A Delay-Centric Dynamic Resource Allocation Algorithm for Wireless Communication Systems based on HNN

Daniel Calabuig, José F. Monserrat, *Student Member, IEEE,* David Gómez-Barquero, *Student Member, IEEE,* and Narcís Cardona, *Member, IEEE*

***Abstract*– This paper proposes a Dynamic Resource Allocation (DRA) algorithm for packet data services in wireless communication systems based on Hopfield Neural Networks (HNN). The resource allocation algorithm assumes a delay-centric approach in that it maximizes resource utilization of the overall system while minimizing the packet delay. The real-time working capability of HNN hardware implementation means that a very powerful scheduling DRA algorithm can be obtained. A generic formulation of the algorithm is presented to establish the optimal bit rate allocation. In addition, some illustrative examples of this formulation are given, considering specific wireless communication systems, such as GPRS or UMTS. To be more precise, the performance of the proposed DRA algorithm is evaluated in a realistic UMTS scenario, considering both real time (RT) and non-real time (NRT) services. In order to obtain the best resource distribution and to fulfill the different Quality of Service (QoS) levels required by RT and NRT services, the new HNN-based delay-centric DRA algorithm is performed twice. Initially only the RT services are considered and following this all the NRT services are taken into account. The results reveal that the proposed DRA algorithm outperforms other reference algorithms not only in terms of average packet delay, but also in terms of the allocated total bit rate.**

***Index Terms*– Dynamic Resource Allocation (DRA), Hopfield Neural Networks (HNN), Wireless Networks.**

I. INTRODUCTION

Mobile wireless communications are in constant evolution due to the continuously evolving requirements and expectations of both users and operators. This is reflected in the spectacular increase in both the quantity and quality of mobile services, especially packet-based data services. Moreover, in the near future, wireless networks will converge towards the sole use of IP-based protocols, meaning that all services will be delivered using IP.

Generally speaking, mobile packet data services can be classified in terms of *real-time* (RT) and *non real-time* (NRT) services. The main difference between these services is the extent to which traffic is delay-sensitive. RT services, such as voice or video calling, are characterized by a strict delay constraint defined by a maximum tolerable delay. Packets exceeding their maximum delay requirement are usually discarded. In contrast, NRT services, such as FTP or web browsing, are characterized by a bursty traffic pattern and non-rigid delay restrictions. Although a maximum delay is usually not considered in NRT services, the service response time, defined as the period of time elapsed since the request instance up to complete message reception, is a satisfactory measure of the quality perceived by the end user, especially for those services referred to as *interactive* NRT services (e.g. web browsing or instant messaging). This measure is important since users expect the message to be delivered within a certain time. As such, shorter delays result in greater user satisfaction. With regards to this, a maximum desirable delay has been defined for all NRT services which the network operator tries to ensure, although it can be exceeded. The main difference between interactive and *background* NRT services (e.g., FTP or email) is that background users are not expecting the data within a certain time, and as such are not aware of the duration of the transmission.

From a Radio Resource Management (RRM) perspective, when demand for a new RT service occurs, a Fixed Resource Allocation (FRA) is typically performed. Thereby, a fixed amount of resources is reserved for the new connection for the entire duration of the session. This policy is inadequate if the traffic generation rate varies significantly or when the traffic source is discontinuous. These are the main features of NRT services which can also apply to some RT services such as the compressed video. Thus, with this approach only active users transmit but both active and inactive users are connected to the system. On the other hand, Dynamic Resource Allocation (DRA) algorithms establish different connections over the same resources and perform a scheduling policy to distribute the resources usage. This *sharing policy* leads to improved resource utilization. However, to achieve this a smart scheduling algorithm which can guarantee user satisfaction in terms of Quality of Service (QoS) is required. The objective of the DRA algorithm is to select the optimal amount of radio resources to be allocated for each user. This is subject to certain restrictions in terms of total available resources, QoS requirements (distinct for each service and user profile), coverage constraints, etc. (for examples see [1]-[9]). DRA algorithms are executed every time a new user enters the system (following acceptance by the admission control algorithm), and during the user sessions.

Certain DRA techniques allocate resources to the users with the best channel quality [1], [2]. This policy can maximize the average throughput of the system, but with an unfair distribution that implies lengthy delays for users with

Manuscript received May 16, 2006; revised August 30, 2006, December 12, 2006, April 25, 2007, September 21, 2007 and February 16, 2008. The associate editor coordinating the review of this paper and approving it for publication was Dr. Phone Lin. This work was partially supported by Spanish Science & Technology Commission (CICYT) under the project TIC2005-08211-C02.

All authors are with the Mobile Communications Group, Institute of Telecommunications and Multimedia Applications (iTEAM), Polytechnic University of Valencia (UPV) (email:{dacaso, jomondel, dagobar, ncardona} @iteam.upv.es).

poor channel quality. Other techniques establish a priority policy for the various types of users and dynamic priorities are assigned to packets belonging to the same class based on packet lifetime [3]. Although these algorithms can enhance the average packet delay, they fail to consider the maximization of resource utilization. Other approaches describe the DRA as an optimization problem and obtain a suboptimal solution with genetic algorithms [4], [5]. These techniques obtain satisfactory results, albeit at the expense of a high computational burden due to their complexity.

To sum up, to date most of the techniques reported in the literature are either incapable of achieving optimum resource allocation or cannot operate in real-time. Hopfield Neural Networks (HNN) have been used as a rapid solution for optimization problems, e.g. the traveling salesman problem [10] and [11], the N-queens problem [12]. In addition, with respect to wireless networks, HNN have been employed for dynamic channel allocation [6] and [7], and for dynamic resource allocation [8] and [9]. HNN have the capacity to find suboptimal solutions in a few microseconds [8], which is fast enough to establish a new resource allocation on a frame by frame basis in current wireless communication systems (e.g., in UMTS the frame period is 10 ms). The first study to introduce an HNN-based algorithm in a wireless system was that presented by Del Re et al. [6] that was built on the research work carried out by Lázaro and Girma [7]. They proposed an algorithm for the dynamic distribution of frequency channels over the cells of a GSM system together with a guard channel technique for handovers. Ahn and Ramakrishna [8] were the first to use HNN for solving the DRA problem. In the main, their algorithm aimed to maximize the allocated resources and to obtain a fair distribution among users. García *et al.* [9] applied this philosophy to the distribution of resources in a CDMA system in a manner which would satisfy the bit rate expectations of users.

These previous studies based on the use of HNNs, have tried to solve the DRA problem from a throughput-centric perspective. In other words, the main focus was on the maximization of the total allocated bandwidth. However, the DRA problem has only been partially solved, failing to guarantee an overall QoS to users since, in the scheduling process, the service response time has not been considered. This paper presents a novel DRA algorithm based on HNN with a delay control technique, called the HNN delay-centric (HNN-DC) algorithm, which tries to maximize resource utilization, while minimizing the packet delay. In order to prioritize RT services and minimize the RT packet dropping rate, the algorithm manages RT and NRT services separately. In the case of NRT services, background services are differentiated from interactive services by setting their maximum desirable delay to infinite. When considering the overall wireless system and the user delay constraints, the HNN-DC algorithm can provide services with an optimized user bit rate allocation together with a controlled delay.

The remainder of the paper is organized as follows. Section II presents the constraints for consideration when establishing the DRA algorithm: maximum system load, user bit rate permissions, and maximum packet delay. Section III introduces the HNN model and presents the HNN-DC formulation. Some illustrative examples of the specific formulation of Radio Access Technologies (RAT) are given in Section IV. Section V compares the HNN-DC algorithm with other scheduling algorithms in a UMTS scenario. Finally, the central conclusions are outlined in Section VI.

II. DRA CONSTRAINTS

The DRA problem centers upon finding the optimal bit rate allocation for all active user given a certain set of limitations or constraints. The actual implication of such constraints must be carefully reflected in the algorithm design. Some of these are *hard* constraints, and need to be satisfied, for example, such as the maximum quantity of resources to be allocated. Others are *soft* constraints and these can be eventually exceeded, such as the maximum packet delay for NRT services.

A. Load Constraint

All wireless communication systems have a limited quantity of resources for distribution among the users. In turn, each bit rate has a corresponding amount of resources for allocation. This not only depends on the actual bit rate, but also on other factors, such as user location, interference conditions, etc. Obviously, greater bit rates demand more resources. The type of shared resources depends upon the RAT under consideration, e.g., time slots in GSM, transmit power and spreading codes in UMTS or the bit rate in WLAN. It should be noted that all the available resources may not be within the control of the DRA algorithm, as some of these may be reserved for other uses (e.g., handovers, interference level reduction, etc.). Besides, some systems may have more than one load constraint, such as occurs for the downlink in UMTS, which operates with a maximum transmit power and a maximum number of spreading codes.

In general, the system load can be defined as the proportion of resources consumed:

$$\eta = \frac{\text{total allocated resources}}{\text{total available resources}}. \quad (1)$$

Typically the load constraint is a hard constraint. As such, the DRA algorithm must ensure that the load factor never exceeds the corresponding limit. However, in some wireless systems limited by interference, as in the uplink in UMTS, the load constraint is soft and can be temporarily exceeded, albeit slightly.

B. Bit Rate Constraint

In this paper it is assumed that each wireless technology has a set of feasible bit rates, $\Re \subset$. In a similar manner, each user can be characterized by a subset of possible bit rates, $\Re_i \subset \Re$, defined by the type of service subscribed to. One of the purposes of the DRA algorithm is to prevent users from transmitting with a non-permitted bit rate. For example, if R_j is the bit rate allocated to user *i*, then the DRA algorithm must verify that $R_j \in \Re_i$.

C. Packet Delay Constraint

In order to introduce the delay in the resource allocation process, a **minimum target bit rate** has been defined for each user - $R_{\min,i}$ - that *guarantees the transmission of all packets in due time*. The mathematical formulation for $R_{\min,i}$ depends on the procedure for packet delivery. Let us define P as the number of packets to be transmitted, β_p and t_p as the number of bits and the time in the buffer of the p-th packet of the buffer, and $t_{\max}$ as the maximum delay for RT services (or the maximum desirable delay for NRT services). For services

with only one data flow per user, i.e. considering a first in - first out queuing policy, then $R_{\min,i}$ can be calculated as:

$$R_{\min,i} = \max_p \left(\frac{\sum_{b=1}^{p} \beta_b}{t_{\max} - t_p} \right), \quad (2)$$

with $p \in \{1, \cdots, P\}$. If packets are simultaneously transmitted, as occurs in web browsing with the existence of multiple data flows per user (i.e. a single data flow for each web page that is downloading, the whole web page representing one packet), $R_{\min,i}$ can be calculated as:

$$R_{\min,i} \le P \cdot \max_p \left(\frac{\beta_p}{t_{\max} - t_p} \right). \quad (3)$$

Hence, (3) provides a bit rate which is greater than the minimum. This inequality is rebalanced to fit exactly with the minimum bit rate when the packet with the most imminent deadline is the shortest one. In any case, as this minimum target bit rate is used to prioritize users, (3) offers a satisfactory close formulation of this parameter.

Finally, if any packet exceeds its maximum delay (i.e. $t_p > t_{\max}$), the minimum target bit rate is set at infinite ($R_{\min,i} = \infty$).

III. HNN-BASED OPTIMIZATION

The DRA is an NP (Non-deterministic Polynomial time) problem rendering it practically impossible to find an analytical solution for several feasible bit rates and a large number of users. As previously mentioned, the main benefit of using an HNN is the speed of hardware implementation which, by taking advantage of the inherent parallelism of the network, facilitates a real-time running of the algorithm. This section describes the proposed HNN-DC algorithm, including the HNN model, the problem formulation and the dynamics of the neural network.

A. HNN Model

An HNN is comprised of a set of interconnected neurons. Neurons dynamically change their outputs until an equilibrium point is obtained. Hopfield showed that an energy function, *E*, can represent the dynamics of the HNN, and that the problem of finding an equilibrium can be solved by finding a local minimum for the energy function [13], [14].

The dynamics of the HNN can be expressed as [8]:

$$\frac{dU_i}{dt} = -\frac{U_i}{\tau} - \frac{\partial E}{\partial V_i}, \quad (4)$$

where U_i and V_i are the input and output of the *i*-th neuron, and τ is the time constant of the circuit. The relationship between the neuron outputs and inputs is non-linear, and is given by the sigmoidal function:

$$V_i = \frac{1}{1 + e^{-\alpha_i U_i}}, \quad (5)$$

where $\alpha_i > 0$ is the shape parameter of the *i*-th neuron, $U_i \in [-\infty, +\infty]$ and $V_i \in [0,1]$.

The minima of the energy function occur at the 2^L corners inside the *L*-dimensional hypercube defined as $V_i \in [0,1]$, *L* being the total number of neurons [10]. Therefore, any optimization problem involves a definition of a suitable energy function for minimization, since the dynamics of the HNN will ensure that neurons evolve to a minimum energy point (equilibrium state). After reaching a stable state, all neurons are either ON (if the output value is greater than or equal to 0.5) or OFF (where the ouput value is lower than 0.5).

B. DRA Problem Formulation

Once a finite set of feasible bit rates has been defined, the DRA problem can be formulated in terms of a 2D-HNN with *N*·*M* neurons, *N* being the number of active users in the system and *M* the number of feasible bit rates. Users are represented in the first dimension of the neural network (by rows), whereas the second dimension represents the set of possible bit rates (in columns). The neuron states indicate the resource allocation, the neuron with indices (*i*,*j*) being ON if the *i*-th user has been allocated the *j*-th bit rate R_j. It should be noted that the remaining neurons in row *i*, corresponding to user *i*, must be OFF. It is important to avoid confusing the neuron states with the neuron outputs V_{ij}. In this study, a neuron is used to account for the case of zero or non-allocation (i.e. allocated bit rate 0 b/s), in order to prevent saturation situations where there are insufficient resources to allocate the minimum bit rate to all users.

Before describing the energy function, let us define $R_{\max}$ as the maximum bit rate of the wireless system, i.e. $R_{\max} = \max\{R_j,\ R_j \in \Re\}$, and $R_{\max,i}$ as the maximum bit rate of user *i*, i.e. $R_{\max,i} = \max\{R_j,\ R_j \in \Re_i\}$.

The proposed HNN-DC energy function for solving the DRA problem is based on the formulation introduced in [8], using the enhancements proposed in [15] to ensure maximum resource utilization while optimizing the neural network convergence. This work includes an improved first term to deal with the delay constraint:

$$\begin{aligned} E = & -\frac{\mu_1}{2}\sum_{i=1}^{N}\sum_{j=1}^{M} B_{ij}V_{ij} - \frac{\mu_2}{2}\sum_{i=1}^{N}\sum_{j=1}^{M}\frac{R_j}{R_{\max}}V_{ij} + \\ & + \frac{\mu_3}{2}\sum_{i=1}^{N}\sum_{j=1}^{M}\frac{R_j}{R_{\max}}\xi_{ij}V_{ij} + \frac{\mu_4}{2}\sum_{i=1}^{N}\sum_{j=1}^{M}\psi_{ij}V_{ij} + \\ & + \frac{\mu_5}{2}\sum_{i=1}^{N}\sum_{j=1}^{M}V_{ij}\left(1-V_{ij}\right) + \frac{\mu_6}{2}\sum_{i=1}^{N}\left(1-\sum_{j=1}^{M}V_{ij}\right)^2. \end{aligned} \quad (6)$$

The constants μ_1 to μ_6 weight the six terms of the energy function, and their value are selected to obtain a fast convergence of the desired solution (see Appendix A). Next, a detailed description of the different terms is given and their influence on the overall behavior of the algorithm is analyzed.

1) First term of the energy function

This introduces benefit function B_{ij}. This function measures the benefit of allocating each bit rate to each user in terms of delay. Here the benefit function is entirely delay-oriented. As greater bit rates entail shorter delays, it follows that the benefit function should be monotonically increasing. In addition, a great difference in the benefit function must exist between the bit rates capable of transmitting the packets in the time due (i.e., those bit rates greater than the minimum target bit rate explained in Section II.C), and those which are unable to do so. Furthermore, this term should not increase uncontrollably. The sigmoidal function is capable of satisfying the previous conditions:

$$S(x,s,r) = \frac{1}{1 + e^{-s(x+r)}}. \quad (7)$$

The benefit function is defined as:

$$B_{ij} = \frac{S(R_j, s_i, r_i) - S(0, s_i, r_i)}{S(R_{\max}, s_i, r_i) - S(0, s_i, r_i)}. \quad (8)$$

With this definition, the benefit function takes values in the interval [0,1] for all the system bit rates (if $R_j = 0$ then $B_{ij} = 0$, and if $R_j = R_{\max}$ then $B_{ij} = 1$). The s_i and r_i parameters, chosen to increase B_{ij} significantly if $R_j \geq R_{\min,i}$, are:

$$s_i = \begin{cases} \dfrac{2\ln(9)}{R_{\min,i}} & R_{\min,i} \leq R_{\max,i}, \\ \dfrac{2\ln(9)R_{\min,i}}{\left(R_{\max,i}\right)^2} & R_{\min,i} > R_{\max,i}, \end{cases}$$

$$r_i = \begin{cases} -\dfrac{R_{\min,i}}{2} & R_{\min,i} \leq R_{\max,i}, \\ -R_{\max,i} + \dfrac{\left(R_{\max,i}\right)^2}{2R_{\min,i}} & R_{\min,i} > R_{\max,i}. \end{cases} \quad (9)$$

Fig. 1 shows some examples of benefit functions with $R_{\max,i} = 300$ kb/s and gives the different values for $R_{\min,i}$. The figure shows how the sigmoidal function is scaled over the bit rate axis from a step function centered on 0 kb/s (for $R_{\min,i} = 0$ kb/s) to another step function centered on 300 kb/s (for $R_{\min,i} = \infty$ kb/s). In such a manner, when the maximum delay is exceeded and $R_{\min,i} = \infty$, the benefit function takes the values $B_{ij} = 0$ for $R_j < R_{\max,i}$ and $B_{ij} = 1$ for $R_j = R_{\max,i}$. Consequently, the allocation which minimizes the energy function is the maximum bit rate possible. Since the sigmoid is a monotonically increasing function, this term also forces the algorithm to maximize the allocated resources. In short, the benefit function assumes its maximum value ($B_{ij} = 1$) for $R_j = R_{\max}$. However, this term is mainly aimed at *guaranteeing the minimum target bit rate to all users*, and not at maximizing resource utilization. This is due to the fact that once the allocated bit rate surpasses the minimum target bit rate, the benefit function no longer increases significantly. This effect can also be observed in Fig. 1, where different bit rates exhibit very similar benefits for the same $R_{\min,i}$.

2) Second term of the energy function

The second term enforces the HNN-DC algorithm to *maximize the allocated bit rates*, and thus the *total resource utilization*. Neurons are proportionally favored towards the corresponding allocated bit rate.

3) Third term of the energy function

This term *penalizes the allocations that imply an excess of the maximum available system resources*. The third term reduces V_{ij} if, when combined with the current neuron outputs, the allocation of R_j to the *i*-th user requires more resources than the maximum allowed. Consequently, only the combinations of allocations which can satisfy the load constraint introduced in section II.A can act as possible equilibrium points of the HNN.

The amount of resources consumed by the users is calculated, for each user *i* and each bit rate *j*, assuming that the rest of users, $k \neq i$, maintain the resource allocation of the current neuron outputs. When H_{ij} is defined as the load factor of this allocation, then ξ_{ij} is:

$$\xi_{ij} = u\left(\frac{H_{ij}}{\eta_{\max}} - 1\right), \quad (10)$$

where $u(\ )$ is the step function and $\eta_{\max}$ is the maximum load factor of the system. Different mathematical formulations

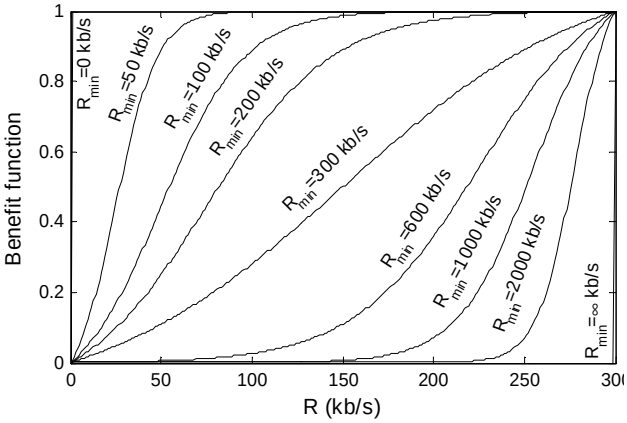

Figure 1. $R_{\min,i}$ effect on the benefit function for $R_{\max} = R_{\max,i} = 300$ kb/s .

for H_{ij} are needed for the various RATs. A specific formulation for GPRS and UMTS is shown in section IV.

4) Fourth term of the energy function

The fourth term *prevents the use of forbidden bit rates*. The $\boldsymbol{\psi}$ matrix represents a permission table describing the feasible bit rate subset for each user, where $\psi_{ij} = 0$ if the *j*-th bit rate is in the subset of user *i*, otherwise $\psi_{ij} = 1$, i.e.:

$$\psi_{ij} = \begin{cases} 0 & R_j \in \Re_i, \\ 1 & R_j \notin \Re_i. \end{cases}$$

Thus, only the bit rates within the user subset can avoid being penalized, and the algorithm can easily cope with the different types and grades of services simply by defining and managing the permission table.

5) Fifth and sixth terms of the energy function

The last two terms were introduced in [8] to ensure a *rapid convergence to correct and stable states of neurons.* The first term forces the neuron outputs to tend towards the extremes 0 and 1. The second term ensures the allocation of only one bit rate per user.

In summary, the first two terms of the energy function are the only ones capable of increasing the value of the neuron outputs. With these first two terms, the QoS is entirely introduced into the energy function, since the first one tries to ensure a minimum bit rate for transmitting all packets within the set time limit and the second one maximizes the total allocated bit rate (system throughput). The third and fourth terms penalize those neurons failing to satisfy the system constraints. In particular, the third term guarantees that the sum of the allocated resources does not exceed the maximum available and the fourth term prevents the use of forbidden bit rates. Finally, the last two terms facilitate system convergence. As only one neuron per user can be active, the neuron that simultaneously satisfies the system constraints, and the first two terms of which display the greater value (more negative), will determine the resource allocated to each user. Hence, striking a balance between delay satisfaction and throughput maximization depends upon the relationship between the first two weights, μ_1 and μ_2.

Finally, the last stage in the design of the HNN is to determine the weighting coefficients (μ_i) of the energy function. This is no easy matter since HNNs present inherent

instability conditions that lead the network to converge to spurious solutions. Nevertheless, by following a well-planned design, an HNN can provide a practical solution. Annex A shows the procedure followed to determine the values of the weighting coefficients.

C. Dynamics of the Hopfield Neural Network

The HNN algorithm begins with all the neurons at approximately 0.5 for the initial state, $V_{ij} = 0.5 + \varepsilon$ where ε is a random variable uniformly distributed in the interval $[-\varepsilon_m, \varepsilon_m]$. Moreover, all the parameters needed for the calculation of B_{ij} and ξ_{ij} must be established in advance, prior to running the neural network. By using these inputs, the neural network is able to reach a stable solution following the dynamics of (4) with no need for any further external interaction. The numerical Euler's technique for solving (4), with $\tau = 1$, in a 2D-HNN is:

$$U_{ij}(t+\Delta t) = U_{ij}(t) + \Delta t\left\{-U_{ij}(t) - \frac{\partial E}{\partial V_{ij}}\right\}, \qquad (11)$$

where Δt is the time interval over which output voltages of neurons are observed and updated. The gradient of the energy function can be calculated as:

$$\frac{\partial E}{\partial V_{ij}} = -\frac{\mu_1}{2}B_{ij} - \frac{\mu_2}{2}\frac{R_j}{R_{\max}} + \frac{\mu_3}{2}\frac{R_j}{R_{\max}}\xi_{ij} + \frac{\mu_4}{2}\psi_{ij} + \frac{\mu_5}{2}(1-2V_{ij}) - \mu_6\left(1-\sum_{l=1}^{M}V_{il}\right). \qquad (12)$$

All the outputs V_{ij} are computed in each iteration using (5) and the solution provided in (11). The equilibrium is reached when the change in neuron output is below a certain tolerance ΔV.

IV. FORMULATION EXAMPLES FOR CONCRETE RATs

The formulation of H_{ij} depends on the particular RAT under study. For example, if a GPRS system with a fixed coding scheme is considered, then it can be calculated as:

$$H_{ij} = \frac{1}{B_T}\left(R_j + \sum_{\substack{k=1\\k\neq i}}^{N}\sum_{l=1}^{M}R_l V_{kl}\right), \qquad (13)$$

where B_T is the maximum total available bit rate, $B_T = R_S N_S$; R_S the bit rate associated with each time slot and N_S the number of available slots for distribution. In this case, the set of feasible bit rates must be a multiple of R_S.

From the power expressions deduced in [16], in downlink UMTS, H_{ij} can be computed as:

$$H_{ij} = \frac{1}{P_{T\max}}\left(\frac{L_{P,i}\dfrac{P_N+\chi_i}{D_{ij}} + \displaystyle\sum_{\substack{k=1\\k\neq i}}^{N}\sum_{l=1}^{M}L_{P,k}\frac{P_N+\chi_k}{D_{kl}}V_{kl}}{1-\dfrac{\rho}{D_{ij}} - \displaystyle\sum_{\substack{k=1\\k\neq i}}^{N}\sum_{j=1}^{M}\frac{\rho}{D_{kl}}V_{kl}}\right), \qquad (14)$$

$$D_{ij} = \rho + \frac{W}{\left(\dfrac{E_b}{N_0}\right)_{ij}R_j}, \qquad (15)$$

where $P_{T\max}$ is the maximum total power available in the base station, $L_{P,i}$ the path loss of the *i*-th user, χ_i the intercell interference observed by the *i*-th user, P_N the thermal noise power, ρ the orthogonality factor ($\rho = 0$ for totally orthogonal codes), *W* the total bandwidth transmission, and $(E_b/N_0)_{ij}$ the target ratio of energy per bit to noise power spectral density specific to each service type and bit rate.

With regard to downlink UMTS, during the scheduling process a consideration of code availability can prove to be of interest. This complex management process can be easily modelled using the correspondence between codes and bit rates. Therefore code management can be transformed into bit rate management. By doing so it is possible to redefine ξ_{ij} as:

$$\xi_{ij} = u\left(\frac{H_{\text{pwr},ij}}{\eta_{\text{pwr},\max}} - 1\right) + u\left(\frac{H_{\text{br},ij}}{\eta_{\text{br},\max}} - 1\right), \qquad (16)$$

where the load factors $H_{\text{pwr},ij}$ and $H_{\text{br},ij}$ are power and bit rate (code)- oriented, defined in (14) and (13) respectively. $\eta_{\text{pwr},\max}$ is the maximum power load factor, and $\eta_{\text{br},\max} = 1$. In the specific case of UMTS B_T represents the maximum total bit rate that can be allocated using the available coding branch.

V. NUMERICAL EVALUATION

In order to evaluate the performance of the HNN-DC algorithm, a downlink UMTS scenario has been considered. For the simulations, video calling has been selected as the RT service, and web browsing and FTP as the interactive and background NRT services respectively. It is worth noting that in this paper the management of voice users is not considered since these are supposed to be served via circuit switching using FRA techniques. The remainder of this section describes the traffic models, the reference DRA algorithms and the simulation scenario. Finally some illustrative results are presented.

A. Traffic Models

The traffic model for the NRT services is an extract from [17]. In particular, for web browsing, a complete modeling of the web page is performed. When a new page is requested, the main object is generated and stored in the buffer pending transmission. After its correct transmission and an additional processing time, the user is able to request the remaining inline objects which are sequentially delivered. The time period between the requests for two consecutive web pages is also modeled. The FTP traffic is similarly implemented, but without inline objects and refers to a different set of statistics.

Regarding the characterization of the RT service, the video calling model is an extract from [18]. This model emulates the real-time H.263 video, employing the VBR H.263 codec which generates instantaneous changes in the output bit rate while maintaining an average constant bit rate, set at 64 kb/s for this study. The model takes into account the three different frame types considered in the H.263 standard, namely I, P and PB. The model describes frame size and duration, the correlation between both parameters for each frame, and the transition probability between different video frames. Modeling is conducted at two levels. The first level establishes the frame type to be generated. I-frames are periodically created, while a Markov chain drives the transition generation between P – and PB – frames. Once the frame type is selected, the model determines the size and duration of the frame to be transmitted. In the H.263 model the traffic source does not wait for the completion of last frame transmission before generating the next one. In this case, the station assumes that the QoS requirements of the video calling service have not been fulfilled, and the older frame is discarded. A video call user is not continuously generating

new data. Hence, it is possible to fail to complete the transmission of a packet and to wait for the next resource allocation period without entailing packet dropping.

Regarding the minimum target bit rate $R_{\min,i}$, for web browsing, this is calculated using (3), assuming that several simultaneous web downloads can exist. For an FTP user, $R_{\min,i}=0$ since, as a background service, its maximum delay is infinite. In contrast to this, $R_{\min,i}$ is calculated using (2) for video calling.

B. Reference DRA Algorithms

The HNN-DC algorithm proposed in this paper is compared to the following five algorithms:

1) Round Robin (RR)

This technique assigns the same priority level to all users. The algorithm creates a list of users to perform a cyclical allocation of the resources. In the first call to the algorithm, the maximum bit rates are allocated, $R_{\max,i}$, to the first n_1 users. n_1 is determined by the power and code restrictions in an attempt to maximize the allocated resources. In the second scheduling period, the algorithm begins with the n_1+1 user and allocates $R_{\max,i}$ to the following n_2. This process is repeated until the end of the list is reached. Then the algorithm returns to the beginning of the list and restarts. In the case of a multi-service scenario, RT users are served before the NRT users. Within this second group, two alternatives have been considered. The first alternative does not differentiate between interactive and background services, and the second option serves interactive users first and only then the background users are served.

2) Weighted Round Robin (WRR)

The WRR is similar to the RR, but with the introduction of weights to prioritize the different services. The weights indicate the bit rate to be allocated to the next user on the list. In this paper, since RT and NRT services are managed separately; the WRR algorithm only differentiates between interactive (web) and background (FTP) users. The weights, and therefore, the bit rates allocated to each type of service are R_{web} for web users and R_{FTP} for FTP users.

3) Optimum Bit Rate (OBR)

This algorithm randomly selects users allocating their $R_{\min,i}$ until either no resources or users remain. Since generally speaking $R_{\min,i} \notin \Re_i$, the algorithm allocates the lowest bit rate greater than $R_{\min,i}$ from the subset $\Re_i$. As for the previous algorithms, RT users are served first and finally NRT users both with or without differentiation.

4) Prioritized Earliest Delay First (PEDF)

This algorithm prioritizes users by deadline. First, the algorithm finds the user with the nearest packet deadline. Subsequently, this user is served with his maximum bit rate $R_{\max,i}$. If $R_{\max,i}$ cannot be reached, then the maximum bit rate possible is allocated. This process of searching and serving is repeated until either no resources or no data pending transmission remain. As before, in the case of multi-service, this process is repeated from the highest priority service, which is, video calling, down to the background service.

5) Descend Bit Rate (DBR)

The DBR algorithm begins the allocation of its maximum allowed bit rate, $R_{\max,i}$ to each user. Then, it reduces the bit rates until the total allocated resources are lower than those available. One user is randomly selected for each iteration. This process is divided into two phases. In the first phase, the algorithm never allocates a bit rate below the target $R_{\min,i}$. If there are insufficient resources to guarantee the $R_{\min,i}$ for all users, then the algorithm enters into the second phase, where the allocated bit rates can be reduced without any limitation. Once again, this algorithm separates RT and NRT services, with the possibility of separately processing interactive and background users.

After the execution of all these algorithms, the resulting allocation is optimized in terms of throughput by including an additional process, the Minimum Noise Rise (MNR). This process increases the bit rates allocated to the users in line with channel quality until either users reach their maximum bit rates or no resources remain. Here the objective is to maximize the total allocated resources as far as possible.

C. Simulation Scenario

The scenario consists of seven cells with a radius of 0.5 km, with the cell under study in the centre. The maximum available power is 43 dBm (20 W), and the maximum power load factor, $\eta_{\text{pwr,max}}$ is set at 0.6 (60%). Conversely, the transmitted power of the interfering cells is 40 dBm (a 50% load factor is considered). The path loss for the *i*-th user is calculated using [19]:

$$L_{P,i}(\text{dB}) = 137.4 + 35.2\log_{10}(d_i), \tag{17}$$

where d_i is the user distance in km to the centre cell. The large-scale fading effect is modeled using the Gudmundson approach [20], assuming a standard deviation of 8 dB. Users are mobile with a constant speed uniformly distributed between 0 and 60 km/h. The thermal noise power level is -102 dBm. The total transmission bandwidth, *W*, is 3.84 Mchips/s. The orthogonality factor, ρ, is set at 0.5. The DRA algorithms are run every 0.1 seconds (scheduling period) [19].

For the B_T calculus, the available number of codes has also been carefully considered .Users are multiplexed over a common Downlink Shared Channel (DSCH), hence, one branch with an SF 256 per user must be reserved for signaling. Finally, five SF 256 are reserved for common and broadcast channels. The remaining available branches establish the total maximum bit rate, B_T, bearing in mind that one SF 8 of the code tree entails a bit rate of 256 kb/s. This calculation is summarized by the following equation:

$$\begin{aligned} N_{FreeCodes\_SF256} &= 256 - N_{Data\_Users} - 5, \\ B_T &= \left\lfloor N_{FreeCodes\_SF256}/32 \right\rfloor \cdot 256 \text{ (kb/s)}. \end{aligned} \tag{18}$$

The set of possible bit rates considered is {256 kb/s, 128 kb/s, 64 kb/s, 32 kb/s, 16 kb/s, 0 kb/s}. The corresponding E_b/N_0 ratios are {5.6 dB, 4.4 dB, 4.62 dB, 4.55 dB, 4.55 dB, -∞ dB} [21]. All bit rates are permitted for web and FTP services, whereas for video calling the maximum bit rate is set at 128 kb/s. The maximum delay desirable is different for the various services: 100 ms for one H.263 frame, 10 seconds for one web object and ∞ for file downloading in FTP. With regard to the WRR algorithm, $R_{\text{web}} = 256$ kb/s and $R_{\text{FTP}} = 64$ kb/s.

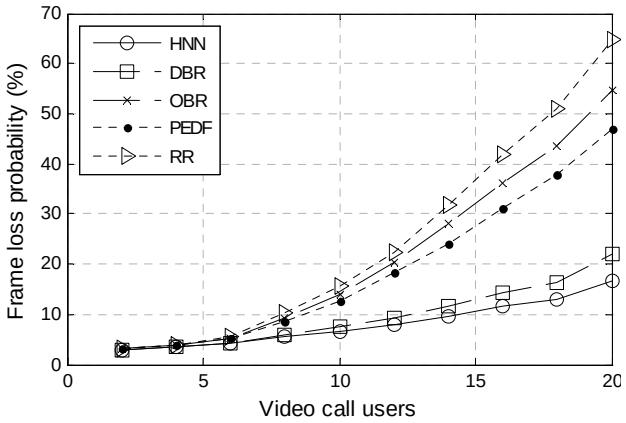

Figure 2. Average frame loss probability with an increasing number of video call users.

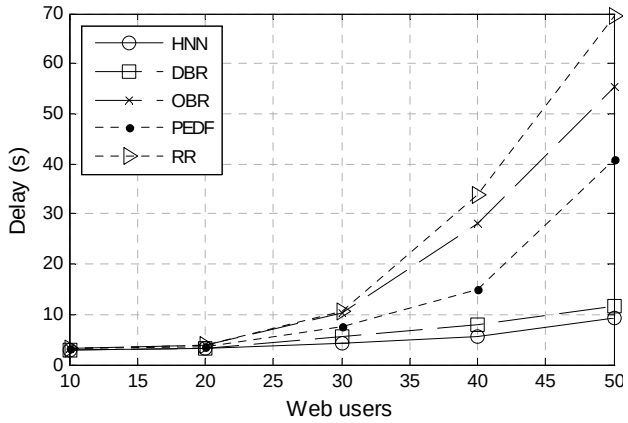

Figure 3. Average web page response time with an increasing number of web browsing users.

By following the rationale outlined in Appendix A, the parameters of the simulated HNN network have been obtained in Appendix B:

$$\mu_1 = 1000 \qquad \mu_2 = 500 \qquad \mu_3 = 17000$$
$$\mu_4 = 11500 \qquad \mu_5 = 15 \qquad \mu_6 = 5000$$

The remaining parameters are selected from [8]:

$$\Delta t = 10^{-4} \qquad \alpha = 1 \qquad \Delta V = 10^{-4} \qquad \tau = 1$$

All results have been obtained averaging over 10 simulations. The simulation time is set at one hour.

D. Simulation Results

The simulations are divided into several stages in order to separately study the QoS delivered by the different services using several scenarios. Firstly, only the RT traffic is considered, studying the performance of video call users. In the second analysis, an increasing number of interactive NRT (web browsing) users are added to a fixed RT traffic load. All the algorithms (including HNN-DC) assign resources to RT users first, and once this has been done resources are then assigned to the NRT users. The performance of the different DRA schemes depends on the strategy for managing the base station transmitted power and the user bit rate allocation. Finally, the third scenario includes background NRT (FTP) users. Two strategies for the reference DRA algorithms are considered. The first strategy involves the DRA algorithms simultaneously handling both interactive and background users. However, in the second strategy, they differentiate between the services, prioritizing the interactive users. In contrast, HNN-DC always allocates resources to both types of service - interactive and background - simultaneously, these being differentiated only by their maximum delay. The results obtained justify this procedure since it makes the maximization of system throughput possible while maintaining the QoS for interactive users.

Throughout the remainder of the section, the improvement in delay achieved by the HNN-DC algorithm is computed as:

$$\text{Delay improvement}(\%) = 100\left(1 - \text{delay}_{\text{HNN-DC}} / \text{delay}_{\text{Ref}}\right). \quad (19)$$

1) Performance with only RT traffic.

The performance of the HNN-DC algorithm when only serving video call users is studied first. Video call users ranging from 2 to 20 are introduced into the cell under study. The mean delay of successfully transmitted frames for video call users is quite similar for all the algorithms and, although the HNN-DC gives the best performance, the improvement is not really significant. However the frame loss probability can vary significantly depending on the algorithm, as shown in Fig. 2. The HNN-DC presents the best performance, improving the frame loss probability of the DBR by up to 25% in the highest load case and by up to 65% when compared to the other algorithms.

2) Performance with RT traffic and an increasing number of web users.

Next, 10, 20, 30, 40 and 50 web users are introduced in the cell under study maintaining 5 video call users. The existence of web users has no effect on the QoS experienced by RT users since both user types are separately handled. As such, the results and conclusions of the previous section remain valid. Fig. 3 represents the average time needed to transmit a web page as a function of the number of web users. The graph reveals that the DBR and the HNN-DC algorithms considerably improve the service response time compared to the other algorithms. Once again, the HNN-DC proves to be the algorithm which performs best. Initially, with only a few users, the behavior of the DRA algorithms is quite similar, but when the number of users increases the HNN-DC can offer a more effective distribution of the available resources. With 50 web users, the HNN-DC improves the performance of the majority of the algorithms by up to 80%. Only the performance of the DBR can approximate that of the HNN-DC, but the HNN-DC still improves the DBR by up to 22%.

Fig. 4 shows the power consumed by the base station, illustrating that optimal performances in web page downloading are due to the aforementioned optimization of the use of available power. Minimization of the consumed power allows the HNN-DC to serve more users satisfying their requirements.

3) Performance in a multi-service traffic scenario.

Finally, the performances of combined background, interactive and RT users are analyzed. Five video call users, 30 web browsing users and also 1, 2, 3, 4 and 5 FTP users are introduced into the system. As previously stated, RT users are managed first and, as such, the NRT users do not influence their service provision.

First of all, the reference DRA algorithms serve FTP and web browsing users simultaneously, only differentiating between these with regard to the different maximum delay, as occurs with HNN-DC. Fig. 5.a depicts the average web page downloading time as the FTP load increases. One of the initial conclusions to be drawn from this study is that, when NRT

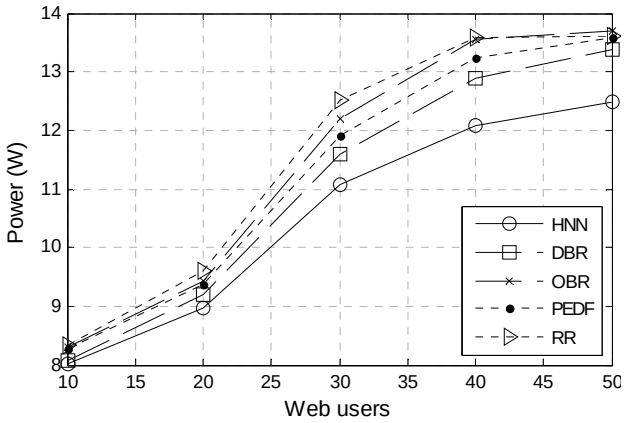

Figure 4. Average Node-B power consumption.

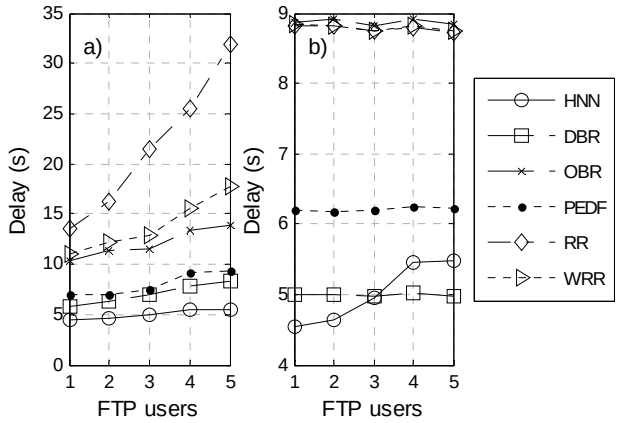

Figure 5. Average web page response time with an increasing number of FTP users.
a) No NRT service differentiation for the reference algorithms.
b) NRT service differentiation for the reference algorithms.

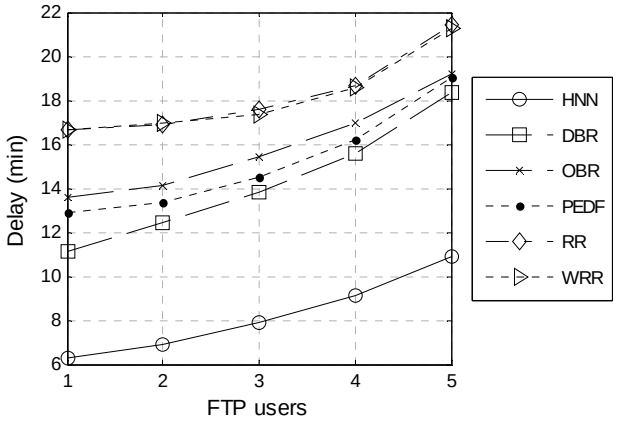

Figure 6. Average response time of FTP service.

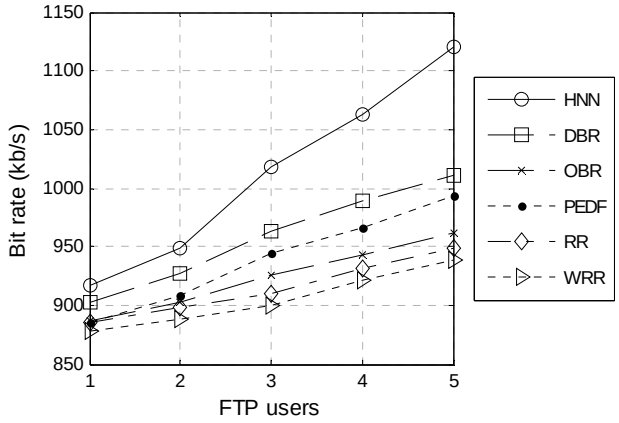

Figure 7. Total allocated bit rate with an increasing number of FTP users.

services are not differentiated, the FTP traffic affects web users producing longer service response times, despite the fact that $R_{\min,i} = 0$ as defined for FTP (i.e. no delay restriction exists). Moreover, the worst performances are obtained by the RR, WRR and OBR algorithms. In the case of both RR and WRR this seems logical, since these are not delay-based algorithms. With Regard to the OBR algorithm, this allocates the minimum target bit rate to all users. This involves assigning 0 kb/s to the FTP users and, in the majority of cases, low bit rates to the web users. Therefore, in general, a significant amount of resources remains after finishing the OBR algorithm. In addition, the MNR process performed immediately afterwards increases the bit rates allocated to the users according to channel quality, albeit without distinguishing among services and actual delay requirements. Consequently, with more FTP users, it is more probable that an FTP user experiences better channel conditions than any other web user, preventing higher bit rates from being allocated to web users, and hence resulting in greater delays. In the particular case of HNN-DC, the web delay presents a negligible increment with the FTP load as compared with all the reference algorithms. Moreover, this is accomplished despite the great load that FTP users introduce in the system.

Next, the same scenario is evaluated but in this case the reference algorithms differentiate between interactive and background users. Fig. 5.b represents the mean web service response time. Note that HNN-DC performance is exactly the same as before, since this algorithm still jointly handles background and interactive users. For the reference DRA algorithms, the average download delay for web page is clearly improved. Service differentiation is reflected in the constant value for web delay obtained by the reference algorithms, while the delay produced by HNN-DC increases slightly with the number of FTP users. In any case, the web service response time obtained with the HNN-DC algorithm does not significantly increase whereas the improvement of the QoS of FTP users is extraordinary, as it can be concluded from the following figures.

Finally, this study ends with the analysis of the FTP service delay. Fig. 6 represents the FTP file downloading time for the reference algorithms with NRT service differentiation and the HNN-DC. It is worth highlighting that the HNN-DC algorithm not only obtains low web download delays, but also considerably improves the FTP service performance. For 5 FTP users, the HNN-DC shows a 40% enhancement when compared with the time required by the other algorithms.

Therefore, for any DRA algorithm, joint allocation improves the performance of background services, while the interactive services are impaired as a result of the service differentiation policy. Nevertheless, any detrimental effect experienced by interactive users is negligible when using the HNN-DC algorithm (note the low increment of web delay with the increasing number of FTP users as shown in Fig. 5. In addition, performance is significantly improved with regard

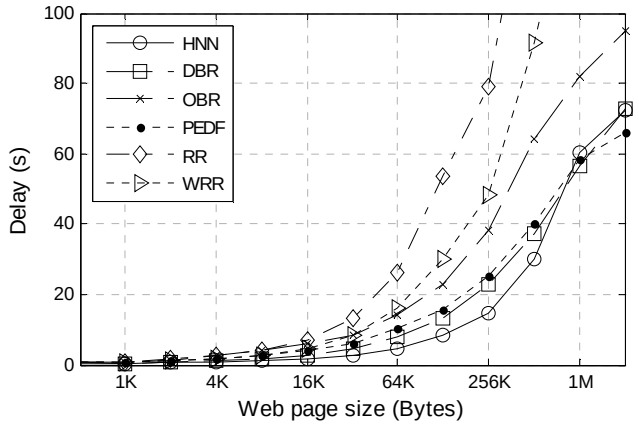

Figure 8. Average service response time vs. web page size.

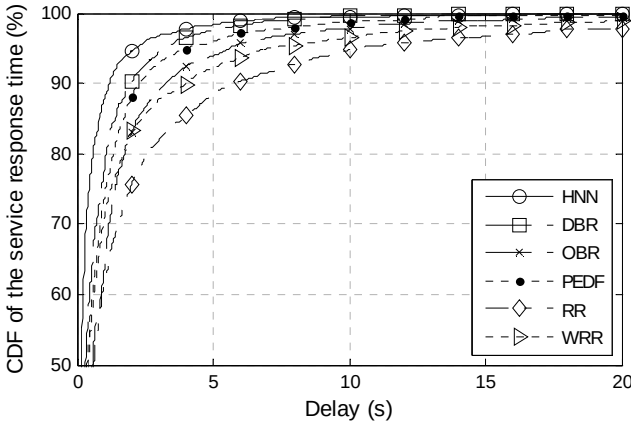

Figure 9. CDF of the service response time for web browsing traffic.

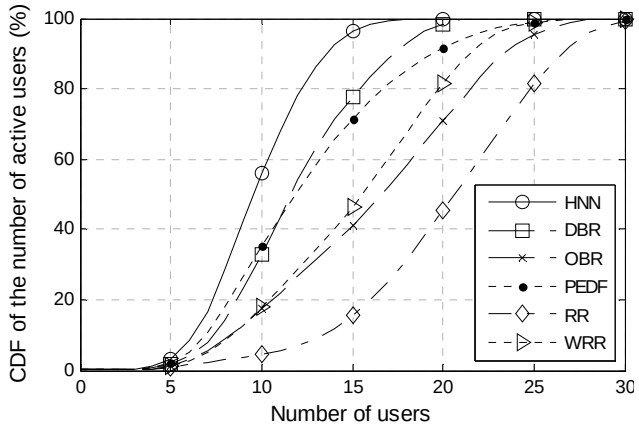

Figure 10. CDF of the number of active users.

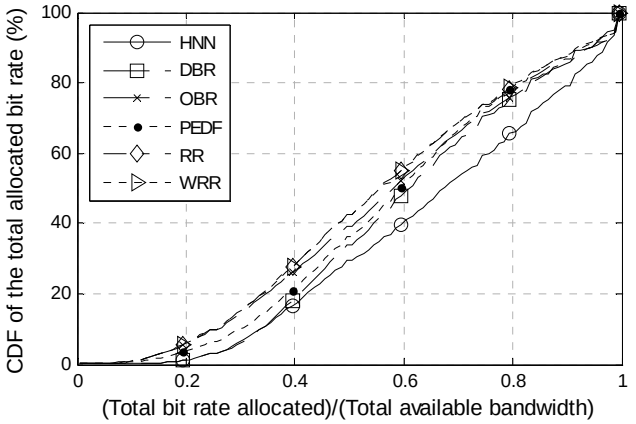

Figure 11. CDF of the total allocated bit rate.

to background users. This results in an improvement in the overall system performance, due to the joint allocation conducted by the HNN-DC algorithm.

Fig. 7 depicts the increment in the average total allocated bit rate as the number of FTP users increases with joint NRT allocation. For HNN-DC, the average allocated bit rate increases faster than witnessed for the other algorithms, growing by approximately 50 kb/s with each new FTP user whereas the reference algorithms cannot reach the 30 kb/s. Besides, the evolution of the total consumed power behaves in the same way as depicted in Fig. 4, i.e. HNN-DC uses less power for bit rate allocation, reflecting the high resource use optimization achieved by this algorithm.

4) Fixed load scenario

After studying the effect of load variation, a scenario with 5 video call users, 30 web users and 3 FTP users without service differentiation is analyzed. Fig. 8 shows the average service response time as a function of the web page download size for web browsing users. Again the HNN-DC algorithm displays the best behavior unless very large web pages are involved, in which case, PEDF and DBR outperform HNN-DC. This is due to the fact that although HNN-DC aims to minimize the delay in the same way PEDF and DBR do, it also maximizes bandwidth usage. The balance between delay satisfaction and throughput maximization depends on the relationship between the first two weights of the energy function, μ_1 and μ_2. Thus, when very large pages are intended for transmission, PEDF and DBR give absolute priority to these web pages, whereas HNN-DC reconsiders this decision by meeting the requirements of the other users. However, since excessively large web pages are unlikely, in most cases HNN-DC provides the best performance, as can be observed in Fig. 5.

Fig. 9 shows the Cumulative Distribution Function (CDF) of the service response time. Once again, it can be seen that the HNN-DC achieves the best performance. Fig. 10 depicts the CDF of the number of active users in the system, which represents the number of users with data pending transmission. Since the HNN-DC has the fastest downloading response time, it also has the lowest number of active users. Fig. 11 represents the CDF of the total bit rate allocated. The RR and WRR algorithms offer the poorest performance with regards to bit rate maximization. Again the HNN-DC provides the best performance, allocating more resources even with fewer users.

VI. CONCLUSIONS

This paper proposes a delay-centric DRA algorithm implemented by means of a Hopfield Neural Network. The HNN-DC algorithm has proven to be an effective resource scheduler for packet data services in a multi-service scenario. Specifically, some illustrative numerical evaluations have been carried out in a downlink UMTS scenario with RT (video calling) and NRT services (web interactive and FTP background).

As a result of the tight delay constraints on RT traffic, RT and NRT services are handled separately. Results have shown that the video calling delay is very similar for all the DRA algorithms, but the frame dropping rate (the main QoS parameter for this service) can be greatly reduced by the HNN-DC algorithm.

After serving RT users, HNN-DC simultaneously distributes the remaining resources among NRT interactive and background users. HNN-DC favors those users with the best channel conditions which allows the power consumption to be reduced. The best power usage, together with an improved throughput maximization, makes the HNN-DC algorithm outperform the other reference algorithms. In this manner, the performance of FTP users is clearly improved, and only a slightly increased delay is incurred for interactive users. Nevertheless, this delay increment does not imply a significant loss in the QoS.

By studying the behavior of the DRA algorithms with and without NRT service differentiation, it can be concluded that joint allocation improves the performance of low priority services at the expense of impairing high priority services. Nevertheless, in the case of the HNN-DC algorithm, the detrimental effect is negligible, while the performance of the low priority services is significantly improved. Hence, the combination of the HNN-DC algorithm and the joint allocation represents the most suitable solution for the DRA process, since the overall system performance is improved.

The strong performance of HNN-DC is also reflected in a reduction in the average number of active users. Therefore, benefits resulting from the greater efficiency in resource distribution are twofold. First of all, this results in faster user transmissions, and secondly, the system load is reduced, thereby permitting either more users to transmit or even faster transmission for current users.

APPENDIX A
THE CALCULATION OF WEIGHTING COEFFICIENTS

In order to obtain the weighting coefficients, the worst cases should be analyzed. For such cases, the chosen weights must ensure the desired behavior of the algorithm. First of all, μ_1 and μ_2 can be selected with certain freedom whereas the remaining weights will depend on these. To correctly select μ_1 and μ_2, it is necessary to decide upon the desired algorithm behavior. If delay satisfaction is more important than throughput maximization, then $\mu_1 > \mu_2$. Furthermore, the greater the difference between these two weights, then the greater the significance of the delay for the algorithm.

A. Fifth term

This term only aims to enhance the convergence speed of the neural network and must not prevent the change in neuron output, from 0 to 1, or vice versa, if the rest of the terms point to this. Let define (i,high) and (i,low) as two neurons belonging to the same user with bit rates R_{high} and R_{low} respectively, $R_{\text{high}} > R_{\text{low}}$, and if neither of these exceeds the maximum resources, the energy gradient of both neurons is:

$$\frac{\partial E}{\partial V_{i,\text{high}}} = -\frac{\mu_1}{2} B_{i,\text{high}} - \frac{\mu_2}{2}\frac{R_{\text{high}}}{R_{\max}} + \frac{\mu_5}{2}\left(1-2V_{i,\text{high}}\right),$$

$$\frac{\partial E}{\partial V_{i,\text{low}}} = -\frac{\mu_1}{2} B_{i,\text{low}} - \frac{\mu_2}{2}\frac{R_{\text{low}}}{R_{\max}} + \frac{\mu_5}{2}\left(1-2V_{i,\text{low}}\right).$$

The optimum allocation is R_{high} since this maximizes the throughput. In the worst case scenario, both bit rates are equally valid for the delay, i.e. $B_{i,\text{high}} = B_{i,\text{low}}$. Assuming that $V_{i,\text{high}} = 0$ and $V_{i,\text{low}} = 1$, to ensure the correct allocation of R_{high}:

$$\frac{\partial E}{\partial V_{i,\text{high}}} < \frac{\partial E}{\partial V_{i,\text{low}}},$$

$$\mu_5 < \frac{\mu_2}{2}\frac{\min\left\{R_{\text{high}} - R_{\text{low}}\right\}}{R_{\max}}.$$

B. Third term

To allocate a bit rate not exceeding the maximum resources, at least one of the correspondent neurons must be favored (either increasing faster or decreasing slower) over the neurons exceeding the maximum resources. Supposing that all bit rates are in the permission table of user i, then in the case of the favored neuron (i,fav), the energy gradient would be:

$$\frac{\partial E}{\partial V_{i,\text{fav}}} = -\frac{\mu_1}{2} B_{i,\text{fav}} - \frac{\mu_2}{2}\frac{R_{\text{fav}}}{R_{\max}} + \frac{\mu_5}{2}\left(1-2V_{i,\text{fav}}\right) - \mu_6\left(1-\sum_l^M V_{il}\right).$$

Whereas the energy gradient of the neurons exceeding the maximum resources (i,exc) would be:

$$\frac{\partial E}{\partial V_{i,\text{exc}}} = -\frac{\mu_1}{2} B_{i,\text{exc}} - \frac{\mu_2}{2}\frac{R_{\text{exc}}}{R_{\max}} + \frac{\mu_3}{2}\frac{R_{\text{exc}}}{R_{\max}} + \frac{\mu_5}{2}\left(1-2V_{i,\text{exc}}\right) - \mu_6\left(1-\sum_l^M V_{il}\right).$$

As such, the condition needed to guarantee the allocation of the correct bit rate is:

$$\frac{\partial E}{\partial V_{i,\text{fav}}} < \frac{\partial E}{\partial V_{i,\text{exc}}},$$

$$\mu_3 > \mu_1\frac{R_{\max}}{R_{\text{exc}}}\left(B_{i,\text{exc}} - B_{i,\text{fav}}\right) + \mu_2\frac{R_{\text{exc}} - R_{\text{fav}}}{R_{\text{exc}}} + 2\mu_5\frac{R_{\max}}{R_{\text{exc}}}\left(V_{i,\text{exc}} - V_{i,\text{fav}}\right).$$

The worst case scenario can be found where $B_{i,\text{exc}} = 1$, $B_{i,\text{fav}} = 0$, $R_{\text{fav}} = 0$, $V_{i,\text{exc}} = 1$ and $V_{i,\text{fav}} = 0$. In this case:

$$\mu_3 > \mu_1\frac{R_{\max}}{R_{\text{exc}}} + \mu_2 + 2\mu_5\frac{R_{\max}}{R_{\text{exc}}}.$$

C. Sixth term

Despite the existence of enough of resources, users should never have more than one bit rate allocated, or in terms of the neural network, more than one neuron ON. The sixth term is minimum when all the neuron outputs of a user sum one. At these points this term and its derivative are zero. As the first two terms continuously increase the neuron outputs and in the event that neither the third nor the fourth term can reduce them, then all neurons begin to increase their value pushing the outputs away from the desired value for the sum of neurons output. Considering δ as the maximum desired distance from the desired sum value, then equilibrium is achieved when $\left|1-\sum_{l=1}^M V_{il}\right| < \delta$. For satisfactory performances, δ should be lower than 1 or even lower than 0.5. With this objective in mind, the following condition needs to be satisfied for the worst case scenario:

$$\left|-\frac{\mu_1}{2} - \frac{\mu_2}{2}\right| < \mu_6\delta,$$

$$\mu_6 > \frac{\mu_1 + \mu_2}{2\delta}.$$

D. Fourth term

This term must decrease the neuron output if $\psi_{ij} = 1$, even if the other terms increase this. The worst case is $B_{ij} = 1$, $R_j = R_{\max}$ and $\xi_{ij} = 0$. Here the energy gradient results in:

$$\frac{\partial E}{\partial V_{ij}} = -\frac{\mu_1}{2} - \frac{\mu_2}{2} + \frac{\mu_4}{2} + \frac{\mu_5}{2}\left(1-2V_{ij}\right) - \mu_6\left(1-\sum_{l=1}^{M} V_{il}\right) > 0,$$

$$\mu_4 > \mu_1 + \mu_2 - \mu_5\left(1-2V_{ij}\right) + 2\mu_6\left(1-\sum_{l=1}^{M} V_{il}\right).$$

Since $\mu_5 < \mu_6$, the worst case for the neuron outputs is $V_{il} = 0, \forall l$. Finally, μ_4 can be obtained as:

$$\mu_4 > \mu_1 + \mu_2 - \mu_5 + 2\mu_6.$$

APPENDIX B
WEIGHTING COEFFICIENTS FOR THE SIMULATIONS

The parameters chosen for the simulation are selected following the description given in Appendix A:

$$\delta = 0.15,$$

$$\mu_1 = 1000,$$

$$\mu_2 = 500,$$

$$\mu_5 < \frac{\mu_2}{2}\frac{\min\left\{R_{\text{high}} - R_{\text{low}}\right\}}{R_{\max}} = \frac{500}{2}\frac{16}{256} \Rightarrow \mu_5 = 15,$$

$$\mu_3 > 1000\frac{256}{16} + 500 + 2\cdot 15\frac{256}{16} \Rightarrow \mu_3 = 17000,$$

$$\mu_6 > \frac{1000+500}{2\cdot 0.15} \Rightarrow \mu_6 = 5000,$$

$$\mu_4 > 1000 + 500 - 15 + 2\cdot 5000 \Rightarrow \mu_4 = 11500,$$

REFERENCES

[1] C. Mihailescu, X. Lagrange, and P. Godlewski, "Performance evaluation of a dynamic resource allocation algorithm for UMTS-TDD systems," in *Proc. IEEE Veh. Technology Conf.*, Tokyo, Japan, 2000.

[2] L. Forkel, T. Kriengchaiyapruk, B. Wegmann, and E. Schulz, "Dynamic channel allocation in UMTS terrestrial radio access TDD systems," in *Proc. IEEE Veh. Technology Conf.*, Rhodes, Greece, 2001.

[3] A. Hernández, A. Valdovinos, and F. Casadevall, "Scheduling with Quality of Service Constraints for real-time and non-real-time traffic in WCDMA," in *Proc. Int. Symp. Wireless Personal Multimedia Communications*, Aalborg, Denmark, 2001.

[4] M. R. Sherif, I. W. Habib, M. Nagshineh, and P. Kermani, "Adaptive allocation of resources and call admission control for wireless ATM using genetic algorithms," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 2, pp. 268–282, Feb. 2000.

[5] S. Papavassiliou, A. Puliafito, O. Tomarchio and J. Ye, "Integration of mobile agents and genetic algorithms for efficient dynamic network resource allocation," *Proc. IEEE Symp. Comput. Commun.*, Hammamet, Tunisia, 2001.

[6] E. Del Re, R. Fantacci and L. Ronga, "A dynamic channel allocation technique based on Hopfield neural networks," IEEE *Trans. Veh. Technology*, vol. 45, no. 1, pp. 26-32, Feb. 1996.

[7] O. Lázaro and D. Girma, "A Hopfield neural-network-based dynamic channel allocation with handoff channel reservation control", *IEEE Trans. Veh. Technology*, vol. 49, no. 5, pp. 1578-1587 , Sept. 2000.

[8] C. W. Ahn and R. S. Ramakrishna, "QoS provisioning dynamic connection-admission control for multimedia wireless networks using a Hopfield neural network," *IEEE Trans. Veh. Technology*, vol. 53, no. 1, pp. 106-117, Jan. 2004.

[9] N. García, R. Agustí, and J. Pérez-Romero, "A user-centric approach for dynamic resource allocation in CDMA systems based on Hopfield neural networks," *Proc. IST Summit*, Dresden, Germany, 2005.

[10] J. J. Hopfield and D. W. Tank, "Neural computation of decisions in optimization problems," *Biological Cybern.*, vol. 52, pp. 141-152, 1985.

[11] K. C. Tan, H. Tang, and S. S. Ge, "On parameter settings of Hopfield networks applied to traveling salesman problems," *IEEE Trans. Circuits Syst.*, vol. 52, no. 5, May 2005.

[12] T.-N. Le and C.-K. Pham, "A new N-parallel updating method of the Hopfield type neural network for N-queens problem," *Proc. IEEE Int. Joint Conf. Neural Netw.*, Montreal, Canada, 2005.

[13] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proc. Natl. Acad. Sci.*, vol. 79, pp. 2554-2558, April 1982.

[14] J. J. Hopfield, "Neurons with graded response have collective computational properties like those of two-state neurons," *Proc. Natl. Acad. Sci.*, vol. 81, pp. 3088-3092, May 1984.

[15] D. Calabuig, J. F. Monserrat, D. Gómez-Barquero, and O. Lázaro, "User bandwidth usage-driven HNN neuron excitation method for maximum resource utilization within packet-switched communication networks," *IEEE Commun. Letters*, vol. 10, no. 11, pp. 766-768, Nov. 2006.

[16] J. Perez-Romero, O. Sallent, R. Agusti, and G. Pares, "A downlink admission control algorithm for UTRA-FDD," *Proc. IEEE Mobile Wireless Commun. Netw.*, Stockholm, Sweden, 2002.

[17] 3GPP2-TSGC5, "HTTP and FTP Traffic Model for 1xEV-DV Simulations."

[18] O. Lázaro, D. Girma, and J. Dunlop, "H.263 video traffic modelling for low bit rate wireless communications", *Proc. IEEE Personal Indoor Mobile Radio Commun.*, Barcelona, Spain, 2004.

[19] H. Holma and A. Toskala, "WCDMA for UMTS," John Wiley & Sons, 3rd edition, 2004.

[20] M. Gudmundson, "Correlation model for shadow fading in mobile radio systems", *Electron. Letters*, vol. 27, pp. 2145–2146, Nov. 1991.

[21] European Project IST-2000-25133, "Advanced radio resource management for wireless services (ARROWS)".

[22] 3GPP TS 23.107 v 5.9.0, "QoS Concept and Architecture."