Self-Organized Resource allocation for LTE Pico Cells: A Reinforcement Learning Approach

Afef Feki, Veronique Capdevielle
Alcatel-Lucent Bell Labs, France
Email: {afef.feki, veronique.capdevielle}@alcatel-lucent.com

Elom Sorsy
Telecom Bretagne
Email: komlan.sorsy@telecom-bretagne.eu

***Abstract*—This article proposes a smart resource sharing algorithm to manage interference in LTE networks. Relying on reinforcement learning theory, the proposed Cyclic Multi Armed Bandit (CMAB) algorithm steers each cell in the choice of the most suitable frequency band portions, in autonomous manner. Thanks to the traffic aware feature, the algorithm adapts as well to the real needs of the cell. Adding to that, a refinement of the decision function is proposed to speed up the convergence time. The proposed approach is tested in LTE compliant simulator and the results shows its efficiency compared to conventional static reuse schemes.**

***Index Terms*—Inter-Cell Interference (ICI), Pico Cell (PC), LTE, Resource allocation, Reinforcement Learning theory**

I. Introduction

The steady increase of traffic demand and the search of more diverse and high quality services poses a real challenge for operators and constructors. In this context, heterogeneous and dense networks are deployed and need to be efficiently operated, within the scarce available frequency resources. We focus mainly on Inter Cell Interference (ICI) which is one of the most inhibiting factor in wireless networks. In Long Term Evolution (LTE) networks, efficient interference management mechanisms are a requirement for reaching the target performances. To this end, 3GPP specified a set of recommendations on Inter Cell Interference Coordination (ICIC) [1]. For example, ICIC concepts suggest coordinating the radio resources use between multiple cells. Thus, eNodeBs (eNBs) can share resource usage status information and interference level over X2 interface through **LOAD INDICATION** procedure [2]. For instance, the following Information Elements (IE) are defined:

- **UL Interference Overload Indication IE** : it indicates the interference level experienced by the indicated cell on all resource blocks, per Physical Resource Block (PRB).
- **UL High Interference Indication IE** : it indicates for a PRB the occurrence of high interference sensitivity.

The receiving eNBs should take this information into account in order to assign PRBs to their attached users. However the amount of information to share should be limited in order to avoid overloading E-UTRAN interfaces.

Adding to that, Pico Cells are gaining increasing interest due to their ability to offload the macro cell and to offer to the end user higher performances in term of data rate and Quality of service. The Pico Cells (PCs) are low-radius cells, characterized mainly by their low cost and their flexibility to deploy. Nevertheless, the expected high density of these PCs, generally deployed in the overlay of macro cells, will for sure increase the problem of ICI. Thus, autonomous ICI management mechanisms are primordial for an efficient co-existence of all the cells, in the same frequency band.

In [3], a reinforcement learning approach has proven its efficiency in steering each cell, in autonomous manner, towards the most appropriate spectral resources, so that interference levels are minimized among all the neighboring cells. This approach inspired from the Multi Armed bandit (MAB) problem resolution was proposed and evaluated in a first version with static users and fixed traffic load conditions. In this study, we propose to enhance this procedure through sequential cells' decision instants and the improvement of the decision function in order to:

1) Accelerate the convergence time of the whole procedure,
2) Take into account the traffic load variation of each cell (and then adapt the chosen resources to the real needs of the cell).

This enhanced approach is tested through an LTE compliant simulator supporting mobility of users.

This paper is organized as follows. Section II describes the proposed approach called as Cyclic Multi Armed Bandit (CMAB) method. In section III, we present the simulation environment and the tested scenario. In section IV, we analyze the performances of our proposed method through the simulation results. Finally, section V yields concluding remarks.

II. Solution overview

An effective ICIC mechanism in dense deployment of cells needs to support the following features:

- **The algorithm should be local to each eNB, with minimal information exchange with the neighboring cells**: In fact, sharing too much information between the eNBs create high load on X2 interfaces. In order to avoid this, each eNB should be able to deduce the interference level and take its decision based only on local indicators.
- **The algorithm must be autonomous** : Since the number of pico cells may be very high, manual operations would be very costly. In consequence self-organizing networks (SON) functionality is necessary.
- **The algorithm must be dynamic** : The algorithm must be able to adapt to the traffic load and the radio condition variations.

To this end, we propose an hierarchical scheduling scheme composed of two levels:

978-1-4673-0990-5/12/$31.00 ©2012 IEEE

- Per cell level: Each cell chooses periodically or upon a trigger, the frequency resources to transmit on it during the upcoming period,
- Per user level: The above chosen resources are then distributed among the attached users, using conventional scheduling algorithm.

This yields to minimal required modification of the current resource management functioning for an implementation of the proposed scheme, as it corresponds only to a limitation of the allowed resources on which the conventional scheduling will be made. The proposed CMAB procedure intervenes at the first level in the form of a deterministic decision procedure that chooses the sub-bands to transmit on it, without any knowledge of the neighboring cells' decision. The main originality of this approach resides in the autonomous operation of each cell to decide on which portion of the frequency band to transmit on, following a predefined set of rules and such that all the neighboring cells, following the same resource sharing procedure, converge to a stable situation.

A. Proposed CMAB algorithm

We consider a network consisting of $\{l\}_{l=1}^{N}$ Pico Cells (PCs), operating on the same frequency band. We propose to divide the total frequency band in a set of sub-bands called $\{SB\}_{j=1}^{K}$. Periodically, each PC_l chooses a SB_j to transmit on it during the upcoming period. The SB corresponds to a set of Resource Blocks (RB). We remind that, in LTE, the RB corresponds to the smallest time-frequency resource that can be allocated to a given user. The proposed CMAB algorithm is a reinforcement learning procedure, functioning in a periodic manner where a set of parameters are updated and a decision function is calculated for each arm (in our case the sub-band). After each period t, each PC_l chooses the best sub-bands to use. These sub-bands are those that maximize the following Decision Function :

$$DF_{j,t}^{l} = \mu_{j,t}^{l} + \sqrt{\frac{2 \times \log(\sum_{j=1}^{K} n_{j,t}^{l})}{n_{j,t}^{l}}} \qquad (1)$$

Where :

- $\mu_{\mathbf{j,t}}^{\mathbf{l}}$ represents the mean reward experienced by the cell PC_l when transmitting on SB_j at time t. It represents an **exploitation** term. It tends to lead the cell to reuse resources that have been already used and that leads to the best performances. The reward corresponds to the resulting gain gathered by the cell (i.e. the mean throughput). A reward model, proposed for CMAB, is detailed in the next section of the article.
- $\mathbf{n}_{\mathbf{j,t}}^{\mathbf{l}}$ corresponds to the number of times PC_l has used SB_j until time t. The lower $\mathbf{n}_{\mathbf{j,t}}^{\mathbf{l}}$ is, the higher is its contribution in the decision function. This has the advantage of promoting the **exploration** of new sub-bands.

Thus, the decision function allows each cell to make a trade-off between exploring the sub-bands and exploiting the cumulated knowledge by choosing the sub-bands identified so far as the best.

It is established in [3] that the proposed first version of the MAB algorithm can converge quickly. In this article, we propose an enhanced CMAB algorithm with sequential decisions among the neighboring cells, in addition to a traffic aware feature.

The state machine (Figure 1) illustrates the global functioning of the proposed CMAB algorithm.

The algorithm begins with an **initialization phase** where each

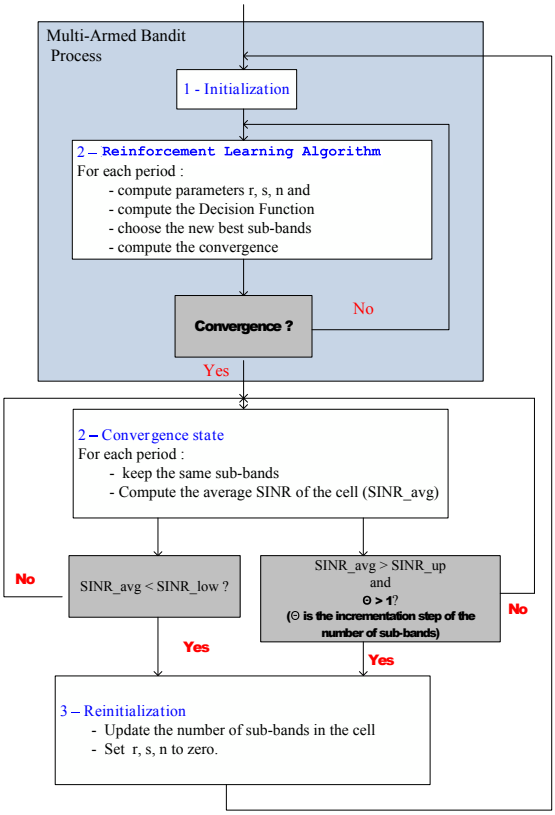

Fig. 1. CMAB Algorithm State Machine

cell uses each existing sub-band at least once in order to set its parameters. For this purpose, a random permutation of the existing sub-bands is generated. At each new period during this initialization phase, the cell chooses the sub-bands according to this permutation.

Thereafter, the cell performs the CMAB Algorithm periodically. At each period t, the cell determines the best sub-bands to use during the next period thanks to the deterministic decision function (equation 1). Here, a convergence test is performed at each period. If the cell reaches a stable state, then a test is made on the average Signal to Interference plus Noise Ratio (SINR) over the attached users and a comparison is made to a predefined threshold. Upon the result of this test, the cell decides to update or not the number of sub-bands to use at each period, which allows better fitting to the real traffic load, as well as channel condition variations. The details of this update will be described in the next section. Finally, a new CMAB process is launched again with a reinitialization phase.

B. Calculation of the reward

In this section, we present the modeling of the instantaneous reward and the mean reward (μ) used in the decision function calculation (equation 1). In the scope of the new algorithm, some modifications are applied to the reward model proposed in [3].

The detailed definitions and formulas are given below.

1) $r^l_{j,t}$: the reward function when PC_l chooses SB_j at time t.

It is calculated as the mean throughput of the cell over all the sub-bands used during the period t and over all the attached users. Let's define :

- m: the index of a given UE attached to the cell at time t (for at least one Transmission Time Interval (TTI)=1 ms);
- M^l_t :the total number of users attached the cell l at time t;
- N_{tti} : the number of TTIs per period;
- tti : the index of a given TTI;
- $tti_{begin} = (t-1) \times N_{tti} + 1$: index of the beginning TTI in the considered period;
- $tti_{end} = t \times N_{tti}$: index of the last TTI in the considered period;
- $thr^j_{m,tti}$: the total throughput of the user m during TTI tti with the assumption that this user was attached to the considered cell at TTI tti and was assigned at least one RB on SB_j. In the case the assumption is not met, the value is zero.
- During the same TTI, a user may be assigned RBs on several sub-bands. Their number is Z^m_{tti}.

Each user m ($m = 1, 2, \ldots, M^l_t$), brings a contribution to the reward $\mathbf{r^l_{j,t}}$. This contribution is :

$$r^{l,m}_{j,t} = \sum_{tti=tti_{begin}}^{tti_{end}} \frac{thr^j_{m,tti}}{Z^m_{tti}}, \tag{2}$$

Where the ratio $\frac{thr^j_{m,tti}}{Z^m_{tti}}$ represents the mean contribution of each sub-band in the total throughput of the user. And the total reward of the sub-band j is :

$$\mathbf{r^l_{j,t}} = (1 \div M^l_t) \times \sum_{m=1}^{M^l_t} r^{l,m}_{j,t} \tag{3}$$

2) $s^l_{j,t}$: the cumulated reward experienced by PC_l when transmitting on SB_j at time t:

$$\begin{cases} \mathbf{s^l_{j,t+1}} = \mathbf{s^l_{j,t}} + \mathbf{r^l_{j,t}} & if\ SB_j\ is\ used\ at\ time\ t; \\ \mathbf{s^l_{j,t+1}} = \mathbf{s^l_{j,t}} & if\ not. \end{cases}$$

With these definitions, the two parameters used to compute the decision function (Equation 1) are given by following expressions :

$$\begin{cases} \mathbf{n^l_{j,t+1}} = \mathbf{n^l_{j,t}} + \mathbf{1} & if\ SB_j\ is\ used\ at\ time\ t; \\ \mathbf{n^l_{j,t+1}} = \mathbf{n^l_{j,t}} & if\ not\ and \\ \mu^{\mathbf{l}}_{\mathbf{j,t+1}} = \frac{\mathbf{s^l_{j,t}}}{\mathbf{n^l_{j,t}}} \end{cases}$$

In order to tune the trade-off between the exploration and the exploitation, a solution is to decrease the order of magnitude of the reward $\mathbf{r^l_{j,t}}$ and consequently of $\mu^{\mathbf{l}}_{\mathbf{j,t}}$. Therefore, we propose to normalize the reward, dividing it by the maximum throughput achievable on a resource block per TTI (Thr_{max}). Thus, the normalized reward is computed as follows :

$\mathbf{Nomalized\ r^l_{j,t}} = \beta \times \frac{\mathbf{r^l_{j,t}}}{\mathbf{Thr_{max}}}$ where β is a tuning factor.

C. Convergence of MAB Process

In this study, the convergence metric is based on the frequency of the sub-bands choices (calculated in terms of percentage). Here, a sliding window is used in order to consider only recent periods. Convergence is decided by comparing the highest percentage (corresponding to the most used sub-band) to a predefined threshold.

Once the CMAB process has converged, the cell stops executing the algorithm during the next periods, as depicted by the state machine (Figure 1). A new CMAB process will be launched depending on the traffic load and radio channel quality.

D. Sequential triggering for resource updating decisions

In the original MAB procedure[3], all the cells perform their resource selection at the same time. For CMAB, we propose a sequential update as depicted in Figure 2. This accelerates the convergence of the algorithm since the resources changes in each cell is taken automatically into account by the neighbor cells. Nevertheless, a minimal information exchange is necessary between the neighboring cells to set up this sequential mechanism. This can be achieved through a simple token passing protocol for example at a very low time periodicity.

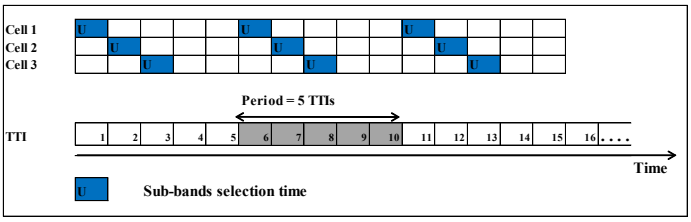

Fig. 2. Concept of Sequential resource updating in CMAB

E. Load aware update

In real networks, the traffic load is subject to changes. Therefore, it is needed to adapt the number of sub-bands selected by the cell to its actual traffic load. This adjustment is performed only after convergence and depends on the overall radio channel quality in the cell. This functionality is inspired from a different algorithm that we proposed in [4].

The number of sub-bands in use in the cell is increased under good radio conditions and decreased in bad conditions. The step of variation for the cell l is proportional to the traffic load which is represented by the number of users. It is defined by:

$$\Theta = E(\alpha \times K \times \frac{M_l}{M}) \tag{4}$$

Where :

- M_l is the number of users attached to the cell PC_l;
- M is the total number of users in the cluster of cells;

- K is the total number of sub-bands;
- α is a tuning factor and
- $E(x)$ is the integer part of x.

This way, cells with high traffic demand preempt the available resources. Two SINR thresholds ($SINR_{low}$ and $SINR_{up}$) are used to estimate the average channel quality in the cell. The objective is to keep the average SINR in the cell between these two bounds. Let's define :

- Z_l the number of sub-bands that are being used by the cell PC_l;
- γ_l,the mean SINR in the cell over the attached users to PC_l. The SINRs are averaged with **MIC (Mean Instantaneous Capacity)** model [5].

Then, the algorithm is the following:

1. If $\gamma_l < SINR_{low}$,
   $Z_l = Z_l - \Theta$ with $Z_l \geq 1$ In this case, many interferers contend for the same resources. Thus, the cell needs to decrease the number of chosen sub-bands, in proportion to their real needs (with the parameter Θ).
2. If $\gamma_l > SINR_{up}$ *and* $\Theta > 1$,
   $Z_l = Z_l + \Theta$;

In our simulations, the used threshold values are ($SINR_{low}$,$SINR_{up}$) = ($10dB$,$20dB$).

III. Simulation overview

The proposed algorithm is implemented in an LTE compliant system level simulator [6] in order to analyze its performances. Table I summarizes the system characteristics and simulation scenarios parameters. Here, we consider homogeneous pico cell networks with 19 omni-directional hexagonal cells. Figure 3 displays the network layout with the coverage areas of the cells.

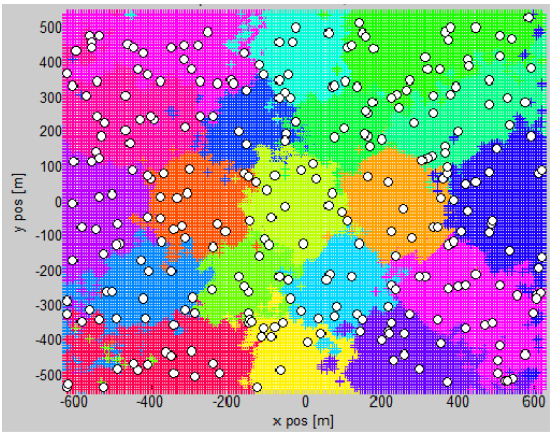

Fig. 3. ICIC Simulation Network Layout

IV. Performance evaluation

A. Convergence and Performance Analysis

1) Impact of reward normalization and sequential updates: In this section, we analyze the convergence duration (the duration until stabilization) of the selected resources as well as the performances in terms of SINR. SINR is evaluated over the RBs used for the transport blocks of scheduled users. We compare three algorithms derived from CMAB: **CMAB-Sync**, which stands for CMAB-Synchronous when update decisions are taken simultaneously by all the cells [3], **CMAB-Sync-Norm** which is an upgrade of CMAB-Sync with normalization of the reward and finally **CMAB-Norm-Seq** with normalized reward and sequential update decisions. As indicated in the previous section, the normalization of the reward is used to tune the trade-off between exploration and exploitation. Decreasing the order of magnitude of the reward increases the duration of the exploration phase towards optimal resource selection. Referring to Figure 4, around 1.5dB SINR gain is achieved at the expense of increase in convergence duration. As a solution in CMAB-Seq-Norm, updating decisions are taken sequentially. In this way, when a cell selects its frequency resources, its surrounding cells potentially impacted by this decision have to trigger new resource selection. As illustrated in Figure 4, the convergence duration of CMAB-Seq-Norm is divided by 2 if compared to CMAB-Sync-Norm and a performance gain up to 3dB is obtained compared to nominal CMAB. The normalized reward jointly with sequencing feature optimize performance while speeding up convergence even in mobility situation of the users.

Network parameters	
Number of cells(L)	19 omnidirectional cells
Cell radius	150m (pico cells)
LTE parameters	
Frequency band	2.0 GHz
Bandwidth	5 MHz
Number of PRBs	25
ENodeB and UE characteristics	
Scheduling	Proportional fair.
Transmission mode	MIMO 2x2, 2 layers
Enode B TX Power	2x125mW over all the 5MHz band
UE distribution	Uniform
Total number of UEs	311
UE speed	3 km/h; straight walking in random direction
Traffic profile	Full buffer, best effort
Radio channel	
Pathloss model	TS36.92
Shadow	Log normal, 5 dB standard deviation.
Fast fading	ITU Pedestrian A.

TABLE I
Simulation environment

2) Comparison of CMAB to static reuse schemes: In this section, we compare CMAB variants: CMAB configured with 3 eligible sub-bands (CMAB 1/3), CMAB configured with 2 eligible sub-bands (CMAB 1/2) with Static Reuse schemes: reuse 1 (all the cells operate on the total bandwidth) and reuse 3 (each cell is allocated fixed 1/3 of the total bandwidth). We analyze the performances in terms of SINR as well as instantaneous throughput experienced by the scheduled mobile users (3km/h) when they are in the central cell of the uniform hexagonal network layout (Figure 3). We also observe the performance of CMAB1/3 considering only post-convergence phase. Simulations have shown that for 90 percent of users, the minimum post-scheduling SINR is about 7dB with CMAB 1/3 after convergence, which is about 9dB higher than with static reuse 1 (The SINR figure is not included for lack of

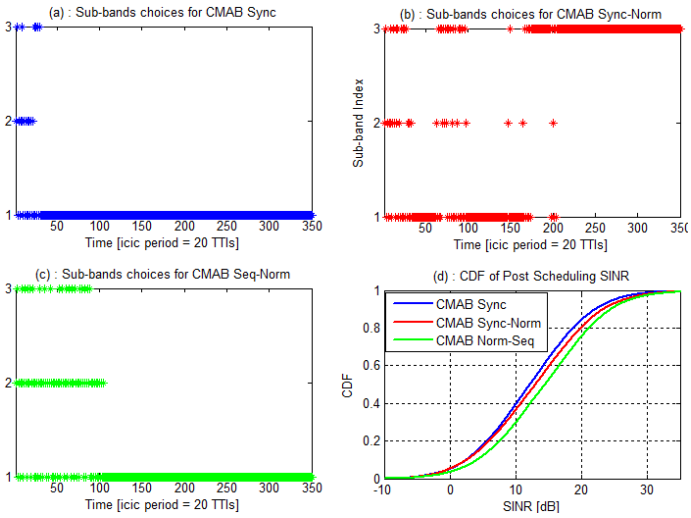

Fig. 4. CMAB convergence and performance analysis

space).

Added to this, in the considered scenario, we observe that performances of CMAB 1/3 (including convergence phase) equals Static reuse 3. It has to be noted that CMAB reaches these performances in a fully autonomous way. Figure 5 depicts the instantaneous throughput performances of the algorithms. It shows that CMAB 1/3 schemes and static reuse 3 outperform as expected the other static reuse 1 and even CMAB 1/2. Indeed, reuse 1 and 2 are sub-optimal in the uniform deployment model. CMAB finds out the optimum frequency reuse 3 scheme in this case, in a self directed way.

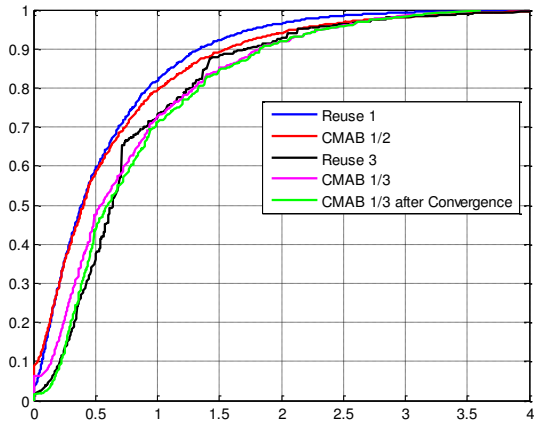

Fig. 5. CMAB vs Static Reuse patterms - Throughput distributions

B. Evaluation of the Traffic load aware functionality

This section illustrates the feature that automatically adjusts the number of resources in use to the current traffic load. The same network parameters are used but here, the users mobility model is changed so that the users converge towards the central cell. Their speed is 50 km/h. We compare the performances of CMAB 1/3 to Static Reuses 1 and 3. For CMAB 1/3, we consider a tuning factor of 2 in Equation 4 for the incrementation step of the number of sub-bands. The simulation duration is 7s. Referring to the left side of Figure 6, we can notice the continuous linear increase of the load. At 6.5s, the number of sub-bands in use by the central cell is doubled. As illustrated in the Figure 6 (right side), thanks to this feature that adjusts the resources volume depending on the actual traffic load, users experience higher throughputs compared to static reuse 1 and 3. This feature is particularly well suited in mobile networks with varying traffic loads.

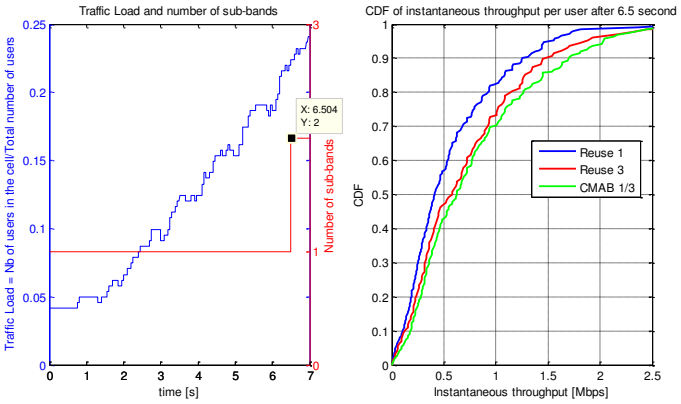

Fig. 6. Performance of traffic load aware

V. Conclusion

In this article, we propose a smart resource sharing algorithm to manage interference in LTE networks. Relying on reinforcement learning theory, the proposed Cyclic Multi Armed Bandit (CMAB) algorithm steers each cell in the choice of the most suitable frequency band portions. The proposed algorithm mainly relies on local information: minimal exchange of information is needed between the cells. The algorithm is autonomous: each cell selects its resources, by its own, without manual intervention which considerably alleviates operational cost. As demonstrated by simulations, thanks to the traffic aware feature, the algorithm can adapt in a dynamic way to varying traffic loads, which is a requisite in mobile radio networks. In addition, we show that a fine tuning of the reward function jointly with sequential resource selection decisions enhances performances while speeding up convergence. The proposed CMAB has been compared to static resource sharing schemes in a dense network and we have shown that it automatically computes the optimum frequency reuse scheme, in a fully autonomous way.

References

[1] 3GPP TS 36.300, *Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN) ; Overall description ; Stage 2 (Release 9)*, 9.5.0 ed., 2010.

[2] 3GPP TS 36.423, *Evolved Universal Terrestrial Radio Access Network (E-UTRAN) ; X2 application protocol (X2AP) (Release 9)*, 9.4.0 ed., 2010.

[3] A. Feki, V. Capdevielle, *Autonomous Resource Allocation for dense LTE networks : A Multi-Armed Bandit formulation*, IEEE. PIMRC, 2011.

[4] A. Feki, V. Capdevielle and E. Temer, *Enhanced Resource Sharing Strategies for LTE Pico cells with Heterogeneous Traffic Loads*, IEEE. VTC Spring, 2011.

[5] Wimax Forum, WiMAX System Evaluation Methodology, v2.0 ed., Dec. 2007.

[6] C. Mehlfuhrer, et all., *The Vienna LTE Simulators - Enabling Reproducibility in Wireless Communications Research*, available at *http : //publik.tuwien.ac.at/files/PubDat_*199104.*pdf*.