Engineering Applications of Artificial Intelligence 55 (2016) 239–249

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

# Cognitive spectrum management in dynamic cellular environments: A case-based Q-learning approach

N. Morozs *, T. Clarke, D. Grace

*Department of Electronics, University of York, York YO10 5DD, UK*

ARTICLE INFO

*Article history:*
Received 28 January 2016
Received in revised form
29 May 2016
Accepted 7 July 2016
Available online 26 July 2016

*Keywords:*
Reinforcement learning
Case-based reasoning
Dynamic spectrum access
Cellular networks

ABSTRACT

This paper examines how novel cellular system architectures and intelligent spectrum management techniques can be used to play a key role in accommodating the exponentially increasing demand for mobile data capacity in the near future. A significant challenge faced by the artificial intelligence methods applied to such flexible wireless communication systems is their dynamic nature, e.g. network topologies that change over time. This paper proposes an intelligent case-based Q-learning method for dynamic spectrum access (DSA) which improves and stabilises the performance of cognitive cellular systems with dynamic topologies. The proposed approach is the combination of classical distributed Q-learning and a novel implementation of case-based reasoning which aims to facilitate a number of learning processes running in parallel. Large scale simulations of a stadium small cell network show that the proposed case-based Q-learning approach achieves a consistent improvement in the system quality of service (QoS) under dynamic and asymmetric network topology and traffic load conditions. Simulations of a secondary spectrum sharing scenario show that the cognitive cellular system that employs the proposed case-based Q-learning DSA scheme is able to accommodate a 28-fold increase in the total primary and secondary system throughput, but with no need for additional spectrum and with no degradation in the primary user QoS.



## 1. Introduction

One of the fundamental tasks of a cellular system is spectrum management. It is concerned with dividing the available spectrum into a set of resource blocks or subchannels and assigning them to voice calls and data transmissions in a way which provides a good quality of service (QoS) to the users. Flexible dynamic spectrum access (DSA) techniques play a key role in utilising the given spectrum efficiently in the face of an ever increasing demand for mobile data capacity. This has given rise to novel wireless communication systems such as cognitive radio networks (Sun et al., 2013) and cognitive cellular systems (Guizani et al., 2015; Sachs et al., 2010). Such networks employ intelligent opportunistic DSA techniques that allow them to access licensed spectrum underutilised by the incumbent users.

The classical and most common application of spectrum sharing in cognitive radio networks to date is the use of the TV white spaces. Such networks reuse the spectrum allocated to TV broadcasters for other wireless communications, whilst eliminating harmful interference to the incumbent TV receivers, e.g. Ghosh et al. (2011) and Gurney et al. (2008). A more recent problem investigated by researchers, mobile network operators (MNOs) and regulators is Long Term Evolution (LTE) and LTE-Advanced spectrum sharing (Matinmikko et al., 2014). In many cases LTE spectrum sharing is required by two or more co-primary MNOs. This can be facilitated by an emerging framework known as licensed shared access (LSA) (Matinmikko et al., 2014). Here, licenses for the use of LTE spectrum are issued upon agreement for a specific geographical area and required time duration. Another type of LTE spectrum sharing actively investigated within the LTE research community, is resource allocation in heterogeneous networks (HetNets) consisting of LTE femto-cells overlapped by a high power macro-cell, e.g. Alnwaimi et al. (2015) and Hamouda et al. (2014). In these scenarios, the problem is often tackled by using game theory or machine learning principles. The LSA method is a static regulatory approach to spectrum sharing, whereas the HetNet problems normally consider a dynamic scenario, where the same LTE channel is used by both the macro-cell and the femto-cells. Both of these spectrum sharing scenarios are investigated in this paper.

*Abbreviations:* 2ON, Second order neighbourhood; CBR, Case-based reasoning; DSA, Dynamic spectrum access; eNB, Evolved node B (LTE base station); ICIC, inter-cell interference coordination; LSA, Licensed shared access; RL, Reinforcement learning; UE, User equipment; UT, User throughput; WoLF, Win-or-learn-fast

* Corresponding author.
*E-mail addresses:* nils.morozs@york.ac.uk (N. Morozs), tim.clarke@york.ac.uk (T. Clarke), david.grace@york.ac.uk (D. Grace).

http://dx.doi.org/10.1016/j.engappai.2016.07.002
0952-1976/

An emerging state-of-the-art technique for intelligent DSA is reinforcement learning (RL); a machine learning technique aimed at building up solutions to decision problems only through trial-and-error, e.g. Malialis and Kudenko ( 2015) and Walraven et al. (2016). It has been successfully applied in a range of wireless network scenarios, such as cognitive radio networks (Jiang et al., 2011), small cell networks (Bennis et al., 2013; Morozs et al., 2016), cognitive wireless mesh networks (Chen et al., 2013), and wireless sensor networks (Chu et al., 2015). The most widely used RL algorithm in both artificial intelligence and wireless communications domains is *Q*-learning (Watkins, 1989). Therefore, most of the literature on RL based DSA focuses on *Q*-learning and its variations, e.g. Chen et al. (2013) and Morozs et al. (2015). The novel algorithm developed in this paper employs distributed *Q*-learning based DSA. The distributed *Q*-learning approach has advantages over centralised methods in that no communication overhead is required to achieve the learning objective, and the network operation does not rely on a single computing unit. It also allows for easier insertion and removal of base stations from the network, if necessary. For example, such flexible opportunistic protocols are well suited to disaster relief and temporary event networks. There, rapidly deployable network architectures with variable topologies are required to supplement the existing wireless infrastructure (Gomez et al., 2016).

The purpose of this paper is to propose an algorithm that combines distributed RL with case-based reasoning (CBR) to improve the stability of intelligent DSA algorithms in realistic, dynamically changing cellular environments, i.e. the type of environments rarely considered in the research literature. The key contributions of this paper are the following:

- First, we present a detailed formulation of the case-based RL framework designed for dynamic learning environments in general.
- We then use this framework to develop the case-based *Q*-learning algorithm for DSA in cellular networks with dynamically changing topologies.
- The proposed algorithm includes a novel network topology based case identification and retrieval mechanism; the two essential components of all CBR systems.
- Finally, we present the results of an extensive empirical evaluation of the proposed scheme using a novel simulation model of a large-scale dynamic wireless environment.

Similar combinations of RL and CBR have already been successfully applied to various decision problems, e.g. dynamic inventory control (Jiang and Sheng, 2009), RoboCup Soccer (Celiberto et al., 2012) and control of a simulated mountain car (Bianchi et al., 2015). For example, Jiang and Sheng (2009) propose an effective case-based RL algorithm, where CBR is used for analysing the similarity between different states of a dynamic multi-agent RL problem. Celiberto et al. (2012) and Bianchi et al. (2015) develop transfer learning algorithms that transfer knowledge between similar learning tasks whilst using CBR to make this process faster. However, the only example of applying this methodology in the wireless communications domain is proposed by us in Morozs et al. (2013). There, a DSA scheme is designed for an unrealistically small and generic cellular network with its own dedicated spectrum, i.e. without secondary spectrum sharing and the presence of the primary users.

The rest of the paper is organised as follows: Section 2 describes the dynamic cellular environments considered in this study, that justify the need for robust intelligent DSA algorithms. Section 3 introduces the classical distributed *Q*-learning approach to DSA. In Section 4 we propose our case-based *Q*-learning algorithm, including novel case identification and case retrieval mechanisms. The results from a number of large-scale simulation experiments are discussed in depth in Section 5. Finally, the conclusions are given in Section 6.

## 2. Dynamic cellular environments

The aim of this paper is to investigate the applications of intelligent DSA in dynamic cellular environments. This section introduces the problem that provides such a challenging learning environment for DSA algorithms.

### 2.1. Heterogeneous temporary event networks

The DSA problem investigated in this paper is currently considered in the EU FP7 ABSOLUTE project. It is designed for a stadium event scenario and involves a temporary cognitive cellular infrastructure that is deployed in and around a stadium to provide extra capacity and coverage to the mobile subscribers and event organisers involved in a temporary event, e.g. a football match or a concert (Reynaud et al., 2014). This scenario is depicted in Fig. 1. There, a small cell network is deployed inside the stadium to provide ultra high capacity density to the event attendees, and an eNodeB (eNB) on an aerial platform is deployed above the stadium to provide wide area coverage.

We consider two different spectrum management cases:

1. The stadium small cell network has access to its own dedicated 20 MHz LTE channel, e.g. it is granted a temporary LSA license for exclusive access to this spectrum for the duration of the event. In this case, its performance is assessed separately, not considering the aerial eNB (AeNB) and the primary eNBs (PeNBs).
2. The cognitive small cells and the AeNB have secondary access to a 20 MHz LTE channel, also used by a network of 3 local PeNBs. This represents a more challenging secondary spectrum sharing task, where, in addition to the performance of the stadium small cells and the AeNB, the primary user QoS guarantees have to be taken into account. We assume that the primary users are those that are served by the local PeNBs depicted in Fig. 1.

A key challenging aspect of the cellular environment considered in this paper is its dynamic nature. We assume that the stadium network is able to dynamically adapt its topology to temporal non-uniform variations in the stadium traffic load. In the secondary spectrum sharing scenario, the dynamic nature of the environment is also caused by periodic deployments of the AeNB. All of these paradigms are explained in more detail in the following subsections.

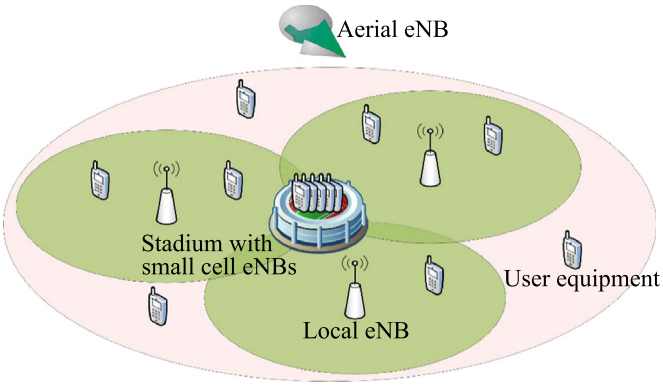



**Fig. 1.** Enhanced cellular network infrastructure during a stadium temporary event.

### 2.2. Dynamic topology management

Topology management is an increasingly popular area of research, especially in green communications, where a trade-off between the QoS provided to the users and the energy savings of the network is achieved by dynamically switching various base stations on/off, e.g. (Marsan et al., 2009; Rehan and Grace, 2013). A simple illustrative example discussed by Marsan et al. (2009) is portrayed in Fig. 2. It involves a classical hexagonal cell layout. There, all base stations surrounding the middle one temporarily enter a sleep mode at times when the traffic load is lower, e.g. night time. The users from all seven cells can then be served by the middle base station that would expand its coverage area accordingly. Employing such topology management schemes can result in significant energy savings, since a major part of energy in telecommunications systems is consumed by base stations (Marsan et al., 2009; Richter et al., 2009).

### 2.3. Dynamic non-uniform traffic load

Another source of the network's dynamic nature considered in this study is the presence of a dynamically moving traffic hotspot area. For example, a rapid increase in the traffic load in a specific part of the stadium small cell network may be observed if a particular event happens close to the given area, e.g. teams walking out at the opening ceremony of the Olympic Games or a goal at a football match etc. In such cases, the topology management algorithm would cause the network to be fully switched on in the hotspot area (left side of Fig. 2), and only partially deployed in other areas of lower traffic intensity (right side of Fig. 2). Furthermore, we assume that the geographical location of this hotspot area varies with time. This makes the wireless environment asymmetric and dynamic in both the offered traffic distribution and the network topology.

### 2.4. Rapidly deployable aerial platform

The second simulation scenario described in Section 2.1 also involves a local primary LTE network and a cognitive eNB on an aerial platform (AeNB) for wide area coverage. The AeNB can be switched on and off several times throughout the duration of the event Reynaud et al. (2014). For example, it can be switched on for providing the event organisers with a dedicated access network when required, and switched off to have its batteries recharged or to minimise the energy consumption in general. Therefore the additional challenge faced by the cognitive stadium network is to adapt to these sudden changes in their radio environment, while not affecting the QoS in the local primary system.

## 3. Distributed Q-learning based dynamic spectrum access

This paper investigates a flexible, distributed approach to DSA based on reinforcement learning (RL). In distributed RL based DSA the learning is performed by a number of individual wireless devices, for example, base stations in a cellular network. There, the task of every base station is to learn to prioritise among the available spectrum resources only through trial-and-error, with no frequency planning involved, and with no coordination with other base stations, e.g. Morozs et al. (2014b). In this way, frequency reuse patterns emerge autonomously using distributed artificial intelligence with no requirement for any prior knowledge of a given wireless environment.

### 3.1. Reinforcement learning

RL is a model-free type of machine learning which is aimed at learning the desirability of taking any available action in any state of the environment only through trial-and error (Sutton and Barto, 1998). This desirability of an action is represented by a numerical value, normally referred to as the $Q$-value—the expected cumulative reward for taking a particular action in a particular state, as shown in the equation below:

$$Q(s, a) = E\left[\sum_{t=0}^{T} \gamma^t r_t\right] \tag{1}$$

where $Q(s, a)$ is the $Q$-value of action $a$ in state $s$, $r_t$ is the numerical reward received $t$ time steps after action $a$ is taken in state $s$, $T$ is the total number of time steps until the end of the learning process or episode, and $\gamma \in (0, 1)$ is a discount factor.

The job of an RL algorithm is to estimate $Q(s, a)$ values for every action in every state, which are then stored in an array known as the $Q$-table. In some cases where an environment does not have to be represented by states, only the action space and a 1-dimensional $Q$-table $Q(a)$ can be considered (Claus and Boutilier, 1998). The job of an RL algorithm then becomes simpler, it aims to estimate an expected value of a single reward for each action available to the learning agent:

$$Q(a) = E[r_t] \tag{2}$$

### 3.2. Stateless Q-learning

One of the most successful and widely used RL algorithms is $Q$-learning. In particular, a simple stateless variant of this algorithm, as formulated by Claus and Boutilier (1998), has been shown to be effective for several distributed DSA learning problems, e.g. (Chu et al., 2015; Morozs et al., 2014b). Fig. 3 shows a flowchart for one file transmission of how distributed stateless $Q$-learning can be applied to DSA in cellular systems.

Each eNB (i.e. LTE base station) maintains a $Q$-table $Q(a)$ such that every subchannel $a$ has an expected reward or $Q$-value associated with it. The $Q$-value represents the desirability of assigning a particular subchannel to a file transmission. Upon each file arrival, the eNB either assigns a subchannel to its transmission or blocks it if all subchannels are occupied. It decides which subchannel to assign based on the current $Q$-table and the greedy

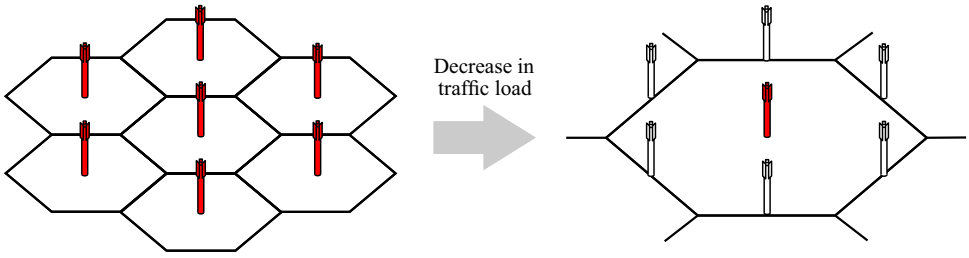



**Fig. 2.** A simple topology management case, where a number of base stations is switched off after a decrease in the overall traffic load.

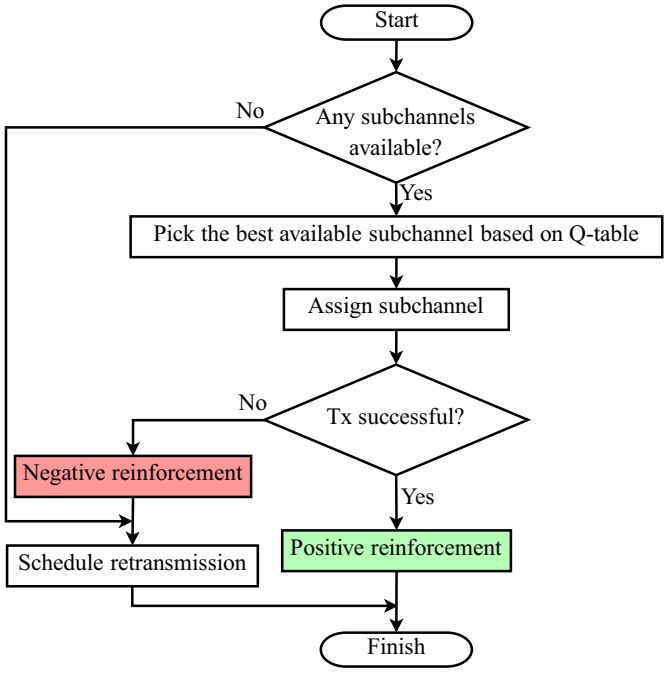



**Fig. 3.** Flowchart of the distributed stateless Q-learning based DSA algorithm.

action selection strategy described by the following equation:

$$\hat{a} = \underset{a}{\operatorname{argmax}}(Q(a)), \quad a \in A', A' \subset A \tag{3}$$

where $\hat{a}$ is the subchannel chosen for assignment out of the set of currently unoccupied subchannels $A'$, $Q(a)$ is the $Q$-value of subchannel $a$, and $A$ is the full set of subchannels.

The values in the $Q$-tables are initialised to zero, so all eNBs start learning with equal choice among all available subchannels. A $Q$-table is updated by the corresponding eNB each time it attempts to assign a subchannel to a file transmission in the form of a positive or a negative reinforcement. The recursive update equation for stateless $Q$-learning, as defined by Claus and Boutilier (1998), is given below:

$$Q(a) \leftarrow (1-\alpha)Q(a) + \alpha r \tag{4}$$

where $Q(a)$ represents the $Q$-value of the subchannel $a$, $r$ is the reward associated with the most recent trial and is determined by a reward function, and $\alpha \in (0, 1]$ is the learning rate parameter which weights recent experience with respect to previous estimates of the $Q$-values.

The reward function, which is generally applicable to a wide range of RL problems and which has been successfully applied to DSA problems in the past (Jiang et al., 2011; Morozs et al., 2015), returns two values:

- $r = -1$ (negative reinforcement), if the file transmission failed due to an insufficient Signal-to-Interference-plus-Noise Ratio (SINR) on the selected subchannel.
- $r=1$ (positive reinforcement), if the file is successfully transmitted, i.e. SINR did not drop below the minimum transmission threshold.

The choice of the learning rate values for this type of distributed $Q$-learning based DSA problems is investigated by us in Morozs et al. (2014a). We found that the best performance is achieved by using the Win-or-Learn-Fast (WoLF) variable learning rate (Bowling and Veloso, 2002) described by the following equation:

$$\alpha = \begin{cases} 0.01 & r = 1 \\ 0.1 & r = -1 \end{cases} \tag{5}$$

There, the learning rate $\alpha$ is 0.01 for successful trials (when $r=1$), and $\alpha = 0.1$ for failed trials ($r = -1$). In this way, the learning agents are learning faster when "losing" and more slowly when "winning".

## 4. Distributed case-based Q-learning

In this paper we investigate case-based RL as the approach for enhancing the stability of RL based DSA algorithms in challenging, dynamic wireless environments, such as those introduced in Section 2. The general principles of the case-based RL methodology are described in the following subsection.

### 4.1. Case-based reinforcement learning

Case-based RL is a combination of RL and case-based reasoning (CBR). CBR is broadly defined as the process of solving new problems by using the solutions to similar problems solved in the past, e.g. Rashedi et al. (2014) and Zhu et al. (2015). Fig. 4 shows a flow diagram of the processes involved in case-based RL. It also demonstrates that it is an extension of classical single-agent RL introduced in Section 3, i.e. the latter can be viewed as a special case of case-based RL.

The unfilled blocks and solid lines in Fig. 4 constitute a flow diagram of a classical RL algorithm. There is an outer output-state-action loop, where the outputs of the environment are sampled to yield the environment state information, and the best action is then chosen for the current state based on the policy of the learning agent. In the context of DSA, the output of interest is whether or not a given transmission is blocked, interrupted or successfully completed, and the action is a spectrum resource allocated to it. There is also an inner learning loop, whose role is to learn a good policy to be used by the learning agent. It achieves this goal by observing the actions taken by the learning agent and their outcomes, and directly estimating the entries in the $Q$-table, e.g. using (4) in the case of stateless $Q$-learning. A policy is then derived from the estimated $Q$-table and used for choosing an action in the current environment state, e.g. as shown in (3).

The highlighted blocks and dotted arrows represent additional functionality afforded by CBR to enable the system to learn several solutions to different cases of the environment at once. It introduces another parallel inner loop which continuously observes the input/output relationship of the environment, and identifies its current model or case. It may also have access to other information supplied from elsewhere to aid the identification process. The idea is that for different cases of the environment the estimated models will be sufficiently different to be detected by

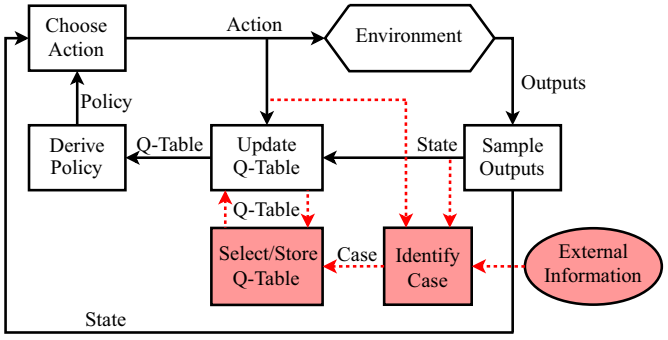



**Fig. 4.** Block diagram of case-based reinforcement learning.

the identification algorithm, and for every identified model of the environment there will be a stored $Q$-table associated with it. In this way, a case-based RL algorithm will always know what phase the environment is currently in and will be able to use a $Q$-table most suitable for it.

### 4.2. Case identification

A crucial part of the case-based RL process is an appropriate mechanism for case identification, such that the dynamically changing environment could be described by a finite number of distinct configurations, i.e. cases. We propose extending an approach developed by us in Morozs et al. (2013) for a simple 9 base station network with three distinct topology configurations, each forming its own case. There, the network topology based case identification approach assumed that every base station was aware of its surrounding network topology, i.e. an information source equivalent to the "external information" node from Fig. 4 was employed. This is also the case with the case identification approach proposed in this subsection. We assume that a small-scale radio environment map (REM) is periodically broadcast to every learning agent, i.e. base station. This is a realistic assumption in the context of DSA, since REM is one of the widely used key features in cognitive wireless networks (McLean et al., 2014).

All changes in the network environment described in Section 2 of this paper involve changes in the network topology, e.g. triggered by the temporally and spatially variable traffic load or the periodically deployed eNB on the aerial platform. Therefore, the network topology based case identification is also applicable in this study. However, the large-scale temporary event networks with time-varying asymmetric traffic loads considered in this paper have a significantly larger number of different topology configurations, compared to a purposely simple illustrative problem investigated by us in Morozs et al. (2013).

In order to limit the potential number of identifiable topology cases and to make this approach scalable and generally applicable to any cellular system, we propose the topology identification process that is localised to the second order neighbourhood (2ON) of a given eNB. We define the 2ON of an eNB as the set of its neighbouring eNBs and all their neighbouring eNBs as illustrated in Fig. 5 for a generic hexagonal cell layout.

The 2ON based topology identification process depicted in Fig. 5 is localised enough to be scalable and generally applicable in arbitrary cellular networks, yet not too limited to disregard valuable information about the radio environment surrounding a given eNB. To use the example in Fig. 5, the spectrum management policy of the middle eNB will be heavily influenced by the on/off configurations of its neighbouring eNBs with their own cognitive spectrum management policies. Equivalently, the latter will be significantly influenced by all of their own neighbouring eNBs, thus potentially having a noticeable impact on the original middle eNB. It is also possible to extend this argument to higher orders of neighbouring eNBs. However, their impact on the original eNB in question is likely to be diminishing. In future adaptations of the approach proposed in this paper further neighbourhoods of eNBs, up to the whole network topology, can also be included in the case identification process without the loss of generality.

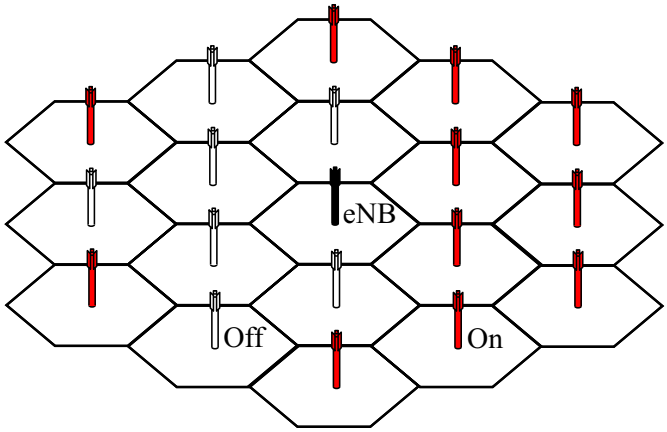



**Fig. 5.** Example of a second order neighbourhood (2ON) used for case identification by the middle eNB.

Similarly to the method used in Morozs et al. (2013), we propose expressing the on/off configurations of a given eNB's 2ON as a binary string, each bit corresponding to a particular eNB in the 2ON. For example, the following binary string would be used to describe the asymmetric topology case surrounding the middle eNB in Fig. 5:

$$T_{2ON} = 1010000100111111111_2 \tag{6}$$

where $T_{2ON}$ is the binary string describing the network topology surrounding the given eNB. The order of the bits in $T_{2ON}$ corresponds to the sequence of the eNBs in the 2ON depicted in Fig. 5 counting from the left-hand column of eNBs downwards and excluding the middle eNB itself.

### 4.3. Case retrieval

Another fundamentally important function that has to be performed by an intelligent CBR agent is case retrieval. It involves selecting a solution, e.g. a $Q$-table, that corresponds to the most appropriate stored case to be used at any given moment as shown in Fig. 4. To facilitate this functionality, a method for comparing a currently identified case with the stored cases and calculating a degree of similarity between them is required. Since every case is expressed in terms of the on/off configuration of the 2ON of a given eNB, we define the similarity measure between any two cases as the number of eNBs in the 2ON with the same on/off status. In order to calculate it, first, the given eNB derives a binary string $T_{same}$ indicating which eNBs in the 2ON are active/idle in both compared cases. It is done by performing a bitwise exclusive NOR operation between the binary strings describing the current case $T_{2ON}^{current}$ and one already stored in the case base $T_{2ON}^{stored}$:

$$T_{same} = \overline{T_{2ON}^{current} \oplus T_{2ON}^{stored}} \tag{7}$$

The similarity measure $\beta$ is then calculated by adding up the bits in $T_{same}$ as follows:

$$\beta = \sum_{n=1}^{N} T_{same}(n) \tag{8}$$

where $T_{same}(n)$ is the $n$'th bit of $T_{same}$, and $N$ is the number of eNBs in the 2ON.

Finally, for any currently identified case our proposed retrieval function will return a stored case using the following principle:

$$\hat{k} = \underset{a}{\operatorname{argmax}}(\beta_k), \quad k \in \{1, 2, \ldots, K\} \tag{9}$$

where $\hat{k}$ is the index of the retrieved case, $\beta_k$ is the similarity measure between the $k$'th stored case and the currently identified case, and $K$ is the total number of stored cases.

### 4.4. Multi-criteria case identification

The case identification and retrieval technique described in this section so far only considers the topology of a homogeneous network. For example, it is applicable to an isolated stadium small

cell network from Fig. 1. However, in the spectrum sharing scenario from Section 2.1, which also involves a dynamically deployable aerial eNB (AeNB), the network environment becomes heterogeneous. This requires an extension to the proposed case identification and retrieval framework.

The presence/absence of an entity such as the wide area coverage AeNB in the network environment can be viewed as a separate major criterion for case identification, in addition to the localised homogeneous topologies depicted in Fig. 5. Therefore, we propose a bias variable $\beta_{bias}$ for the case similarity assessment formula given in (8), such that the cases with the same AeNB status are recognised as more similar to each other, than those with a different AeNB status. The presence/absence of the AeNB is chosen to be a primary criterion for case identification and retrieval, since it represents a significantly more substantial change in the radio environment than changes in the active/idle mode of an eNB's local 2ON from Fig. 5. Therefore, the proposed extended multi-criteria similarity measure formula is the following:

$$\beta = \sum_{n=1}^{N} T_{same}(n) + \beta_{bias} \tag{10}$$

where the bias variable $\beta_{bias} > N$, i.e. a value higher than the maximum possible unbiased similarity measure, when the AeNB status of the two given cases is the same, and $\beta_{bias} = 0$ otherwise.

### 4.5. Case-based Q-learning algorithm

Algorithm 1 summarises the steps of our proposed case-based $Q$-learning approach to DSA in dynamic cellular environments. The extra functionality specific to CBR is described by steps 5, 6, 7 and 11. If these steps are taken out, the algorithm simplifies down to classical stateless $Q$-learning described in Section 3.

**Algorithm 1.** Subchannel assignment using case-based $Q$-learning in dynamic cellular environments

1: **for** every new file arrival
2: **if** all subchannels are occupied
3: Block transmission
4: **else**
5: Identify current case $k$
6: Find most similar stored case $\hat{k}$ using (9)
7: Retrieve $Q$-table $Q(a)$ associated with $\hat{k}$
8: Assign a subchannel using $Q(a)$ and (3)
9: Observe outcome, calculate reward $r = \pm 1$
10: Update $Q(a)$ using (4)
11: Store $Q(a)$ in case base, associate it with $k$
12: **end if**
13: **end for**

## 5. Simulation results

This section presents the results from a number of simulation experiments that assess the performance of our proposed case-based $Q$-learning approach to DSA. The event-driven system-level simulation model was custom-built in C++ to simulate the temporary event network scenario introduced in Section 2. This simulation model was used by us in a number of simulation studies in the past, e.g. Morozs et al. (2014b, 2015).

This section is organised as follows:

- Section 5.1 describes the parameters and assumptions used in our simulation model in order to make our study reproducible.
- Sections 5.2 and 5.3 describe the spectrum management policies we simulate in the primary and the secondary network, including the DSA schemes we use for baseline comparison.
- Section 5.4 explains how we implement traffic load based topology management, introduced in Section 2.2, in our simulation experiments.

Afterwards, the rest of the section covers the results from three separate simulation experiments that correspond to the three different sources of the dynamic nature of wireless environments introduced in Section 2:

- In the experiment in Section 5.5 we add a dynamically moving traffic hotspot area to the stadium network to see how well our proposed case-based $Q$-learning DSA algorithm copes with small, frequent changes in the radio environment.
- The experiment in Section 5.6 simulates a time-varying network-wide traffic load in the stadium network. In contrast with the moving hotspot area scenario, this experiment introduces large, infrequent changes in the radio environment.
- Finally, the experiment in Section 5.7 involves dynamic spectrum sharing among the stadium network, the dynamically deployable aerial eNB and the primary network.

The three simulation experiments summarised above provide a diverse set of simulation scenarios for a thorough empirical evaluation of the performance of the case-based $Q$-learning algorithm in different dynamic radio environments. The experiments involve small, frequent environment changes as well as large, less frequent changes. Furthermore, the network topology is symmetric in some cases, but asymmetric in others. Also, the first two experiments involve a standalone homogeneous stadium network (i.e. the first spectrum management scenario from Section 2.1), whereas the last simulation experiment investigates dynamic spectrum sharing in the presence of incumbent users (the second spectrum management scenario from Section 2.1).

### 5.1. Simulation model parameters and assumptions

The temporary event network scenario introduced in Section 2.1 involves an aerial eNB (AeNB) and a network of small cell eNBs inside a stadium, both of which coexist with a local network of primary eNBs (PeNBs) operating in the area.

The stadium small cell network architecture is shown in Fig. 6. There, the users are located in a circular spectator area 53.7–113.7 m from the centre of the stadium. It is covered by 78 eNBs arranged in three rings at 1 m height, e.g. with antennas attached to the backs of the seats or to the railings between different row levels. The seat width is assumed to be 0.5 m, and the space between rows is 1.5 m, which yields the total capacity of 43,103 seats. 25% of the stadium capacity is filled with randomly distributed wireless subscribers, i.e. $\approx$ 10,776 user equipments (UEs). In the secondary spectrum sharing scenario 500 primary UEs are randomly distributed outside the stadium in a circular area from the stadium boundary out to 1.5 km from the stadium centre point, producing an overall offered traffic of 20 Mb/s. The AeNB is located above the stadium centre point at 300 m altitude. The coordinates of the PeNBs are $(-600, -750)$, $(100, 750)$ and $(750, -800)$ m with respect to the centre point of the stadium. The other parameters and assumptions of the simulation model are listed in Table 1.

### 5.2. Spectrum management in the primary network

The primary system is assumed to employ a dynamic inter-cell interference coordination (ICIC) scheme typical for conventional

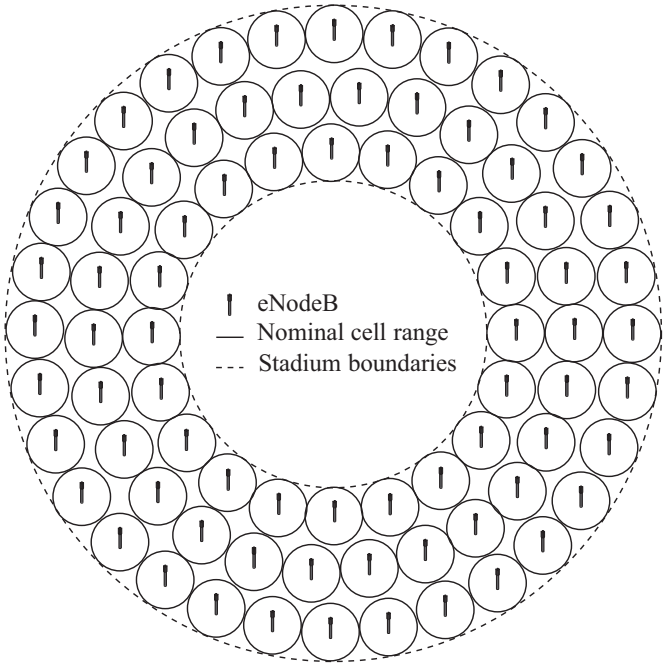



**Fig. 6.** Stadium network architecture (Morozs et al., 2014b).

**Table 1**
Simulation model parameters and assumptions.

| Parameter | Value |
|---|---|
| Channel bandwidth | 20 MHz: 100 LTE virtual resource blocks (VRBs) |
| Subchannel bandwidth | 4 VRBs: 4 × 180 kHz (3GPP, 2013) |
| Frequency band | 2.6 GHz |
| UE receiver noise floor | 94 dBm (290 K temperature, 20 MHz bandwidth, 7 dB noise figure) |
| Stadium propagation model | WINNER II B3 (Kyösti et al., 2008) |
| Outdoor propagation model | WINNER II C1 (Kyösti et al., 2008) |
| Stadium-outdoor propagation model | Combined WINNER II C4 with C1 term (Kyösti et al., 2008) |
| AeNB-ground propagation model | Free space + 8 dB log-normal shadowing |
| Traffic model | 3GPP FTP Traffic Model 1 (3GPP, 2010), file size - 4.2 Mb ( ≈ 0.5 MB) |
| Retransmission scheduling | Uniform random back-off between 0 and 960 ms |
| Link model | 3GPP Truncated Shannon Bound model (3GPP, 2012) |
| Primary eNB Tx power | 10 dBW |
| **Assumptions** | |
| UEs inside the stadium are associated with a small cell or aerial eNB with a minimum estimated downlink pathloss, based on the Reference Signal Received Power (RSRP) | |
| UEs outside the stadium are associated with a primary or aerial eNB based on the strongest RSRP. The reference signal Tx power of the primary eNB is 13 dB higher than that of the AeNB | |
| Cognitive small cell and aerial eNBs employ open loop power control, using a constant Rx power of -74 dBm (20 dB Signal-to-Noise Ratio) | |
| The minimum Signal-to-Interference-plus-Noise Ratio (SINR) allowed to support data transmission is 1.8 dB | |
| One subchannel (4 VRBs) is allocated to every data transmission | |

LTE networks (Fraimis et al., 2010; Sesia et al., 2011). There, all three PeNBs exchange their current spectrum usage as Relative Narrowband Transmit Power (RNTP) messages every 20 ms, and exclude the subchannels currently used by the other two eNBs from their available subchannel list. We assume that they always try to assign an available subchannel with the lowest index if any, e.g. they always scan the availability of the subchannels in the same order from the 1st subchannel to the last. In this way, the primary network would make its spectrum usage less random and more appropriate for the cognitive cellular system to share, which is in the interests of both the primary and the secondary system. However, the cased-based *Q*-learning scheme proposed in this paper does not assume this and would also work regardless of the spectrum management strategy of the primary system.

### 5.3. Spectrum management in the secondary network

In addition to implementing the proposed case-based *Q*-learning algorithm in the secondary network, its performance is compared with the following baseline schemes (also implemented in the secondary network):

- "*Dynamic ICIC*" - all systems use ICIC signalling as described above for the primary system. The stadium eNBs receive ICIC messages from the AeNB and from their neighbouring small cells. They only report subchannels used at a Tx power above −3 dB with respect to the average power in the cell, and choose randomly among the subchannels deemed "safe". The AeNB randomly assigns subchannels not used by the primary system, based on the ICIC messages of the latter.
- "*Q-learning*"—the AeNB and the stadium small cells run the distributed *Q*-learning algorithm described in Section 3.2.

The "dynamic ICIC" approach represents a heuristic baseline DSA scheme, typical for LTE networks (Fraimis et al., 2010; Sesia et al., 2011), whereas the "*Q*-learning" approach represents a pure RL based approach without the CBR functionality added to it.

### 5.4. Topology management

Fig. 7 shows how the principle of traffic load dependent dynamic topology management described in Section 2.2 is adapted to the stadium small cell network used in simulation experiments in this paper. The following relationship between the network-wide offered traffic density (*OTD*) and the topology patterns from Fig. 7 is used:

- all eNBs are active if $OTD > 27$ Gbps/km$^2$
- 5/6 eNBs are active if $OTD \in (21, 27]$ Gbps/km$^2$
- 2/3 eNBs are active if $OTD \in (15, 21]$ Gbps/km$^2$
- 1/3 eNBs are active if $OTD \in (8, 15]$ Gbps/km$^2$
- 1/6 eNBs are active if $OTD \leq 8$ Gbps/km$^2$

In this way the stadium network is able to provide adequate QoS to the users across a wide range of traffic loads, whilst achieving significant energy savings when the offered traffic is low by employing these partial small cell network deployments.

### 5.5. Simulation experiment 1: stadium network with a moving traffic hotspot area

In addition to the network-wide traffic load variations, another feature of the simulation scenario investigated in this paper is the presence of a traffic hotspot area within the stadium that changes its geographical location with time. An example of such a hotspot area and its effect on the topology of the stadium network is shown in Fig. 8. If an increased user activity in the 60 degree sector is observed, while the offered traffic density is lower elsewhere, the topology management algorithm detects the possibility of deploying all available eNBs in the hotspot area and keeping a number of them switched off according to one of the partial deployment patterns from Fig. 7.

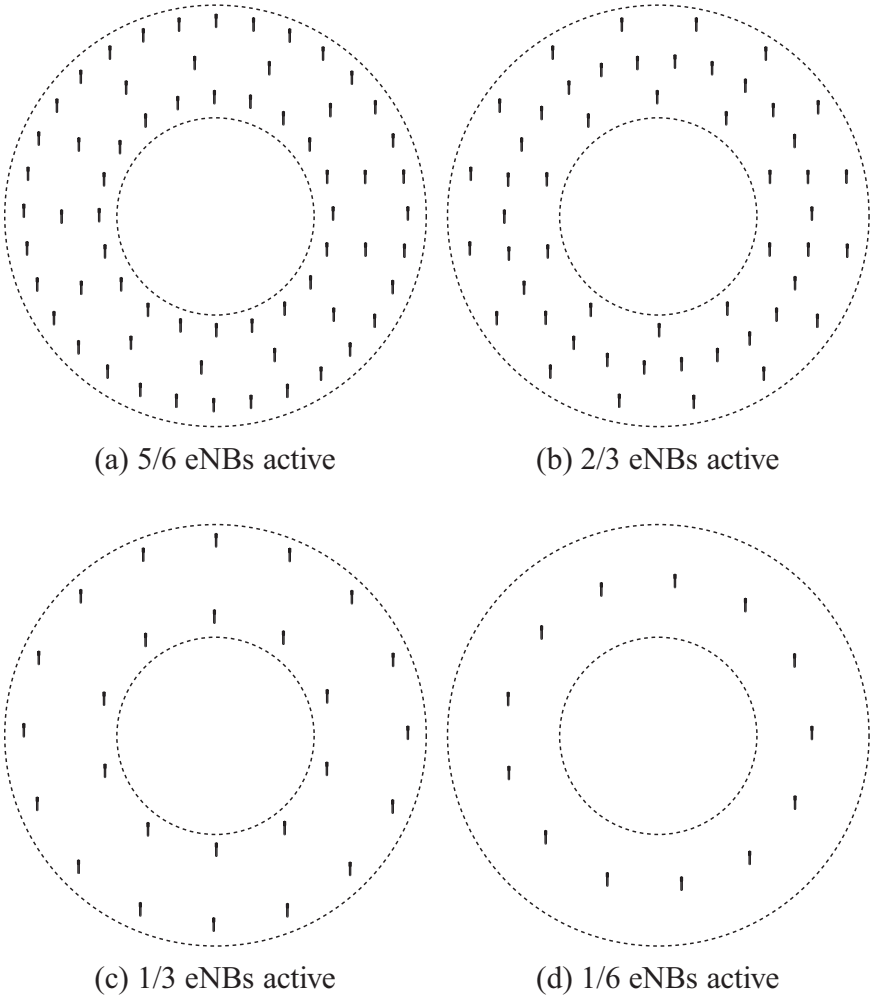



**Fig. 7.** Traffic load based partial deployments of the stadium network (centralised topology management)

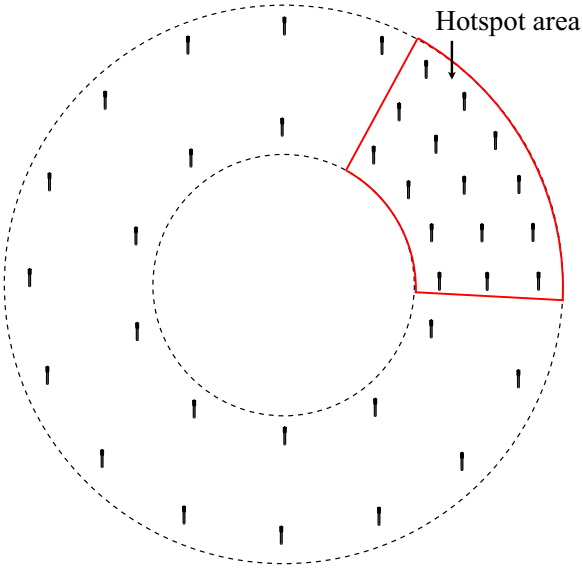



**Fig. 8.** Asymmetric network topology due to a localised increase in offered traffic.

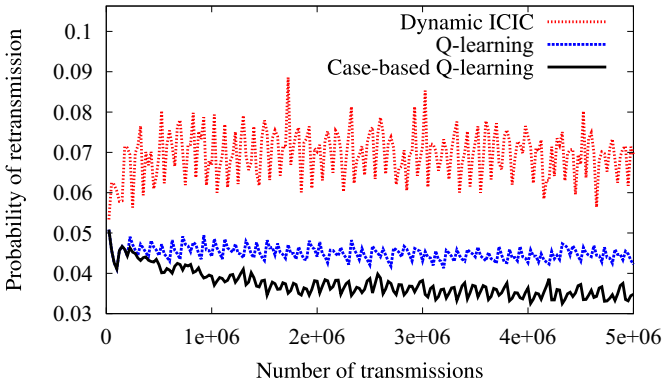



**Fig. 9.** Probability of retransmission in the small cell stadium network with a dynamically moving traffic hotspot.

Fig. 9 shows the probability of retransmission time response in the stadium small cell network inspected individually with its own dedicated spectrum (20 MHz LTE channel). The location of the 60° hotspot area is randomly changed every 100,000 transmissions to one of its six possible locations—{0°, 60°, 120°, 180°, 240°, 300°}. The offered traffic density within the hotspot is 34 Gbps/km$^2$, and 13 Gbps/km$^2$ elsewhere. The topology management algorithm is assumed to detect a change in the offered traffic distribution with a delay of 5000 file transmissions. The plots are obtained by averaging every data point using the results from 50 simulations with different randomly generated UE locations and initial traffic. $P$ (*retransmission*) is calculated as follows:

$$P(retransmission) = \frac{N_r}{N_r + N_s} \quad (11)$$

where $N_r$ is the number of retransmissions and $N_s$ is the number of successfully completed tranmissions during a given sampling period.

Firstly, both Q-learning based schemes significantly outperform the dynamic ICIC approach. This demonstrates the effectiveness of applying distributed RL to DSA in cellular networks. Secondly, although the classical Q-learning and case-based Q-learning schemes start at the identical QoS level, the latter goes on to gradually improve its performance in the dynamic environment faced by it. In contrast, the classical Q-learning process is disturbed by the environment changes frequently enough not to show any notable performance improvement over time. As a result, by the end of the simulation the proposed case-based Q-learning scheme shows an $\approx 22\%$ reduction in the number of retransmissions compared to the classical Q-learning alternative.

### 5.6. Simulation experiment 2: temporal network-wide traffic variations in the stadium network

A further challenge introduced into the simulation experiments hereafter is the variable network-wide traffic load shown in Fig. 10. These variations in the offered traffic density trigger the network topology changes according to the topology management scheme described in Section 5.4. Fig. 11 a shows the probability of retransmission time response of the stadium network with such uniform temporal variations in the network-wide traffic load. Due to the uniform nature and a lower number of possible topology cases compared to the dynamic traffic hotspot scenario from the previous subsection, the difference in performance between case-based Q-learning and classical Q-learning is larger than that observed in Fig. 9. It is especially pronounced at times shortly after the network topology transitions. There, incorporating CBR into the learning process often results in as much as a two-fold reduction in the probability of retransmission.

Fig. 11 b shows the probability of retransmission time response of the stadium network both with uniform variations in the offered traffic density and with the dynamically moving traffic hotspot area. There, in contrast to the results in Fig. 11 a, the increase in the complexity of the problem and the number of network topology cases reduces the magnitude of the performance improvements gained by case-based Q-learning. Nevertheless, the CBR functionality is still able to provide a consistent noticeable decrease in the number of retransmissions experienced by the UEs in the stadium network.

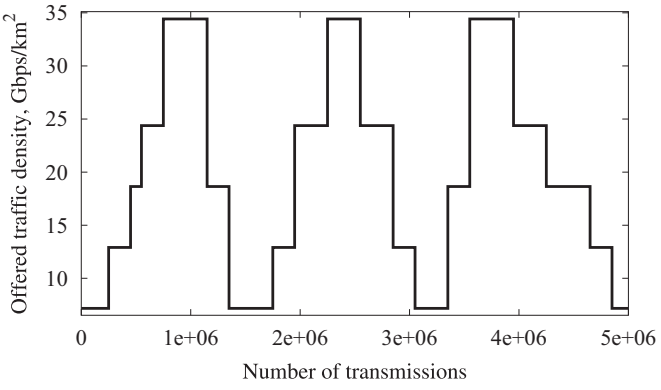



**Fig. 10.** Temporal variations in the stadium network-wide offered traffic density.

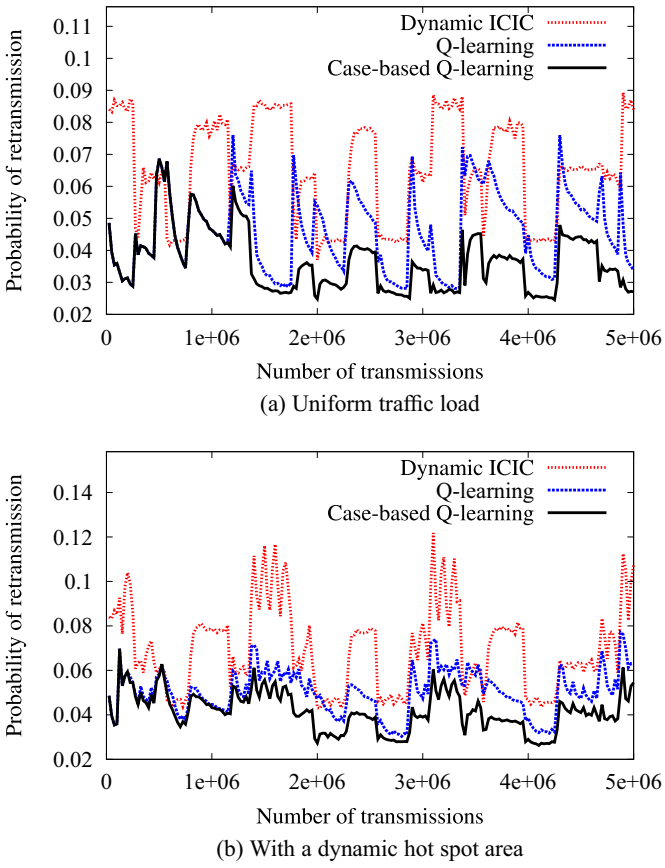



**Fig. 11.** Probability of retransmission in the stadium network with temporal variations in the network-wide offered traffic.

### 5.7. Simulation experiment 3: spectrum sharing with dynamic aerial platform deployment

The last set of simulation results discussed in this paper considers the performance of both the primary and the secondary network in the full spectrum sharing scenario described in Section 2.1. In addition to the dense stadium small cell network, it involves an aerial eNB (AeNB) and a local network of primary eNBs (PeNBs), all sharing the same 20 MHz LTE channel. The stadium small cell network includes both dynamic environment features investigated in the previous subsections:

- a dynamically moving 34 Gbps/km$^2$ offered traffic density area depicted in Fig. 8
- an updated version of the temporal variations in the network-wide traffic load shown in Fig. 12

The variable network-wide traffic loads are slightly lower than those used in the previous experiments, since the 20 MHz LTE channel is no longer fully dedicated to the stadium network, but is shared with the primary system and the cognitive AeNB. The latter is running a classical Q-learning algorithm described in Section 3.2 and is periodically deployed and redeployed into the network.

Fig. 13 shows how the probability of retransmission changes over time in the two independent secondary systems involved in the spectrum sharing scenario—the stadium small cell network and the AeNB. All simulations start with the AeNB switched off, and the vertical dash-dot lines in Fig. 13 a mark the times when it is switched on and off again. It shows that the performance gap between case-based and classical Q-learning in the stadium network is further reduced due to an even more complicated

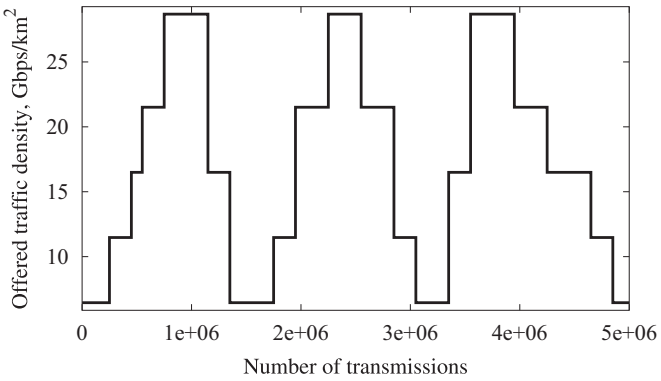
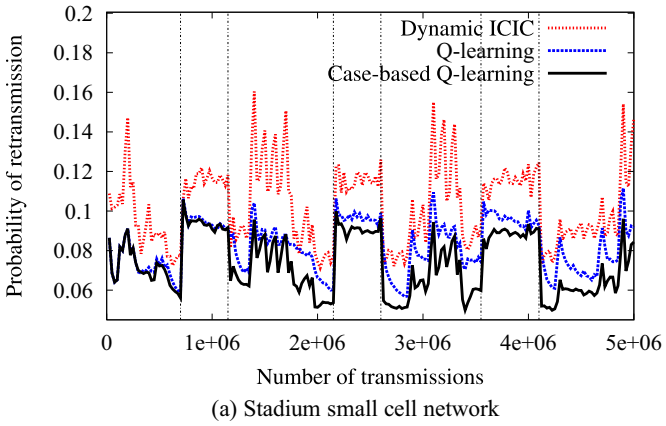
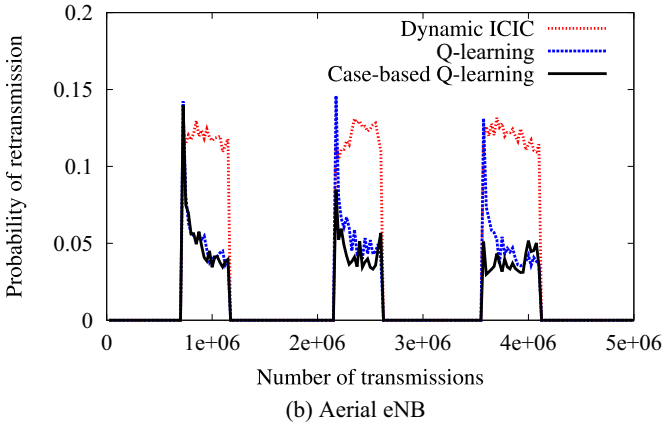



**Fig. 12.** Temporal variations in the stadium network-wide offered traffic density in the full spectrum sharing scenario.




(a) Stadium small cell network




(b) Aerial eNB

**Fig. 13.** Probability of retransmission in the stadium network and the Aerial eNB in a dynamically changing radio environment.

scenario, the presence of an interfering primary network and a higher number of possible network topologies. However, Fig. 13 b shows that employing the case-based *Q*-learning approach in the stadium network dramatically improves the QoS achieved by the AeNB shortly after it is switched on for the second and third time. This is due to the capability of cognitive small cell networks to distinguish between various network topologies, including whether or not the AeNB is switched on. In this way, the stadium small cells are able to revert their *Q*-learning DSA policies to those most appropriate for the AeNB to share spectrum with them, resulting in the QoS improvement in both of these secondary access networks.

An essential requirement for cognitive cellular systems is to ensure that they do not have a harmful effect on the QoS in the primary system. Table 2 compares the QoS provided to the users outside of the stadium with and without the presence of the stadium users and the secondary network. It describes the statistical distribution of user throughput (UT) achieved by the primary network. The equation for calculating UT for any given UE, as defined by 3GPP (2010), is given below:

**Table 2**
Primary user quality of service (QoS) with and without the presence of the secondary network (SN).

| QoS metric (Mb/s) | No SN | With SN |
|---|---|---|
| Mean user throughput (UT) | 3.03 | 3.07 |
| 95th percentile UT | 3.16 | 3.16 |
| 5th percentile UT | 2.76 | 2.90 |
| Mean UT 0–100 m from stadium | 2.95 | 2.93 |

$$UT = \frac{\sum_{f=1}^{F} S_f}{\sum_{f=1}^{F} T_f} \tag{12}$$

where $F$ is the number of files downloaded by the given UE, $S_f$ is the size of the $f^{th}$ file, and $T_f$ is the time it took to download it.

Table 2 shows that the introduction of the secondary stadium network and the AeNB results in an insignificant degradation in the average probability of retransmission and the mean UT provided to the primary users in the 100 m vicinity of the stadium. Interestingly, it even achieves an improvement in the 5th percentile UT, which represents the minimum UT provided to at least 95% of the users and which is an important metric for ensuring fair QoS distribution across the whole network. This is because the AeNB manages to provide higher quality opportunistic links to some primary users than those that could be provided by the local eNBs. The results in Table 2 emphatically show that it is possible to develop a temporary heterogeneous cognitive network that is capable of servicing a dramatic increase in the mobile data capacity (546 Mb/s overall throughput compared to 19.8 Mb/s in the primary system only), but with no need for additional spectrum and with no degradation in the primary user QoS.

## 6. Conclusion

The case-based *Q*-learning technique proposed in this paper is an effective and feasible approach to DSA in dynamic cellular environments. Large-scale system level simulations of a stadium small cell network with an asymmetric time-variant topology show that augmenting classical *Q*-learning with the CBR functionality in this way results in increased adaptability of the cognitive cellular system to changes in its radio environment. For example, it is capable of achieving a two-fold reduction in the number of retransmissions, compared to a classical *Q*-learning approach, shortly after transitions between different network topologies. However, as the complexity of the dynamic environment and the possible number of network topologies increase, the performance gap between classical and case-based *Q*-learning decreases. Nevertheless, the proposed case-based *Q*-learning approach achieves a consistent improvement in the system QoS and its stability in the dynamic cellular environment considered. Both case-based and classical *Q*-learning DSA methods also dramatically outperform a heuristic dynamic ICIC approach typical for current LTE systems.

Simulations of a spectrum sharing scenario, where the stadium small cell network shares the same LTE channel with a cognitive aerial eNB and a local primary network, show that the proposed approach achieves a significant improvement in the reliability of the aerial eNB, whilst maintaining a small yet consistent QoS

improvement inside the stadium, compared to the classical *Q*-learning algorithm. Furthermore, these simulations show that the cognitive cellular system that employs the case-based *Q*-learning DSA scheme with only secondary access to an LTE channel, is able to accommodate a 28-fold increase in the total primary and secondary system throughput, but with no need for additional spectrum and with no degradation in the QoS of the primary users.

Finally, one of the directions for further work on the case-based *Q*-learning approach to DSA presented in this paper is developing a theoretical framework for it, e.g. using game theory (Alnwaimi et al., 2015) or Bayesian networks (Morozs et al., 2016). While our study provides a thorough empirical evaluation of the proposed algorithm using large scale simulations, a theoretical model could help gain additional, deeper insight into the potential performance gains and limitations of case-based *Q*-learning.

## Acknowledgement

This work has been funded by the ABSOLUTE Project (FP7-ICT-2011-8-318632), which receives funding from the 7th Framework Programme of the European Commission.

## References

3GPP, 2010. Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Further Advancements for E-UTRA physical layer aspects (3GPP TR 36.814 version 9.0.0 Release 9), December.

3GPP, 2012. LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Frequency (RF) system scenarios (3GPP TR 36.952 version 11.0.0 Release 11), December.

3GPP, 2013. LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures (3GPP TS 36.213 version 11.5.0 Release 11), December.

Alnwaimi, G., Vahid, S., Moessner, K., 2015. Dynamic heterogeneous learning games for opportunistic access in LTE-based macro/femtocell deployments. IEEE Trans. Wirel. Commun. 14 (April (4)), 2294–2308.

Bennis, M., Perlaza, S., Blasco, P., Han, Z., Poor, H., 2013. Self-organization in small cell networks: a reinforcement learning approach. IEEE Trans. Wirel. Commun. 12, 3202–3212.

Bianchi, R., Celiberto Jr., L., Santos, P., Matsuura, J., Lopez de Mantaras, R., 2015. Transferring knowledge as heuristics in reinforcement learning: a case-based approach. Artif. Intell. 226, 102–121.

Bowling, M., Veloso, M., 2002. Multiagent learning using a variable learning rate. Artif. Intell. 136, 215–250.

Celiberto, L., Matsuura, J., Lopez de Mantaras, R., Bianchi, R., 2012. Reinforcement learning with case-based heuristics for robocup soccer keepaway. In: 2012 Brazilian Robotics Symposium and Latin American Robotics Symposium (SBR-LARS). pp. 7–13.

Chen, X., Zhao, Z., Zhang, H., 2013. Stochastic power adaptation with multiagent reinforcement learning for cognitive wireless mesh networks. IEEE Trans. Mob. Comput. 12, 2155–2166.

Chu, Y., Kosunalp, S., Mitchell, P., Grace, D., Clarke, T., 2015. Application of reinforcement learning to medium access control for wireless sensor networks. Eng. Appl. Artif. Intell. Part A 46, 23–32.

Claus, C., Boutilier, C., 1998. The dynamics of reinforcement learning in cooperative multiagent systems. In: Proceedings of the Fifteenth National/tenth Conference on Artificial Intelligence/Innovative Applications of Artificial Intelligence.

Fraimis, I., Papoutsis, V., Kotsopoulos, S., 2010. A decentralized subchannel allocation scheme with inter-cell interference coordination (ICIC) for multi-cell OFDMA systems. In: IEEE Global Telecommunications Conference (GLOBECOM).

Ghosh, C., Roy, S., Cavalcanti, D., 2011. Coexistence challenges for heterogeneous cognitive wireless networks in TV white spaces. IEEE Wirel. Commun. 18, 22–31.

Gomez, K., Kandeepan, S., Vidal, M.M., Boussemart, V., Ramos, R., Hermenier, R., Rasheed, T., Goratti, L., Reynaud, L., Grace, D., Zhao, Q., Han, Y., Rehan, S., Morozs, N., Jiang, T., Bucaille, I., Wirth, T., Campo, R., Javornik, T., 2016. Aerial base stations with opportunistic links for next generation emergency communications. IEEE Commun. Mag. 54 (April (4)), 31–39.

Guizani, M., Khalfi, B., Ghorbel, M., Hamdaoui, B., 2015. Large-scale cognitive cellular systems: resource management overview. IEEE Commun. Mag. 53, 44–51.

Gurney, D., Buchwald, G., Ecklund, L., Kuffner, S., Grosspietsch, J., 2008. Geo-location database techniques for incumbent protection in the TV white space. In: IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN).

Hamouda, S., Zitoun, M., Tabbane, S., 2014. Win–win relationship between macrocell and femtocells for spectrum sharing in LTE-A. IET Commun. 8, 1109–1116.

Jiang, C., Sheng, Z., . 2009. Case-based reinforcement learning for dynamic inventory control in a multi-agent supply-chain system. Expert Syst. Appl. 36 (3), 6520–6526.

Jiang, T., Grace, D., Mitchell, P.D., 2011. Efficient exploration in reinforcement learning-based cognitive radio spectrum sharing. IET Commun. 5, 1309–1317.

Kyösti, P., Meinilä, J., Hentilä, L., Zhao, X., Jämsä, T., Schneider, C., Narandzić, M., Milojević, M., Hong, A., Ylitalo, J., Holappa, V., Alatossava, M., Bultitude, R., de Jong, Y., Rautiainen, T., 2008. IST-4-027756 WINNER II Deliv. D1. 1. 2: WINNER II channel Model.

Malialis, K., Kudenko, D., 2015. Distributed response to network intrusions using multiagent reinforcement learning. Eng. Appl. Artif. Intell. 41, 270–284.

Marsan, M., Chiaraviglio, L., Ciullo, D., Meo, M., 2009. Optimal energy savings in cellular access networks. In: IEEE International Conference on Communications Workshops (ICC Workshops).

Matinmikko, M., Okkonen, H., Palola, M., Yrjola, S., Ahokangas, P., Mustonen, M., 2014. Spectrum sharing using licensed shared access: the concept and its workflow for LTE-advanced networks. IEEE Wirel. Commun. 21, 72–79.

McLean, R., Silvius, M., Hopkinson, K., Flatley, B., Hennessey, E., Medve, C., Thompson, J., Tolson, M., Dalton, C., 2014. An architecture for coexistence with multiple users in frequency hopping cognitive radio networks. IEEE J. Sel. Areas Commun. 32, 563–571.

Morozs, N., Grace, D., Clarke, T., 2013. Case-based reinforcement learning for cognitive spectrum assignment in cellular networks with dynamic topologies. In: Military Communications and Information Systems Conference (MCC).

Morozs, N., Clarke, T., Grace, D., Zhao, Q., 2014a. Distributed *Q*-learning based dynamic spectrum management in cognitive cellular systems: Choosing the right learning rate. In: IEEE International Symposium on Computers and Communications (ISCC).

Morozs, N., Grace, D., Clarke, T., 2014b. Distributed *Q*-learning based dynamic spectrum access in high capacity density cognitive cellular systems using secondary LTE spectrum sharing. In: International Symposium on Wireless Personal Multimedia Communications (WPMC).

Morozs, N., Clarke, T., Grace, D., 2015. Heuristically accelerated reinforcement learning for dynamic secondary spectrum sharing. IEEE Access 3, 2771–2783.

Morozs, N., Clarke, T., Grace, D., 2016. Distributed heuristically accelerated *Q*-learning for robust cognitive spectrum management in LTE cellular systems. IEEE Trans. Mob. Comput. 15 (4), 817–825.

Rashedi, E., Nezamabadi-pour, H., Saryazdi, S., 2014. Long term learning in image retrieval systems using case based reasoning. Eng. Appl. Artif. Intell. 35, 26–37.

Rehan, S., Grace, D., 2013. Combined green resource and topology management for beyond next generation mobile broadband systems. In: 2013 International Conference on Computing, Networking and Communications (ICNC). pp. 242–246.

Reynaud, L., Allsopp, S., Charpentier, P., Cao, H., Grace, D., Hermenier, R., Hrovat, A., Hughes, G., Ioan, C., Javornik, T., Munari, A., Vidal, M., Strother, J., Valcarce, R., Zaharia, S., 2014. FP7-ICT-2011-8-318632-ABSOLUTE/D2.1 Use cases definition and scenarios description.

Richter, F., Fehske, A., Fettweis, G., 2009. Energy efficiency aspects of base station deployment strategies for cellular networks. In: IEEE Vehicular Technology Conference (VTC-Fall).

Sachs, J., Maric, I., Goldsmith, A., 2010. Cognitive cellular systems within the TV spectrum. In: IEEE Symposium on New Frontiers in Dynamic Spectrum.

Sesia, S., Baker, M., Toufik, I., 2011. LTE-The UMTS Long Term Evolution: From Theory to Practice. John Wiley & Sons, Chichester.

Sun, H., Nallanathan, A., Wang, C.-X., Chen, Y., 2013. Wideband spectrum sensing for cognitive radio networks: a survey. IEEE Wirel. Commun. 20, 74–81.

Sutton, R., Barto, A., 1998. Reinforcement Learning: An Introduction. MIT Press, Cambridge, MA.

Walraven, E., Spaan, M.T., Bakker, B., 2016. Traffic flow optimization: a reinforcement learning approach. Eng. Appl. Artif. Intell. 52, 203–212.

Watkins, C., 1989. Learning from Delayed Rewards. Ph.D. thesis, University of Cambridge, England.

Zhu, G.-N., Hu, J., Qi, J., Ma, J., Peng, Y.-H., 2015. An integrated feature selection and cluster analysis techniques for case-based reasoning. Eng. Appl. Artif. Intell. 39, 14–22.