Abstract—Beyond 3G systems is usually the term used to refer to the new scenarios in the wireless arena where different Radio Access Technologies (RATs) will coexist and operate in a coordinated way. This cooperation must indeed be regarded as a new challenge to offer services to the users over an efficient and ubiquitous radio access. In this way, the user can be served through the RAT that fits better to the terminal capabilities and service requirements, and also a more efficient use of the radio resources can be achieved. This challenge calls for the introduction of new Radio Resource Management (RRM) algorithms operating from a common perspective that take into account the overall amount of resources offered by the available RATs. In this context, this paper presents the framework for developing RRM algorithms in the B3G scenarios, including some possible approaches.

I. INTRODUCTION

In parallel with the development of the 3G cellular systems, other wireless access technologies for Wireless Local Area Networks (WLAN) and Personal Area Networks (PAN), like IEEE 802.11 and Bluetooth, have been developed, standardised and have experienced a significant growth, arriving to the mass market. In turn, in the field of cellular systems, the extension of GSM (Global System for Mobile communications) to GPRS (General Packet Radio Service) including packet transmission capabilities in the radio interface has been a first milestone in the evolution path of 2G cellular systems towards UMTS (Universal Mobile Telecommunications System). The co-existence and interactions between UMTS and GSM/GPRS technologies constitute one of the key points for the success of 3G technologies. As a matter of fact, GSM/GPRS has also followed its independent path with the development of an improved radio access technology that allows higher bit rates thanks to the use of more efficient modulation schemes. The term EDGE (Enhanced Data rates for GSM Evolution) is used to refer to this improved system and the term GERAN (GSM/EDGE Radio Access Network) is the name of the evolved radio access network including these new capabilities.

As a result of the above, the scenarios where UMTS will be deployed will probably differ from those for which it was initially thought, and it will have to co-exist not only with previous 2G and 2.5G systems but also with WLAN and other emerging technologies. These new scenarios where different Radio Access Technologies (RATs) will coexist and will operate in a coordinated way are often referred as beyond 3G (B3G) systems. Fig. 1 shows an example of such a heterogeneous networks scenario. It is constituted by several radio access networks (RAN) interfacing a common core network. Radio access networks include cellular networks, e.g. UTRAN (UMTS Terrestrial Radio Access Network) with the two modes FDD (Frequency Division Duplex) and TDD (Time Division Duplex), and GERAN. These networks may in turn be subdivided into different cellular layers (e.g. macro, micro or picocells) depending on the expected coverage area, and also other public non-cellular access networks (e.g. WLAN). The core network infrastructure is typically subdivided in the circuit switched (CS) and packet switched (PS) domains providing access to external networks, e.g. PSTN (Public Switched Telephone Network) or Internet. The scenario assumes the existence of multi-mode terminals, providing connectivity to multiple access networks either in different time instants or even simultaneously.

The availability of several access networks must indeed be regarded as a new challenge to offer services to the users over an efficient and ubiquitous radio access thanks to coordinating the available RATs. In this way, not only the user can be served through the RAT that fits better to the terminal capabilities and service requirements, but also a more efficient use of the available radio resources can be achieved [1]. This calls for the introduction of new Radio Resource Management (RRM) algorithms operating from a common perspective that take into account the overall amount of resources available in the existing RATs, and therefore are referred to as CRRM (Common Radio Resource Management) algorithms [2][3]. Furthermore, for a proper support of such algorithms, suitable network architectures and procedures must ensure the desired interworking capabilities between the different technologies [4]-[6]. The interworking architecture enhanced with CRRM functionality will pave the way for the extension of these heterogeneous networks to include also new 4G radio access technologies.

This paper provides the framework for coping with the radio resource management problem in beyond 3G scenarios, describing first the differences in how the problem is dealt up
to 3G systems and then how the CRRM functionalities are introduced in beyond 3G systems. Also some specific approaches to deal with the problem are analysed. The rest of the paper is organised as follows. Section II and Section III discuss the RRM problem for 3G and beyond systems, respectively. Section IV presents the CRRM functional model devised in 3GPP (Third Generation Partnership Project) and the functionalities are discussed in Section V. Finally, Section VI presents some CRRM solutions and the conclusions are summarised in Section VII.

II. Radio Resource Management up to 3G Systems

A Radio Resource Unit (RRU) can be defined by the set of basic physical transmission parameters necessary to support a signal waveform transporting end user information corresponding to a reference service. These physical transmission parameters depend on the multiple access technique being used. In particular, in FDMA (Frequency Division Multiple Access), a radio resource unit is equivalent to a certain bandwidth within a given carrier frequency. In TDMA (Time Division Multiple Access), a radio resource unit is equivalent to a pair of a carrier frequency and a time slot. In CDMA (Code Division Multiple Access), a radio resource unit is defined by a carrier frequency, a code sequence and a power level. The main difference arising in CDMA is that the required power level necessary to support a user connection is not fixed but depends on the interference level. Thus, the amount of transmitted power resources will vary along time according to multiple elements of the scenario, such as propagation conditions, interference, cell load level, etc. In addition to the main physical dimensions (frequency, time slot, code sequence and power level), there are other physical transmission elements such as the modulation scheme, channel coding scheme, etc. Clearly, depending on the exploitation of the basic dimensions in terms of the former elements different spectral efficiencies may follow.

The objective of a network operator is the deployment of a network able to support its customers with the required QoS under the target coverage area. Focusing on the radio part, the output of the radio network planning will be the provision of RRUs along the service area by means of a certain radio network topology and a given configuration of the cell sites. As long as e.g. service penetration and service usage vary along time and space, the amount of radio resource units to be provisioned also varies and, consequently, the radio network planning is an evolving process. After the RRUs have been provisioned in the service area during the planning phase, RRM functions are in charge of allocating and managing the provisioned RRUs.

Cellular mobile communications are dynamic in nature. Dynamism arises from multiple dimensions: propagation, traffic generation, interference, etc. Thus, the dynamic network evolution calls for a dynamic management of the available RRUs, which is carried out by means of RRM mechanisms with an associated number of parameters that need to be chosen, measured, analyzed and optimized. Besides, RRM mechanisms may overcome at some extent the long term reactivity in radio network planning and deployment, which otherwise would prevent the network operator to accommodate sudden and transient traffic increases [2][7].

RRM functions should then take into account the constraints imposed by the radio interface in order to make decisions regarding the setting of the different elements and parameters influencing the air interface behaviour. Some of these are the number of active users, the number of simultaneous users transmitting, the corresponding transmission rates for each user, the transmitted power levels corresponding to every simultaneous user, etc.

Clearly, the number of parameters to be controlled as well as their different nature claims for a set of several RRM functions, whose joint behaviour should lead to an overall radio access network optimisation. The need of a controlled air interface usually requires a strong centralised component in the RRM operation. Nevertheless, the higher signalling load associated to a centralised operation usually requires having some decentralised elements located in the terminal side. RRM algorithms are fed by measurements, some of them provided from the terminal (e.g. serving and neighbouring cells received power levels) and some others from the Radio Access Network (RAN) side (e.g. serving and neighbouring cells load level). Decisions taken from centralised RRM algorithms are transferred to the terminal through layer 3 protocol procedures.

RRM functions need to be consistent for both uplink and downlink, although the different nature of these links introduces some differences in the followed approach. Since the different RRM functions will target to track different radio interface elements and effects, RRM functions can be classified according to the time scales they use to be activated and executed. Then, the set of RRM functions with the corresponding typical time scales between consecutive activations of the involved algorithm would be [2]:

- Order of 1ms: Inner loop power control in CDMA
- Order of 10ms: Packet scheduling, MAC (Medium Access Control) algorithms
- Order of 100ms - 1s: Admission control, Handover, Congestion control, Outer loop power control in CDMA

RRM functions can be implemented in many different algorithms, this impacting on the overall system efficiency and on the operator infrastructure cost. Additionally, RRM strategies are not subject of standardisation, so that they can be a differentiation issue among manufacturers and operators. The specific RRM strategies are leading to an increase in competitiveness among the manufacturers. RRM strategies of legacy networks (e.g. GSM/GPRS) are of rather low dimensionality, i.e. only a few parameters are needed to tune their optimality. On the contrary, in the case of UTRAN (UMTS Terrestrial Radio Access Network), the required flexibility to accommodate different service requirements together with the more sophisticated nature of the WCDMA technique turned into more dimensions in the optimisation.
problem thus increasing the complexity of the RRM problem. Finally, WLANs are also expected to play an important role in the provision of high data rate services, so that RRM should be targeted if QoS needs to be provided.

III. RADIO RESOURCE MANAGEMENT IN B3G SYSTEMS

In a B3G where several RATs coexist the management of the provisioned RRUs can be seen as a problem with multiple dimensions. Every RAT is based on specific multiple access mechanism exploiting in turn different orthogonal dimensions, such as frequency, time and code. Then, local RRM mechanisms are needed for every considered RAT: GERAN (GSM/EDGE Radio Access Network), UTRAN and WLAN, as shown in Fig. 2 (a). In addition to that, a proper coordination among the different RATs can be achieved with development of Common RRM (CRRM) algorithms. CRRM is based on the picture of a pool of radio resources, belonging to different RATs but commonly managed, as shown in Fig. 2 (b). Then, the additional dimensions introduced by the multiplicity of RATs available provide further flexibility in the way how radio resources can be managed and, consequently, overall improvements may follow.

Notice that the CRRM vision allows also considering different amounts of radio resources spatially available, because in terms of current network deployment deployment, different spatial availabilities are found for the existing RATs. For example, GERAN tends to be the most widespread RAT, while UTRAN is not yet everywhere deployed and in turn WLAN hotspots with reduced coverage areas are also widespread around cities.

IV. CRRM FUNCTIONAL MODEL

The functional model assumed in 3GPP for CRRM operation considers the total amount of resources available for an operator divided into radio resource pools. Each radio resource pool consists of the resource units available in a set of cells, typically under the control of a RNC (Radio Network Controller) in UTRAN or a BSC (Base Station Controller) in GERAN. Two types of entities are considered for the management of these radio resource pools [4][5], as shown in Fig. 3.

- The RRM entity, which carries out the management of the resources in one radio resource pool of a certain radio access network. This functional entity involves different physical entities in the RNS (Radio Network Subsystem) or BSS (Base Station Subsystem) depending on the specific considered functions, although for representation purposes it is usual to assume the RRM entity residing in the RNC or the BSC. Notice that different RRM entities do not necessarily belong to different radio access technologies.

- The CRRM entity, which is involved in the coordinated management of the resource pools under different RRM entities. In this way, decisions on radio resources usage may take into account the resource availability in several RRM entities. Each CRRM entity controls a number of RRM entities and may communicate with other CRRM entities as well, thus collecting information from other RRM entities that are not under its direct control.

The interactions between RRM and CRRM entities involve mainly two types of functions:

a) Information reporting function

The information reporting function allows the RRM entity to report relevant information to its controlling CRRM. The reporting can be performed periodical or event-triggered, or even at a given instant, and it is totally up to CRRM entity’s request. The exchange of information is also possible between different CRRM entities in order to know the status of their corresponding RRM entities. There are mainly two types of information to be reported to the CRRM entity:

- Dynamic common measurements on cells controlled by a given RNC or BSC entity. These measurements include the current cell loads, transmitted carrier power, the received total wideband power, interference measurements, etc.
- Static information on cells controlled by a given RNC or BSC entity. This includes the knowledge about the cell relations (e.g. if they are overlapped or if they belong to different HCS layers), the cell capabilities (e.g. whether a cell supports GPRS, EDGE,...) the cell capacities (e.g. the number of available time slots) or the available QoS (e.g. maximum bit rate for a given service or average buffer delay)

b) RRM decision support function

This function describes the way how the CRRM and RRM entities interact for taking decisions. For example, it is possible that the CRRM simply advises the RRM entity, so that the RRM remains as the master of the decisions, and, on the contrary, it is also possible that the CRRM is the master so that its decisions are binding for the RRM entity.
V. CRRM FUNCTIONS

As explained in section II, the main RRM functionalities arising in the context of a single RAN are: admission and congestion control, horizontal (intra-system) handover, packet scheduling and power control. When these functionalities are coordinated between different RANs in a heterogeneous scenario, they can be denoted as “common” (i.e. thus having common admission control, common congestion control, etc.) as long as algorithms take into account information about several RANs to make decisions. In turn, when a heterogeneous scenario is considered, a specific functionality arises, namely RAT selection (i.e. the functionality devoted to decide to which RAT a given service request should be allocated).

After the initial RAT selection decision, taken at session initiation, vertical (inter-system) handover is the procedure that allows switching from one RAN to another. The successful execution of a seamless and fast vertical handover is essential for hiding to the user the underlying enabling infrastructure. Issues related to vertical handover comprise scanning procedures for the terminal to discover available RANs, measurement mechanisms to capture the status of the air interface in the different RANs, vertical handover triggers (i.e. the events occurring in the heterogeneous network scenario that require the system to consider whether a vertical handover is actually required or not), vertical handover algorithm (i.e. the criteria used to decide whether a vertical handover is to be performed or not) and protocol and architectural aspects to support handover execution.

Vertical handover procedures from one RAN to another may be useful to support a variety of objectives, such as avoiding disconnections due to lack of coverage in the current RAT, blocking due to overload in the current RAN, possible improvement of QoS by changing the RAT, support of user’s and operator’s preferences in terms of RANs usage or load balancing among RATs. Thus, the vertical handover procedure enables another dimension into the CRRM problem and provides an additional degree of freedom in rearranging traffic [3][8]-[10].

VI. CRRM SOLUTIONS

According to the framework presented in the previous sections in the following some specific solutions are addressed coping with the RAT selection problem in heterogeneous scenarios.

A. Policy-based RAT selection schemes in UTRAN/GERAN

Policy-based management has been the subject of extensive research during the last years in IP-based multiservice networks [11], and may also be considered as a possibility for CRRM design. A policy can be defined as a high-level declarative directive that specifies some criterion to guide the behaviour of a network responding to some network operator preferences. In the following several policies for RAT selection in a heterogeneous scenario including UTRAN and GERAN are discussed.

a) Service Based RAT selection

A service-based RAT selection policy is based on a direct mapping between services and RATs [12]. As an example, in a scenario including voice and interactive service assuming all terminals have multi-mode capabilities (i.e. they can work either with UTRAN or GERAN), two possibilities would be:

- VG (voice GERAN) policy: This policy allocates voice users into GERAN and interactive services into UTRAN.
- VU (voice UTRAN) policy: This policy allocates voice users into UTRAN and interactive services into GERAN.

If no capacity is available in the primary RAT, the other RAT is selected instead. If no capacity is available in the alternative RAT, the service request gets blocked (at service set-up) or dropped (during service life-time).

Table I compares the performance in terms of aggregated throughput (i.e. including both voice and www users) when basic policies VU and VG are considered in a scenario with seven omnidirectional cells with radius 1 km for UTRAN and GERAN assuming that the cells of both systems are co-sited. Notice that, in all the cases, VG policy outperforms VU, revealing the suitability of allocating voice users in GERAN. The main reasons are two-fold. First, with respect to www users, a higher throughput can be obtained in UTRAN as long as DCH channels are used while in GERAN www users are subject to a scheduling algorithm. In turn, from the voice users’ point of view, if the distance between cell radius was set to 500m, no significant differences would be observed between VU and VG (the results are not shown for the sake of brevity), but when increasing the radius, a higher degradation is observed in VU because UTRAN users at the cell edge experience some erroneous transmissions due to power limitations and the interference-limited nature of WCDMA.

Table I

<table>
<thead>
<tr>
<th>Users</th>
<th>VU</th>
<th>VG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voice</td>
<td>UL DL</td>
<td>UL DL</td>
</tr>
<tr>
<td>400</td>
<td>2.08 2.17</td>
<td>2.14 2.22</td>
</tr>
<tr>
<td>600</td>
<td>2.88 3.09</td>
<td>2.95 3.15</td>
</tr>
<tr>
<td>1000</td>
<td>3.64 3.96</td>
<td>3.76 4.08</td>
</tr>
</tbody>
</table>

b) Load Balancing-based RAT selection

Load balancing (LB) is another possible guiding principle for resource allocation in which the RAT selection policy will distribute the load among all resources as evenly as possible. Specifically, the selected RAT will be the one having the lowest load. Therefore, an influential run-time parameter in a load balancing decision-making procedure is the load metric.

For UTRAN, an average of the cell load factor can be used, while in GERAN a useful way to measure the data load is to distribute the load among all resources as evenly as possible. If no capacity is available in the alternative RAT, the service request gets blocked (at service set-up) or dropped (during service life-time).

Fig. 4 compares the load balancing decision-making procedure for resource allocation in which the RAT selection policy will distribute the load among all resources as evenly as possible. Specifically, the selected RAT will be the one having the lowest load. Therefore, an influential run-time parameter in a load balancing decision-making procedure is the load metric.

For UTRAN, an average of the cell load factor can be used, while in GERAN a useful way to measure the data load is to measure the average amount of time slots utilized by GSM/EDGE services.

Fig. 4 compares the LB RAT selection algorithm against the service-based VG policy in a scenario with seven omnidirectional cells for UTRAN and GERAN co-sited with cell radius 500m. Particularly, it shows the voice call dropping probabilities (in %) for both algorithms. Up to 600 voice users, dropping values are kept sufficiently low. For 800 voice users however, VG reveals higher dropping values than policy LB. The higher dropping rates experienced by VG policy is
explained bearing in mind the load distribution in GERAN induced by VG and LB policies. In particular, for VG the load is at its maximum value most of the time which implies a lack of flexibility in order to accommodate handover users being redirected to GERAN. Therefore, VG may incur in more potential dropping situations than in the case of LB policy appliance which presents more fluctuations in the load values and can provide resources to incoming handover users if necessary.

B. Fuzzy-Neural based CRRM in UTRAN/GERAN/WLAN

From the above solutions shown in the previous section it has been observed that the CRRM operation can be guided by very different principles, related with technical aspects as well as with operator policies. Furthermore, one of the problems that CRRM algorithms must face is the existence of uncertainties when comparing different measurements belonging to different RATs that are necessarily of a different nature together with subjective criteria that have to do with techno-economic issues. As a result, the use of fuzzy logic as a robust decision making procedure becomes another possible solution for CRRM algorithm development. The fuzzy subset methodology has been proved to be good at explaining how to reach the decisions from imprecise information by using the fuzzifier and defuzzifier rules and the inference engine concept [14][15]. On the other side, the use of neural networks, which are good in recognizing patterns by means of learning procedures, could also be considered and, as a matter of fact, they have been proposed to be used in hybrid fuzzy-neural based systems [16][17]. Taking these considerations into account, in the following the characteristics of a fuzzy neural CRRM solution are discussed in a scenario with UTRAN, GERAN and WLAN access technologies.

Fig. 6 depicts the block diagram of the Fuzzy-Neural CRRM. Two main blocks are identified, named fuzzy neural and reinforcement learning. In addition to them, other blocks including techno-economical aspects could also be included [18]. A brief description of these blocks is detailed in the following. For details the reader is referred to [18][19].

The purpose of the fuzzy neural algorithm is to obtain for each RAT a numerical indication (denoted as Fuzzy Selected Decision: FSD) between 0 and 1 of the suitability to select it. The decision is obtained from a set of input linguistic variables (LVi), reflecting technical measurements. This decision is taken in three steps, as depicted in Fig. 6.

Step 1.- Fuzzification. This process assigns, for each input linguistic variable, a value between 0 and 1 corresponding to the degree of membership of this input to a given fuzzy subset. A fuzzy subset is a linguistic subjective representation of the input variable. The considered input variables are the signal strength (SS) and the resource availability (RA) for each of the considered RATs and the mobile speed. Just as an example, a fuzzy subset for the resource availability RA (e.g. for the number of available time slots in GERAN) could be formed by the possibilities H (high), M (medium) and L (low). There exists one membership function for each one of the three terms...
H, M and L, reflecting the degree of membership of the RA value to each term.

Step 2.- Inference Engine. For each combination of fuzzy subsets from step 1, the inference engine makes use of predefined fuzzy rules to indicate, for each RAT, the suitability of selecting it. So, at the output of this step there will be a combination of three output linguistic variables D (D_{UTRAN}, D_{GERAN}, D_{WLAN}) each one with four fuzzy subsets: Y(Yes), PY (probably yes), PN (probably not) and N (not), with different degrees of membership for each of them. Together with the suitability of selecting each RAT, the inference engine can also determine a level of allocated bandwidth in it with the fuzzy subsets H (high), M (medium) and L (low).

Step 3.- Defuzzification. This procedure converts the outputs of the inference engine into a number ranging between 0 and 1, named Fuzzy Selected Decision (FSD) for each RAT that reflects the suitability of select it. The selected one will be then the one having the highest FSD. This procedure also provides the allocated bandwidth in the selected RAT.

The reinforcement learning procedure is used to suitably tune the parameters of the different and rather subjective functions involved in the fuzzy logic controller [16] in order to ensure a certain target value of a QoS parameter, like e.g. the ratio of non-satisfied users (i.e. the users that receive a bandwidth below a certain desired value), the ratio of blocked users, dropping calls, etc.

An example of the performance that can be obtained by means of the fuzzy-neural CRRM strategy operating with reinforcement learning is shown in Fig. 7. It presents the time evolution of the percentage of non-satisfied users in a simple scenario with three concentric cells: one UTRAN cell, one GERAN cell and one WLAN access point. In this example the simulation time is measured in periods of 100 ms. Two values of the target ratio of non-satisfied users are considered, namely P*=1% and P*=3% and it can be noticed that the algorithm is able to converge to the desired value under variable traffic and mobility conditions.

![Fig. 7 Evolution of the probability of non satisfied users](image)

VI. CONCLUSIONS

This paper has focused on the RRM problem in heterogeneous wireless networks where different RATs coexist. The framework for developing CRRM strategies including the functional model and a description of the major functionalities has been provided. Also different approaches for the RAT selection procedure have been presented, making use of policy-based strategies and fuzzy-neural methodology.

REFERENCES

[4] 3GPP TR 25.881 v5.0.0 “Improvement of RRM across RNS and RNS/BSS”
[5] 3GPP TR 25.891 v0.3.0 “Improvement of RRM across RNS and RNS/BSS (Post Rel-5) (Release 6)”