Hindawi Publishing Corporation
EURASIP Journal on Wireless Communications and Networking
Volume 2008, Article ID 817676, 9 pages
doi:10.1155/2008/817676

Research Article

A New OFDMA Scheduler for Delay-Sensitive Traffic Based on Hopfield Neural Networks

Nuria García,[1] Jordi Pérez-Romero,[2] and Ramón Agustí[2]

[1] *Grup de Recerca en Tecnologies i Estratègies de les Telecomunicacions, Departament de Tecnologies de la Informació i les Comunicacions, Universitat Pompeu Fabra, Passeig de Circumval.lació 8, 08003 Barcelona, Spain*
[2] *Grup de Recerca en Comunicacions Mòbils, Departament de Teoria del Senyal i Comunicacions, Universitat Politècnica de Catalunya, C/ Jordi Girona 31, 08034 Barcelona, Spain*

Correspondence should be addressed to Jordi Perez-Romero, jorperez@tsc.upc.edu

Received 1 May 2007; Revised 6 November 2007; Accepted 4 January 2008

Recommended by Luc Vandendorpe

This paper introduces a novel joint channel and queuing-aware OFDMA scheduler for delay-sensitive traffic based on a hopfield neural network (HNN) scheme. It allows providing an optimum OFDMA performance by solving a complex combinational problem. The algorithm is based on distributing the available subcarriers among the users depending, on the one hand, on the time left for the transmission of the different packets in due time, so that packet droppings are avoided. On the other hand, it also accounts for the available channel capacity in each subcarrier depending on the channel status reported by the different users. The different requirements are captured in the form of an energy function that is minimized by the algorithm. In that respect, the paper illustrates two different algorithms coming from two settings of this energy function. The algorithms have been evaluated for delay-sensitive traffic and they have been compared against other state-of-the-art algorithms existing in the literature, exhibiting a better behavior in terms of packet-dropping probability.

1. INTRODUCTION

Orthogonal frequency division multiple access (OFDMA) has emerged as one of the most promising schemes for broadband wireless networks. By using multiple parallel low-rate subcarriers, OFDMA can offer satisfactory high-speed data rate, robust wireless transmission, and flexible radio resource management, among other remarkable features, as it is widely documented in the open literature. In fact, current standards like DVB-T, wireless LAN IEEE.802.11a, and fixed-mobile broadband access system IEEE 802.16 have adopted OFDMA scheme. In addition to that, OFDMA has also been selected as access technology for the future 3G long-term evolution (LTE) in the evolved universal telecommunication radio access (EUTRA) [1], and most of the 4G initiatives also consider OFDMA as a prime access strategy. As a result of this current trend and from the radio resource allocation point of view, there has recently been a lot of attention to manage dynamically the inherent flexibility offered by OFDMA in an optimal and still practical· way, either in isolated [2] or in multicell OFDMA systems [3].

Concerning packet data transmission, most of the subcarrier allocation strategies proposed in OFDMA-based wireless multimedia networks intent somehow to maximize the system throughput or minimize the overall transmitted power while achieving the terminal bit rate requirements [4]. A recent good survey on these topics can be found in [5]. Unfortunately, the traffic-related queuing impact when considering dynamic resource allocation (DRA) scheduling schemes is not covered at the same extent. That is particularly relevant for interactive services and in general terms for delay-bounded services, in which packets should be delivered within specified deadlines. In that respect, there has been little work on these relevant performance measures such as the delay bound and the delay violation probability, which are indicative of the worst-case delay behavior. To the best of our knowledge, [6, 7] are among the first papers to face the constrained delay issue in managing the OFDMA system resources using for this purpose a heuristic approach based on utility and priority functions when assigning resources to users.

OFDMA scheduling should actually include both joint subcarrier and power allocations. This is a rather complex problem, and usually it is simplified by separating these two allocations. Subcarrier allocation provides more gain than power allocation in [8], and in fact it is shown in [9, 10] that waterfilling allocation only brings marginal performance improvement over fixed power allocation with adaptive code and modulation (ACM). Then, in this paper, we focus on a subcarrier allocation strategy that is aware of the queuing state per each user and that retains a fixed power allocation as well as adaptive quadrature amplitude modulation (QAM).

Subcarrier allocation in OFDMA can be seen as a combinational problem where there are plenty of possible combinations associated to a given user. A user can be granted with many subcarriers at a given point of the time. In turn, each subcarrier provides a given channel capacity depending on the current fading and interference, so that multiuser diversity can be exploited. Also, a given subcarrier can only be assigned to one user. This is a natural choice, based on [9], that proves that the optimum OFDMA performance is reached by assigning each subcarrier to one user only in a cell among the many users trying to get access. In queuing-aware OFDMA systems like the one considered here, the information about queuing and channel status is exploited to efficiently allocate resources through proper cross-layer designs of the data scheduler. As a matter of fact, like in general OFDMA, DRA proposals, heuristic algorithms are usually selected to circumvent the fact that NP-hard algorithms would be necessary to obtain the optimum solutions. This is the case of [11], where a heuristic two-step algorithm is proposed to first allocate a number of subcarriers to each user and then to assign the specific subcarrier to each terminal. Similarly, in [12], an alternative approximate asymptotic mechanism is exploited. It relies on the fact that in a heavy traffic scenario minimizing delay violation is approximately equivalent to minimizing mean waiting time. Similarly, in [13], an allocation strategy depending on the queue size of each terminal relative to the overall data queued at the access point is presented for video streams. It is shown that this allocation achieves significant improvements with respect to static allocation methods, in spite of not including in the allocation neither channel gain information nor any stream specific knowledge. In [14], a cross-layer DRA strategy is presented that combines the channel status information together with the queue status and quality requirements in order to maximize power efficiency and ensure user fairness using a virtual clock scheduling algorithm, an adaptive subcarrier, and power allocation.

In this paper, due to the fact that the typical DRA in OFDMA turns out to be actually a combinational problem among all subcarriers involved, we have devised the collective computation property featured by the hopfield neural networks (HNN) which provide an optimal solution for many combinational problems [15, 16], as a very suitable approach for the problem addressed here. In fact, the HNN approach provides feasible solutions to complex optimization problems, like the NP-hard algorithms mentioned above. Under the proper conditions we can take advantage of the fact that the so-called HNN energy evolves toward a minimum value [17] providing a final neuron state that includes, in a natural way, the optimal subcarrier combination to be allocated. Consequently, this optimal allocation can be obtained by properly including different constraints (i.e., channel and queue status for the different users) in the definition of the HNN energy. From an implementation point of view, HNN methodology can be carried out either by solving iteratively a numerical differential equation based on the Euler technique or by means of hardware implementations (HNN is derived with an initial hardware implementation in mind) such as the field-programmable gate array (FPGA) chip [18] that has been proved practically for implementation purposes.

Under this framework, this paper proposes a novel HNN-based joint channel and queue-aware scheduling strategy for downlink OFDMA systems suitable for delay-bounded services. A multiuser scenario with statistically independent fading channels and an isolated cell is considered. Then, the subcarrier allocation is directly related to the remaining time before the agreed bounded delay service per user is violated for each packet as well as to the channel state. The proposed algorithm is compared against other approaches existing in the literature [11] and against a heuristic algorithm, also proposed in this paper as a first simple step in the provision of a joint channel and queue-aware strategy, which simply prioritizes the users according to their remaining packet lifetime and assigns subcarriers until they are exhausted.

The rest of the paper is organized as follows. Section 2 describes the considered system model, including the queueing behavior and the OFDMA considerations. Section 3 describes the proposed HNN-based algorithm with two different possibilities depending on the definition of the energy function. The proposed algorithm will be compared against the reference schemes presented in Section 4. Results are given in Section 5 and finally, conclusions are summarized in Section 6.

2. SYSTEM MODEL

The considered DRA problem assumes a set of N users, $i = 1, \ldots, N$, with their corresponding queues located at the base station of the access network which contains the packets pending to be transmitted in the downlink direction of an OFDMA system, as illustrated in Figure 1. It is considered that nonshaped traffic is arriving to the queues, so that all the incurred packet delay is introduced at the network level. Also the model allows for differentiating among different classes of traffic (services classes) as will be discussed later. When a packet cannot be delivered within this bounded delay it is dropped and therefore, a dropping probability appears as a key performance indicator of the scheduling behavior.

The envisaged HNN-based scheduling algorithm operates in frames of duration T and allocates a certain bit rate to each user by assigning to him a set of subcarriers. Multiple transmissions of different users in parallel are allowed by making use of different subcarrier combinations. A granularity of one subcarrier is considered in the assignment process.

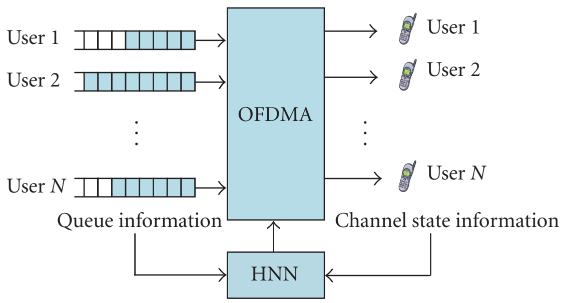

Figure 1: System model.

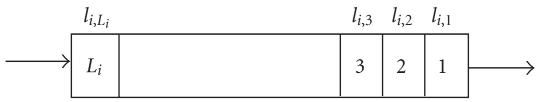

Figure 2: Queue of the ith user (for simplicity, the dependency with the number of frame k has been omitted).

The bit rate allocation will be executed by means of an optimal mechanism based on HNN, through the minimization of a properly defined energy function which includes a main function associated to the eligible bit rate per user according to the queue status and service-class requirements, as well as other terms to include the OFDMA downlink network restrictions. These considerations related with queue status and OFDMA model are explained in the following.

2.1. Queuing model

With respect to the queue model, let us assume that at the beginning of the kth frame, the ith user has L_i packets in the queue as depicted in Figure 2. $l_{i,m}(k)$ denotes the number of bits of the mth packet of the ith user in the kth frame.

Assuming a first input first output (FIFO) policy for the queue of each user, the amount of bits that should be transmitted until the transmission of the mth packet of the complete ith user is given by

$$B_{i,m}(k) = \sum_{n=1}^{m} l_{i,n}(k). \tag{1}$$

On the other hand, the delay constraint is given by $D_{\max,i}$, measured as the maximum packet delay measured in frames specified in the contract of each user. Let $f_{i,m}(k)$ be the elapsed time at the beginning of the kth frame since the arrival of the mth packet in the queue of the ith user. Then, the maximum timeout left for transmission of this packet is

$$\mathrm{TO}_{i,m}(k) = D_{\max,i} - f_{i,m}(k). \tag{2}$$

Consequently, the minimum bit rate required to guarantee the transmission in due time of this packet is given by

$$v_{i,m}(k) = \frac{B_{i,m}(k)}{\mathrm{TO}_{i,m}(k)}. \tag{3}$$

We define the *optimum bit rate* (OBR) for the ith user in the kth frame as the one that allows transmitting all the packets in due time, given by

$$R_{b,i,\mathrm{opt}}(k) = \max_{m=1,\ldots,L_i} \{v_{i,m}(k)\} \cdot (1+\theta), \tag{4}$$

where θ (≥ 0) is a safety empirical factor introduced to face fluctuations in the packet generation of the successive frames. This OBR should be provided to each user by the OFDMA scheduler. To this end, it will perform the most suitable aggregation of a given number of subcarriers. Notice that a continuous transmission at the OBR would avoid packet losses for this user. However, it cannot always be guaranteed for all the users because of the total bandwidth restrictions.

2.2. OFDMA system model

The system model assumes a total of S subcarriers with separation Δf (Hz) to be allocated to the N users. It is assumed that the transmitter knows the channel state at the terminal side, and in particular, the receiver signal-to-noise ratio of the ith user in the jth subcarrier $\rho_{ij}(k)$ in the kth frame. This value should be transmitted regularly by the mobile to the base station via a feedback channel, as illustrated in Figure 1, being the elapsed time lower than the channel coherence time. Then the actual capacity $c_{ij}(k)$ of the jth QAM modulated subcarrier with Gray bit mapping in the kth frame for the ith user can be approximated by [18]

$$\begin{gathered} c_{ij}(k) = \log_2\left(1+\beta\rho_{ij}(k)\right) \text{ bits/s/Hz} \\ \beta = -1,6/\ln(5\ \mathrm{BER}), \end{gathered} \tag{5}$$

where BER is the target bit-error rate. Then, the throughput of the OFDMA system in the kth frame is given by

$$R(k) = \sum_{i=1}^{N}\sum_{j=1}^{S} \chi_{ij}(k) c_{ij}(k)\,\Delta f, \tag{6}$$

where $\chi_{ij}(k)$ is set to 1 when the jth subcarrier is assigned to the ith user and is set to 0, otherwise. Finally, as not all the $c_{ij}(k)$ values are allowed in a QAM modulation, the value obtained in (5) will be rounded to the highest integer lower than or equal to $c_{ij}(k)$ from the set $\{0, 1, 2, 4, 6\}$ bits/s/Hz.

3. HNN-BASED SCHEDULING MODEL

This section presents the proposed HNN-based scheduling algorithm to be executed in each frame k in order to determine the subcarrier allocation in accordance with the channel status for each user captured in the capacity $c_{ij}(k)$ seen by the ith user in the jth subcarrier in the kth frame, and the buffer status captured in the value of the OBR for the ith user in the kth frame, $R_{b,i,\mathrm{opt}}(k)$. For simplicity in the notation, the explicit dependency with the number of frame k will be omitted in the following.

The above DRA problem subject to the mentioned restrictions can be formulated in terms of a two-dimensional neural network with $L = N \times S$ neurons [15]. The output values of the neurons, denoted by V_{ij}, will be equal to 1, if the jth subcarrier is assigned to the ith user and 0, otherwise.

In a 2D HNN, each neuron is modeled as a nonlinear device with a sigmoid monotonically increasing function defined by the logistic function

$$V_{ij} = f(U_{ij}) = \frac{1}{1 + e^{-\alpha U_{ij}}}, \tag{7}$$

where U_{ij} and V_{ij} are the input and output, respectively, of the (i, j) neuron, and α is the corresponding gain of the amplifier of the neuron.

Each neuron receives resistive connections from other neurons and these connections are fully described by the interconnection matrix $T = [T_{ij,pq}]$, where $T_{ij,pq}$ is the interconnection weight from the (i, j) neuron to the (p, q) neuron. Each neuron also receives an input bias current I_{ij} that is an adjustable parameter. The dynamics of the HNN are represented by [15]

$$\frac{dU_{ij}}{dt} = -\frac{U_{ij}}{\tau} + \sum_{p=1}^{N}\sum_{q=1}^{S} T_{ij,pq}V_{pq} + I_{ij}, \tag{8}$$

where τ is a time constant. Furthermore, the quadratic energy function is defined as

$$E = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{S}\sum_{p=1}^{N}\sum_{q=1}^{S} T_{ij,pq}V_{ij}V_{pq} - \sum_{i=1}^{N}\sum_{j=1}^{S} I_{ij}V_{ij}. \tag{9}$$

Then, taking into account the derivative of the energy function E in (9), the HNN dynamics represented by (8) can be formulated in a more compact way by the following differential equation:

$$\frac{dU_{ij}}{dt} = -\frac{U_{ij}}{\tau} - \frac{\partial E}{\partial V_{ij}}. \tag{10}$$

It is shown in [20] that, for a symmetric matrix T and sufficiently high gain α, neurons in HNN evolve along a trajectory over which the energy function decreases monotonically to a minimum occurring at the $2^{N\times S}$ corners inside the N × S-dimensional hypercube defined on $V_{ij} \in \{0, 1\}$, thus providing the allocation of subcarriers to users.

It is worth noticing that by selecting a suitable expression for the energy function E, a queuing-aware OFDMA embedded optimization can be achieved. The optimization process of the HNN is carried out on a frame-by-frame basis and relies on minimizing the energy function through the convergence of the above differential equation. In the following, two suitable expressions for the energy function compliant with the definition in (9) are introduced, which will give rise to two different scheduling HNN-based algorithms.

3.1. *HNN1 algorithm*

A first expression proposed for the energy E follows as

$$\begin{aligned} E = {} & \frac{\mu_1}{2}\sum_{i=1}^{N}\left(1 - \frac{\sum_{j=1}^{S} c_{ij}\, V_{ij}\Delta f}{\overline{m_i}}\right)^2 + \frac{\mu_2}{2}\sum_{i=1}^{N}\sum_{j=1}^{S}\psi_{ij}V_{ij} \\ & + \frac{\mu_3}{2}\sum_{i=1}^{N}\sum_{j=1}^{S} V_{ij}(1 - V_{ij}) + \frac{\mu_4}{2}\sum_{j=1}^{S}\left(1 - \sum_{i=1}^{N} V_{ij}\right)^2. \end{aligned} \tag{11}$$

The first term is a cost function intended to be minimized by a proper setting of V_{ij}. It includes the expression

$$\overline{m_i} = \left(\left[\frac{R_{b,i,\mathrm{opt}}}{\Delta f}\right] + 1\right)\cdot \Delta f, \tag{12}$$

where $[\cdot]$ denotes the integer part, so that (12) is actually a quantification of the OBR value $R_{b,i,\mathrm{opt}}$ in multiples of Δf. The minimum value of the energy E would be achieved for specific combinations of V_{ij} that minimize each summand, so that each user tends to be allocated with its OBR. Notice that OBR can be changed at each frame depending on the traffic dynamics and the packets evolution in the queues. Similarly, the channel fading also impacts OBR dynamics as the capacity of the different subcarriers can be changed on a frame basis.

The second summand in (11) simply penalizes the allocated subcarriers with bit rates equal to zero. That is, when the ith user is considered with a subcarrier jth in which $c_{ij} = 0$, $\psi_{ij} = 1$, thus increasing their contribution to the energy function. In this way, the corresponding subcarriers are brought out of the energy minima and become available for other users. Otherwise, it is set to $\psi_{ij} = 0$.

The third summand of (11) was introduced in [21] in order to force convergence toward $V_{ij} \in \{0, 1\}$ and the fourth term is introduced to reflect the physical OFDMA constraint that a given subcarrier can only be allocated to one user. The relationship between the energy function (11), the HNN interconnection matrix $T = [T_{ij,pq}]$, and the input bias current I_{ij}values in the general expression of the energy function in (9) can be obtained according to the details shown in the appendix.

The terms μ_1, μ_2, μ_3,and μ_4 are constants to be set.

The numerical iterative solution of (10) is obtained following the Euler technique as

$$U_{ij}(n+1) = U_{ij}(n) + \Delta\left[-\frac{U_{ij}(n)}{\tau} - \frac{\partial E}{\partial V_{ij}}\right], \tag{13}$$

where Δ denotes the discrete step and neuron's voltage is updated at each nth iteration using (7). After reaching a final state, each neuron is either ON (i.e., V_{ij} is set to 1 if $V_{ij} \geq 0,5$) or OFF (i.e., V_{ij} is set to 0 if $V_{ij} < 0,5$).

Then, once the final V_{ij} values are achieved as solution of (7) and (10), the final bit rate assigned to the ith user after the execution of the algorithm in one frame follows:

$$R_{b,i} = \sum_{j=1}^{S} V_{ij}\, c_{ij}\, \Delta f. \tag{14}$$

3.2. *HNN2 algorithm*

A second expression for the energy function that captures additional features concerning both users and channel subcarrier status not considered in algorithm HNN1 is

$$\begin{aligned} E = {} & \frac{\mu_1}{2}\sum_{i=1}^{N}\omega_i\left(1 - \frac{\sum_{j=1}^{S} c_{ij}\, V_{ij}\Delta f}{\overline{m_i}}\right)^2 \\ & + \frac{\mu_3}{2}\sum_{i=1}^{N}\sum_{j=1}^{S} V_{ij}(1 - V_{ij}) + \frac{\mu_4}{2}\sum_{j=1}^{S}\left(1 - \sum_{i=1}^{N} V_{ij}\right)^2. \end{aligned} \tag{15}$$

In this case, the first term has been modified with respect to the HNN1 algorithm with the inclusion of a new coefficient ω_i introduced with a two-fold objective. First, it should favor the users with high OBR that the former formulation of the term could not capture. Second, it should favor the allocation of subcarriers to the users with the best channel capacity thus making a better exploitation of the multiuser diversity. For that purpose, the users are first ordered in decreasing value of their OBR $R_{b,i,\text{opt}}$, and order_i is defined as the position of the ith user in this ordered list. Then, the coefficient ω_i is empirically defined as

$$\omega_i = \left(2 - \frac{1}{\text{order}_i}\right) \cdot \left(1 - \frac{\sum_{j=1}^{S} c_{ij}}{\sum_{i'=1}^{N} \sum_{j=1}^{S} c_{i'j}}\right). \qquad (16)$$

Notice that, with this definition, users with either a high OBR (i.e., a low value of order_i) or a good channel status in the different subcarriers will tend to have smaller values of ω_i and consequently, the minima of the energy function will tend to occur in V_{ij} values, so that a certain number of subcarriers is allocated to these users.

On the other hand, notice also that the effect of coefficient ω_i already captures to some extent the avoidance to allocate the subcarriers to the users with a bad channel status, which was intended by the second summand in the energy function of the HNN1 algorithm in (11), and consequently, this summand is not included in the definition of the energy function for the HNN2 algorithm in (15). The Appendix shows the relationship between the energy function in (15) and the interconnection matrix and input bias currents.

4. REFREENCE SCHEDULING SCHEMES

The proposed HNN-based algorithms described in the previous section have been compared against other approaches. First, a simple heuristic reference scheduling algorithm has been considered which exploits the OBR concept, but does not take into consideration the optimization in accordance with the HNN procedure. This algorithm, denoted in the following as reference scheduling scheme 1(RSS1), simply tries to allocate to each user its optimum bit rate OBR as defined in (4). The algorithm operates then in the following steps in each frame.

Step 1. Order the users in the increasing value of $R_{b,i,\text{opt}}$.

Step 2. Allocate sequentially to each user ith the necessary number of subcarriers, so that its final scheduled bit rate is higher than or equal to $\overline{m_i}$ from (12). This allocation is carried out by ordering first all the available subcarriers still pending to be allocated in the increasing value of c_{ij}.

Step 3. Once all the S available subcarriers have been allocated, assign a bit rate equal to 0 kb/s (i.e., no transmission) to the remaining users.

Notice that, as far as the OFDMA capacity can satisfy the required OBR per user and frame, the reference system would lead to a quite satisfactory scheduling approach from a delay point of view.

Furthermore, focusing on the existing approaches in the literature, the proposed algorithm has also been compared against the recent proposal introduced in [11], which will be denoted in the following as reference scheduling scheme 2 (RSS2). This algorithm also focuses on delay-sensitive traffic and operates on two different steps.

The first step is the subcarrier allocation algorithm which determines the number of subcarriers to be allocated to each user. For that purpose, it accounts for different average channel conditions in all subcarriers as well as for the delay requirements of the different packets in the queue. After an initial computation, the algorithm executes several iterations in order to ensure that the total number of allocated subcarriers equals the number of available subcarriers S.

In the second step, the subcarrier assignment algorithm is executed which decides the specific subcarriers allocated to each user. This is done by creating a priority list ordering the different users in accordance with the history of packet droppings experienced by each one, so that users with a higher number of droppings have a higher priority. In turn, for users with equal number of droppings, the priority is computed in accordance with the channel quality (i.e., users with better quality have a higher priority). Then, according to the priority list, each user selects the best available subcarriers up to the number of subcarriers computed in the first step.

For details of the algorithm, the reader is referred to [11]. It is worth mentioning that this algorithm was in turn compared in [11] against other previous references, such as [13], exhibiting better performance. Consequently, this algorithm has been retained here for comparison purposes as an appropriate reference representative of the state-of-the art in OFDMA dynamic resource allocation algorithms for delay-sensitive traffic.

5. RESULTS AND DISCUSSION

A single cell scenario has been considered to assess the proposed HNN-based DRA strategy for a downlink OFDMA wireless access. We consider $S = 128$ subcarriers and $\Delta f = 15$ kHz. In the simulation, the scheduling algorithm operates in frames of $T = 10$ milliseconds. We will also assume that the coherence time is larger than the frame time, so within a frame it is assumed that the channel impulse response does not vary. In our simulation, each user channel suffers from multipath Rayleigh fading with a delay profile characterized by a time variant impulse response following the pedestrian model of [22] with a mobile speed of 5 km/h and an average signal-to-noise ratio equal to 17 dB. We let a target BER $= 10^{-4}$ and assume a set of possible transmission bit rates: 15 m kb/s per subcarrier (m = 0,1,2,4,6) by properly adjusting the modulation levels of a 2^m QAM-adopted signalling format.

The selected parameters appearing in the formulation of the HNN are $\mu_1 = 4000$, $\mu_2 = 30000$, $\mu_3 = 800$, $\mu_4 = 18000$, $\tau = 1$, and $\alpha = 1.0$. Simulations not shown here for the sake of brevity concerning the variation of these parameters have revealed that they are actually robust values, so that changing them to a certain extent (i.e., variations as large as 50% have been tested) does not impact significantly the final results.

The only conditions are that these parameters should be positive and satisfy $\mu_3 < \mu_4$, as it is shown in the appendix.

On the other hand, the iterative numerical solution in (13) is finalized when iterations n and $n-1$ satisfy $\|\mathbf{V}^n - \mathbf{V}^{n-1}\|_2 < \varepsilon$, where $\| \|_2$ is the Euclidean norm and $\mathbf{V}$ is a matrix which includes all the elements V_{ij}. We have set $\Delta = 10^{-4}$ in (13) and $\varepsilon = 10^{-5}$. The convergence to a stable value is attained in practice in most of the situations between 1000 and 1500 iterations. As a result of that, a maximum of 2000 iterations has been used to stop the iterative process. If all these conditions are fulfilled, we decide that the process converges and the values V_{ij} provide us the inputs to calculate the total bit rate allocated to each user in each frame $R_{b,i}$ according to (14).

An interactive service following the WWW traffic model from [22] has been considered as a representative of a delay-sensitive service. Specifically, WWW sessions are composed of an average of 5 pages with an average time between pages of 30 seconds. In each page, the average number of packets is 25 with an average time between packets of 0.0277 second. The packet length follows a Pareto with cutoff distribution with parameters alpha = 1.1, minimum packet size 81.5 bytes and maximum packet size 6000 bytes. The average time between WWW sessions is 0.1 second (i.e., it is assumed that a user is continuously generating sessions). Two interactive user classes, namely, Class 1 and Class 2, have been included, as representatives of two different user profiles, with maximum allowed delays of 120 milliseconds and 60 milliseconds, respectively; 60% of the users belong to class 1 and 40% to class 2.

By setting the parameter $\theta > 0$ for the OBR in (4), queues are forced to be emptied faster than for $\theta = 0$, which is particularly true for low-loaded systems. However, there is not an optimum θ setting unique for all the loads. Then, from the obtained results, $\theta = 0.6$ has been retained as a satisfactory value in all the studied cases. Let us notice that, in general, high values of θ could end up at assigning bandwidth in excess to some users in detriment of others. This is clearly pointed out in the RSS1 scheme, where the first-ordered users could be provided with an excessive bandwidth (and actually not required), which would prevent the allocation to other users in the ordered list.

Figure 3 plots the comparison between the considered strategies in terms of packet dropping probability for class-1 users as a function of the total number of users in the scenario (similar results not shown here for the sake of brevity would be observed for class-2 users). It can be observed that the worst performance is obtained with the RSS1 scheme, and that the two approaches based on HNN are able to outperform both RSS1 and RSS2 strategies, thanks to the consideration of both queuing time constraints and channel status in the optimization carried out by hopfield neural networks. Notice that, for low dropping probability values, the reduction achieved by HNN-based strategies is in around one order of magnitude with respect to both RSS1 and RSS2. In that respect, notice also that the energy function from HNN2 is able to achieve always a lower dropping probability than the energy function from HNN1. Equivalently, the performance in terms of dropping probability can be translated into a certain system capacity (i.e., maximum number of users that the system can handle for a certain maximum dropping probability of, for example, 1%). Specifically, while RSS2 would exhibit a capacity of around 1200 users, in the case of HNN2 the capacity is increased up to around 1350 users (i.e., a capacity gain of 13%).

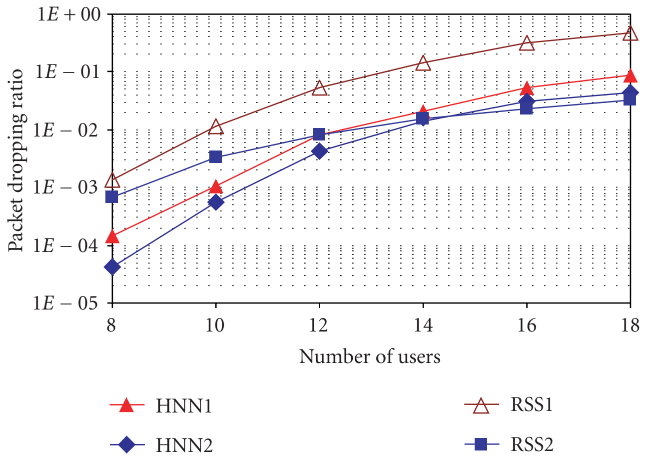

Figure 3: Packet dropping ratio for class-1 users as a function of the number of users in the scenario.

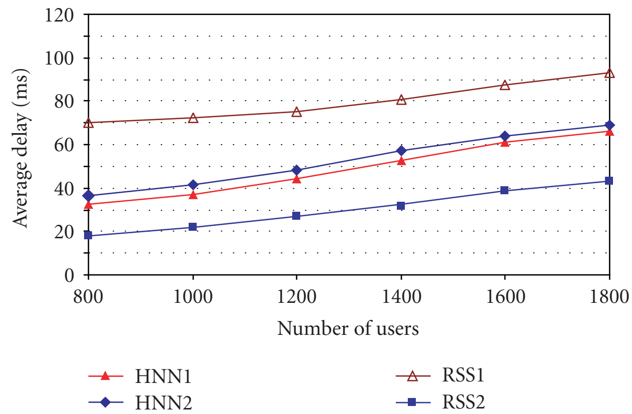

Figure 4: Average packet delay for class-1 users as a function of the number of users in the scenario.

With respect to the performance on average terms, Figure 4 compares the average packet delay measured for class-1 users with different approaches. In this case, the comparison reveals that HNN-based approaches achieve an average delay that lies between the RSS1 and RSS2 schemes. However, Figure 5 which plots the ratio between standard deviation and average delay for each strategy indicates that RSS2 is actually the strategy with the highest dispersion in terms of delay, which eventually justifies that, in spite of having a good performance on average terms, the packet dropping ratio is higher than with the HNN-based algorithms. Consequently, whenever delay-sensitive traffic is considered, the performance should not be optimized only on average terms but also specific conditions in terms of maximum allowed delays should be considered. Finally, it is worth mentioning

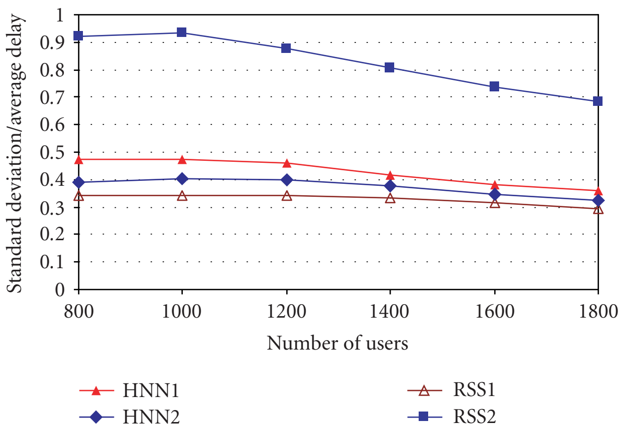

Figure 5: Ratio between the standard deviation and the average of the packet delay for class-1 users.

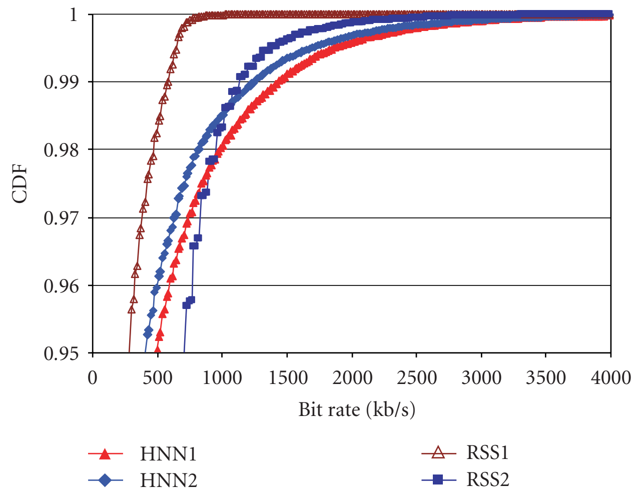

Figure 6: CDF of the allocated bit rate for class-1 users for the different strategies with 1000 users in the scenario.

that actually both HNN1 and HNN2 provide just an upper bound of the dropping probability due to the above-mentioned truncation of the iterative Euler technique and the existence of a minimum in the energy function. So, even better results could be expected by exploring other numerical solutions or alternative improved energy function definitions, what is left for future work.

As an illustrative result of how the different algorithms operate, Figure 6 plots the cumulative distribution function (CDF) of the bit rate allocated per user with the different approaches for a situation with 1000 users in the scenario (for illustrative purposes, only the bit rate of class-1 users is presented, but the performance for class-2 users would be similar). Actually, only the most relevant part of CDF relative to the highest percentile is stressed in Figure 6 to better differentiate the reference and the HNN-based scheduler algorithms operation. It can be observed that both HNN-based strategies are able to make allocations of higher bit rates, thanks to the HNN-optimization accounting for the joint queue and channel status, which ensures that the subcarriers are allocated to the most suitable users. In that respect, the main difference between HNN1 and HNN2 would be for the lowest bit rates (i.e., below 30 kb/s, not shown in the graph, and where the crossing point between the HNN1 and HNN2 curves occurs), in which HNN1 would exhibit a higher probability than HNN2 of allocating low bit rates.

Finally, Figure 7 plots the comparison in terms of the CDF of the total allocated bandwidth obtained with HNN1 with respect to the total requested bandwidth (i.e., the sum of all the OBRs of the different users) for the cases with 1200 users and 1600 users. It can be observed how the total requested bandwidth increases with the number of users, but the total allocated bandwidth remains approximately the same; meaning that the system has reached its maximum capacity. However, in spite of the fact that the total requested bandwidth is higher than the total allocated bandwidth, the algorithm carries out a smart allocation that keeps the packet dropping probability at low values, as illustrated in Figure 3. Similar results are obtained with HNN2.

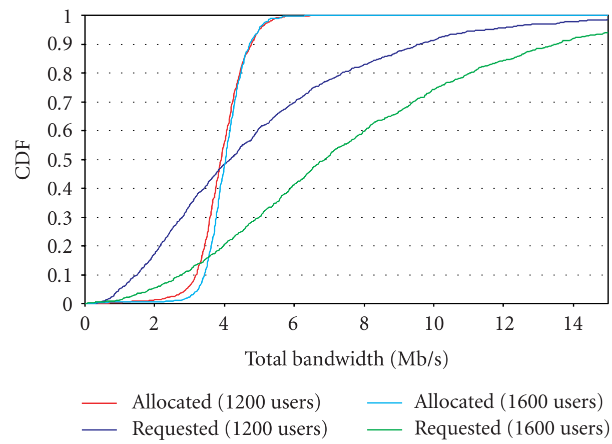

Figure 7: Cumulative distribution function of the total allocated and requested bandwidth for the HNN1 algorithm with 1200 and 1600 users.

6. CONCLUSIONS

This paper has presented a novel strategy to carry out the dynamic resource allocation of subcarriers to users in OFDMA systems with delay-sensitive service in which packets should be transmitted within a specific maximum delay bound. It is based on hopfield neural network methodology which is a powerful optimization technique and takes into account both service-class constraints in terms of maximum allowed delay as well as channel capacity limitation in each subcarrier. Actually, HNN methodology has been carried out by solving iteratively a numerical differential equation having a hardware implementation in mind, and the different delay requirements are captured in the form of an energy function that is minimized by the algorithm. In that respect, two different energy functions have been analyzed by means of simulations and compared against two reference schemes revealing a better behavior in terms of packet-dropping probability, which eventually turns into system capacity increase. Specifically, capacity gains of around 13% for a maximum dropping probability of 1% have been observed with respect to a representative state-of-the-art algorithm existing in the literature.

APPENDIX

INTERCONNECTION MATRIX AND INPUT BIAS CURRENT FOR THE PROPOSED HNN MODEL

This appendix presents the relationship between the energy functions considered in the two HNN-based algorithms and the general expression of the energy function for an HNN given in (9), so that both the interconnection matrix $T_{ij} = [T_{ij,pq}]_{\substack{p=1,\dots,N\\ q=1,\dots,S}}$ and the input bias current I_{ij} can be obtained.

In order to make the derivation valid for both HNN1 and HNN2 algorithms, let us consider the following common definition of the energy function E:

$$E = \frac{\mu_1}{2}\sum_{i=1}^{N}\omega_i\left(1-\frac{\sum_{j=1}^{S}c_{ij}V_{ij}\Delta f}{\overline{m_i}}\right)^2+\frac{\mu_2}{2}\sum_{i=1}^{N}\sum_{j=1}^{S}\psi_{ij}V_{ij} + \frac{\mu_3}{2}\sum_{i=1}^{N}\sum_{j=1}^{S}V_{ij}(1-V_{ij})+\frac{\mu_4}{2}\sum_{j=1}^{S}\left(1-\sum_{i=1}^{N}V_{ij}\right)^2. \tag{A.1}$$

Notice that, the energy function of HNN1 in (11) is obtained by taking $\omega_i = 1$ in (A.1), while the energy function of HNN2 in (15) is obtained by taking $\mu_2 = 0$ in (A.1).

For the energy function defined as (15) and a given $V_{i^*j^*}$ neuron we obtain

$$\begin{aligned}\frac{\partial E}{\partial V_{i^*j^*}} &= \frac{\mu_1}{2}\sum_{i=1}^{N}2\omega_i\left(1-\frac{\sum_{j=1}^{S}c_{ij}V_{ij}\Delta f}{\overline{m_i}}\right)\\ &\quad\times\left(-\sum_{j=1}^{S}\frac{c_{ij}\Delta f}{\overline{m_i}}\frac{\partial V_{ij}}{\partial V_{i^*j^*}}\right)+\frac{\mu_2}{2}\sum_{i=1}^{N}\sum_{j=1}^{S}\psi_{ij}\frac{\partial V_{i,j}}{\partial V_{i^*j^*}}\\ &\quad+\frac{\mu_3}{2}\sum_{i=1}^{N}\sum_{j=1}^{S}\left[\frac{\partial V_{ij}}{\partial V_{i^*j^*}}(1-V_{ij})+V_{ij}\frac{\partial(1-V_{i,j})}{\partial V_{i^*j^*}}\right]\\ &\quad+\frac{\mu_4}{2}\sum_{j=1}^{S}2\left(1-\sum_{i=1}^{N}V_{ij}\right)\left(-\sum_{i=1}^{N}\frac{\partial V_{ij}}{\partial V_{i^*j^*}}\right).\end{aligned} \tag{A.2}$$

Furthermore, since $\partial V_{ij}/\partial V_{i^*j^*} = 1$ for $i = i^*$, $j = j^*$, and $\partial V_{ij}/\partial V_{i^*j^*} = 0$ for $i \neq i^*$, $j \neq j^*$ (A.2) can be expressed as

$$\begin{aligned}\frac{\partial E}{\partial V_{i^*j^*}} &= -\mu_1\frac{c_{i^*j^*}\Delta f\omega_{i^*}}{\overline{m_{i^*}}}\left(1-\frac{\sum_{j=1}^{S}c_{i^*j}V_{i^*j}\Delta f}{\overline{m_{i^*}}}\right)+\frac{\mu_2}{2}\psi_{i^*j^*}\\ &\quad+\frac{\mu_3}{2}(1-2V_{i^*j^*})-\mu_4\left(1-\sum_{i=1}^{N}V_{ij^*}\right).\end{aligned} \tag{A.3}$$

By substituting (A.3) into (10) it is obtained that

$$\begin{aligned}\frac{\partial E}{\partial U_{i^*,j^*}} &= -\frac{U_{i^*j^*}}{\tau}+\mu_1\frac{c_{i^*j^*}\Delta f\omega_{i^*}}{\overline{m_{i^*}}}\left(1-\frac{\sum_{j=1}^{S}c_{i^*j}V_{i^*j}\Delta f}{\overline{m_{i^*}}}\right)\\ &\quad-\frac{\mu_2}{2}\psi_{i^*j^*}-\frac{\mu_3}{2}(1-2V_{i^*j^*})+\mu_4\left(1-\sum_{i=1}^{N}V_{ij^*}\right).\end{aligned} \tag{A.4}$$

By identifying the coefficients in (A.4) with the corresponding coefficients in (8), it is possible to obtain the interconnection weights and bias currents as

$$\begin{aligned}T_{ij,pq} &= -\mu_1\frac{c_{ij}c_{iq}(\Delta f)^2\omega_i}{(\overline{m_i})^2}\delta_{ip}+\mu_3\delta_{ip}\delta_{jq}-\mu_4\delta_{jq},\\ I_{ij} &= \mu_1\frac{c_{ij}\Delta f\omega_i}{\overline{m_i}}-\frac{\mu_2}{2}\psi_{ij}-\frac{\mu_3}{2}+\mu_4,\end{aligned} \tag{A.5}$$

where function δ_{ip} is 1 if $i = p$ and 0 otherwise.

It is worth mentioning that a solution for the selection of the optimal bit rate per user can be easily performed simply by changing the input bias current I_{ij} and the interconnections values $T_{ij,pq}$ at a frame basis.

In order to have minimum points with respect to output voltages V_{ij} of neurons, it is necessary that the second derivatives be positive, or equivalently

$$\frac{\partial^2 E}{\partial V_{ij}^2} > 0 \Longleftrightarrow \mu_1\frac{c_{ij}c_{iq}(\Delta f)^2}{(\overline{m_i})^2}-\mu_3+\mu_4 > 0, \tag{A.6}$$

Condition (A.6) is satisfied if we ensure always that $-\mu_3 + \mu_4 > 0$, which yields the following relationship between the parameters of the energy function

$$\mu_3 < \mu_4. \tag{A.7}$$

ACKNOWLEDGMENTS

This work has been partially funded by the European Network of Excellence NEWCOM (Contract no. 507325) and by the Generalitat de Catalunya under Contract no. AGAUR 2005SGR00197.

REFERENCES

[1] 3GPP, R1-050779, Texas Instruments, Throughput Evaluations in EUTRA OFDMA Downlink, 2005.

[2] C. Y. Wong, R. S. Cheng, K. B. letaief, and R. D. Murch, "Multiuser OFDM with adaptive subcarrier, bit, and power allocation," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 10, pp. 1747–1758, 1999.

[3] G. Li and H. Liu, "Dowlink radio resources allocation for multi-cell OFDMA systems," *IEEE Transactions on Wireless Communications*, vol. 5, no. 12, pp. 3451–3459, 2006.

[4] S. Kittipiyakul and T. Javidi, "Resource allocation in OFDMA with time-varying channel and bursty arrivals," *IEEE Communications Letters*, vol. 11, no. 9, pp. 708–710, 2007.

[5] J. Gross and M. Bohge, "Dynamic mechanisms in OFDM wireless systems: a survey on mathematical and system engineering contribution's," TKN Technical Report Series TKN-06-001, Telecommunication Networks Group, Technische Universität Berlin, Berlin, Germany, May 2006, http://www.tkn.tu-berlin.de/publications/papers/TKN_http:Report_06_001.pdf.

[6] K. Kim, H. Kang, and K. Kim, "Providing quality of service in adaptive resource allocation for OFDMA systems," in *Proceedings of the 59th IEEE Vehicular Technology Conference (VTC '04)*, vol. 3, pp. 1612–1615, Milan, Italy, May 2004.

[7] A. Pandharipande, M. Kounrouris, H. Yang, and H. Park, "Subcarrier allocation schemes for multiuser OFDM systems," in *Proceedings of the International Conference on Signal Processing and Communications (SPCOM '04)*, pp. 540–544, Bangalore, India, December 2004.

[8] Z. Shen, J. G. Andrews, and B. L. Evans, "Optimal power allocation in multiuser OFDM systems," in *Proceedings of IEEE Global Telecommunications Conference (GLOBECOM '03)*, vol. 1, pp. 337–341, San Francisco, Calif, USA, December 2003.

[9] J. Jang and K. B. Lee, "Transmit power adaptation for multiuser OFDM systems," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 2, pp. 171–178, 2003.

[10] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: information theoretic and communications aspects," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2619–2692, 1998.

[11] A. K. F. Khattab and K. M. F. Elsayed, "Opportunistic scheduling of delay sensitive traffic in OFDMA-based wireless," in *Proceedings of the IEEE International Symposium on a World of Wireless Mobile and Multimedia Networks (WoWMoM '06)*, pp. 279–288, Buffalo, NY, USA, June 2006.

[12] G. Song, L. J. Cimnini Jr., and H. Zheng, "Joint channel-aware and queue-aware data scheduling in multiple shared wireless channels," in *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC '04)*, vol. 3, pp. 1939–1944, Atlanta, Ga, USA, March 2004.

[13] J. Gross, J. Klaue, H. Karl, and A. Wolisz, "Subcarrier allocation for variable bit rate video streams in wireless OFDM systems," in *Proccedings of the 58th IEEE Vehicular Technology Conference (VTC '03)*, vol. 4, pp. 2481–2485, Orlando, Fla, USA, October 2003.

[14] Y. J. Zhang and K. B. Letaief, "Adaptive resource allocation and scheduling for multiuser packet-based OFDMA networks," in *Proceedings of IEEE International Conference on Communications (ICC '04)*, vol. 5, pp. 2949–2953, Paris, France, June 2004.

[15] C. Wook 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, 2004.

[16] N. García, R. Agustí, and J. Pérez-Romero, "A user-centric approach for dynamic resource allocation in CDMA systems based on Hopfield neural networks," in *Proceedings of the 14th IST Mobile & Wireless Communications Summit*, Dresden, Germany, June 2005.

[17] S. Abe, *Neural Networks and Fuzzy Systems: Theory and Applications*, Kluwer Academic Publishers, Norwell, Mass, USA, 1997.

[18] M. Wakamura and Y. Maeda, "FPGA implementation of Hopfield neural network via simultaneous perturbation rule," in *Proceedings of the 41st SICE Annual Conference (SICE '03)*, vol. 2, pp. 1272–1275, Fuki, Japan, August 2003.

[19] S. T. Chung and A. J. Goldsmith, "Degrees of freedom in adaptive modulation: a unified view," *IEEE Transactions on Communications*, vol. 49, no. 9, pp. 1561–1571, 2001.

[20] J. J. Hopfield, "Neurons with graded response have collective computational properties like those of two-state neurons," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 81, no. 10, pp. 3088–3092, 1984.

[21] E. Del Re, R. Fantacci, and L. Ronga, "A dynamic channel allocation technique based on Hopfield neural networks," *IEEE Transactions on Vehicular Technology*, vol. 45, no. 1, pp. 26–32, 1996.

[22] UMTS 30.03 v3.2.0 TR 101 112, "Selection procedures for the choice of radio transmission technologies of the UMTS," ETSI, April 1998.