

# Managing Pi-resources in 4G Wireless Systems: the Opportunistic Way

Pedro Sobral<sup>1</sup>, Luis Bernardo<sup>2</sup>, and Paulo Pinto<sup>2</sup>

<sup>1</sup>Faculdade de Ciência e Tecnologia, Universidade Fernando Pessoa, Porto, Portugal,

<sup>2</sup>Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Caparica, Portugal,  
pmsobral@ufp.pt    {lflb,pfp}@uninova.pt

## Abstract

*Integration of different radio access networks will become a reality in a near future. Our interworking architecture assumes a ubiquitous primary network (the cellular one) and secondary networks used on an availability basis. Secondary networks form islands of coverage providing partial and intermittent resources with greater Quality of Service as the user moves along. This paper describes an application architecture that manages these resources in an opportunistic way. It makes use of the Open Service Access (OSA) to control the core network and the OSA application servers to host the new applications that can take advantage of the pi-resources. These applications create a context for secondary networks and the context can survive losses of coverage.*

## 1. Introduction

The wireless architecture known as the 4<sup>th</sup> generation (4G) will integrate several radio access networks (RAN) with the cellular core to produce a feeling to the users of a common environment towards the wireless technology. However, the different RANs will have different characteristics regarding its resources (bandwidth, ownership, price, etc.), and the ability of the users, or the applications, to manage the appropriate RAN will be one of the issues for the success of the 4G networks. The definition of the exact interworking framework will also influence the possibilities users and applications have by creating an application model. The typical approach in the literature assume that every RAN is identical and autonomous and the core can chose one using a concept of always best connected (ABC) [1]. In a previous work [2] we have developed an interworking framework where the GPRS (General Packet Radio Service) network is ubiquitous and forms the primary RAN. All other RANs (802.11, HyperLan,

etc.) are secondary networks, not ubiquitous, and used on an availability basis (figure 1 shows two distinct secondary RANs). The main idea is that applications can be launched to use one specific secondary RAN: when the user is under coverage it uses the chosen RAN, otherwise the user can use the GPRS (either to continue the application, or simply to synchronize it waiting for a subsequent appearance in a coverage area). Therefore, secondary RANs can be seen as partial and intermittent (Pi)-resources to be managed at application level.

The user equipment (UE) is connected to each RAN using a RAN-specific (IP) address— so, the UE is multiple connected. Even in dark areas of the RAN the address is valid and data is exchanged but the application is aware that the traffic is conveyed via the GPRS.

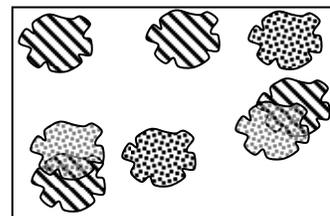


Figure 1. WLANs form islands

This paper describes a framework to manage the use of the secondary RANs as pi-resources and to provide an environment for the easy construction of user applications for the 4G systems. It begins with a characterization of the scope of the problem followed by a description of the supporting interworking architecture. Section 3 describes the protocol stack, section 4 presents the integration with OSA and section 5 addresses the application model. Before the conclusions some considerations about related work are done.

## 2. Scope of the problem and Supporting Architecture

The system must provide a seamless continuation of services across RANs. This implies short handover times; integration of authentication, authorization and

accounting (AAA) procedures to allow for one-time sign in; and mobility management. The fact that the UE has several IP addresses (at least one per RAN) introduces a problem different from the multi-homing [3] or the multi-connection [4] problems. These latter problems are mainly related with redundancy of networks connections and the choice of network paths are mainly done by the network without any intervention of the applications. With our problem the application must be in control of the usage of the connections. It can even happen that the application can decide not to exchange information if the secondary RAN is not available. Therefore, the Always Best Connected can be not connected at all via secondary RAN for a while. Finally, it is also a different problem from disconnected computing and state coherence establishment addressed by systems such as CODA [5]. In our system there is always an active connection (GPRS) to maintain state between the parts and applications just seek the opportunity to use the secondary RAN for heavy duty transport.

In our architecture the user has a multi-radio terminal and is always connected to the GPRS network. Each WLAN technology forms a Hotspot Network (HN). The topology of a Hotspot Network can be compared to islands of hotspots. They are controlled by a new component in the 3G core, named HNAC (Hotspot Network Area Controller). Figure 2 shows the overall architecture and a brief description is given on the sequel. A thorough description of this mode of integrating WLANs with the cellular network, called core-level approach, can be found in [6]. Figure 2 shows the data paths for a terminal connected both to GPRS and to one WLAN. The main characteristics are:

Applications can use up to three routable IP addresses: IP1 (using 3G); IP2 (using the WLAN and

the new core part); and IP3 (using local facilities of the WLAN). Each address uses a certain network path by default. When IP2 is used, WLAN traffic goes via the 3G core network but uses new links (thicker lines) and not the ones existing today. I.e, there is no overload of the current 3G core. The interconnection between the SGSN (Serving GPRS Supporting Node) and HNAC (line A in the figure) makes it possible to send packets through another RAN than the default one transparently to the application (for instance, traffic coming to IP2 address can be sent through GPRS).

A UE is known at the core by its IMSI (if the core wants to use UTRAN), or by the address IPa (if the WLAN is used). The IM (island manager) is a router with functionality of AAA proxy. The GHSN (Gateway Hotspot Network) has an identical functionality to that of GGSN (Gateway GPRS Support Node) (Internet context creation).

HN contexts remain valid during disconnection periods from secondary network. Finally, this solution has all the advantages of the tightly coupled approach [7] for radio network interconnection without the drawbacks of exposing internal 3G interfaces. The upper part of figure 2 (the open service access part) is the subject of this paper.

### 3. HN Protocols

In order to take advantage of this architecture, applications must receive information about the status of the radio connections to the UE (the WLAN link is on or off); localization of the users (if there are HN islands in the vicinity); user movement pattern (he is moving so fast that the WLAN contact periods will be very short); etc. This information is obtained from the HNAC and from the core network. The protocol stack showed in figure 3 refers to the HNAC way of

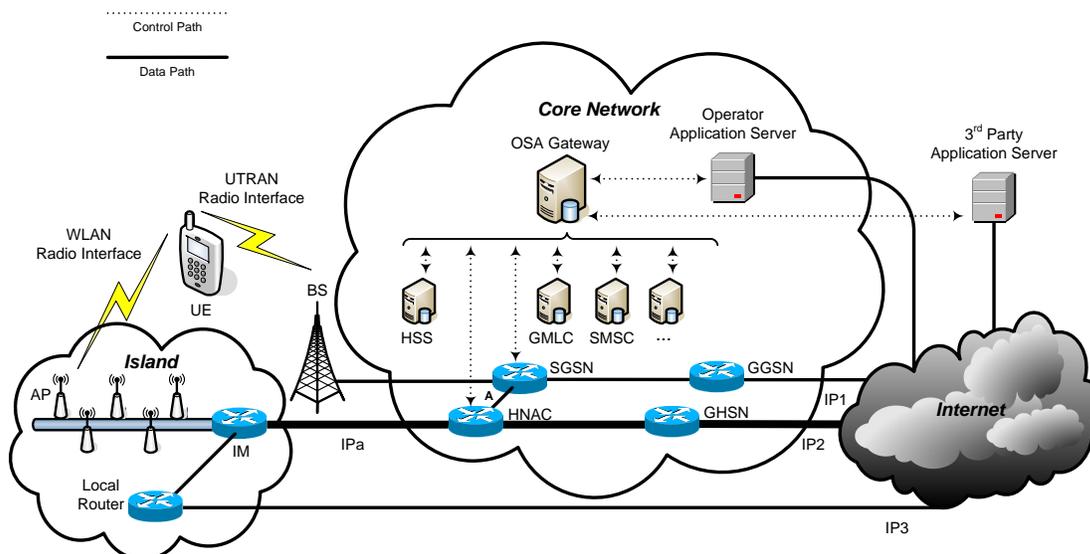


Figure 2. Overall architecture

obtaining the information and the exchange of data packets. A thorough description of the protocol stack can be found in [6]. Basically, the HN Attach Management contains the WLAN attachment procedures including the AAA procedures, and provides this information upwards to the applications. Each time an UE enters into an area covered by a WLAN Access Point it performs an attachment, or reattachment, procedure. When the user walks away there is, most of the times, simply loss of coverage and not a detachment procedure. The Delivery Service is used to overcome these attach/detach situations. The Delivery Service is a combined layer between the HNAC and the SGSN at the core side and the UE at the other side. It implements a confirmed data packet delivery mechanism over the WLAN just below the IP layer. The purpose is to manage the availability of the WLAN radio connection and to transfer a packet using the other radio interface if necessary. It uses probe packets during silent periods in the data part just to test the WLAN link. Whenever there is a change in the link condition a signal is sent upwards to the HN Attach Management for processing.

With such a setting any transport layer above the IP can work normally. Packets are sent via the preferred radio link depending on the IP address unless otherwise stated (UTRAN if IP1 is used, and WLAN if IP2 is used). Packets are always delivered to the UE (using the UTRAN if necessary) because the Delivery Service recovers from any failure on the WLAN link and sends the packet via UTRAN.

Different islands may be managed by different HNACs. So, HN reattachments may trigger HNAC handovers. The process of changing HNACs involves several procedures: the IP2 address does not change; the user profile must be retrieved and updated at the HSS (Home Subscriber Server); the new HNAC must

establish the same contact the old HNAC had with the SGSN and the GHSN; all the service level information must be transferred to the new HNAC; and the context with the old HNAC must be held and resumed with the new HNAC. This handover can also happen during WLAN dark areas (there is a trigger condition from the information of the SGSN).

#### 4. Integration with Open Service Access

Applications run on OSA – application servers either belonging to the operator or not (see fig. 2) [8]. To accomplish their purpose they have access to core information regarding specific UEs and subscribe the reception of events triggered by changes in their execution environment. The standard and protected way to do this is using the Service Capability Functions (SCFs) implemented on a component called OSA Gateway [9]. Figure 4 shows some of the SCFs: the Framework SCF authenticates the applications and provides access to network resources, load management and service discovery; User Location SCF provides geographical location of a mobile subscriber (either by replying to a single request, or by sending periodic reports, or still by sending reports after some triggering condition); the User Status SCF reports the status of a user terminal (also single report, periodic report or triggered report). To enable integration with the proposed architecture this SCF must be extended in order to provide applications with UE WLAN status. It can also include some algorithms to raise the semantic level of the information to facilitate the decisions at the application layer. Some examples are:

- An algorithm that can infer the speed of the UE from the cell handover rate or from the type of cell the UE is in (pico, micro, or macro-cell).

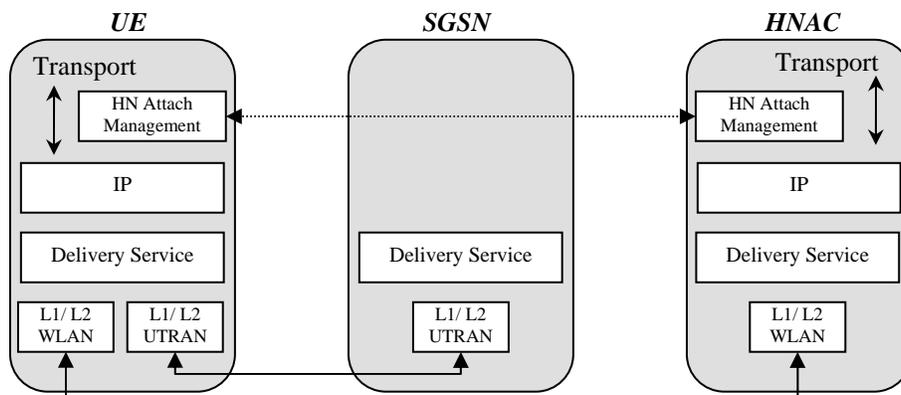
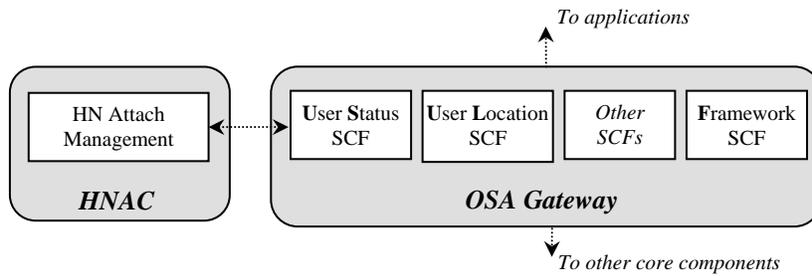


Figure 3. Protocol Stack for the HN Operation



**Figure 4.** Providing HN status and network control through OSA API

- An algorithm that can categorize the UE mobility pattern in fast moving, slow moving or stationary to help applications decide whether to use the WLAN or the GPRS.

- An algorithm that using the information about (a) the WLAN link status provided by the User Status SCF, (b) the information about the CellID and Routing Area provided by the SGSN, and (c) the information about availability of islands in the surrounding areas provided by some kind of database, implements a power management service to place a HN session in one of three states: standby – meaning that the HN connection is OFF and no island is known to exist in the vicinity of the UE (the HN radio interface can be turned off); searching – meaning that the HN connection is OFF but there are islands near the UE’s current location; or connected – the HN connection is ON.

OSA defines some additional SCFs as the Multiparty Call Control SCF, for instance, which allows the setting up of multiparty connections using SIP (Session Initiation Protocol), but they are not central to our application model.

## 5. Application Model

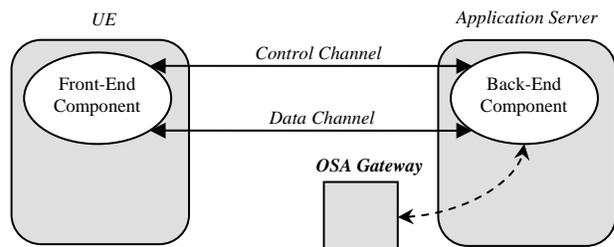
Legacy applications using IP2 address run unaltered. If the UE is inside an island the WLAN RAN is used, otherwise packets are delivered via the UTRAN. Therefore, applications do not experience network disconnections.

Opportunistic applications are different because they have access to the OSA Gateway and can use control information to drive their behavior. When an UE leaves an island the application can receive a signal and halt the transfer of data packets. Note that while the terminal is in dark areas the Transport connection of IP2 is still on and keep alive packets are transported via UTRAN.

The general structure of the application model for opportunistic applications is pictured in figure 5. An

application can be seen as split in two: a front-end component (running on the UE) and a back-end component (running on the application server) connected by the wireless links. Communication is always possible between them due to the UTRAN. However, they can decide not to transfer data in dark areas waiting for the appearance in an island. An easy way to structure these applications is to define two (virtual) channels between the components: a control channel (with modest bandwidth requirements) and an ordinary data channel. Although they both can be used all the time it is better to think on the control channel as a mean to enforce state coherence (checkpoints on data transferred, expectations on when data will be exchanged, etc.) and used regardless of the available network connection; and the data channel as a channel to be used when both sides agree. Two final observations on this model are the following:

- In order to take advantage of pi-resources the purpose of the application cannot be a real-time or synchronous task. Otherwise, the UTRAN will have to be used all the time and the problem falls into a Quality of Service (QoS) problem and not so much a pi-resource problem. The typical type of applications that can take advantage of these pi-resources are non-real time bulk information transfer (be it data, video, voice, etc.), or extra facilities to a baseline application (for instance a video stream on a voice call).



**Figure 5.** Opportunistic Application Model

- A sub-layer responsible to manage the appearances of the UE in islands (pi-resources) or to decide the immediate UTRAN use because certain conditions are met can have common features towards different applications. Behaviors such as “I want the information in 30 minutes”, or “I do not care about the delay but I want the information via the cheapest path” can be implemented by a general purpose model with input parameters.

### 5.1. Opportunistic Web Proxy

The first application example refers to asynchronous transfer of data to/from the UE. A typical usage scenario may be the download of a multimedia catalogue from the company website by a salesman. Using his UE’s web browser he finds the link and orders the downloading of the file. As the size of the catalogue is huge and assuming that the salesman has configured his local proxy settings to download objects larger than 1Mbyte using only the WLAN connection, the download will not start immediately since only 3G network is available there. However, using the 3G connection, the local proxy instructs the back-end component, running on the OSA-AS, to save (possibly part of) the catalogue in cache. When the salesman stops in the next traffic light, his UE senses a WLAN network with a roaming agreement. A reattachment procedure is initiated and the WLAN interface is enabled. Part of the catalogue is downloaded to the mobile station. Possibly, during his trip to the next client, the whole catalogue will become available on the UE. If not, the salesman can force the download of the missing parts using the 3G network.

The scenario presented above has the following components (see fig. 6): the front-end component acts as local proxy server for the UE’s WEB browser (the user configures his browser to use a proxy whose

address is the loop back interface and designated port number) communicating using the HTTP protocol. The back-end component interacts with the OSA Gateway receiving notifications on subscribed events related to the UE, and acting according to a certain behavior. The user is able to influence the application behavior by customizing a profile. He can define what RANs will be used to download objects based on their size. For example, he may define a maximum object size to be synchronously downloaded using the 3G connection in places without WLAN coverage. Larger objects will wait until a WLAN hotspot becomes available. The (virtual) control channel is used to process data requests and to manage overall application status (urgent messages). The (virtual) data channel is used to transfer WEB objects. Whenever a download request arrives at the front-end it is immediately forwarded to the back-end independently of the available network. The back-end performs the download of the requested objects from their web servers. Depending on the current execution environment conditions (user profile, available WLAN connections, hotspot hit probability, location, speed) the back-end component decides what to do with the data and the UE. For example, if the user is found to be “fast moving”, it might be better to wait for a while and instruct the UE not to look for hotspots. When the proper conditions are met, objects are delivered to the front-end component and become available to the user. Upload operations follow a similar pattern.

### 5.2. WLAN videophone

The purpose of the “WLAN videophone” application is to enable the establishment of multimedia calls (video + voice) between users attached to the HN with a fall back to voice only calls when one party loses WLAN connectivity. The

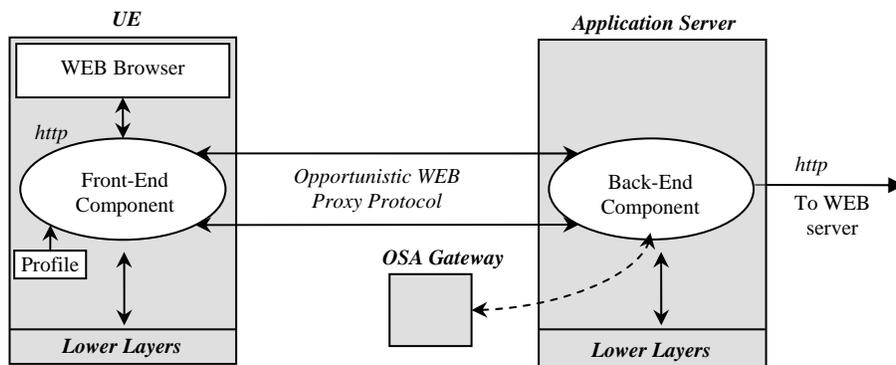


Figure 6. Opportunistic WEB Proxy

establishment of the call is assisted by the back-end component. This component is responsible for the establishment of the voice and the video paths and supervises the call. If a user moves out of the HN island the video path is torn down. The back-end component knows the information about the UE by subscribing the User Status SCF (running in the OSA gateway). When the user leaves the island the packets (both video and audio) are still transmitted using the UTRAN. However, the Delivery Service signals upwards and the back-end component receives the event from the SCF. At that moment the video is torn down (the back-end uses the control channel to the front-end to stop the video) but the voice is still transmitted via UTRAN. When the user walks in again the video call can be reactivated. Voice calls made in the IMS world (Voice over IP) can be upgraded with video if the user steps into a HN island if the proper signaling for interworking is defined.

## 6. Related Work

The increasing importance of mobile computing, where the available computing resources may vary widely during application execution, has led to many developments on application adaptation [10]. One of the most widely known examples of software adaptation is the process used in TCP to perform window management in response to apparent network congestion. However this is just an example of “parameter adaptation”. “Compositional adaptation” is a different strategy in which applications modify their algorithmic behavior to adjust to their current execution environment. There are mainly three enabling technologies for compositional adaptation. The “separation of concerns” that allows the separate development of an application functional behavior and the code for “crosscutting concerns”, such as QoS, energy consumption, etc. A second approach is the “computational reflection” that refers to an application’s ability to reason about, and possibly alter, its own behavior [11]. A third mechanism is the “component-based design”. Software components may be developed by different parties and they collaborate through defined interfaces to implement the application behavior. In static composition the developer can combine several components at compile time to produce an application. In dynamic composition the developer can add, remove, or reconfigure components within an application at runtime. Yet another approach to implement adaptation can be played by the middleware. An adaptive middleware is a set of context aware services that help

applications adjust to changes in the execution environment. Examples of frameworks implementing these adaptation techniques can be found in [12]. Regardless of the approach, the extent of application adaptation is tightly connected with the sensor information provided by the environment. Opportunistic applications closely interact with the OSA gateway and subscribe the reception of the events. For the adaptation itself, depending on the application purpose, one (or more) of the models presented above can be used.

CME [13] “Connectivity Management Engine” is an example of a middleware architecture for network aware adaptive applications. CME performs network channel management and adaptation using a connection controller running on both client and server sides of the application. The connection controllers also interact with a context server that maintains static network information for each interface (type, bandwidth, cell size, etc.), user profiles related to network management (default interface, interface selection policies, etc.) and end host information (energy status, velocity, cell dwell time, etc.). However, nothing is said about the concrete way of obtaining such end host information. In CME applications running in collaborative operation mode manage their interactions based on end-to-end parameters. We optimize the interactions using the secondary network by placing the back-end component near the core and preparing interactions with the front-end component in advance using the control channel. This option enables the opportunistic model in our system. In CME it seems very difficult to take advantage from the highly intermittent connection between a moving UE and the secondary network islands.

## 7. Conclusions and further work

Network operators are seeking for wireless network integration in order to allow interoperation between scattered hotspots of several radio technologies (e.g. WLAN) and the 3G network. 3GPP has defined six scenarios with increased integration requirements [14]. However, implementation details for the most demanding scenarios (service integration and continuity) are still an open issue. We have presented a framework to seamlessly integrate different RANs, providing applications with a network environment where disconnection periods due to vertical handovers do not exist. On top of this network abstraction we have defined the opportunistic application model. Opportunistic applications closely interact with the

core network (via an improved OSA framework) in order to gather end user context information and adapt their operation accordingly. We have presented examples that show the operation of this application model in the asynchronous transfer of large bulks of data and in synchronous adaptation to available network resources in multimedia calls. Opportunistic application model is effective for providing mobile users with context aware services in an integrated wireless environment.

As further work subjects there is the convergence of the fixed and wireless networks. Hotspot operators from the fixed network world can also take advantage of a system like this provided the necessary roaming agreements are done and an architecture focused on their views is defined. There is also the development of the most used behaviors to take advantage of the opportunistic concept.

## 8. References

- [1] Gustafsson, E., Jonsson, A, “*Always Best Connected*”, IEEE Wireless Comm, vol. 10, pp. 49-55, Feb 2003
- [2] Pinto, P., Bernardo, L., Sobral, P., “*UMTS-WLAN Service Integration at Core Network Level*”, LNCS, Springer-Verlag Heidelberg, Volume 3262/2004, October 2004, DOI: 10.1007/b101785.
- [3] Abley, J., et al, “*Goals for IPv6 Site-Multihoming Architectures*”, rfc3582, August 2003.
- [4] Caro, A., et al, “*SCTP: A Proposed Standard for Robust Internet Data Transport*”, *Computer*, pp. 56-63, Nov 2003.
- [5] Mummert, L., et al “*Exploiting Weak Connectivity for Mobile File Access*”, 15th ACM Symp on Operating Systems Principles, Copper Mountain Resort, pp. 307-319, Dec. 95.
- [6] Pinto, P., Bernardo, L., Sobral, P., “*Seamless Continuity of PS-Services in WLAN/3G Interworking*”, Technical Report TR-2004-1, Tele-DEE-UNL, <http://tele1.dee.fct.unl.pt/tech-reports/TR-2004-01.pdf>
- [7] Salkintzis, A., et al., “*WLAN-GPRS Integration for Next-Generation Mobile Data Networks*”, IEEE Wireless Comm., vol. 9, pp. 112-124, Oct 2002
- [8] 3GPP, “3GPP, TS Group Core Network; Open Service Access (OSA); Application Programming Interface (API); Part 1: Overview (Release 6)”, TR 29198-01-631, Dec 04.
- [9] Musa R. Unmehopa, et. al., “*The support of Mobile Internet Applications in UMTS Networks through the Open Service Access*”, Bell Labs Technical Journal 6(2), pp. 47-64, 2002.
- [10] McKinley, K., et al., “*Composing Adaptive Software*”, IEEE Computer Magazine, July 2004.

[11] Grace, P., et. al., “*Middleware Awareness in Mobile Computing*”, Proc. of the 23<sup>rd</sup> Int. Conf. on Distributed Computing Systems Workshops (ICDCSW’03), 2003.

[12] Aksit, M., et al., “*Dynamic, Adaptive and Reconfigurable Systems Overview and Prospective Vision*”, Proc. 23<sup>rd</sup> Int. Conf. on Distributed Computing Systems Workshops (ICDCSW’03), 03.

[13] Sun, J., et. al., “*CME: A Middleware Architecture for Network-Aware Adaptive Applications*”, Proc. 14<sup>th</sup> IEEE 2003 Int. Symp. on Personal, Indoor and Mobile Radio Communication, 2003.

[14] 3GPP, “*Group Services and System Aspects; Feasibility study on 3GPP system to WLAN interworking; (Release 6)*”, TS 22.934.v.6.1.0, Dec. 2002