

Hopfield Neural Network – based Approach for Joint Dynamic Resource Allocation in Heterogeneous Wireless Networks

David Gómez-Barquero*, Daniel Calabuig*, José Monserrat*,
Nuria García**, and Jordi Pérez-Romero***

**Mobile Communications Group – Polytechnic University of Valencia, Spain*

***University Pompeu Fabra, Spain*

****Radio Communication Research Group – Polytechnic University of Catalonia, Spain*

{dagobar, dacaso, jomondel}@iteam.upv.es, **nuria.garcia@upf.edu, ***jorperez@tsc.upc.edu

Abstract—This paper presents a comprehensive approach to solve the problem of Joint Dynamic Resource Allocation (JDRA) in heterogeneous wireless networks using a Hopfield Neural Network (HNN). A generic formulation for packet services with delay constraints is proposed to decide the optimal bit rate and Radio Access Technology (RAT) allocation. Some illustrative simulation results in a basic scenario are presented to evaluate performance of the proposed algorithm.

I. INTRODUCTION

One of the main challenges of future wireless telecommunications systems will be the ability to provide high bit rate multimedia services with Quality of Service (QoS) guarantees across heterogeneous wireless networks. Different Radio Access Technologies (RATs) will co-exist and will have to interwork in an optimum way, with the objective of providing the end users with the requested services and corresponding QoS requirements [1].

The provision of services in heterogeneous wireless networks is conceptually a very attractive notion. The fundamental goal is to make the heterogeneous network transparent to the users, combining all available RATs into a single system, being possible to deliver the services through the most suitable network (“*Always Best Connected*” paradigm [2]). A key issue in wireless heterogeneous networks is how QoS can be provisioned and managed in a flexible and affordable way over different RATs [3].

Radio Resource Management (RRM) plays a major role in QoS provisioning for wireless communication systems. RRM techniques control the amount of resources (frequency and/or time channels, codes, transmitted power, etc.) assigned to each user subject to some QoS constraints (bit rates, error probabilities, delays, etc.) for a given objective.

of RRM can be maximization network throughput, number of satisfied users, etc.

Multi Radio Resource Management (MRRM) is a relatively new concept of coordinating in a unified manner the radio resources over a set of different RATs (either belonging to the same operator, or to several operators with collaboration agreements) [4]. MRRM fulfils a key role in heterogeneous wireless networks for providing services with improved capacity, coverage and quality. MRRM functionalities will depend on the network architecture and the coupling scheme. For very tight coupling schemes, MRRM and local RRM functionalities may tend to merge into a single unit, being possible to perform a joint admission and congestion control, and a joint packet scheduling [5].

In this paper the functionality that decides the most suitable bit rate and RAT for each user is called *Joint Dynamic Resource Allocation* (JDRA). JDRA will lead to significant benefits in terms of efficiency in the resource utilization. In particular, JDRA will play a crucial role maximizing the number of simultaneous packet-switched connections, and consequently the overall system capacity.

The problem of dynamic resource allocation within a single system is a well known topic in the literature. However, not many approaches to the MRRM problem aiming at finding the optimal resource allocation can be found so far. Very few specific algorithms have been published to evaluate the potential benefits of different JDRA strategies even in basic scenarios. Some illustrative examples are [6]–[9].

This paper presents a comprehensive approach to solve the JDRA problem in heterogeneous wireless networks for packet services with delay constraints using a Hopfield Neural Network (HNN). HNN are considered very good candidates to design dynamic allocation algorithms, since they can provide feasible solutions to very complex optimization problems within a very short time [10]–[12]. HNN are an efficient approach to solve the JDRA problem, as HNN allow for a hardware implementation that can work in real-time.

© 2006 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

¹ Supported by a Ph.D. scholarship from Generalitat de Valencia (Spain).

Moreover, HNN are recurrent networks that operate in an unsupervised mode, requiring no training.

The rest of the paper is organized as follows: Section II introduces the JDRA problem. Section III describes the HNN-based approach, considering a generic formulation for a heterogeneous wireless network. Section IV presents the scenario used for numerical evaluations, the reference JDRA algorithm used for comparison purposes, and shows some relevant results. Conclusions and future work are summarized in Section V.

II. JDRA PROBLEM DEFINITION

The JDRA entity manages dynamically the allocation and de-allocation of radio resources in a heterogeneous wireless network. By considering all resources available in all RATs as a whole, a more efficient utilization can be obtained. The objective of the JDRA entity is to select for each user the optimum RAT and radio resources allocated, subject to certain restrictions in terms of total available resources (that might vary over time, as e.g., the capacity not used by real-time connections), QoS requirements (distinct for each service and user profile), coverage constraints, etc., as shown in Fig. 1. JDRA algorithms are executed every time a new user enters the system (after being accepted by the joint admission control algorithm), and during the users sessions.

It should be pointed out that in heterogeneous wireless networks not all the RATs have necessarily the same coverage area (e.g., WLANs in hot spots), and not all users might be able to connect to all RATs (e.g., not all terminals with multi-mode capabilities). Also, some users might not be allocated all possible bit rates within a RAT (e.g., users far away from the base station in UMTS). Other information that could be used in the decision is measurements from the mobile terminals, users speed, users and operators preferences, etc.

The JDRA problem adds a new dimension to the classical resource allocation problem within a single system, which is the selection of the appropriate RAT, increasing considerably the number of allocations possible. This problem can be classified as a multi-objective optimization problem with constraints, with the objectives of maximizing the number of satisfied users and the resource utilization subject to certain QoS requirements. The QoS performance indicators for packet services with delay constraints are the packet delay and the packet dropping ratio (assuming that packets that exceed their maximum delay are dropped). The JDRA algorithm shall thus try to guarantee a maximum contracted packet delay and a maximum packet dropping ratio.

For the sake of clarity and simplicity, it is assumed that the JDRA problem consists in finding the optimal bit rate and RAT allocation for all active users given a certain set of constraints in terms of available bandwidth in each RAT, QoS requirements, and traffic and coverage conditions. Note that some RATs might not be strictly limited by bandwidth (e.g., UMTS is typically limited by transmitted power in downlink

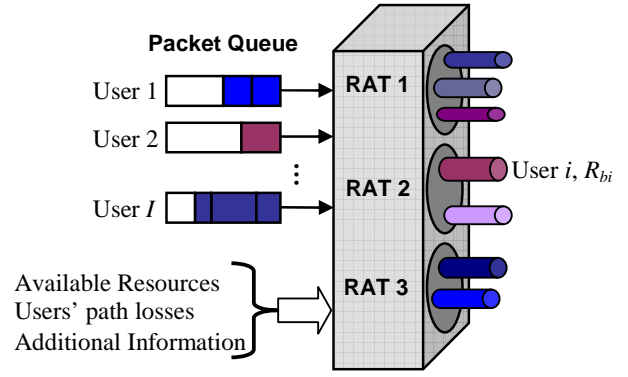


Fig. 1. Joint Dynamic Resource Allocation (JDRA) entity.

and by interference level in uplink), but somehow making simple approximations a maximum bandwidth can be usually computed for any RAT. The JDRA algorithm thus simply allocates to each user a certain bit rate and RAT every frame. In this paper a very tight coupling architecture among the different RATs has been assumed.

To formulate the JDRA problem we consider I active users in the system, J feasible bit rates in each RAT, and K RATs in the network. We define the bit rate allocation vector $\mathbf{r} = (r_1, \dots, r_I)$ and the RAT allocation vector $\mathbf{t} = (t_1, \dots, t_I)$, where r_i and t_i denote the bit rate and RAT allocated to the i th user, $r_i \in [0, J]$, $t_i \in [0, K]$. In \mathbf{r} , the index 0 denotes no allocation, whereas the index 1 and J denotes the minimum and the maximum bit rate considered. Users not allocated any bit rate are denoted by $r_i = 0$ and $t_i = 0$. Available resources in each RAT are given by the bandwidth vector $\mathbf{b}_T = (b_{T1}, \dots, b_{TK})$, in b/s. It represents the total available radio resources in the network (capacity constraint).

The maximum packet delay QoS requirement can be used with the packet queue information to compute a minimum target bit rate for each user $R_{b, \text{Target}, i}$ (b/s). Assuming a FIFO policy for the packets in the queue of each user, the minimum bit rate required to guarantee the transmission in due time of the j th packet of the i th user is given by:

$$R_{b,j}^i = \frac{\sum_{p=1}^j l_{i,p}}{D_{\max} - t_{i,j}} \quad (1)$$

where $l_{i,p}$ is the number of bits of the p th packet in the queue, D_{\max} is the maximum contracted packet delay (in seconds), and $t_{i,j}$ is the time in the queue of the j th packet. A minimum target bit rate that guarantees transmit all packets in due time can be defined for each user as:

$$R_{b, \text{Target}, i} = \max_j \left(R_{b,j}^i \right). \quad (2)$$

Note that a continuous transmission at the target bit rate would avoid packet losses. The target bit rate represents thus both QoS requirement and traffic conditions.

The aim of the JDRA algorithm is to find the best \mathbf{r} and \mathbf{t} possible vectors, so as to satisfy the design objectives.

III. HNN-BASED JDRA ALGORITHM

A. Optimization based on HNN

The use of HNN to solve optimization problems was initiated by Hopfield and Tank in [13]. Since then, many researchers have applied the HNN model to diverse optimization problems, including dynamic resource allocation [10]–[12].

Hopfield showed that neurons evolve into their stable states by gradient descent of an energy function E . The dynamics of the HNN can be written as [11]:

$$\frac{dU_i}{dt} = -\frac{U_i}{\tau} - \frac{\partial E}{\partial V_i} \quad (3)$$

where U_i and V_i are the input and output of the i th neuron, $V_i \in \{0,1\}$, and τ is the time constant of the neural network. The relationship between the outputs and the inputs of the neurons is non-linear, and can be approximated by the sigmoid function:

$$V_i = \frac{1}{1 + e^{-\lambda_i U_i}} \quad (4)$$

where λ_i is the gain scaling parameter of the i th neuron.

The minima of the energy function occur at the $2L$ corners inside the L -dimensional hypercube defined on $V_i \in [0,1]$, being L the total number of neurons [13]. Therefore, any optimization problem turns into defining a suitable energy function to be minimized, since the dynamics of the HNN will make neurons evolve to a minimum energy point (equilibrium state). After reaching a stable state, all neurons are either ON (if $V_i \geq 0.5$) or OFF (if $V_i < 0.5$).

Designing a suitable energy function is not a trivial task, since HNNs present inherent instability conditions that make the network converge to spurious solutions. Nevertheless, with a careful design a HNN can become a practical solution [14]. HNNs usually include an additional function that performs a local search with a greedy algorithm once the network reaches its equilibrium state, since HNNs might find a local optimum located near the global optima [11].

The dynamics of the HNN can be simulated solving (3) numerically (e.g., using first order Euler's technique):

$$U_i(t + \Delta t) = U_i(t) - \Delta t \cdot \left(\frac{U_i(t)}{\tau} + \frac{\partial E}{\partial V_i} \right) \quad (5)$$

where Δt is the time step. A stable state is reached when the output value of any neuron does not change by more than a threshold value ΔV_{\max} between consecutive updates.

B. HNN Formulation

The JDRA problem described in the previous section can be formulated in terms of HNN using a 3-D HNN with $I \times (J+1) \times K$ neurons, where I is the number of users, J is the number of feasible bit rates in each RAT (an additional neuron is used to account for the case of no allocation), and K is the number of RATs in the network.

The allocation matrix is given by the output value of

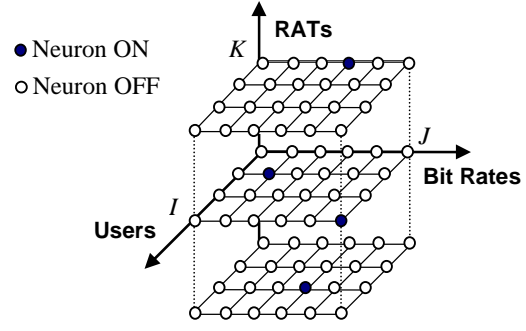


Fig. 2. 3-D HNN as the JDRA allocation.

neurons, denoted by $\mathbf{V} = [V_{ijk}]$, $i \in [1, I]$, $j \in [0, J]$ and $k \in [1, K]$, and indicates the bit rate and RAT allocated to each user (i.e., bit rate \mathbf{r} and RAT \mathbf{t} allocation vectors). Each neuron output V_{ijk} is associated a bit rate (in b/s) given by the bit rate matrix $\mathbf{R}_b = [R_{b,ijk}]$. Each neuron output V_{ijk} characterizes the allocation of user i , bit rate j , and RAT k , as can be seen in Fig. 2:

$$V_{ijk} = \begin{cases} 1 \text{ (ON)}, & \text{if } r_i = j; t_i = k \\ 0 \text{ (OFF)}, & \text{otherwise.} \end{cases} \quad (6)$$

Note that if a neuron V_{ijk} is ON, all neurons corresponding to user i (V_{imn} , $m \neq j$, $n \neq k$) must be OFF. A user is not allocated any bit rate when $V_{i0k} = 1$, $k \in [1, K]$.

In order to take into account the coverage constraints, it is introduced an allocation indicator matrix $\Psi = [\psi_{ijk}]$. This matrix indicates the bit rate and RAT allocations possible for each user. Neurons that represent unfeasible allocations (i.e., cannot be ON) are denoted by $\psi_{ijk} = 1$, whereas feasible allocations are denoted by $\psi_{ijk} = 0$. The allocation indicator matrix could be also used in case some RATs had less number of feasible bit rates than J , or to consider different grades of services with different sets of allowed bit rates.

For the sake of clarity, it is introduced a cost function to maximize the overall resource utilization in the network, and favour or penalize certain allocation conditions. The costs are given by the cost matrix $\mathbf{C} = [C_{ijk}]$, where C_{ijk} denotes the cost associated with user i , bit rate j , and RAT k . The cost function is defined as:

$$C_{ijk} = \frac{R_{b,ijk}}{R_{b,\max}} + \alpha_{ijk} + \beta_{ijk} \quad (7)$$

where $R_{b,\max}$ is the maximum bit rate considered, and α_{ijk} stresses the attainment of the target bit rate for each user:

$$\alpha_{ijk} = \begin{cases} 1 & R_{b,ijk} \geq R_{b,\text{Target},i} \\ 0 & R_{b,ijk} < R_{b,\text{Target},i} \end{cases} \quad (8)$$

The term β_{ijk} is used to prioritise the different RATs (it could also be used to consider user and operator preferences).

Maximization of the cost function will tend to maximize the resource utilization, while trying to guarantee a maximum packet delay by guaranteeing a minimum bit rate.

In order to avoid that some users obtain all resources, the allocation indicator matrix is used to disable bit rates higher

than the first available bit rate higher or equal than $R_{b,Target,i}$ in each RAT. In case all users get their target bit rate and there is bandwidth left, a Greedy algorithm will try to allocate higher bit rates until exhausting all resources.

The proposed 3-D HNN energy function for solving the JDRA problem is:

$$E = -\frac{\mu_1}{2} \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K \frac{C_{ijk}}{C_{\max}} V_{ijk} + \frac{\mu_2}{2} \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K \xi_{ijk} V_{ijk} + \frac{\mu_3}{2} \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K \psi_{ijk} V_{ijk} + \frac{\mu_4}{2} \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K V_{ijk} (1 - V_{ijk}) + \frac{\mu_5}{2} \left(1 - \sum_{j=1}^J \sum_{k=1}^K V_{ijk} \right)^2 \quad (9)$$

where μ_i are the weighting coefficients, C_{\max} is introduced to normalize the cost function to unity (depends on the value of β_{ijk}), and ξ_{ijk} is the capacity constraint matrix, defined as:

$$\xi_{ijk} = u \left(\frac{\eta_{ijk}}{b_{Tk}} - 1 \right). \quad (10)$$

Being $u(\bullet)$ the unit step function and η_{ijk} an indicator of the bandwidth utilization. It represents the bandwidth used at RAT k if the j th bit rate is allocated to the i th user, and the rest of users are assigned their current bit rate given by the neurons that are ON V_{jk}^{ON} (neurons OFF are not considered):

$$\eta_{ijk} = R_{b,ijk} + \sum_{\substack{m=1 \\ m \neq i}}^I \sum_{n=1}^J R_{b,mnk} V_{mnk}^{ON}. \quad (11)$$

The first term maximizes the cost function, driving the HNN toward higher resource utilization while trying to guarantee a minimum bit rate for each user. The second term ensures that the sum of allocated resources in each RAT does not exceed the resources available. The third term prevents the use of forbidden bit rates and RATs due to lack of coverage, or when the system provides different grades of services. The remaining terms are auxiliary factors which ensure rapid convergence to correct stable states of neurons. Specifically, the fourth forces all neurons to a stable state (either 0 or 1), whereas the fifth term assures that only one neuron is ON for each user.

Finally, the last step to design the HNN is to determine the weighting coefficients (μ_i) of the energy function. It is not a trivial task since several criteria must be taken into account, and all terms shall be weighted correctly [14]. Considering similar criteria as the ones proposed in [11], the following values have been decided:

$$\mu_1 = 4000; \quad \mu_2 = 5000; \quad \mu_3 = 6000; \quad \mu_4 = 800; \quad \mu_5 = 8000;$$

The set of system parameters for simulating the proposed HNN considered are:

$$\tau = 1; \quad \gamma_i = 1; \quad \Delta t = 10^{-4}; \quad \Delta V_{\max} = 10^{-4};$$

IV. NUMERICAL EVALUATION

A. Reference JDRA Algorithm

The reference JDRA algorithm considered simply allocates the target bit rate given by (2) to each user until exhausting resources. The algorithm sorts users according to their target bit rate in ascending order, and sequentially allocates to each user the first available bit rate higher or equal than his target bit rate if possible. The algorithm checks which RATs each user can be connected to, and gives preference to the RAT with the highest capacity and lower coverage, and so forth. If the target bit rate cannot be reached, the algorithm allocates the maximum bit rate possible.

B. Simulation Scenario

Initial evaluations of the proposed JDRA algorithm have been performed in a basic scenario, considering only three concentric cells (each cell corresponding to one RAT), with cell radii of 150 m, 650 m, and 1 km. The total bandwidth available in each RAT is 10 Mb/s, 2.5 Mb/s, and 625 kb/s. The set of feasible bit rates in each RAT considered are: {384, 512, 640, 768, 1024} kb/s, {32, 64, 128, 192, 256} kb/s, and {16, 32, 64, 80, 92} kb/s.

The values of the term β_{ijk} in (7) considered for each RAT are $\beta_{ij1} = 1$, $\beta_{ij2} = 2/3$ and $\beta_{ij3} = 1/3$ ($j \geq 1$), to prioritise the RATs according to their bandwidth and coverage. The choice of these values should be done considering the different values the cost function might take.

Users can be allocated the whole set of bit rates in each RAT if they are in the coverage area. Users are distributed in such a way that one third of users are within the coverage area of the first RAT, and two thirds within the coverage area of the second RAT. Users move according a random walk with an average velocity of 3.6 km/h.

The traffic model considered is based on the traffic model for packet service of 3GPP [15]. It is assumed that users are always active (i.e., infinite number of packet calls). The average reading time between packet calls is 30 s. The average number of packets within a packet call is 25, with an average time between packets of 30 ms. The packet length follows a Pareto with cut-off distribution with shape parameter 1.1, minimum packet size 81.5 bytes, and maximum packet size 6000 bytes. These parameters give an average requested file size of 9.15 kbytes, and an average bit rate at the source level of 90 kb/s.

Two user profiles, namely Class 1 and Class 2, have been considered with maximum packet delays of 100 ms and 200 ms respectively. The JDRA algorithm is executed every 10 ms to re-allocate bit rates and/or RATs to all active users.

C. Numerical Results

Fig. 3 shows the average packet delay as a function of the number of users in the network with a traffic mix of 50% per traffic class, for the HNN and the reference JDRA algorithms. Fig. 4 shows the average packet dropping ratio. It can be

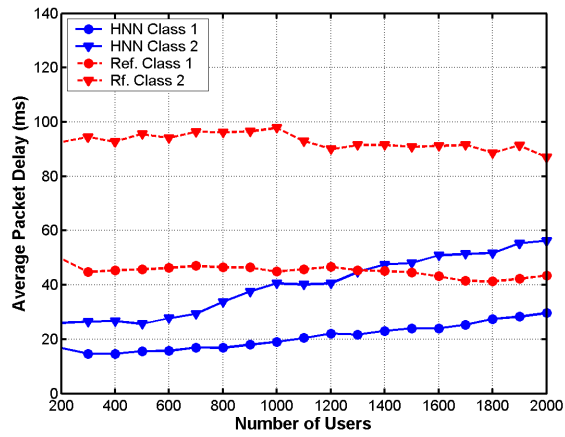


Fig. 3. Average packet delay (ms) vs. number of users. Traffic mix 50% class 1 users and 50% class 2 users.

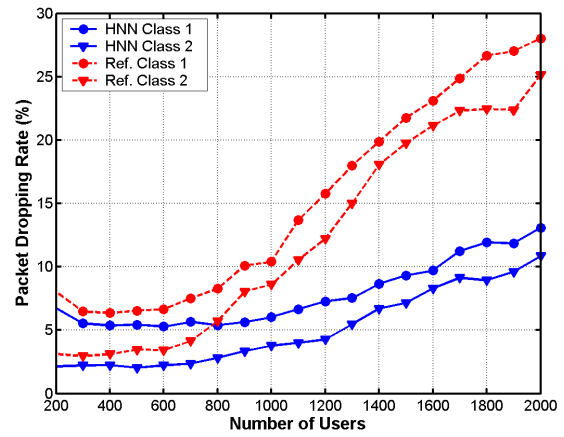


Fig. 4. Packet dropping ratio (%) vs. number of users. Traffic mix 50% class 1 users and 50% class 2 users.

noticed that the HNN algorithm provides a lower average delay and dropping ratio than the reference algorithm for all traffic loads considered, revealing that the HNN algorithm is able to adapt the resource allocation to the specific traffic conditions. Note that the reference algorithm provides a nearly constant average packet delay. However, the packet dropping ratio increases drastically once a certain load in the system is achieved. The HNN algorithm does not experience this effect, due to its better efficiency in the resource allocation process, increasing both packet delay and dropping ratio slightly with the number of users in the system.

V. CONCLUSIONS AND FUTURE WORK

This paper has presented a novel JDRA algorithm for packet-switched services with delay constraints in heterogeneous wireless networks using a HNN, where the decision variables are the bit rate and RAT allocated. The algorithm has been evaluated through simulations in a basic scenario, showing the potential of the algorithm due to its high capability to adapt itself to the scenario conditions.

In the future work, we will consider a beyond 3G network comprising GERAN, UTRAN and WLAN, modifying the proposed generic formulation to each specific RAT.

ACKNOWLEDGMENT

This research has been conducted within the NEWCOM Network of Excellence in Wireless Communications funded through the EC 6th Framework Programme, and has been partially supported by the Spanish Science & Technology Commission (CICYT) under the project TIC2005-08211-C02.

REFERENCES

- [1] N. Niebert *et al.*, "Ambient Networks: an architecture for communication networks beyond 3G," *IEEE Wireless Communications*, vol. 11, no. 2, pp. 14-22, April 2004.
- [2] E. Gustafsson and A. Jonsson, "Always best connected," *IEEE Wireless Communications*, vol. 10, no. 1, pp. 49-55, Feb. 2003.

- [3] G. Fodor, A. Eriksson, and A. Tuoriniemi, "Providing quality of service in always best connected networks," *IEEE Communications Magazine*, vol. 41, no. 7, pp. 154-163, June 2003.
- [4] F. Berggren *et al.*, "Multi-radio resource management for Ambient Networks," *IEEE Personal, Indoor and Mobile Radio Communications (PIMRC)*, Berlin, Germany, 2005.
- [5] J. Pérez-Romero *et al.*, "Common radio resource management: functional models and implementation requirements," *IEEE Vehicular Technology Conference (VTC) Spring*, Stockholm, Sweden, 2005.
- [6] A. Tölli, P. Hakalin, and H. Holma, "Performance evaluation of Common Radio Resource Management (CRRM)," *IEEE International Conference on Communications (ICC)*, New York, USA, 2002.
- [7] F. Malavasi *et al.*, "Traffic control algorithms for a multi access network scenario comprising GPRS and UMTS," *IEEE Vehicular Technology Conference (VTC) Spring*, Seoul, Korea, 2003.
- [8] A. Furuskär and J. Zander, "Multiservice allocation for multiaccess wireless systems," *IEEE Trans. Wireless Communications*, vol. 4, no. 1, pp. 174-184, Jan. 2005.
- [9] A. Baraev, L. Jorguseski, and R. Litjens, "Performance evaluation of radio access selection procedures in multi-radio access systems," *Wireless Personal Multimedia Communications (WPMC)*, Aalborg, Denmark, 2005.
- [10] O. Lázaro and D. Girma, "A Hopfield neural-network-based dynamic channel allocation with handoff channel reservation control," *IEEE Transactions on Vehicular Technology*, vol. 49, no. 5, pp. 1578-1587, Sept. 2000.
- [11] C. W. Ahn and R. S. Ramakrishna, "QoS provisioning dynamic connection-admission control for multimedia wireless networks using a Hopfield Neural Network," *IEEE Transactions on Vehicular Technology*, vol. 53, no. 1, pp. 106-117, Jan. 2004.
- [12] N. García, R. Agustí, and J. Pérez-Romero, "A user-centric approach for dynamic resource allocation in CDMA systems based on Hopfield Neural Networks," *IST Mobile and Wireless Communications Summit*, Dresden, Germany, 2005.
- [13] J. J. Hopfield and D. W. Tank, "Neural computations of decisions in optimization problems," *Biological Cybernetics*, vol. 52, pp. 141-152, 1985.
- [14] S. Abe, "Global convergence and suppression of spurious states of the Hopfield Neural Networks," *IEEE Transactions on Circuits and Systems*, vol. 40, no. 4, pp. 246-257, April 1993.
- [15] 3GPP TR 101 112 v3.2.0, "Selection procedures for the choice of radio transmission technologies of the UMTS," (1998-04).