. The AG is similar to GIF ([48]) but it differs from it since it incorporates PEP and IP functionalities. In addition to the aforementioned characteristics, this approach offers the capability of exchanging policies between different domains through the Master PDP as in [39].

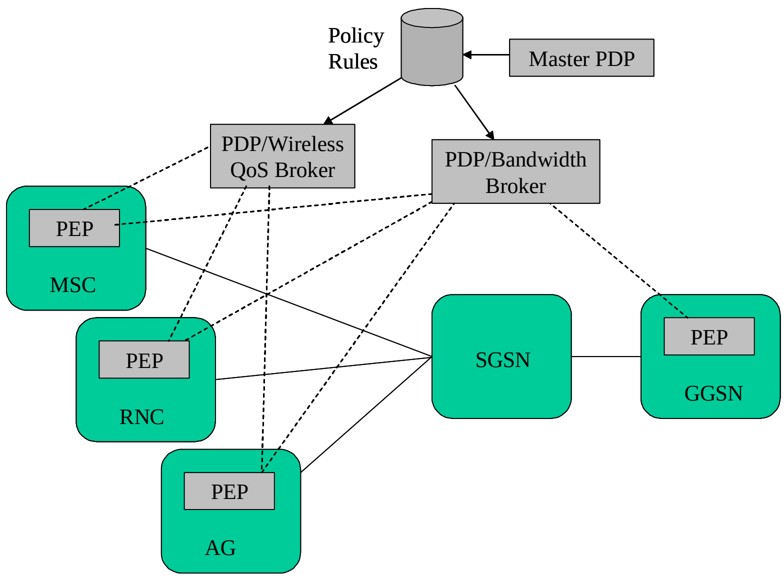

Figure 8. EVEREST approach for tight coupling with policy considerations

In [52], the proposed policy-based scheme is not based on the IETF Policy Framework as in [9]. Moreover, the inter-operability between different providers' networks is not specified. This approach proposes the connection of WLAN with UMTS through an interworking unit (referred to as ERNC). This unit is directly connected to SGSN (as in [48]). The main advantage of this proposal compared to [9] is that it incorporates policy-based decision entities (similar to PDPs) both at the RNC and the mobile terminal. In this way, the mobile terminal manages to handover its data or voice connections according to different parameters such as user, terminal, service profiles, as well as the availability of alternative networks. This results in more flexible and efficient use of the resources for both the WLAN and the cellular network. Since, the architecture is similar to that in [48], most of the UMTS infrastructure is reused, while at least service continuity (scenario 4) is provided.

Interworking at the SGSN or the GGSN is also possible with an adaptation layer [7] placed below RRC in the control plane and below UMTS Packet Data Convergence Protocol (PDCP) layer [54] in the user plane. Simplifications of UMTS functionality are required in order to properly adjust complex link layer features and RRC procedures onto the WLAN protocol stack. The proposed solution provides flexibility to use both radio interfaces seamlessly, but several modifications are needed in RRC functionality. The idea of the adaptation layer is very similar to the WAF found in [48]. Both try to adapt protocols of the same level over WLAN, but this solution concerns UMTS networks and not GPRS networks as in [48]. Another difference is that in [48], more attention is paid to addressing and routing issues between the two networks, while in this solution the mapping of logical channels and transfer modes, as well as the changes in the RRC states/modes are mainly described.

5.2.3. Tight coupling solutions' characteristics

Tight coupling schemes include architectures that either use part of the Mobile IP functionality or base the interworking only on UMTS procedures carefully adapted to comply with their respective ones in WLAN. In both cases the UMTS protocols are enhanced or new ones are added. Whether coupling is done at GGSN or SGSN, integration without using Mobile IP is considered to provide seamless handover functionality. However, this asks for such interworking entities that result in a significant increase on the implementation complexity. AAA problems can be solved reusing UMTS AAA mechanisms and thus decreasing the deployment cost. This kind of interworking is tailored-made for WLAN deployment by cellular operators. On the contrary, integration at the IP layer requires that Mobile IP and/or SIP are used. The inclusion of these protocols slows down the handover process, but the implementation of dual mode terminals is quite easier.

Moreover, different advantages and shortcomings characterize the two levels of tight coupling (i.e., at the GGSN and at the SGSN). In solutions that propose coupling at the GGSN level, data traffic can bypass a large part of the UMTS infrastructure. Therefore, less congestion applies in the UMTS CN than in coupling solutions at the SGSN level. Signaling, however, goes through the UMTS CN and requires a close co-operation between the operators of UMTS and WLAN. Another advantage of coupling at the GGSN level is that it requires less complicated modifications in the UMTS architecture, since changes affect higher level protocols. On the other hand, in coupling at the SGSN level, performance during handover is expected to be better because fewer components located at the lower levels of the UMTS architecture are involved.

In general, all tight coupling solutions are expected to offer less handover latency than loose coupling ones, since the interworking takes place in a point closer to the mobile terminal. Initial performance evaluation efforts give an initial view of what can be expected when the interworking point moves towards the mobile terminal. In [5], simulation results indicate that loose coupling solutions based on Mobile IP may face undesirable handover delays for real-time applications, in contrast to tight or very tight coupling solutions. Moreover, in [8], the importance of minimising the path that packets follow during handover is highlighted. For example, in tight coupling, if an interworking unit connects the WLAN to SGSN, the delay between the RNC and the interworking unit should be kept minimal. If this delay is high, then larger processing power should be placed at the interworking unit in order to keep the handover delay in acceptable limits. By providing significant computational resources, the handover delay could reach the value of 250 milliseconds, which is the preferred delay for real-time data flows as also specified in [55].

5.3. Very tight coupling solutions

Very tight coupling schemes focus on interworking at the UTRAN level and, more precisely, on incorporating RNC or lower UMTS entities' functionality into WLAN components. These solutions can perform soft handover and take the handover decision as in the tight coupling case. Very tight coupling is considered quite complicated and only few solutions have been proposed in this area.

5.3.1. Coupling at the RNC level

A solution in this category is described in [56]. Integration of WLAN with UMTS networks is accomplished at the RNC level with the aid of an InterWorking Unit (IWU). Figure 9 presents the proposed scheme. The coupling is possible in two different ways, depending on the interface with the RNC. If the interworking takes place at the Iub interface, WLAN signaling can be exchanged either over the WLAN or the UMTS radio interface. In case bidirectional data transfer for WLAN data is assumed and WLAN signaling is carried over UMTS, the integration is performed below the MAC layer of UMTS. This introduces a WLAN Interworking Layer located at the terminal and the IWU, which is responsible for protocol translation and signaling exchange with the RNC and the APs. When interworking at the Iur interface is considered, the IWU functions similar to a RNC. The main difference with Iub interworking is that call establishment procedures are not supported over the Iur, therefore this interface must be upgraded in order to allow the WLAN to operate independently.

Both cases require significant enhancements in UMTS functionality and result in increased complexity. Modifications mainly affect RNC and more specifically the RRC. The decision for selecting the proper radio interface, the WLAN management, the handover control, and the RRC state model are some points that should be changed in RRC functionality. The proposed architecture can offer minimal handover delay, but at the expense of considerable implementation complexity. Moreover, it is ideal only for WLAN deployed by cellular operators.

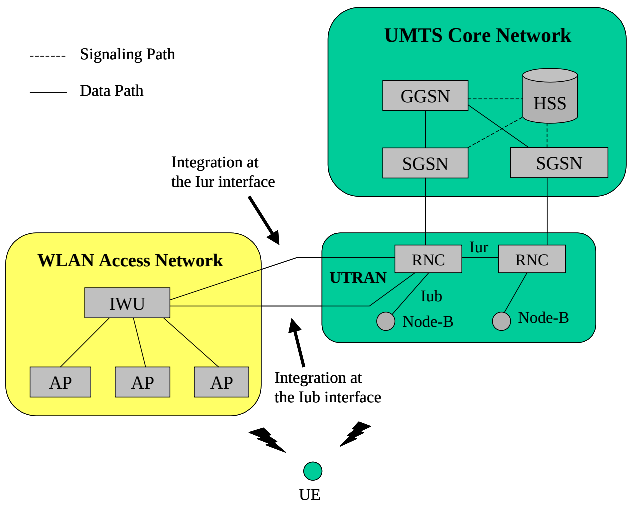

Figure 9. Very tight coupling using different interfaces with RNC

In contrast to the previous solution, a direct interconnection with IP is introduced in [57]. The main difference in the proposed architecture is the introduction of a Radio Gateway (RG) with IP routing capabilities that acts as a Node-B plus a subset of RNC (Figure 10). Moreover, SGSN and GGSN are replaced by IP routers. In order to achieve smooth functionality between UMTS protocols and IP, a Generic Radio Access Adaptation Layer (referred to as GRAAL) is used between IP and RRC in the control plane and between IP and PDCP in the user plane. In this way, only a subset of the standard RRC functionality is used for the transfer of IP messages. An important characteristic of this proposal is that a direct adaptation of IP over the UMTS protocol stack is considered, while in [9] the IP protocol was incorporated only as a transport protocol in the RNC stack.

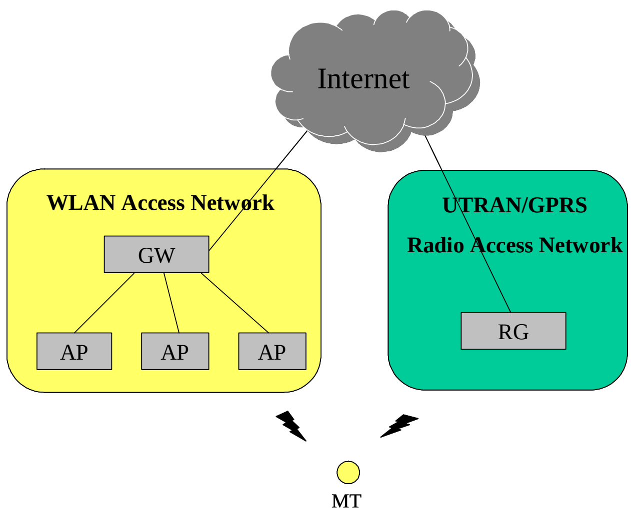

Figure 10. Very tight coupling with IP-capable Radio Gateway

The main advantage of this solution is the simplicity and flexibility offered by the use of IP protocol. At the same time, this means that handover is implemented with Mobile IP mechanisms, which definitely slows down the handover process and results in worse performance than in the previous solution. However, in contrast to other solutions that use Mobile IP mechanisms (loose or tight coupling solutions), this solution provides faster handover as the path between RNC and GGSN does not exist and the RG can be connected directly to the WLAN IP Gateway or the Internet.

5.3.2. Very tight coupling solutions' characteristics

Very tight coupling schemes ask for major modifications in UMTS functionality and more strenuous implementation efforts. The critical point in the interworking architectures is RNC, while most changes are required in the functionality of RRC. In case WLAN functions as another RNC or Node-B, handover latencies are expected to be minimal. This is verified by simulation work in [5], where very tight integration solutions perform better than any other interworking scheme in terms of handover delay. When an IP Radio Gateway is used, handover delay is increased due to the Mobile IP procedure, but performance is expected to be better than any other Mobile IP coupling solution. In any case, the price to pay for the enhanced handover performance is the high implementation complexity. Very tight solutions are mainly tailored to cellular operators deploying own WLANs, as UMTS infrastructure is mostly re-used. However, in case of the Radio Gateway approach, WISPs could assist the deployment of interworking environments by connecting directly their IP infrastructure. Such deployment would affect significantly the cellular operators. Unlike loose coupling solutions where the IP interconnection does not require major changes, direct interconnection of UMTS network with IP nodes in very tight coupling requires a considerable reorganization of the existing cellular infrastructure.

6. Summarization and qualitative evaluation

Different perspectives are revealed after an overall walk-through. The key difference between the proposed solutions is the point of system of integration. Loose coupling is the dominant and most widely used among the proposed architectures. This is depicted in the unbalanced amount of proposed solutions among the three different types of integration. It manages to interconnect different technologies in an independent way using Mobile IP. It offers an easily deployed network infrastructure with significant attention paid to handover initiation/decision and execution for sophisticated handover management. This can be accomplished due to special purpose entities or with the adoption of the IETF Policy Framework. Moreover, loose coupling offers both mobile-assisted and mobile-controlled handovers. All these make it adequate for service continuity, but not for tight performance guarantees. Furthermore, enhancements in the functionality of interconnected networks are needed. These include the introduction of common AAA entities and the inclusion of Mobile IP functionality in WLANs and cellular networks. These minor modifications needed make it appealing for the WISPs, which see in this type of interworking possible co-operation with cellular operators and new areas for profits.

	Loose Coupling	**Tight Coupling**	**Very tight Coupling**
Types of handover supported	Mobile-assisted, Mobile-controlled	Network-controlled, Mobile-assisted	Network-controlled, Mobile-assisted
Mobile IP functionality	Yes	Yes/No	Yes/No
Consideration of Policies into Handover Management	Independent solutions, IETF Policy Framework	Independent solutions, IETF Policy Framework, Standard UMTS handover management procedures	Standard UMTS handover management procedures
Handover Latency	Low	Low/Medium	Medium/High
Development difficulties	Common AAA entities, Inclusion of Mobile IP functionality in WLAN and cellular networks	Integrated UMTS/WLAN terminals, Modifications in UMTS CN functionality	Integrated UMTS/WLAN terminals, Modifications in UTRAN functionality
Implementation Complexity	Low	Medium	High
Implementation Preference	Wireless Internet Service Providers	Cellular Operators	Cellular Operators

Table 1. Integration types and their characteristics.

On the other hand, the philosophy under tighter types of interworking is closer to the cellular operators' concept. Integration types like tight and very tight coupling offer network-controlled and mobile-assisted handovers. They are characterized in general by seamless service continuity, although the usage of Mobile IP drives practically in non-seamless operation. The handover performance is improved even more when very tight coupling is used. Despite the performance enhancement, the standard UMTS handover management procedures are followed in most cases, not taking into account all available network context information in the network. In any case, the development of tighter coupling solutions asks for combined UMTS/WLAN capability at the terminals as well as significant changes in specific UMTS nodes in order to support transparently WLAN functionality. Therefore, the implementation complexity is considerably increased and this phenomenon is more evident as the interworking point moves from the UMTS CN to the UTRAN. The re-use of part of cellular functionality makes the application of such solutions more appropriate for cellular operators that deploy their own WLANs.

The aforementioned characteristics of each category are summarized in Table 1. From this table, it is clear that no single type of interworking can satisfy both cellular operators and WISPs. It is very interesting to forecast which type of interworking solution will dominate in the market, since the selection of architecture is not based only on performance criteria, but on its cost and its respective profits as well.

Useful deductions can be made if some representative solutions from different levels of integration are compared in terms of policy-based handover characteristics. From Table 2 it is can be observed that load balancing has not been considered from all types of integration solutions. In [40] and [52] such functionality has been incorporated as both are policy-based solutions. In [46] and [48], load balancing capabilities can be added if their architectures are fully exploited. For example, in [46], uplink traffic can bypass the UMTS infrastructure and result in less congestion, while in [48] users can be shifted to either WLAN or GPRS at the SGSN level.

As for the roaming capabilities between different operators, only [40], [46] and [57] offer such functionality. In [57], the least co-operation between the operators is required as it is a Mobile IP-based solution. The cellular operator is considered to be the master of the integrated network in [40], which is not always true in 4G environments. In [46], roaming capabilities are provided under the hypothesis of strong co-operation between the operators, although it is a tight coupling solution. Concerning the QoS capabilities, [40] incorporates mechanisms for QoS management at various levels, while [48] includes scheduling mechanisms in WAF. QoS support is also included in [52], where various profiles are taken into account. Both in [40] and [52] handover decision mechanisms are incorporated (as policy-based schemes). On the other hand, [46] is difficult to be extended with such capability, because of its very dynamic treatment of data traffic. This is more feasible with [48] and [57], where the IETF Policy Framework could be adopted.

	CREDO [40] (Loose Coupling)	VGSN [46] (Tight Coupling at the GGSN level)	WAF [48] (Tight Coupling at the SGSN level)	ERNC [52] (Tight Coupling at the SGSN level)	RNC-IP Gateway [57] (Very Tight Coupling)
Load balancing	Yes	Yes (Bypass part of the UMTS CN)	Yes (Shifting of users to WLAN)	Yes	No
Roaming capabilities between different providers	Yes (Cellular operator is the owner)	Yes (Close co-operation required)	No	No	Yes (Simple Mobile IP solutions)
QoS capability	Yes	No	Yes	Yes	No
Incorporation of advanced handover decision entities	Yes	No (Difficult to be added)	No (IETF Policy Framework can be adopted)	Yes	No (IETF Policy Framework can be adopted)

Table 2. Comparison between some representative solutions from different interworking levels

Finally, in our opinion, as network architectures evolve and different philosophies merge, it is important for integrated networks to take into account the aforementioned policy-based characteristics whatever the interworking point of coupling might be.

7. Related Standardization Efforts

Several efforts from different standardization organizations and bodies have recently been made in the area of 3G/WLAN interworking. 3GPP is the main active contributor in this field. Its standardization work in UMTS/WLAN interworking has been considered by TSG SA WG1 (Services) [58]. As already mentioned, six scenarios have been proposed for the evolvement of integration work. In Release 6, the focus has been only on the first three scenarios. Efforts towards service continuity and seamless service provision are left for Release 7, which started in the second half of 2004.

In addition to 3GPP, IEEE 802.21 Working Group (WG) [59] is also actively contributing to the integration process. A different perspective of 3G/WLAN interworking is emerging from its work. The Group's objective is to enable interoperability between heterogeneous networks supporting different types of 802 networks (802.x) and cellular technologies. Concerning 802.x/Cellular interworking, attention is paid to the innovation of mechanisms for seamless service continuity according to the scenarios specified by 3GPP [4]. More precisely, the scope is to introduce a media independent handover mechanism above L2 in order to optimise handovers and add intelligence in detection and selection of new network attachment points. Group's intention is to cooperate with 3GPP and to complete the first draft of the standard by October 2005 and the final version by March 2006.

On the other hand, IETF SEAMOBY Working Group [60] has quite recently finished its activities since it accomplished its goals. This group worked on specific issues regarding seamless mobility. It focused on context transfer between edge mobility devices in order to allow real-time services to function with minor interruptions. The related draft [61] proposes a Context Transfer Protocol whose objectives are to minimize latency and packet losses, and to avoid re-establishing signaling connections at the new access point. Moreover, the problem of pinpointing the candidate access points for handover was also considered in [62]. According to that work, identifying the IP addresses and the capabilities of these candidates is essential for a seamless IP-layer handover.

8. Conclusion

The necessity to provide telecommunication services at any time, anywhere and with the best possible quality, has created the need to deploy a required set of mechanisms that integrate different systems. Therefore, a plethora of solutions has recently focused on interworking between WLANs and cellular networks in order to identify the most important issues of this area.

Handover has been a critical process in both WLANs and cellular systems. This functionality is more difficult to be performed in an efficient manner when user's connections are handed over from one technology to another. Handover management architectures in WLAN/Cellular networks have been surveyed in this paper, aiming at providing a comprehensive summary of interworking solutions mainly about 802.11-based WLANs and GPRS/UMTS cellular networks. The different frameworks have been described and classified based on the functional point of integration at the UMTS architecture. Three main categories have been identified for the inclusion of the solutions and the respective solutions have been compared within each category. Moreover, a general comparison among all the proposed schemes indicated that it is difficult to combine the characteristics of all the solutions as these often contradict each other. Another important outcome of the study is the trade-off between the complexity of the implementation and the performance of the handover procedure. This fact is critical for sensitive real-time applications.

Policy-based architectures appear to influence future handover management schemes. Along with that, tight integration seems to be the next logical step towards the implementation of seamless handover in integrated WLAN/Cellular network environments. These trends and the intense efforts from individuals and standardization bodies will play a significant role in the evolution towards 4G networks.

9. Acknowledgements

The authors would like to thank the anonymous reviewers for their constructive recommendations and comments that helped us to improve the quality of this paper.

10. References

[1] S. Hui and K. Yeung, “Challenges in the migration to 4G Mobile Systems”, IEEE Communications Magazine, December 2003.

[2] ETSI, “Requirements and Architectures for Interworking between HIPERLAN/2 and 3rd Generation Cellular Systems”, Tech. rep. ETSI TR 101 957, Aug. 2001.

[3] 3GPP TS 23.234 V6.2.0, “3GPP system to Wireless Local Area Network (WLAN) interworking; System description (Release 6)”, September 2004.

[4] 3GPP TR 22.934 V6.2.0, “Feasibility study on 3GPP system to Wireless Local Area Network (WLAN) interworking (Release 6)”, September 2003.

[5] S.-L. Tsao, C.-C. Lin, “Design and evaluation of UMTS/WLAN interworking strategies”, Vehicular Technology Conference, VTC 2002-Fall, 2002.

[6] V. Varma, S. Ramesh, K. D. Wong, M. Barton, G. Hayward and J. Friedhoffer, “Mobility Management in Integrated UMTS/WLAN Networks”, Proceedings of ICC 2003, May 2003.

[7] D. Wong, M. Barton, B. Kim, V. Varma, S. Ramesh, G. Hayward, J. Friedhoffer, “UMTS Signaling over 802.11 Wireless LAN”, VTC Fall 2003.

[8] R. Samarasinghe, V. Friderikos, A.H. Aghvami, “Analysis of Intersystem Handover: UMTS FDD & WLAN”, London Communications Symposium, 8th-9th September 2003.

[9] IST-2002-001858 EVEREST project, http://www.everest-ist.upc.es/

[10] F. Di Cola, P.M.L. Chan, R.E. Sheriff and Y.F. Hu, “Handover and QoS Support in Multi-Segment Broadband Networks”, Proceedings of the 4th European Workshop on Mobile and Personal Satellite Communications, London, UK, 2000.

[11] I.F. Akyildiz, J. McNair, J.S.M. Ho, H. Uzunalioglu, and W. Wang, “Mobility Management in Next Generation Wireless Systems”, IEEE Proceedings Journal, Vol. 87, No. 8, pp. 1347-1385, August 1999.

[12] N.D. Tripathi, J.H. Reed and H.F. VanLandingham, “Handoff in Cellular Systems”, IEEE Personal Communications, December 1998.

[13] 3GPP TS 23.002 V6.5.0, “Technical Specification 3rd Generation Partnership Project; Technical Specification Group Services and Systems Aspects; Network architecture (Release 6)”, June 2004.

[14] H. Kaaranen et. al., “UMTS networks: Architecture, Mobility and Services”, Wiley 2001.

[15] 3GPP TS 25.331 V6.2.0, “Radio Resource Control (RRC); Protocol Specification (Release 6)”, June 2004.

[16] Y.-B. Lin, Y.-R. Haung, Y.-K. Chen, and I. Chlamtac, “Mobility Management: From GPRS to UMTS”, Wireless Communications and Mobile Computing, 1(4): 339-360, 2001.

[17] 3GPP TS 24.008 V6.5.0, “Mobile radio interface Layer 3 specification; Core network protocols; Stage 3 (Release 6)”, June 2004.

[18] C. Perkins, ed., “IP Mobility Support for IPv4”, RFC 3344, Aug. 2002.
[19] D. Johnson, C. Perkins and J. Arkko, “Mobility Support in IPv6”, RFC 3775, June 2004.
[20] P. Reinbold and O. Bonaventure, “IP Micro-Mobility Protocols”, IEEE Communications Surveys, Third Quarter 2003, Volume 5, No. 1.
[21] IEEE Std. 802.11f-2003, “Recommended Practice for Multi-Vendor Access Point Interoperability via an Inter-Access Point Protocol Across Distribution Systems Supporting IEEE 802.11 Operation”, January 2003.
[22] N. Passas, A. Salkintzis, G. Nikolaidis, M. Katsamani, “A New Technique to Expedite RSVP Path Re-establishments in 802.11 Wireless LANs”, Kluwer Wireless Personal Communications Journal, 2005, accepted for publication.
[23] I. F. Akyildiz, J. Xie, and S. Mohanty, “A survey of mobility management in next-generation all-IP-based wireless systems”, IEEE Wireless Communications, vol. 11, no. 4, pp. 16-28, August 2004.
[24] S.C. Lo, G. Lee, W.T. Chen and J.C. Liu, “Architecture for mobility and QoS support in all-IP wireless networks”, IEEE J. on Selected Areas in Communications, vol. 22, no. 4, May 2004, pp. 691-705.
[25] J. McNair and F. Zhu, “Vertical Handoffs in Fourth Generation (4G) Multi-network Environments”, IEEE Wireless Communications Magazine, vol. 11, no. 3, pp. 8-15, June 2004.
[26] K. Pahlavan, P. Krishnamurty, A. Hatami, M. Ylianttila, J. Mäkelä, R. Pichna, J. Vallström, “Handoff in hybrid data networks”, IEEE Personal Communications, April 2000.
[27] A. Duda, C.J. Sreenan, “Challenges for Quality of Service in Next Generation Mobile Networks”, Proceedings of Information Technology & Telecommunications Conference (IT&T), October 2003.
[28] R. Yavatkar, D. Pendarakis, and R. Guerin, “A Framework for Policy-Based Admission Control,” RFC 2753, Jan. 2000.
[29] B. Moore et al., “Policy Core Information Model — Version 1 Specifications,” RFC 3060, Feb. 2001.
[30] D. Durham et al, “The COPS (Common Open Policy Service) Protocol”, RFC 2748, January 2000.
[31] S. Balasubramaniam, J. Indulska, “Vertical Handover Supporting Pervasive Computing in Future Wireless Networks”, Computer Communication Journal, Special Issue on 4G/Future Wireless networks. Vol 27/8, pp.708-719, 2003.
[32] Q. Wei, K. Farkas, P. Mendes, C. Prehofer B. Plattner, N. Nafisi, “Context-aware Handover Based on Active Network Technology”, Fifth Annual International Working Conference on Active Networks (IWAN 2003), Kyoto, Japan, Lecture Notes in Computer Science, Springer Verlag, Berlin Heidelberg New York, 10-12 December, 2003.
[33] A. Kassler, A. Schorr, L. Chen, M. Talanda, C. Meyer, M. Helbing, “Multimedia Communication in Policy based Heterogeneous Wireless Networks”, Proceedings of the VTC-Spring 2004, Milan, Italy, May 2004.
[34] M-Zones project, http://www.m-zones.org
[35] K. Murray, R. Mathur, and D. Pesch, “Intelligent access and mobility management in heterogeneous wireless networks using policy”, ACM 1st International Workshop on Information and Communication technologies, pages 181-186, 2003.
[36] K. Murray, R. Mathur, D. Pesch, “Network Access and Handover Control in Heterogeneous Wireless Networks for Smart Space Environments”, First International Workshop on Management of Ubiquitous Communications and Services, MUCS 2003, Waterford, Ireland, December 11, 2003.
[37] H. Chaouchi, G. Pujolle, I. Armuelles, M. Siebert, C. Bader, I. Ganchev, M. O’Droma, N. Houssos, “Policy Based Networking in the Integration Effort of 4G Networks and Services”, IEEE Semiannual Vehicular Technology Conference, VTC spring, May 17-19, 2004, Milan, Italy.
[38] H. Chaouchi, G. Pujolle, “Pre Handover Signaling for QoS aware mobility control”, Anwire Workshop, Athens, May 2004.
[39] W. Zhuang, et al., “Policy-Based QoS Management Architecture in an Integrated UMTS and WLAN Environment”, IEEE Communications Magazine, November 2003.
[40] IST 2001-33093 CREDO project, http://www.credo.nal.motlabs.com/
[41] M. Buddhikot, G. Chandranmenon, S. Han, Y.-W. Lee, S. Miller, L. Salgarelli, “Integration of 802.11 and Third-Generation Wireless Data Networks”, Proceedings of the 22nd Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE INFOCOM 2003), San Francisco, CA, USA, March 30 - April 3, 2003.

[42] R. Chakravorty, P. Vidales, K. Subramanian, I. Pratt, J. Crowcroft, “Performance Issues with Vertical Handovers: Experiences from GPRS Cellular and WLAN hot-spots Integration”, Proceedings of the IEEE Pervasive Communications and Computing Conference (IEEE PerCom 2004), March 2004.
[43] “TIA/EIA/IS-835B cdma2000 Wireless IP Network Standard”, 3GPP2, 2000.
[44] J. B. Postel, “Transmission Control Protocol”, RFC 793, Sept. 1981.
[45] Q. Zhang, C. Guo, Z. Guo, and W. Zhu, “Efficient mobility management for vertical handoff between WWAN and WLAN”, IEEE Communications Magazine, vol. 41, no. 11, November 2003.
[46] S.-L. Tsao, C.-C. Lin, “VGSN: A Gateway Approach to Interconnect UMTS/WLAN Networks”, IEEE Personal, Indoor, Mobile Radio Communications, 2002.
[47] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, E. Schooler, “SIP: Session Initiation Protocol”, IETF RFC 3261, June 2002.
[48] A. Salkintzis, C. Fors, R. Pazhyannur, “WLAN – GPRS Integration for next-generation mobile data networks”, IEEE Wireless Communications, October 2002.
[49] A. K. Salkintzis, “Interworking techniques and architectures for WLAN/3G integration toward 4G mobile data networks”, IEEE Wireless Communications, vol. 11, no. 3, Jun 2004, pp. 50 – 61.
[50] T. L. Mann, “A Network System Level Simulator for Investigating the Interworking of Wireless LAN and 3G Mobile Systems”, MSc Thesis, Virginia Polytechnic Institute and State University, April 2003.
[51] M. Jaseemuddin, “An Architecture for Integrating UMTS and 802.11 WLAN Networks”, Proceedings of IEEE Symposium on Computers and Communications (ISCC 2003), Antalya, Turkey, pp. 716-723, 2003.
[52] A. Kaloxylos, G. Lampropoulos, N. Passas, “A Flexible Mechanism for Service Continuity in 4G Environments”, Elsevier Computer Communications Journal, special issue on “End-to-end QoS Provision Advances”, accepted for publication.
[53] 3GPP TS 23.207 V6.4.0, “End-to-end Quality of Service (QoS) concept and architecture (Release 6), September 2004.
[54] 3GPP TS 25.323 V6.0.0, “Packet Data Convergence Protocol (PDCP) Specification (Release 6)”, December 2003.
[55] 3GPP TS 22.105 V6.3.0, “Service aspects; Services and service capabilities (Release 6), March 2005.
[56] N. Vulic, I. Niemegeers and S. Heemstra de Groot, “Architectural options for the WLAN integration at the UMTS radio access level”, VTC Spring, May 17-19, 2004, Milan, Italy.
[57] M. Wetterwald et. al., “An original adaptation of the UMTS protocols for a Direct Interconnection with IPv6”, 13th IST Mobile and Wireless Communications Summit 2004, 27-30 June 2004, Lyon, France.
[58] 3GPP TSG SA WG1 (Services), http://www.3gpp.org/TB/SA/SA1/SA1.html
[59] IEEE 802.21 WG, http://www.ieee802.org/21/
[60] IETF SEAMOBY Working Group, http://www.ietf.org/html.charters/OLD/seamoby-charter.html
[61] J. Loughney, M. Nakhjiri, C. Perkins, R. Koodli, “Context Transfer Protocol”, Internet Draft, August 2004.
[62] M. Liebsch, A. Singh, H. Chaskar, D. Funato, E. Shim, “Candidate Access Router Discovery”, Internet Draft, September 2004.