Joint Dynamic Resource Allocation for Coupled Heterogeneous Wireless Networks based on Hopfield Neural Networks

Daniel Calabuig, Jose F. Monserrat, David Martín-Sacristán and Narcís Cardona
Institute of Telecommunications and Multimedia Applications, Polytechnic University of Valencia (UPV)
Valencia, Spain
`{dacaso,jomondel,damargan,ncardona}@iteam.upv.es`

Abstract — **This paper proposes an algorithm to solve the problem of Joint Dynamic Resource Allocation in heterogeneous wireless networks. The algorithm is based on Hopfield Neural Networks to achieve fast and suboptimal solutions. The generic formulation is particularized and evaluated in an HSDPA and 802.11e WLAN coupled networks. Some illustrative simulations results are presented to evaluate the performance of the new algorithm as compared with other strategies. The obtained results confirm the validity of the proposal.**

I. INTRODUCTION

The notion of being always best connected, introduced in [1], is an extension for heterogeneous systems of the notion of being always connected. Now, users not only should be connected anywhere, anytime, but also they should be served with the best available connection, what can be only accomplished with the interworking of the different technologies. For that reason, the standardization bodies are doing their best to make the interworking possible. For instance, the 3GPP not only allows UMTS to interwork with GPRS (two 3GPP Radio Access Technologies (RATs)) but also establishes the basis for a WLAN interworking (a non-3GPP RAT). In addition, the IEEE Standards Association is working in the 802.11u standard (scheduled for 2009) which gives WLAN the capacity of interworking with external networks. Likewise, the IPv6 network mobility management protocol (NEMO) enables a fast handover between different RATs, which is currently paving the way for a real interworking between RATs [2].

The multihoming concept provides multiple radio access for a single terminal in order to allow the terminal to maintain simultaneous links with the RATs [3]. Considering the higher level of coupling, the user could receive packets simultaneously from all links, but this approach entails an increasing User Equipment (UE) and network complexity. The traffic must be split into several flows what is not a trivial issue, and the UE must have several receivers and/or transmitters. A simpler solution consists in dynamically reconfiguring the active connection thus transmitting only through the best link. Thus the traffic is not split and the UE needs only one receiver and transmitter. This paper is focused on this last scenario for multihoming.

If the Dynamic Resource Allocation (DRA) in a single RAT is a hard optimization problem, when dealing simultaneously with multiple RATs, what is usually referred to as Joint DRA (JDRA), the problem becomes unmanageable, unless real-time sophisticated optimizers are employed.

Hopfield Neural Networks (HNNs) have proven useful in solving optimization problems in a short time (see e.g. [4]). HNNs have the capability of finding suboptimal solutions in few microseconds [4], what is fast enough to establish a new resource allocation in a frame-by-frame basis in current wireless communication systems.

This paper proposes a JDRA HNN-based that decides on which RAT serves each user in the next time interval and also on the distribution of resources among users to fulfill their QoS. The proposed algorithm follows a user-centric approach, since bit rates are not only allocated by network constraints, but also by users' expectations and requirements. This feature allows the system to maximize the utilization of the radio resources as a function of the user service profile.

II. DRA CONSTRAINTS

This paper assumes a set of feasible bit rates per RAT, $\Re_k$. Without loss of generality and in order to simplify the formulation, the sets $\Re_k$ are supposed to have the same number of elements.

The DRA algorithm shall find the optimal bit rate R_{jk}, i.e. the j-th bit rate in the k-th RAT, for each user satisfying the following constraints.

A. Resources constraint

The total allocated resources must not exceed the maximum available ones. The effective throughput is a function of the channel quality perceived by each user and the resources allocated to him. This dependence is supposed to be known by the algorithm. Hence, if $SNIR_{ik}$ is the Signal to Noise and Interference Ratio (SNIR) perceived by the i-th user in the k-th RAT, then his effective throughput by resource unit (r.u.) in that RAT is $Q_k\left(SNIR_{ik}\right)$, where function Q_k is known. Resource units can be time slots in GSM, or seconds of channel occupancy in WLAN.

B. Bit rate constraint

Each user is characterized by a subset of possible bit rates in each RAT, defined by the type of service he is subscribed to. The algorithm must only allocate to each user one of the bit rates predefined in the associated subsets. The minimal bit rate ensured is thus the minimal bit rate in the subsets.

C. Delay constraint

In order to introduce the delay in the resource allocation process, it is defined a minimum target bit rate for each user in each RAT, $R_{\min,ik}$, that guarantees the transmission of all packets in due time. $R_{\min,ik}$ can be defined as:

$$R_{\min,ik} = \begin{cases} P \max\limits_{p} \left\{ \dfrac{\beta_p}{t_{\max} - t_p - t_{\text{change},ik}} \right\}, & t_{\max} > t_p + t_{\text{change},ik} \\ \infty, & t_{\max} \le t_p + t_{\text{change},ik} \end{cases} \qquad (1)$$

where P are the number of data units (d.u.) in the buffer, β_p is the size of the p-th d.u., $t_{\max}$ is the maximum delay, t_p is the time the p-th d.u. is in the buffer and $t_{\text{change},ik}$ is the time needed for the i-th user to change to the k-th RAT. Data units depend on the service, for example for web browsing, FTP service and video calling, the d.u. is a web page, a file and a frame respectively. The formulation of (1) assumes that if several d.u. are stored in a buffer then the allocated bit rate is equally divided to transmit all the d.u. simultaneously. Therefore, (1) reflects the actual behavior of web browsers, since several opened web pages or file downloads are transmitted all together.

III. HNN-BASED JDRA ALGORITHM

A. HNN model

A HNN is composed by a set of interconnected neurons. Neurons will change dynamically their state until reaching an equilibrium point. Hopfield showed that an energy function E can represent the dynamics of the HNN, and that the problem of finding an equilibrium state of the neurons can be solved by finding a local minimum of the energy function [5],[6].

The dynamics of the HNN can be expressed as:

$$\frac{dU_i}{dt} = -\frac{U_i}{\tau} - \frac{\partial E}{\partial V_i} \qquad (2)$$

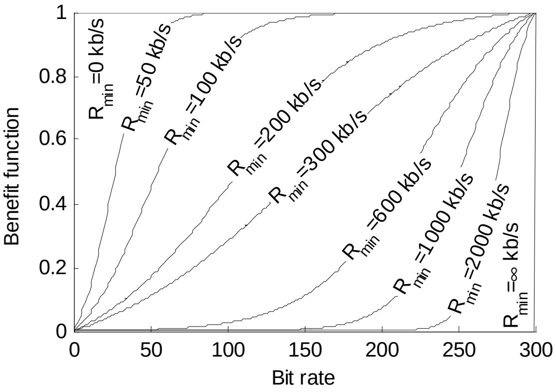

Fig. 1 $R_{\min,ik}$ effect in the benefit function for $R_{\max} = 300$ kb/s

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 outputs and the inputs of the neurons is non-linear, and is given by the sigmoid function:

$$V_i = f\left(U_i\right) = \frac{1}{1 + e^{-\alpha U_i}} \qquad (3)$$

where α is the gain of the neurons.

B. HNN formulation

Provided the sets of feasible bit rates, $\Re_k$, the JDRA problem can be formulated in terms of a 3D-HNN with $L = IJK$ neurons, being I the number of users in the system, J the number of elements in each set $\Re_k$ and K the number of RATs. The neuron states indicate the resource allocation, being the neuron with indices (i, j, k) ON if the i-th user has the j-th bit rate of the k-th RAT, R_{jk}, allocated. Note that the rest of neurons of user i must be OFF. It is important not to confuse the neuron states with the neuron outputs V_{ijk}. A neuron is ON if $V_{ijk} \ge 0.5$ and is OFF if $V_{ijk} < 0.5$.

The energy function is a quadratic function whose terms make the system converge to the expected solution. For the JDRA problem the energy function proposed in this paper is:

$$\begin{aligned} E = & -\frac{\mu_1}{2}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} B_{ijk}V_{ijk} - \frac{\mu_2}{2}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} \frac{R_{jk}}{R_{\max}} V_{ijk} \\ & + \frac{\mu_3}{2}\sum_{k=1}^{K}\left(\eta_k - \bar{\eta}\right)^2 + \frac{\mu_4}{2}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} \xi_{ijk}V_{ijk} \\ & + \frac{\mu_5}{2}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} \psi_{ijk}V_{ijk} + \frac{\mu_6}{2}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} V_{ijk}\left(1 - V_{ijk}\right) \\ & + \frac{\mu_7}{2}\sum_{i=1}^{I}\left(1 - \sum_{j=1}^{J}\sum_{k=1}^{K} V_{ijk}\right)^2 \end{aligned} \qquad (4)$$

where μ_n, $n = 1 \cdots 7$, are coefficients that weight each term.

The first term of the energy function introduces the benefit function B_{ij} which measures the benefit of allocating each bit rate to each user in terms of delay. This function is defined as:

$$B_{ijk} = \frac{S\left(R_{jk}, s_{ik}, r_{ik}\right) - S\left(0, s_{ik}, r_{ik}\right)}{S\left(R_{\max}, s_{ik}, r_{ik}\right) - S\left(0, s_{ik}, r_{ik}\right)} \tag{5}$$

$$S\left(x, s, r\right) = \frac{1}{1 + e^{-s(x+r)}} \tag{6}$$

$$s_{ik} = \begin{cases} \dfrac{2\ln(9)}{R_{\min,ik}} & R_{\min,ik} \le R_{\max} \\ \dfrac{2\ln(9) R_{\min,ik}}{\left(R_{\max}\right)^2} & R_{\min,ik} > R_{\max} \end{cases} \tag{7}$$

$$r_{ik} = \begin{cases} -\dfrac{R_{\min,ik}}{2} & R_{\min,ik} \le R_{\max} \\ -R_{\max} + \dfrac{\left(R_{\max}\right)^2}{2R_{\min,ik}} & R_{\min,ik} > R_{\max} \end{cases} \tag{8}$$

where $R_{\max} = \max_{j=1\cdots J, k=1\cdots K}\left\{R_{jk}\right\}$ is the maximum allocable bit rate and s_{ik} and r_{ik} are selected to increase B_{ijk} highly if $R_{jk} > R_{\min,ik}$ reflecting the high benefit of selecting a bit rate higher than the minimum target one. Fig. 1 shows the effect of $R_{\min,ik}$ in the benefit function. The benefit is scaled from a step function centered in 0 kb/s to another step function centered in $R_{\max}$. If the maximum delay is exceeded and $R_{\min,ik} = \infty$, then the only allocation that reduces the energy is $R_{\max}$

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

The third term favors those RATs with lower resource consumption. The term η_k is the load factor of the k-th RAT and $\bar{\eta}$ is the average load factor of all RATs, mathematically:

$$\eta_k = \max\left\{\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{R_{jk}}{Q_k\left(SNIR_{ik}\right)\rho_k}V_{ijk}, 1\right\} \tag{9}$$

$$\bar{\eta} = \frac{\sum_{k=1}^{K}\eta_k}{K} \tag{10}$$

where ρ_k are the total amount of available resources in the k-th RAT.

The fourth term penalizes the allocations that imply an excess of the maximum available system resources. Consequently, only the allocations combinations that satisfy the resource constraint introduced in section II.1 are possible equilibrium points of the HNN. ξ_{ijk} is defined as:

$$\xi_{ijk} = u\left(\frac{H_{ijk}}{\rho_k} - 1\right) \tag{11}$$

$$H_{ijk} = \sum_{\substack{l=1 \\ l\neq i}}^{I}\sum_{m=1}^{J}\frac{R_{mk}}{Q_k\left(SNIR_{lk}\right)}V_{lmk} + \frac{R_{jk}}{Q_k\left(SNIR_{ik}\right)} \tag{12}$$

where u is the step function. Note that H_{ijk} are the total needed resources to allocate to the i-th user the j-th bit rate in the k-th RAT.

The fourth term prevents the use of forbidden bit rates. Therefore, ψ_{ijk} is a permission table with $\psi_{ijk} = 1$ if the j-th bit rate in the k-th RAT is forbidden for the i-th user, and $\psi_{ijk} = 0$ otherwise. Therefore, it is possible to define different bitrates for different services. Moreover, in order to prevent the undesirable ping-pong effect, after a RAT reselection, the bit rates of the rest of RATs can be temporally banned.

The last two terms ensure a rapid convergence to correct and stable neuron states. The first one forces the neuron outputs to tend to the extremes 0 and 1, whereas the second one makes users have only one bit rate allocated in only one RAT.

C. HNN dynamics

The HNN algorithm starts with random neuron outputs uniformly distributed between 0 and 1. The numerical Euler's technique to solve (2) with $\tau = 1$ in a 3D-HNN is:

$$U_{ijk}\left(t + \Delta t\right) = U_{ijk}\left(t\right) + \Delta t\left\{-U_{ijk}\left(t\right) - \frac{\partial E}{\partial V_{ijk}}\right\} \tag{13}$$

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:

$$\begin{aligned} \frac{\partial E}{\partial V_{ijk}} &= -\frac{\mu_1}{2}B_{ijk} - \frac{\mu_2}{2}\frac{R_{jk}}{R_{\max}} + \mu_3\frac{R_{jk}}{Q_k\left(SNIR_{ik}\right)\rho_k}\left(\eta_k - \bar{\eta}\right) \\ &+ \frac{\mu_4}{2}\xi_{ijk} + \frac{\mu_5}{2}\psi_{ijk} + \frac{\mu_6}{2}\left(1 - 2V_{ijk}\right) - \mu_7\left(1 - \sum_{m=1}^{J}\sum_{n=1}^{K}V_{imn}\right) \end{aligned} \tag{14}$$

All the outputs are calculated each iteration using (3) and the solution of (13). The equilibrium state is reached when the outputs V_{ijk} changes less than a tolerance ΔV.

IV. APPLICATION TO HSDPA AND WLAN NETWORKS

In order to introduce a RAT in the algorithm, function Q_k, the quantity of available resources ρ_k and the set of feasible bit rates $\Re_k$ must be defined.

In this work, function Q_k is obtained supposing an optimal link adaptation, where, for a given SNIR, the optimum modulation and coding scheme are always selected. Therefore, if S_k Transmission Modes (TMs) exist, i.e. modulation and coding scheme pairs, function Q_k is:

$$Q_k\left(SNIR\right) = \max_{s=1\cdots S_k}\left\{\frac{L}{L + C_s}Br_s^k\left(1 - Er_s^k\left(SNIR\right)\right)\right\} \tag{15}$$

where Br_s^k and Er_s^k are respectively the bit rate per r.u. and error rate of the s-th TM of the k-th RAT, L is the payload length and C_s is the header length.

A. 802.11e WLAN

802.11e WLANs use a Cyclic Redundant Check (CRC) to protect data from errors. Therefore, the Packet Error Rate (PER) depends not only on the channel quality but also on the payload length L . Assuming a Viterbi decoding at the receiver, the PER of the s -th TM is [7]:

$$PER_s\left(L, SNIR\right) = 1 - \left(1 - P_u^s\left(SNIR\right)\right)^L \tag{16}$$

where P_u^s is the bit error probability of the s -th TM. The optimum payload length that maximizes the throughput for each TM is [7]:

$$L_s^*\left(SNIR\right) = -\frac{C_s}{2} + \frac{1}{2}\sqrt{C_s^2 - 4\frac{C_s}{\ln\left(1 - P_u^s\left(SNIR\right)\right)}} \tag{17}$$

Finally, function Q_k for WLAN is:

$$Q\left(SNIR\right) = \max_{s=1\cdots S}\left\{\frac{L_s^*\left(SNIR\right)Br_s\left(1 - P_u^s\left(SNIR\right)\right)^{L_s^*\left(SNIR\right)}}{L_s^*\left(SNIR\right) + C_s}\right\} \tag{18}$$

The available resources in WLAN are the seconds of channel occupancy which are supposed to be collision free thanks to the use of the HCCA mechanism. The set of feasible bit rates is $\Re \equiv$ {16384 kb/s, 4096 kb/s, 1024 kb/s, 512 kb/s, 256 kb/s, 0 kb/s}.

B. HSDPA

HSDPA uses turbo codes instead of a CRC to protect data from errors. The Block Error Rate (BLER) depends also on the block size and on the channel quality. Nevertheless, each TM has a fixed block size and, hence, the BLER for a specific TM depend only on the channel quality.

HSDPA have a wide range of possible TMs, from which 30 have been defined in the standard as Channel Quality Indicators (CQIs). Only these 30 TMs are used in this work. The BLER of the s -th CQI can be approximated as [8]:

$$BLER_s\left(SNIR\right) = \left(10^{\frac{2\left(SNIR - 1.03s + 17.3\right)}{\sqrt{3} - \log(s)}} + 1\right)^{-\frac{1}{0.7}} \tag{19}$$

Users are supposed to be time multiplexed. Thus, the 15 available codes are allocated to a unique user each 2 ms. Therefore, the available resources are the transmitted blocks of a maximum of 500 in one second. The size of blocks depends on the TM.

This assumption also implies that the actual BLER differs from the one obtained in [8], since the BLER is a function of the SNIR per code. If more codes are allocated then more SNIR is needed to maintain the same SNIR per code. Therefore, if 15 codes are allocated, the BLER is:

$$BLER_s\left(SNIR\right) = \left(10^{\frac{2\left(SNIR - 10\log\left(\frac{15}{N_s}\right) - 1.03s + 17.3\right)}{\sqrt{3} - \log(s)}} + 1\right)^{-\frac{1}{0.7}} \tag{20}$$

where N_s is the number of codes of the s -th CQI, shown in Table I. The set of feasible bit rates for HSDPA is set to $\Re \equiv$ {4096 kb/s, 1024 kb/s, 256 kb/s, 128 kb/s, 64 kb/s, 0 kb/s}.

Table I. Number of codes of each CQI.

N_s	1	2	3	4	5
CQI	1-6	7-9	10-12	13-14	15-22
N_s	7	8	10	12	15
CQI	23	24	25	26	27-30

Table II. Probability of exceeding the maximum web delay (%).

HNN	*WLANp-HNN*	*WLANp-RR*
0.84	1.63	48.79

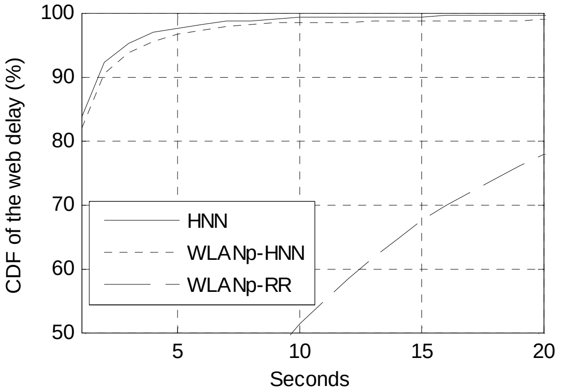

Fig. 2 CDF of the service response time for web service

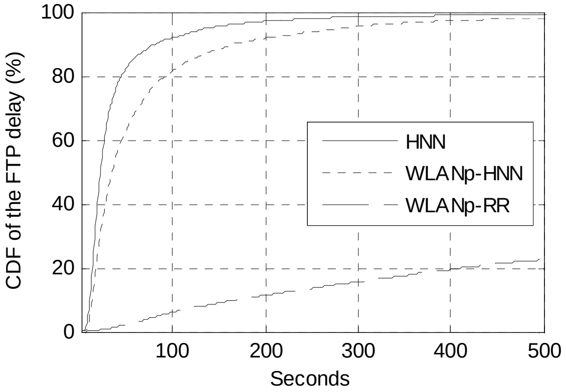

Fig. 3 CDF of the service response time for FTP service

V. NUMERICAL RESULTS

The employed simulator models seven cells with radius of 500 m and with the cell of interest in the center. The HSDPA Base Station (BS) is in the center cell. The 802.11e WLAN Access Point (AP) is in the centre of a hotspot separated 250 m from the center cell. In these two areas, the hotspot and the cell, two types of users are generated: pedestrian users moving at 1 km/h in the hotspot and vehicular users moving at 50 km/h in the cell. The ratio pedestrian/vehicular users is 4.

250 web browsing and 100 FTP downloading users are introduced in the simulation. The traffic models are extracted from [9] for both services. The maximum delay is set to 10 s for web browsing and to ∞ for FTP, i.e. the FTP service is supposed to be a background service with no maximum delay. All the bit rates in both RATs are allowed for both services.

The path losses between the i-th user and each RAT is calculated using:

$$L_{iHSDPA} = 137.4 + 35.2\log\left(d_{iHSDPA}\right) \quad (21)$$

$$L_{iWLAN} = 145 + 35\log\left(d_{iWLAN}\right) \quad (22)$$

where d_{iHSDPA} and d_{iWLAN} are the distances in km between the i-th user and the HSDPA BS and the WLAN AP respectively. The HSDPA BS transmit with 43 dBm and the WLAN AP with 20 dBm. BSs and APs of the interfering cells are supposed to be half loaded and, hence, they transmit with half the maximum power. The thermal noise power level is -132 dBm.

The parameters of the HNN network considered are the following:

$$\mu_1 = 1000 \quad \mu_2 = 1000 \quad \mu_3 = 0.1 \quad \mu_4 = 2500$$
$$\mu_5 = 16000 \quad \mu_6 = 2 \quad \mu_7 = 7000 \quad \alpha = 1$$
$$\Delta V = 10^{-4} \qquad \tau = 1$$

The proposed JDRA algorithm (HNN) has been compared with other two strategies. Both techniques select the RAT before the resource allocation. If the user is in the coverage area of the WLAN AP, then WLAN is selected, and HSDPA otherwise. This RAT selection is called WLAN preference (WLANp). After the RAT selection, resources are allocated separately in each RAT with a Round Robin (RR) and a HNN-based technique. The HNN-based algorithm is the same explained in this paper but for only one RAT. This two resource allocation techniques lead to two reference algorithms: the WLANp-HNN and the WLANp-RR.

Fig. 2 and 3 represent the Cumulative Distribution Function (CDF) of the service response time for web and FTP users respectively. The new algorithm proposed in this paper achieves the best performances, reducing the delay for all services. Moreover, Table II shows that the HNN algorithm satisfies more users since the maximum delay is exceeded with less probability.

Although the HNN algorithm is better than the WLANp-HNN, the main improvement is the use of an HNN-based resource allocation algorithm since the WLANp-RR has much worse performances than the other two algorithms.

Furthermore, HNN algorithms not only improve the performances of RR algorithms but also can perform a JDRA which benefits are shown in the differences between the HNN and the WLANp-HNN algorithms. The JDRA has the capability of reallocating users in other RATs if necessary, whereas, if a pre-RAT selection is performed, some RATs can be saturated while others are empty.

VI. CONCLUSIONS

This paper has presented a novel HNN-based JDRA algorithm for packet-switched services with delay constraints in heterogeneous wireless networks. The algorithm has been evaluated through simulations in a basic HSDPA-WLAN scenario with mobile users. As compared with other strategies, the HNN-based JDRA algorithm is always preferred since users are served with less delay. Moreover, this work has also depict the benefit of JDRA against pre-RAT selection schemes.

ACKNOWLEDGMENT

This work has been partially funded by CICYT (Spanish National Science and Technology Council) and the FEDER program of the European Commission under the Project TEC2005-08211-C02.

REFERENCES

[1] E. Gustafsson and A. Jonsson, “Always best connected,” IEEE Wireless Communications, vol. 10, no. 1, pp. 49-55, 2003.

[2] Multinet, Enabler for Next Generation Service Delivery. Project IST-27437 STP.

[3] J. Luo, R. Mukerjee, M. Dillinger, E. Mohyeldin and E. Schulz, “Investigation of radio resource scheduling in WLANs coupled with 3G cellular network”, IEEE Communications Magazine, vol. 41, no. 6, pp. 108-115, June 2003.

[4] C. W. Ahn and R. S. Ramakrishna, “QoS provisioning dynamic connection-admission control for multimedia wireless networks using Hopfield neural networks,” IEEE Trans. on Vehicular Technology, vol. 53, no. 1, pp. 106-117, Jan. 2004.

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

[6] 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.

[7] S. Choudhury and J. D. Gibson, “Payload length and rate adaptation for throughput optimization in wireless LANs,” IEEE Vehicular Technology Conference (VTC), Melbourne, Australia. Spring 2006.

[8] F. Brouwer, I. de Bruin, J. C. Silva, N. Souto, F. Cercas and A. Correia, “Usage of link-level performance indicators for HSDPA network-level simulations in E-UMTS,” IEEE International Symposium Spread Spectrum Techniques and Applications (ISSSTA), Sydney, Australia. Fall 2004.

[9] 3GPP2-TSGC5, “HTTP and FTP traffic model for 1xEV-DV Simulations,” 2001.