Managing Resources in Heterogeneous Wireless Networking Environments

S. Kyriazakos and G. Karetsos¹, Telecommunications Laboratory, National Technical University of Athens, Greece F. Di Giandomenico, ISTI/CNR, Italy F. Casadeval, UPC, Spain M. Jubera, TELEFONICA I&D, Spain I. Mura, MOTOROLA, Italy N. Tselikas, Ericsson Hellas, Greece K. Vlahodimitropoulos, COSMOTE S.A, Greece

Abstract

The always-increasing number of the subscribers of the mobile networks and the evolving demand for advanced services pose new requirements for the underlying networking infrastructure. On the other hand radio resources are scarce and already overexploited and the existing architectures lack the flexibility that would allow easy and efficient introduction of the emerging services. These problems led us to investigate the possibility for real-time Radio Resource Management in heterogeneous wireless networking environments. In this paper we present a methodology and an approach for designing a hierarchical system that is augmenting the functionality of wireless network architectures by enforcing smooth co-operation and is capable to react when resource shortcomings appear.

1 Introduction

The main goal in the wireless telecommunication world can be briefly summarized as: "communication anywhere, anytime, any-media and principally at high-data rates". However, this goal is in conflict with the co-existence of a plethora of different running and emerging wireless systems covering almost the whole world, each one following its own architecture and based on its particular bedrocks.

The development of wireless systems evolved in an unimaginable way during the last two decades. In cellular wireless systems the so-called First Generation (1G) is already antiquated. The dominant Generations, which are nowadays in the limelight, are 2G, 2.5G and 3G. In Europe their representatives are GSM (Global System for Mobile Communication), GPRS (General Packet Radio Service) and UMTS (Universal Mobile Telecommunications System) respectively and belong in the terrestrial wide area cellular systems. The circuit-switched GSM provides very slow data rates (9.6 - 14.4 kbps) to satisfy the burst applications, even after the appliance of High Speed Circuit Switched Data (HSCSD), it doesn't overcome the limit of 40 kbps. Packet-switched networks, based on the access network of GSM with actual changes only in the core network (GPRS), appeared with the promise of higher bit rates (theoretically 172 kbps), but in practice the maximum bit rate achieved is about 45 kbps.

The UMTS access network follows a different approach, in comparison to GSM and GPRS, making the achievement of higher data rates more feasible. UMTS offers data rates up to 384 kbps, even if in theory the 2 Mbps transfer rate is possible. Nevertheless, the actual performance of UMTS has still to be verified during real operation conditions with heavy network loads.

On the other hand, there are various wireless systems such as global area systems (e.g. satellite systems), wireless personal area networks (WPANs), which are formed by wireless communications between devices using technologies such as Bluetooth or IEEE 802.15 [1] and wireless networks used for limited local area coverage called Wireless Local Area Networks (WLANs) (e.g. IEEE 802.11a, IEEE 802.11b or HIPERLAN). These kind of networks provide incomparably high data rates. For example the 802.11b WLAN provides throughput up to 5 Mbps, while the data rates in 802.11a can be up to over 25 Mbps, with the perspective to reach in the future the inconceivable limit of 155 Mbps.

The co-existence of these technologies results in a heterogeneous set of wireless communications systems. Its active components are based on different theoretical backgrounds and are optimized for

¹ Corresponding author: e-mail <u>karetsos@cs.ntua.gr</u>, Tel. +302107721511, Fax. +302107722534

different ranges, exposing a great challenge for potential co-operation of all existing and emerging systems in a complementary way, in the concept of a 3GB (beyond 3rd Generation) system [2]. The objective of 3GB systems is not to abandon the existing systems, but exploit efficiently their functionality by combining them with the emerging ones. Such a hybrid 3GB system is examined in the CAUTION++ framework containing the interworking of four different access networks (GSM, GPRS, UMTS and WLAN) under a unified and hierarchical resource management model [3]. At this point it has to be mentioned that a similar system has been already validated for 2nd Generation networks (GSM/GPRS), showing network services increase of up to 40%, while in this paper an evolution of this system for heterogeneous environments is presented.

Enhanced IP networking technologies are used to integrate current and future systems to a unified super-network, enabling a truly seamless mobile Internet, beyond the simple wireless access to the Internet, thus extending the scope of a monolithic system. The Internet Protocol version 6 (IPv6) does not offer only virtually unlimited address space, but also constitutes the technical foundation for evolutionary networking, offering also interoperability and interconnectivity with respect to security, mobility and Quality of Service (QoS) [4].

This paper is organized in 5 sections. Section 1 is the introduction where the motivations for this work are discussed. In section 2, a heterogeneous environment is presented, as well as its features from the view of the service provider, the network operator and the end-user. In section 3, the requirements of such a Radio Resource Management (RRM) system are listed. These are the starting-points for the system architecture that is described in the next section. In section 4, the CAUTION++ approach for resource management in heterogeneous wireless networking environments is given. Finally, we sum up with the conclusions in section 5.

2 Features of heterogeneous mobile networks

Before describing the RRM system, a clear definition of a heterogeneous environment is necessary. There are currently many interpretations of a heterogeneous wireless environment; therefore, we try to focus on a realistic situation, where a number of service providers utilize different access technologies for offering their services to the subscribers.

The current trends in Mobile Networking aim at a change of the ownership of the mobile customer, from the network operators hands' to the content and applications providers' ones. Due to the constraints of 3G in bandwidth and mobility, it is necessary to rely on several mobile standards, in order to choose the system, which best suits to the application employed by the customer. The driver of this process is the provision of the best possible QoS to the user, wherever they are, in a cost-efficient way, thus making an attractive service offering. We are going to see this future scenario, with all integrated systems, from three points of view: service providers, network operators and end-users.

Service providers will play a vital role in this future scenario and are considered to have a certain degree of independence from network operators. Several service providers are expected to co-exist, which will act as intermediaries between the network operator and the final customer. They will provide the users their requested service according to the needed QoS levels.

It is necessary to consider existing networks (GPRS, UMTS, WLAN) as part of a wider infrastructure, where network operators will be able to avoid rejecting traffic by redirecting the request in order to be served by an affiliated network operator. Additionally, the operator shall provide intersystem handovers that will have to be able to avoid call losses and maintain QoS levels when switching. This seamless system will increase the network complexity, but will lead to cut costs and to improve QoS levels. Access networks are expected to be connected to the Internet using probably an IP core network. Quality of service requirements from the network layer can be directly applied along the end-to-end path, placing IP routers at the very edge of the terrestrial network [5].

The future subscribers are likely to use a device, not necessarily a cellular phone, which will be able to interface with multiple radio standards. This device will be used to offer a multitude of services, including voice, messaging, Internet and corporate access, entertainment, etc. The most prominent services that are likely to become the "killer applications" in the future, however, are services such as net games, video streaming, e-mail (including SMS, EMS), internet or e-commerce on the move These

applications are expected to provide significant revenues for the operators but on the other hand, they will require a considerable amount of bandwidth delivered to the subscriber.

3 User and system requirements in heterogeneous networking environments

The starting point for attaining the objectives of resource management as proposed in this paper is to state the user and system requirements. This encompasses addressing individually all the components included in the foreseen architecture for resource management, as well as their interconnections and the interfaces between the architecture's components and the network elements. As it is typical in system engineering disciplines, requirements pertain to two main categories: functional requirements and non-functional ones. Functional requirements are associated with specific functions, tasks or behaviors the system must support, while non-functional requirements are constraints on various attributes of these functions or tasks. Herein our focus is on the functional requirements for our system mainly refer to those that are absolutely necessary in order to guarantee efficient Radio Resource Management (RRM) in a heterogeneous networking environment and they can be summarized as follows:

Real time network monitoring: The system should be able to monitor each network separately and respond to any traffic overload situation. For that purpose it is needed to define a large number of Key Performance Indicators (KPIs) that characterize the stable (or not) network behavior. Some of these KPIs are: *Blocking Rate, Drop Call Rate, Reliability, Mean/Average Throughput, Handover Failure*, etc. Whenever one of these KPIs reaches a predefined threshold, the monitoring system should generate an alarm to a higher level of the RRM hierarchy, in order for the congestion problem to be faced.

Efficient Local RRM: The Local RRM is the functionality of the platform that is responsible for executing those resource management techniques that can solve the traffic overload problem within the same Radio Access Network (RAN). Depending on the wireless technology (GSM, GPRS, UMTS, WLAN) some of them are: *Cell Breathing, Admission control, Handover Adjustment, Dynamic Cell Reconfiguration, Power Control, etc.*

Efficient Global RRM: The Global RRM differentiates from the local one, since it exploits the total available heterogeneous infrastructure to provide efficient RRM. Whenever the local RRM entity is not able to reconfigure the system in such a way to solve the problem, it escalates to the highest hierarchy level that is responsible for the Joint RRM. This involves vertical handover, as well as roaming of users from one service provider to another, in the case that two service providers have such a service level agreement (SLA).

Information retrieval: Obviously the system, as it is envisaged up to now, should contain components that possess useful information, such as real-time network monitoring data. This data can be exploited from different user categories, such as: subscribers, technicians and authorities. For each one of these categories the KPIs can be interpreted from different viewpoints, such as pricing, network availability, QoS, etc.

Emergency situation: Since the proposed RRM system should be able to reconfigure the system onthe-fly, it should prioritize users in emergency situations and guarantee service provision.

Manual re-configuration: Another feature for the RRM platform should be the manual system reconfiguration. It has been proved that in several cases the operator would like to change network parameters without considering the overall impact on the performance.

Black spot: Wireless systems of 3G have the feature of cell-breathing, which is a result of constant power control in the downlink. This might result in black spots in the area of coverage. For that, the platform should be able to detect any black spot and offer to it adequate resources by utilizing resources from the rest of the network segments that it is controlling.

Seamless service provision: Finally, seamless service provision is required and this can be achieved by utilizing both the monitoring part of the system and the global RRM entity. Whenever a traffic overload is detected, this can also be interpreted as a degradation in QoS, for that the global RRM entity can force-handover the user to adjacent cells or networks, in order to efficiently fulfill his

requirements. We are going now to present a suitable resource management system that corresponds to the presented requirements.

4 Hierarchical radio resource management system

The CAUTION++ system provides the framework for unified resource management of the four network technologies GSM, GPRS, UMTS and WLAN, based on the "monitor and manage" concept [3]. Resources at the air-interface are real-time monitored and a centralized hierarchical system receives alarms from the distributed monitoring components, so that a set of management techniques is selected and applied where and when needed. The goal is to keep up a satisfactory level of the available resources, of the QoS and of the interoperation between the different wireless systems. Forced vertical handover of the user (i.e. from WLAN to GPRS) can be performed in normal congestion [6,7]. But when the operator of the super-set of access networks is unable to decongest a phenomenon in its network, the CAUTION++ system provides also the capability to perform a forced vertical-vertical handover to another operator's super-network (with available resources), in order to finally avoid the user's drop.

The hierarchical Radio Resource Management system in the CAUTION++ approach is composed by three main components, linked by means of dedicated wired lines or an IP based backbone network, as illustrated in figure 1.



Figure 1: Hierarchical Radio Resource management architecture for the CAUTION++ approach

These components are:

- ITMU (Interface Traffic Monitoring Unit); it extracts the status information from each one of the access networks by providing the traffic information and the alarm message type towards the corresponding RMU entity.
- **RMU** (Resource Management Unit); it is the core element for each network separately, where the management techniques are decided and executed.
- **GMU** (Global Management Unit); it is the centralized CAUTION++ core component that enables the coexistence of different access networks, providing a global resource management to them and further assistance for the correct operation of the RMU tasks.

Moreover, due to the fact that different access networks are located in the same area with either cells or access points, a Location Server (LS) is also used in order to provide location information of the mobile users to the GMU for maximizing the efficiency of its operations. Furthermore, a database (MGIS) containing map information for the network can assist the decisions of the management elements of the system. Finally, although it is not part of the architecture, terminals that are CAUTION++ enabled are named IMT (Interactive Mobile Terminal) and they are capable of communicating with the platform in order to get assistance for the call/session setup procedure. In the next paragraphs, the main functional characteristics of the relevant radio resource management entities are described in detail. We have to note also here that the presented approach does not affect the interfaces of the various access networks since this a strong requirement for smooth system evolution [8].

4.1 Interface Traffic Monitoring Unit

For every radio access technology, the ITMU is responsible to monitor the corresponding network, collecting reports, surveying these reports to recognize a congestion situation and reporting them to the corresponding RMU. So, each ITMU collects the various RTT (Real Time Traffic) reports from each network and calculates the appropriate Key Performance Indicators² (KPIs). Each KPI has a critical threshold. If the calculated value overcomes this critical threshold, the ITMU sends the appropriate alarm message to the RMU. This knowledge helps us to understand the system's situation at any time (problems, congestion situation, etc.). Examples of KPIs are: Blocking probability, delay, peak and mean throughput etc.

When the ITMU identifies an overloaded resource, that is a KPI parameter is over the threshold, an alarm towards the RMU it is attached to is generated. The ITMU entity sends all the data it has that are related to the congested cell(s). So, the RMU is provided with a snapshot of the overall congestion situation. The alarm is periodically sent and deactivated only when the resource comes back to the normal operation (non congestion situation).

4.2 Resource Management Unit

The RMU is a centralized element within each individual radio network (GSM, GPRS, UMTS, and WLAN) that is able to manage the alarms generated by several ITMUs and to react properly to the congestion situations originated in the network it belongs to. Since each network type has its own features and characteristics, a specific RMU must be designed for each of them. The RMU receives from the ITMU the information on congestion situations with the related list of alarms or KPIs, and:

- a. Determines the Traffic Load Scenario that best suits to the current congestion situation;
- b. Selects the best Resource Management Strategy to cope with the congestion situation, when it is possible;
- c. Applies the selected strategies, through the Network Management System (NMS) of the underlying wireless network.

In order to fulfill the tasks listed at the previous points, the RMU is composed by the following modules:

• The *Traffic Load Scenario Recogniser* (TLSR), which is in charge of analysing the alarm messages coming from the ITMU and, by inquiring the internal knowledge database, identifies the appropriate traffic load scenario that corresponds to the current real traffic conditions.

The *Strategy Selector* (SS), which is responsible to select the most appropriate Resource Management Technique (RMT) as well as the values of the corresponding RMT parameters to efficiently manage the identified traffic load scenario. Once an RMT has been chosen to deal with the identified scenario, a model-based approach is used to perform the fine-tuning of the RMT parameters exploiting also past observed congestion cases and applied techniques that are stored in an appropriate Knowledge Base Manager (KBM).

• The *Strategy Actuator* (SA), which is the module responsible for the execution of the resource management techniques selected during the selection phase.

The architecture for the RMU entity is shown in figure 2.

² Each network (GSM, GPRS, UMTS, WLAN) has its own character and features. Thus different KPIs exist for each network type.



Figure 2: RMU architecture

4.3 Global Management Unit

The RMU may, in some cases, be unable to completely manage a congestion situation. Then the Global Management Unit (GMU) takes the control and tries to find a solution. Conversely to the RMU, which is a network dependent unit, the GMU is a centralized entity, being attached to all four possible networks (GSM, GPRS, UMTS and WLAN). That is, the role of the GMU is to decongest the network at a higher level, using not only the resource management techniques that pertain to the RMU but also the so-called heterogeneous resource management techniques that can be applied horizontally in more than one of the access networks.

The main tasks of the GMU in order to cope with a congestion situation in an efficient way are:

- a. Split traffic between all access networks
- b. Maintain QoS to all sessions
- c. Resolve "black spots" areas

From a high level system architecture point of view, the GMU unit is divided into three segments, shown in figure 3, and named: the interface modules, the triggered modules and the core modules.

Interface Modules

The interface modules are divided into:

RMU interface, which is in charge to provide all the input and output information between the two CAUTION++ modules, the GMU and the RMU. Since a GMU may be connected to up to four different RMUs, according to the access network technology, all incoming information from the RMUs should be decoded so as to allow efficient handling.

LS/MGIS interface. This module is responsible for performing all needed encoding and decoding actions in order to convert the information into the desired format for further internal process.

GMU interface. This module performs also decoding/encoding to all input and output information. Furthermore, it stores the list of all attached GMUs.

IMT interface. Since the GMU has the capability to connect with the SMS Centre or Web Server, this module communicates with both the two external modules to provide all the updated information towards the IMTs. In addition, this interface also provides a direct communication between the queried IMT an the GMU.



Figure 3: GMU System Architecture

Triggered Modules

The triggered modules act as input to the SS Core Module and are divided into:

Self-Triggered Module (STM) is the entity in charge of resolving the black spot areas, which is one of the GMU's roles. Since the RMU could modify some network parameters, which potentially could affect the cell or access point coverage, the GMU monitors these updates for consistency purposes. Input information from the ITMU component is passed through RMU, and it is processed in this module.

Traffic Resource Map Builder (TRMB) module is the main module that triggers a decision making process when an *RMU ALARM* is present. This module, with the help of the GMU Interface Module and the RMU Interface Module, creates a "Traffic Resource Map" (TRM) for the affected area. At the same time, it communicates with the LS/MGIS Interface Module and updates the traffic map with coverage info, in order to identify possible candidate networks – from the same operator. The output of the TRMB module is a TRM report, which is fed to the SS Module. A Traffic Resource Map is a combination of ITMU data resulted from monitoring together with radio resource information, Quality of Service information, scale of the affected area and coverage information queried from the MGIS component.

Information Exchange Module (IEM). This module enables, through the TRMB, the exchange of coverage and other network information between the GMU and the ITMU. The information is exchanged by periodic access to either the SMS Centre or Web Server and the ITMU entities. The IEM module also provides past information of the congestion affected area.

Core Modules

The Core Modules correspond to the main functionalities and processes of the GMU. These modules are quite similar to those that are implemented in the RMUs. However now, because the GMU is located at a higher architectural level (the highest in the CAUTION++ system approach), it can evaluate more accurately a congestion situation by comparing the results and inputs from different RMUs. The entities allocated in the core network are the following:

Strategy Selector (SS) module. Similarly to the RMU-SS, it is the Core Module for the decision making process based upon the provided TRM.

Strategy Actuator (SA) module. This module is responsible for constructing the correct format of the congestion control action that is forwarded to the RMU Interface. The SA communicates also with the LS/MGIS Interface Module and transforms the GMU location information to the corresponding Cell ID or Access Point number.

Knowledge Base Manager (KBM) module. It stores information from all internal processes of the GMU, like successful actions from SS and TRMs from the TRMB Module to be used in future similar situations. At the same time, the KBM stores all the predefined intra_GMU RMTs and makes them available to the SS Module.

Considering the inter-networking of the systems in the framework of the CAUTION++ architecture, we face the situation of multiple network operators, as well as multiple service providers. In absence of overload traffic scenarios, when a user is roaming within an area with overlapping coverage (e.g. an area that is covered by both WLAN and GPRS), it sounds very reasonable to desire the exploitation of the capabilities of the most attractive one. In the above example, the usage of WLAN instead of GPRS would be not a need, but just a user's "desire" for achieving better results. In presence of overload traffic scenarios (whichever be the cause originating the congestion, such as bad bandwidth allocation and management, or because on New Year's Eve everyone is trying to communicate with his/her familiars) the choice of a specific technology may be no more a "desire", but a vital and absolute necessity. For example, in presence of scarce resource availability, using the radio resource according to the packet switching pattern instead of the circuit-switching one, allows more users to share the available radio resources by exploiting the principle of statistical multiplexing. A gain in efficiency is derived, since a large number of users can potentially share the same bandwidth and be served from a single cell/access point. However, even in this case, congestion situations in signaling or data overload cases cannot be avoided. For example, a GPRS user may request, via the Packet Random Access Channel (PRACH), one or more Packet Data Traffic Channel (PDTCH), but due to the lack of resources, no dedicated channels are available. In a worse situation, the user cannot even access a PRACH in order to request a PDTCH[7]. Consequently, it is clear that there is an imperative need for super-system's management, as provided by the CAUTION++ framework, in order to avoid congestion situations within it.

4.4 Hierarchy in Decision-Making

As shown in figure 1, in CAUTION++ the congestion treatment process can be computed at three levels, which actually constitute a hierarchy:

- RMU level: the selected RMT impacts only the wireless network connected to the RMU which operated the reconfiguration;
- Intra-GMU level: the selected RMT impacts (a subset of) the wireless networks belonging to the same operator;

• Inter-GMU level: the selected RMT impacts wireless network belonging to different operators. Some criteria have to be defined, to regulate the escalation process among these three levels. The general principle is that alarms about overload conditions, issued by ITMU components, have to be managed as soon as possible, from the bottom (i.e., RMU) to the top (inter-GMU) of the decision making hierarchy. Therefore, reconfiguration locally to the overloaded wireless network is attempted first; should the RMU be not able to identify a proper RMT, then the alarm is signaled at GMU level, which first attempts an intra-GMU RMT and resorts to an inter-GMU RMT if the previous attempt failed. There is no specific obstacle in splitting the congestion treatment process between more than one level, that is the RMU may perform a partial reconfiguration which accommodates locally part of the traffic overload, and demands the treatment of the remaining congestion to the GMU.

According to the description of the CAUTION++ components, the RMU decision-making process is triggered by the ITMU each time KPI thresholds are exceeded. For the triggering of the GMU decision-making process, the following scenarios are possible:

- **GMU triggered by the RMU:** When the GMU is triggered by any of its subordinate RMUs, this means that this RMU is experiencing a congestion situation that cannot be satisfactorly solved by the local decision making of the RMU.
- **GMU self-activation:** In this scenario the GMU starts resource management because it has been able to identify a particular situation that is happening in the underlying wireless networks, and that is likely to be effectively handled at its level.

• **GMU triggered by another GMU:** When a GMU decides to escalate a part of its offered workload to another GMU, this means it is experiencing a congestion situation that cannot be satisfactorily treated with the resources of the network operator.

At any rate, it is reasonable to assume that each network operator will hide details of its network to the competitors, and thus a GMU being requested to accommodate some traffic currently attached to another operator's network will not be informed of the actions already taken to deal with the congestion situation, but will only receive the information required to take a decision on whether to accept, totally or partially, the support request. Once this decision has been taken, it is forwarded to the requesting GMU, and a set of actions performed by the two GMUs to actually realize the traffic movement from the networks of an operator to the others' ones.

5 Conclusions

In this paper we have presented the structure of a future heterogeneous wireless networking environment as well as an architecture for the efficient management of resources within it. Based on this structure, we have listed the system requirements for the RRM system that aims in increasing the performance of such a next generation wireless networking environment. Furthermore, we presented in detail the components together with their functionalities of the CAUTION++ approach and more importantly the factors that enable the triggering of the RRM operations. The potential of this approach is already known, since a first version of this platform is validated in a real networking environment of a GSM/GPRS network, showing important performance increase. This is the major motivation for this work that is expected to provide a powerful platform that can be utilized by network operators and service providers in order to increase system performance and user satisfaction in wireless networks of the future.

Acknowledgement

This work has been performed in the framework of the IST research project CAUTION++ (IST-2001-38229), which is partly funded by the European Community. The authors would like to acknowledge the contributions of their colleagues from CAUTION++ project.

References

- [1] Glyn Roberts, STMicroelectronics, "IEEE 802.15 Overview of WG and Task Groups", available at http://grouper.ieee.org/groups/802/15/pub/Tutorials.html
- [2] Nikos Tselikas et al, "Architectural framework for resource management optimisation over heterogeneous wireless networks", in Proc. ITCom (Information Technologies and Communications) SPIE conference, Orlando, Florida, USA, 7-11 September 2003.
- [3] CAUTION++ IST project (IST-2001-38229) http://www.telecom.ece.ntua.gr/CautionPlus
- [4] W. Stallings, "IPv6: The New Internet Protocol," IEEE Communications Magazine, pp. 96--108, July 1996.
- [5] Jukka Manner, Louise Burness, Eleanor Hepworth, Alberto López, Enric Mitjana, "Provision of QoS in Heterogeneous Wireless IP Access Networks", PIMRC 2002, Lisboa, Portugal, September 2002.
- [6] Josep Bada, Ferran Casadevall, Javier Jimenez, Eduardo Medrano, "Radio resource management with QoS in GPRS", IST Mobile Communications Summit, Thessaloniki, Greece, June 2002.
- [7] Lin, Y.-B., Haung, Y.-R., Chen, Y.-K., and Chlamtac, I. "Mobility Management: From GPRS to UMTS", Wireless Communications and Mobile Computing, Vol. 1, Issue 4, pp. 339-360, 2001.
- [8] Uskela Sami, "Key concepts for evolution toward beyond 3G networks", IEEE Wireless Communications Magazine, Volume 10, Issue 1, Feb 2003.