

A DRA scheme based on Hopfield Neural Networks Methodology

N. García,
Universitat Pompeu Fabra (UPF)
Barcelona, Spain, E-mail : nuria.garcia@upf.edu

J. Pérez-Romero
Universitat Politècnica de Catalunya (UPC)
Barcelona, Spain, E-mail : jorperez@tsc.upc.edu

Abstract. This paper proposes a dynamic resource allocation (DRA) algorithm that makes use of the Hopfield Neural Network (HNN) methodology, which provides a fast way of finding the optimum resource when formulated as combinational problem. The proposed algorithm is applied to both schedule the downlink transmissions in a CDMA scenario with delay-oriented services, and real time services, although by a proper modification of the constraints imposed in the energy function, it could be easily extended to other services or access technologies. The algorithm is evaluated by means of simulations, revealing its ability to adapt to the specific service and traffic conditions. Finally, the convergence issues related to HNN methodology are considered in the light of the obtained results.

I. INTRODUCTION

The Dynamic Resource Allocation (DRA) problem in multi-user and multi-service wireless environments can be considered as a combinational problem where radio resources (e.g. bandwidth) have to be allocated to several users in each frame subject to certain restrictions in terms of Quality of Service (QoS) and of total amount of available resources. In that sense, it is known that a Hopfield Neural Network (HNN) searches for the combinational solution that minimizes an Energy Function under specific constraints [1]. In addition to that, it is claimed that real time operation at a very low time scale could be provided either by means of Hardware HNN implementations or by means of numerical algorithms. In [2] a HNN is proposed to manage the multirate and multiservice structure of an ATM wireless network. In turn, [3] shows the good HNN behavior when managing different service profiles for a user centric approach in the uplink of a CDMA system, in which each user has a range of bit rates that can be allocated depending on its service profile, defined by a minimum satisfaction bit rate. Consequently, the use of HNN can be regarded as a firm candidate for implementing the DRA problem in different scenarios whenever a high number of possible combinations of bandwidth allocations to the users exist, thanks to the ability of HNN to find the optimum combination that fits in the specific constraints. Also the great deal of possible frequency-time slots combinations in e.g. OFDMA systems presents a very challenging dynamic resource allocation problem in a multi user environment that could be faced with HNN.

Taking into account the above framework, in this paper the basis of the HNN is captured for a downlink CDMA system where both real time and delay-sensitive services are to be provided to the users sharing the access. With respect to the real time services, the user requirements are given in terms of a

specific bit rate contracted with the mobile operator. In that sense, the main overall goal is that the system can continue providing an optimized throughput, while at the same time delivering to each real time user a bit rate at least equal to the satisfactory bit rate specified in his contract. In turn, with respect to delay-sensitive services, the main overall goal is not necessarily to achieve the maximum throughput nor to guarantee a certain bit rate, but to deliver each packet without exceeding a specific time deadline. This can be applicable usually to interactive data services. The optimum solution for the DRA problem in this scenario is found by solving a nonlinear constrained optimization problem by means of HNN. Unlike most of the previous and current research in wireless scheduling, which focus on the scheduling issue of one single server, our downlink CDMA HNN-based approach corresponds to a multiple server case where multiple concurrent packets can be transmitted by the base station simultaneously and where there is not much related work available [4]. Finally the convergence issues related to the HNN have been identified and properly circumvented in the DRA algorithm introduced.

The rest of the paper is organised as follows. Section II presents the DRA problem formulation. Section III presents the proposed HNN model and discusses the convergence issues. Section IV provides some results obtained through simulation and finally conclusions are summarised in Section V.

II. PROBLEM FORMULATION

The considered DRA problem assumes a set of N_r real time (RT) users and N_d delay sensitive users with their corresponding queues located at the base station of the access network, operating with CDMA. It is considered that non-shaped traffic is arriving to the queues so that all the incurred packet delay is introduced at the network level. The scheduling algorithm operates in frames of T (s) and allocates a certain bit rate to each user from a set of M possible bit rates, $j=1, \dots, M$. Multiple transmissions of different users in parallel are allowed by making use of multi-code transmissions. Similarly, variable bit rates are achieved by varying the spreading factor. The bit rate allocation is executed by means of an optimal mechanism based on HNN through the minimisation of a properly defined Energy Function, which includes a cost function per each service associated to each Bit Rate usage and a proper formulation for the downlink network restrictions.

In particular, in the downlink of CDMA, the total available power at the base station is shared by all the transmissions. Then, as shown in [5], the power devoted to the i -th user

assuming that the base station is transmitting at maximum power P_{Tmax} , i.e. assuming that the base station operates at the capacity limit, is given by:

$$P_{Ti} \approx \frac{\rho}{W/R_{b,i}} \frac{P_{T,max}}{\left(\frac{E_b}{N_o}\right)_i} \quad (1)$$

where W is the transmission bandwidth after spreading, $R_{b,i}$ is the bit rate of the i -th user, $(E_b/N_o)_i$ is the target requirement of the i -th user to assure a certain bit error rate and ρ is the orthogonality factor. A downlink power control is considered as usual in operative CDMA systems to achieve the target (E_b/N_o) requirements. The approximations given in (1) assume, for the sake of simplicity, a single isolated cell and that background noise is negligible with respect to intracell interference. However, the presented formulation could be easily extended to consider the multi-cell case. To properly formulate the proposed DRA algorithm, first some constraints and a Cost functions are introduced in the following.

The sum of the powers devoted to the $N=N_r+N_d$ users must be below a fraction β of P_{Tmax} (the remaining fraction $1-\beta$ is usually devoted to common control channels). Then, the power constraint can be translated into a bit rate constraint given by:

$$\sum_{i=1}^N R_{b,i} \leq \frac{\beta W}{\rho \left(\frac{E_b}{N_o}\right)} = B_T \quad (2)$$

where it has been assumed for simplicity the same E_b/N_o for all the users. Then, the efficient exploitation of the total available bandwidth B_T in terms of the allocated bit rate $R_{b,i}$ to the i -th user can be obtained by trying to fulfil (2) with equality.

The Cost function for the real time users is the one proposed in [3], which depends on the desired maximum non-satisfaction probability (defined as the probability of receiving a bit rate below the satisfaction bit rate specified in the contract). In turn, for the delay-sensitive users a dynamic cost function $C_{i,j}^k$ associated to the i -th user in the k -th frame and for the j -th possible bit rate to be allocated $R_{b,i,j}^k$, is defined as:

$$C_{i,j}^k = \begin{cases} 1 & \text{if } R_{b,i,j}^k < R_{b,i,opt}^k (1-a) \\ \frac{R_{b,i,j}^k TO_i^k - B_{i,left}^k (1-a)}{B_{i,left}^k (1+a) - B_{i,left}^k (1-a)} & \text{if } R_{b,i,opt}^k (1-a) \leq R_{b,i,j}^k \leq R_{b,i,opt}^k (1+a) \\ 0 & \text{if } R_{b,i,opt}^k (1+a) < R_{b,i,j}^k \end{cases} \quad (3)$$

where TO_i^k and $B_{i,left}^k$ are the remaining time-out and the number of bits to be transmitted, respectively, corresponding to the most restrictive packet in the queue of the i -th user. In turn, a is a parameter to be set.

The cost functions and the system constraints are included in the HNN Energy function, defined by [2][3]:

$$E = \frac{\mu_1}{2} \sum_{i=1}^N \sum_{j=1}^M C_{i,j}^k V_{i,j} + \frac{\eta^c \mu_2}{2} \left| 1 - \sum_{i=1}^N \sum_{j=1}^M \frac{R_{b,i,j}}{B_T} V_{i,j} \right| + \frac{\mu_3}{2} \sum_{i=1}^N \sum_{j=1}^M \psi_{i,j} V_{i,j} + \frac{\mu_4}{2} \sum_{i=1}^N \sum_{j=1}^M V_{i,j} (1-V_{i,j}) + \frac{\mu_5}{2} \sum_{i=1}^N \left(1 - \sum_{j=1}^M V_{i,j} \right)^2 \quad (4)$$

The minimisation of this energy function provides the allocation of the i -th user given by $V_{i,j}$, which takes the value 1 for the j -th bit rate (i.e. $R_{b,i,j}$) and 0 for the rest of bit rates. The system tries to ensure (2) by means of the second summand in (4), which includes the following exponent to penalize the undesired situations in which the total allocated bandwidth is above B_T , and, at the same time, drives (2) to the equality:

$$\zeta = u \left(\sum_{i=1}^N \sum_{j=1}^M \frac{R_{b,i,j}}{B_T} V_{i,j} - 1 \right) \quad (5)$$

where $u(\cdot)$ is the unit step function.

The third summand in (4) simply penalizes non allowed bit rates according to the contractual user profile. That is, for the allowed bit rates $R_{b,i,j}^k$ of the i -th user, $\psi_{i,j}=0$, while for the non allowed bit rates $\psi_{i,j}=1$.

III HNN MODEL AND CONVERGENCE ISSUES

In a HNN each neuron is modelled as a nonlinear device with a sigmoid monotonically increasing function f_i defined by the logistic function [6]:

$$V_i = f_i(U_i) = \frac{1}{1 + e^{-\alpha_i U_i}} \quad (6)$$

where U_i is the input of the i -th neuron, V_i is the output of the i -th neuron and α_i is the gain of the amplifier of the i -th neuron.

Each neuron receives resistive connections from other neurons and these connections are fully described by the interconnection matrix $T = [T_{i,j}]$. Here $T_{i,j}$ is the interconnection weight from the j -th neuron to the i -th neuron. Each neuron also receives an input bias current I_i that is an adjustable parameter. The dynamics of the HNN are represented by

$$\frac{dU_i}{dt} = \sum_{k=1}^L T_{ik} V_k - \frac{U_i}{\tau} + I_i \quad (7)$$

where τ is the time constant of the circuit and L the number of neurons.

Hopfield provided [7] that for a symmetric interconnection matrix $T_{i,j}$ and sufficiently high gain of the amplifiers, neurons evolve by gradient descent of the quadratic energy functions as follows:

$$E = -\frac{1}{2} \sum_{i=1}^L \sum_{k=1}^L T_{ik} V_i V_k - \sum_{i=1}^L I_i V_i \quad (8)$$

then, from the above expression (7), it results:

$$\frac{dU_i}{dt} = -\frac{U_i}{\tau} - \frac{\partial E}{\partial V_i} \quad (9)$$

The minima of the Energy occur at the 2L corners inside the L-dimensional hypercube defined on $V_i \in \{0,1\}$ [7].

The above equations (8) and (9) can be rewritten in our two dimensional DRA problem by substituting T_{ik} , V_i and V_j by $T_{ij,ki}$, V_{ij} and U_{ij} respectively. Then, the above equation (7) can be rewritten for a two dimensional HNN deployment [2] as

$$\frac{dU_{ij}}{dt} = -\frac{U_{ij}}{\tau} - \frac{\partial E}{\partial V_{ij}} \quad (10)$$

In order to formulate the precise temporal algorithm making HNN to evolve along a trajectory over which the energy function decreases monotonically, for neuron outputs and inputs given by V_{ij} and U_{ij} respectively, a numerical iterative solution for (10) is accomplished following the Euler technique as:

$$U_{ij}(n+1) = U_{ij}(n) + \Delta \left[-\frac{U_{ij}(n)}{\tau} - \frac{\partial E}{\partial V_{ij}} \right] \quad (11)$$

The neuron's voltage is updated at the n -th iteration with the setting $\Delta=10^{-4}$. After reaching a stable state each neuron is either ON (1 if V_{ij} greater or equal than 0.5) or OFF (0, if V_{ij} lower than 0.5).

The iterative numerical solution in (11) converges when the iterations n and $n-1$ satisfy

$$\|V^n - V^{n-1}\|_2 < \varepsilon \quad (12)$$

being $\| \cdot \|_2$ the Euclidean Norm, provided that only one element of each row of the V matrix is greater than 0.5. We have set $\varepsilon=5 \cdot 10^{-6}$. If all these conditions are fulfilled we decide the process converges and the matrix V_{ij} provides us the $R_{b,ij}$ values.

It is worth mentioning that, due to the constraints in the considered DRA problem, the 2nd summand in the expression (4) actually violates the above Energy expression formulated in (8) and, consequently, to the author's knowledge, the convergence would not be mathematically guaranteed. Keeping this in mind, and from a practical point of view, a maximum number of iterations $N_{\max}=1000$ is considered in the iterative solution of (11). If this maximum is reached it is assumed that the procedure has not converged and the $R_{b,ij}$ values are decided to be the same present in the last frame. Under this strategy only a 0.005% of the frames simulated in the results presented in the next Section failed to converge.

IV. RESULTS

This section presents some illustrative results obtained with the proposed allocation strategy. The CDMA wireless network scenario consists of one isolated circular cell with radius 0.5 km. Users are randomly distributed. The transmission bandwidth is $W=3.84$ Mchips/s, the orthogonally factor is $\rho=0.4$ and $\beta=0.95$. The maximum power available at the base station is 43 dBm. Furthermore, the E_b/N_0 target is 5 dB which

leads to an available bandwidth $B_T=2.88$ Mb/s. The available set of possible bit rates is given by: {384 kb/s, 256 kb/s, 128 kb/s, 64 kb/s, 32 kb/s and 16 kb/s}. The selected parameters appearing in the formulation of the HNN are $\mu_1=1000$, $\mu_2=4000$, $\mu_3=8000$, $\mu_4=800$, $\mu_5=6000$, $\tau=1$, $\alpha=1.0$ and $\eta=10$. These values have been set based on [2]. The frame period T is set to 10 ms and the parameter a in the cost function (3) takes the value 0.1. An interactive service has been considered for simulation purposes as representative of delay-sensitive users with the www traffic model given in [8]. Similarly, for real time users, sessions are generated according to a Poisson arrival process with exponential call duration. Two delay-sensitive user classes, namely Class 1 and Class 2, have been considered, as representative of two different user profiles, with maximum allowed delays of 120 ms and 60ms, respectively. In turn, for the real time services, two user profiles are considered: Class 1 users have a maximum bit rate of 256 kb/s and class 2 users of 64 kb/s. In turn, the satisfaction bit rate is 192kb/s for class 1 and 64 kb/s for class 2. The maximum non-satisfaction probability for RT users is 0.2.

The results presented here consider two different case studies:

- case study 1: only www users exist in the scenario. The user profile mix is 60% of class 1 and 40% of class 2.
- case study 2: a variable number of www users coexist with 10 RT users. For both services, the profile mix is 60% class 1 and 40% class 2.

Fig. 1 plots the Cumulative Distribution Function (CDF) of the packet delay in case study 1 for the two classes of users when the total number of www users is 600. The results are compared against a reference scheduling scheme (RSS) in which users are served at the bit rate that allows keeping the delay constraint without taking into consideration the status of the rest of users. It can be noticed that the HNN approach significantly improves RSS, exhibiting delay values much lower than the maximum delay bound for each traffic class.

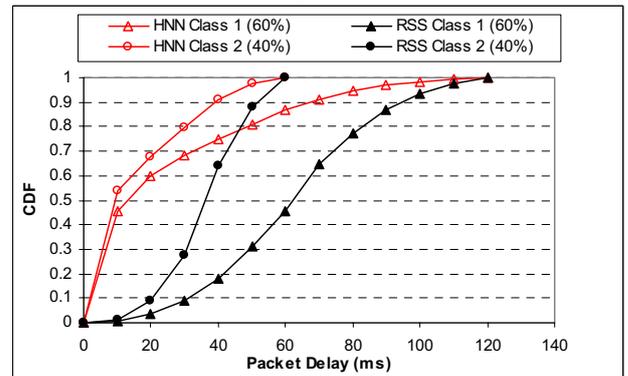


Fig. 1 Cumulative Distribution Function of the packet delay of www users in case study 1, with 600 www users in the scenario

Fig. 2 plots the packet dropping ratio (i.e. the ratio of discarded packets due to the expiration of their deadline) for delay-sensitive users in the two case studies. Clearly, class 1 users, having a higher delay bound, exhibit a lower dropping rate than class 2 users. In turn, the inclusion of RT users in the scenario (i.e. case study 2), introduces a certain increase in the dropping

ratio of both classes because the algorithm reduces the bit rate allocated to www users. This can be also observed in Fig. 3, which plots the histogram of the allocated bit rates to the different services in the two considered case studies. Notice that in case 1, most of the www users receive a bit rate equal to 384 kb/s, while in case 2 the number of allocations of this bit rate are reduced. Similarly, notice in case 2 that the algorithm is able to keep the bit rate of both RT user classes above the satisfaction bit rate.

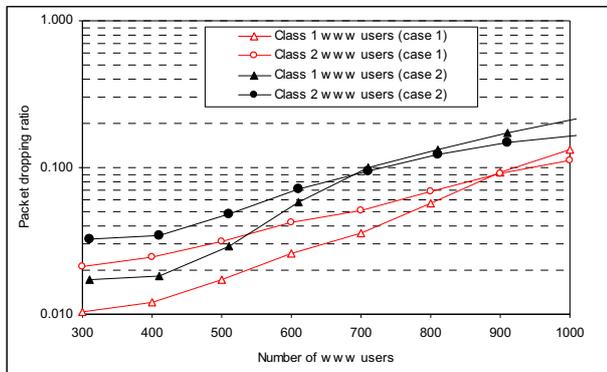


Fig. 2 Packet dropping ratio for delay-sensitive for the two considered case studies

Regarding the performance of RT users, Fig. 4 plots the satisfaction probability for the two user classes when increasing the number of RT users in the scenario and there are a total of 300 www users. The profile mix is 60% users of class 1 and 40% of class 2. It can be observed that, for the considered load levels, the algorithm is able to keep the satisfaction probability above the limit of 0.8 (i.e. corresponding to the maximum non-satisfaction probability 0.2) for both user classes. Notice also that the satisfaction is higher for class 1 users than for class 2 users because for the latter the satisfaction bit rate is equal to the maximum possible bit rate, so the algorithm has fewer combinations that fulfil the satisfaction rate.

V. CONCLUSIONS

This paper has proposed a novel DRA strategy based on Hopfield Neural Networks. It has been applied in the downlink of a CDMA system with both real time and delay-oriented services. The different requirements of each service have been included in the formulation of the cost function. Results have shown that the proposed strategy is able to adapt to the specific traffic conditions and service requirements.

ACKNOWLEDGEMENTS

This work is partially funded by the European Network of Excellence NEWCOM (Contract number 507325), and by the COSMOS grant (ref. TEC2004-00518, Spanish Ministry of Science and Education and European Regional Development Fund).

REFERENCES

[1] Shigeo Abe, *Neural Networks and Fuzzy Systems. Theory and Applications*, Kluwer Academic Publishers, 1997
 [2] C.W.Ahan, R.S.Ramakrishna, "QoS Provisioning Dynamic Connection-Admission Control for Multimedia Wireless Networks Using Hopfield Neural Networks", *IEEE Transactions on Vehicular Technology*, Vol 53, Jan 2004

[3] N. García, R. Agustí, J. Pérez-Romero, "A User-Centric Approach for Dynamic Resource Allocation in CDMA systems based on Hopfield Neural Networks", *IST Summit Dresden June 2005*
 [4] Y.Cao, V.O.K. Li "Scheduling Algorithms in Broadband Wireless Networks", *Proceedings of IEEE*, January 2001
 [5] J. Pérez-Romero, O.Sallent, R.Agustí, M.A.Díaz-Guerra, *Radio Resource Management Strategies in UMTS*, John Wiley & Sons, 2005.
 [6] S.Haykin, *Neural Networks. A Comprehensive Foundation*, New York: Mac Millan, 1994
 [7] J. Hopfield, "Networks with graded response have collective computational properties like those of two state neurons", *Proc. Nat. Acad. Sci.*, vol 81, May 1984
 [8] UMTS 30.03 v3.2.0 TR 101 112 "Selection procedures for the choice of radio transmission technologies of the UMTS", ETSI, April, 1998.

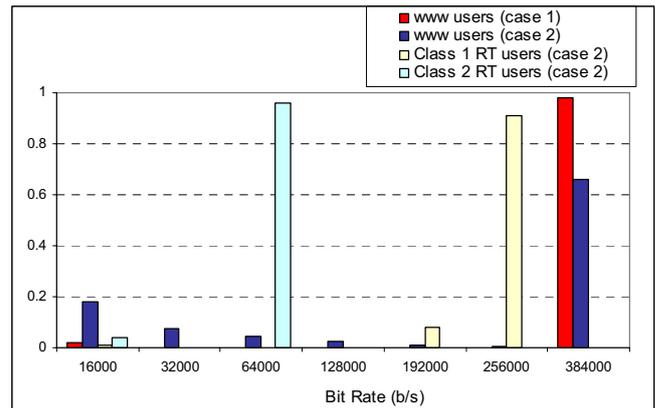


Fig. 3 Histogram of the utilisation of the different bit rates for www and RT users in the two considered case studies with a total of 300 www users in the scenario

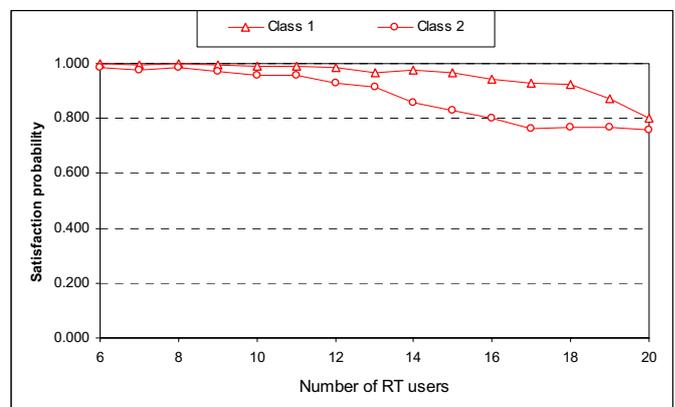


Fig. 4 Satisfaction probability of RT users as a function of the number of RT users in a scenario with 300 www users