Integrated Project

ABSOLUTE - Aerial Base Stations with Opportunistic Links for Unexpected & Temporary Events

Contract No. 318632

Deliverable

FP7-ICT-2011-8-318632-ABSOLUTE/D3.3.3

Performance Evaluation of Cognitive Dynamic Spectrum Access

Contractual date:	**M32 31/05/2015**
Actual date:	**15/06/2015**
Authors/Editors:	**Yunbo Han, Nils Morozs, David Grace, Tim Clarke**
Participants:	**UoY**
Work package:	**WP3**
Security:	**PU**
Nature:	**Report**
Version:	**1.0**
Total number of pages:	**66**

Abstract

This document is the third deliverable of task 3.3 'Cognitive Spectrum Access' which presents additional improvements on algorithms developed and reported on in earlier deliverables to deliver dynamic spectrum access for Public Safety and Disaster Relief (PPDR) and temporary event scenarios. The performance of transfer learning based K-means clustering algorithms to deliver improved cognitive spectrum management is investigated for ABSOLUTE public safety scenario. Knowledge transfer and processing, and knowledge reward are the major issues in transfer learning. Data mining techniques are selected to solve this problem. A scheme combining Reinforcement Learning (RL) and case based reasoning (CBR) is then proposed to improve the stability of RL based dynamic spectrum access (DSA) for temporary event networks with dynamic topologies that use secondary LTE spectrum sharing. A novel heuristically accelerated reinforcement learning (HARL) approach is also designed to overcome the common disadvantage of RL algorithms, which is the need for many learning iterations to converge on an acceptable solution. The HARL algorithm is developed to speed up RL algorithms, using a radio environment map (REM) to mitigate the problem of poor temporal performance of RL algorithms applied to DSS problem.

Keywords

Cognitive Spectrum Assignment, Reinforcement Learning, Transfer Learning, Data Mining, K-means clustering, LTE/LTE-A, secondary LTE spectrum sharing, Case Base Reasoning, Radio Environment Map, Convergence Speed.

Table of Contents

List of Figures

List of Tables

Abbreviations

ABSOLUTE	Aerial Base Stations with Opportunistic Links For Unexpected & Temporary Events
AeNB	Aerial Evolved Node B
BS	Base Station
CPE	Customer-Premises Equipment
CR	Cognitive Radio
CRA	Channel Reservation Access
CSMA/CA	Carrier Sense Multiple Access with Collision Avoidance
DCA	Dynamic Channel Assignment
DFE	Digital Front-End
DFT	Discrete Fourier Transform
DIAQ	Distributed ICIC Accelerated Q-learning
DSA	Dynamic Spectrum Access
DSM	Dynamic Spectrum Management
DSS	Dynamic Spectrum Sharing
DTV	Digital TV
EUTRA	Evolved UMTS Terrestrial Radio Access
FBSE	Filter Bank Spectrum Estimation
FCA	Fixed Channel Assignment
FDD	Frequency Division Duplex
FRU	Flexible Reuse
GDB	Geo-location Data Base
GRAN	GSM EDGE Radio Access Network
HARL	Heuristically Accelerated Reinforcement Learning
ICIC	Inter-cell Interference Coordination
IEA	Interference Exclusion Area
LTE	Long-Term Evolution
LTE-A	Long-Term Evolution –Advanced
MCD	Measurement Capable Device
OFDM	Orthogonal Frequency Division Multiplexing
PCA	Prioritized Connection Access
PMSE	Programme Making and Special Event

PRB	Physical Resource Block
PU	Primary User
PUSCH	Physical Uplink Shared Channel
QoS	Quality of Service
QP	Quiet Period
RB	Resource Block
RBG	Resource Block Group
REM	Radio Environment Map
REM-SA	REM data Storage and Acquisition
RL	Reinforcement Learning
RLSS	Registered Location Secure Server
RNTP	Relative Narrowband Transmit Power
RSRP	Reference Signal Received Power
SEA	SNR Exclusion Area
SINR	Signal to Interference plus Noise Ratio
SU	Secondary User
TeNB	Terrestrial Evolved Node B
TVWS	TV White Space
UE	User Equipment
UT	User Throughput
VBR	Virtual Resource Block
WoLF	Win or Learn Fast
WRAN	Wireless Regional Area Network
WSD	White Space Devices

Executive Summary

This document is the third deliverable of task 3.3 ‘Cognitive Spectrum Access’ scheduled in month 32. The purpose of this deliverable is to propose detailed performance evaluation of cognitive dynamic spectrum access and improvements to algorithms reported on in the earlier deliverables in this task. Two scenarios are discussed here: the Pubic Protection and Disaster Relief (PPDR) in Callania and the Temporary Event in Bastian. In the PPDR scenario (Callania), seven use cases are included based on the work defined in D2.1 [1]: ABS.UC.01, ABS.UC.02, ABS.UC.03, ABS.UC.04, ABS.UC.05, ABS.UC.17 and ABS.UC.18. Three transfer learning based algorithms are designed for different scenarios in order to reduce the convergence speed for newly deployed eNBs and improve the system performance (e.g. retransmission probability and average delay). In the temporary event scenario (Bastian), four use cases are considered: ABS.UC.01, ABS.UC.02, ABS.UC.19 and ABS.UC.20. Two algorithms are developed to guarantee the high throughput density and reduce the convergence speed.

In the first part of this deliverable, an overview and introduction of this document is presented.

The second part shows an approach for cognitive spectrum resource management in the PPDR scenario. The main requirement is to develop a rapidly deployable and flexible network solution with broadband service in a large scale area (e.g. rural areas) [1] A transfer learning based cognitive spectrum assignment approach for PPDR scenario is introduced to achieve this goal , which includes the knowledge transfer and processing and the knowledge reward. A data mining technique, K-means clustering, is selected to deal with these problems. The details of the system model, the K-means learning model and the K-means learning algorithm are all described. Three transfer learning based cognitive spectrum management schemes are designed for different conditions in order to provide a flexible network solution, including newly deployed eNBs, existing eNBs looking for extra spectrum and existing eNB assigning spectrum.

The third part shows a case based cognitive cellular system for the temporary event scenario (Bastian) using the ABSOLUTE system. The main target of temporary event scenario is to provide network capacity enhancement in urban areas [1]. A combination of RL and case based reasoning (CBR) is investigated to improve the stability of reinforcement learning (RL) based dynamic spectrum access (DSA) algorithms for temporary event networks with dynamic topologies that use secondary LTE spectrum sharing, aimed at providing a high throughput density service and highly agile reconfiguration through reduced convergence speeds of the learning algorithm.

Next, a heuristically accelerated reinforcement learning (HARL) approach is proposed to speed up RL algorithms, particularly in the multi-agent domain. A radio environment map (REM) is used to improve the performance of RL algorithms applied to Dynamic Spectrum Sharing (DSS) problem.

Finally, a recommendation and conclusion chapter is given to present how to use all the material and recommended schemes in ABSOLUTE system. Then, it provides a summary of the work in the task.

1 Introduction

The main goal of ABSOLUTE is to '*design and validate a holistic and rapidly deployable mobile network to provide broadband services based on a flexible, scalable, resilient and secure network design*' [1]. Innovative concepts like cognitive mechanisms are of significant importance to ABSOLUTE's future success. Dynamic cognitive spectrum management techniques aim to enable a seamless network reconfiguration as well as efficient self-organizing networking, ensuring maximum system coverage, capacity and reliability.

In the ABSOLUTE project, cognitive spectrum management has been chosen as the solution to achieve the project's main goal. Under the first phase of work task T3.3, a thorough state-of-the-art summary has be presented, mainly dealing with spectrum assignment issues tailored to the heterogeneous requirement of ABSOLUTE. Meanwhile, preliminary recommendations are given as a basis for subsequent developments. The contributions have been shown in detail in the first deliverable of T3.3, "D3.3.1: Initial Approaches for Cognitive Spectrum Assignment using Distributed Artificial Intelligence". In the second phase of T3.3, the initial performance of a transfer learning based K-means clustering algorithm to deliver improved cognitive spectrum management is investigated for the PPDR scenario. The integration of the spectrum awareness techniques with the spectrum management approaches has also been discussed. Meanwhile, a cognitive dynamic spectrum management (DSM) scheme combined with distributed Q-learning and inter-cell interference coordination (ICIC) is proposed to provide better and more robust Quality of Service (QoS) than purely heuristic ICIC approach and a pure distributed Q-learning approach in the LTE downlink for temporary event scenario. Moreover, the possible interference between the AeNB and satellite is investigated for S-band. The research has been presented in the second deliverable of T3.3, "D3.3.2: Interim Performance Results of Cognitive Dynamic Spectrum Access".

The purpose of this deliverable is to propose the Performance Evaluation of Cognitive Dynamic Spectrum Access in 4G LTE-A ABSOLUTE system. This work is produced in month 32 of the ABSOLUTE project, which includes additional simulation results and analysis of ABSOLUTE cognitive dynamic spectrum management compared to the work in D3.3.2. The Callania Public Safety and Disaster Relief (PPDR) scenario and the Bastian temporary event scenario are considered in this document. The PPDR scenario requires a rapidly deployable and flexible network solution with broadband service [1]. The related use cases include ABS.UC.01, ABS.UC.02, ABS.UC.03, ABS.UC.04, ABS.UC.05, ABS.UC.17 and ABS.UC.18. The temporary event scenario aims to provide network capacity enhancement in the urban area. The following use cases are considered: ABS.UC.01, ABS.UC.02, ABS.UC.19 and ABS.UC.20. Meanwhile, the simulation experiments of approaches incorporating T3.1 awareness technique and the distributed reinforcement learning scheme for dynamic secondary spectrum sharing are also investigated.

Chapter 2 presents the further discussion of the transfer learning algorithms based on K-means clustering techniques for cognitive spectrum management in ABSOLUTE for PPDR scenario. Cognitive Spectrum Assignment techniques are highly desirable for LTE-A in ABSOLUTE because of their self-organizing and self-optimizing features, which are considered to improve the system Quality of Service (QoS) and reduce the convergence speed compared to traditional Q-learning algorithms. Knowledge transfer and Knowledge reward are the major issues in transfer learning. Data mining techniques are used to overcome these challenges, where a K-means clustering algorithm is considered to find the hidden structure based on the information exchanged. Three transfer learning based schemes are designed to work for different conditions.

Chapter 3 proposes a way of improving the stability of reinforcement learning (RL) based dynamic spectrum access (DSA) algorithms for temporary event networks with dynamic topologies that use secondary LTE spectrum sharing. A combination of RL and case based reasoning (CBR) is investigated to solve this problem, which has been successfully applied to decision problems. A distributed Q-learning based DSA is also explained, and then the details of a case-based Q-learning scheme for dynamic secondary spectrum sharing are provided.

In Chapter 4, a heuristically accelerated reinforcement learning (HARL) approach is designed to overcome the common disadvantage of RL algorithms described in Chapter 3, which is the need for many learning iterations to converge on an acceptable solution. The goal of HARL is to speed up RL algorithms, particularly in the multi-agent domain. Here, the HARL algorithm uses a radio environment map (REM) to mitigate the problem of poor temporal performance of RL algorithms applied to DSS problem.

The final chapter provides a discussion on all the material and competing schemes investigated in this deliverable, in order to show how to make use of these schemes in ABSOLUTE system. A summary and conclusion of the work is also provided here.

2 Advanced Transfer Learning based Cognitive Spectrum Management in 4G LTE ABSOLUTE with K-means Clustering

2.1 Introduction

In the work on spectrum management for disaster recovery and event servicing in the LTE-A ABSOLUTE [1] system, the main challenges are the highly dynamic feature of the ABSOLUTE system and the unpredictable post-disaster and temporary event scenarios, which makes spectrum management extremely complex. Traditional fixed frequency assignment is unlikely to perform well in such scenarios because firstly ABSOLUTE has AeNBs as well as a variety of different TeNBs. The coverage areas associated with the ABSOLUTE eNBs are significantly different, and there is likely to be potentially considerable overlapping of coverage areas. Secondly, network planning is unlikely to be performed in a post-disaster scenario so that the ABSOLUTE base stations are unlikely to be deployed in optimal locations or even sub-optimal locations. Thirdly, the ABSOLUTE system is designed to cope with a changing environment which means that the topology of the system changes with time. The spectrum assignment aspect of the ABSOLUTE system is required to be able to adjust itself with any changes form the environment and the system itself. Cognitive spectrum management techniques therefore are highly desirable for ABSOLUTE because of its self-organizing and self-optimizing features.

The purpose of this chapter is to investigate further cognitive spectrum management techniques for LTE-A ABSOLUTE system, which is the advanced work of Chapter 4, D3.3.2. The Callania PPDR scenario is assumed here based on the work in D2.1. The use cases include ABS.UC.01, ABS.UC.02, ABS.UC.03, ABS.UC.04, ABS.UC.05, ABS.UC.17, and ABS.UC.18. By using reinforcement-based learning, eNBs will assess the success level of a particular action. Transfer learning techniques tailored to ABSOLUTE are proposed to improve the system performance. By allowing neighbouring entities to exchange information gained through reinforcement learning at the minimum level, transfer learning approaches are proven to be significantly effective in reducing convergence time [2]. The further approaches of the transfer learning within the context of the LTE-A network are also discussed. Three transfer learning based schemes are designed to work for different conditions.

The remaining part of this chapter is structured as follows. Section 2.2 forms the system model and the role of reinforcement learning and transfer learning in CSM. The major issues for transfer learning in LTE-A ABSOLUTE network are then given in Section 2.3. The basic process of a powerful data mining technique (K-means clustering algorithm) is also introduced here to solve the knowledge transfers and process problem. Section 2.4 shows three TL based CSM algorithms for different conditions in details. Simulation results are presented in Section 2.5. Finally, conclusions are given in Section 2.6.

2.2 System Model

An urban/suburban/rural area with M AeNBs and N TeNBs is considered. Figure 2.1 is an example of how transfer learning based cognitive spectrum management with K-means clustering works in the ABSOLUTE network. It is assumed that TeNB 1 is the centre target TeNB, TeNB 2-7 are the adjacent TeNBs of TeNB 1. The reinforcement learning is used to develop knowledge of the local agents based on their own radio environments. The role of transfer learning is to exchange knowledge (The Q-values and received signal strength of the learning agents) from neighbouring TeNBs (TeNB 2-7), and then the exchanged information is processed locally (TeNB 1) to extract useful information. Here, K-means clustering algorithm plays an important role in processing exchanged information, which helps to find the more/less frequently used spectrum bands by adjacent TeNBs.

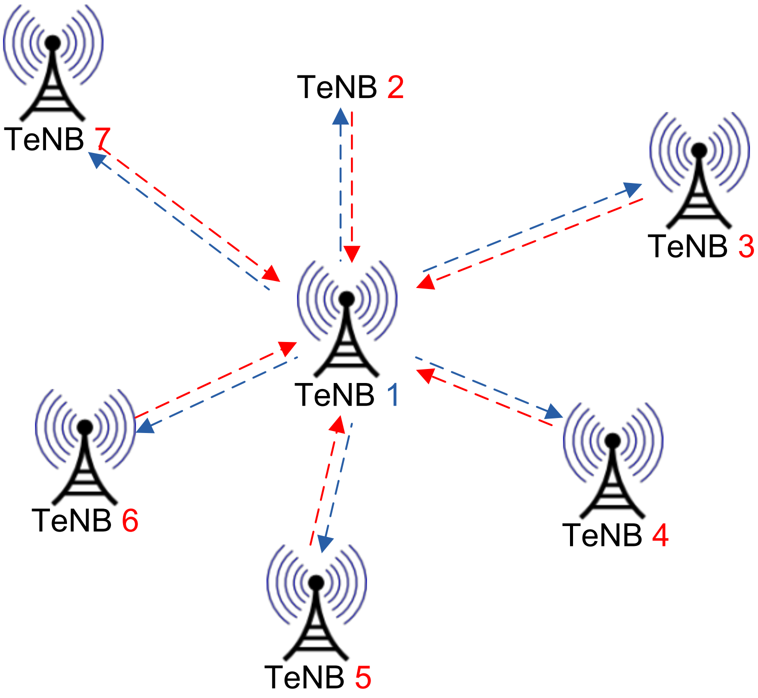


Figure 2.1 Transfer Learning Example

The transfer learning aims to help the target TeNB learn the spectrum assignment 'habits' from its adjacent TeNBs. The purpose of this is to avoid using the spectrum bands which are frequently used by adjacent TeNBs. In this chapter, the information received from neighbouring TeNBs are the Q-values of source tasks and the received signal strength, which tell us the overall spectrum allocation 'habits' at a specific service area and the 'weight' of adjacent TeNBs to be obtain. In other words, the frequently used spectrum bands on adjacent TeNBs with high received signal strength will have more impact on the target TeNB than the remaining adjacent TeNBs.

2.3 Major issues in Cognitive Spectrum Management

There are three major issues for the transfer learning based cognitive spectrum management in the ABSOLUTE system:

- Knowledge transfer and processing.
- Knowledge reward.

- Coping with transfer learning in LTE networks.

2.3.1 Knowledge Transfer and Processing

In this report, the knowledge being transferred is assumed related to two attributes: Q-Tables and received signal strength from adjacent TeNBs. Data mining is selected to find hidden structure of the knowledge transferred from neighbouring eNBs. The definition of Data Mining is given in [3]: *Data Mining, an interdisciplinary subfield of computer science, is the computational process of discovering patterns in large data sets, involving methods at the intersection of artificial intelligence, machine learning, statistics, and database system.*

For Q-Tables, the patterns we would like to discover in the cognitive spectrum management are the Virtual Resource Blocks (VRBs) which are frequently used by adjacent eNBs. For received signal strength, the impact of neighbouring eNBs is distinguished. Some eNBs will be considered to have higher weights than others for the target eNB. Specifically, the cluster analysis of data mining is used to achieve these goals. Clustering is concerned with grouping together objects that are similar to each other and dissimilar to the objects belonging to other clusters [4].

2.3.1.1 Data Pre-processing

Data pre-processing is a data mining technique that involves transforming raw data into an understandable format. Data have quality if they satisfy the requirements of the intended use. There are many factors comprising data quality, including accuracy, completeness, consistency, timeliness, believability, and interpretability. Even when the data is in the standard form it cannot be assumed that it is error free. In real-world datasets erroneous values can be recorded for variety of reasons, including measurement errors, subjective judgements and malfunctioning or misuse of automatic recording equipment. As shown in Figure 2.2, the major work of data pre-processing are summarized as follows [5]:

- **Data cleaning** – To clean the data by filling missing values, smoothing noisy data, identifying or removing outliers and resolving inconsistencies.
- **Data integration** – To combine data residing in different sources and providing users with a unified view of data.
- **Data reduction** – To obtain a reduced representation of the data set that is much smaller in volume, yet produces the same or similar analytical results.
- **Data transformation** – Normalization, data discretization and concept hierarchy generation.

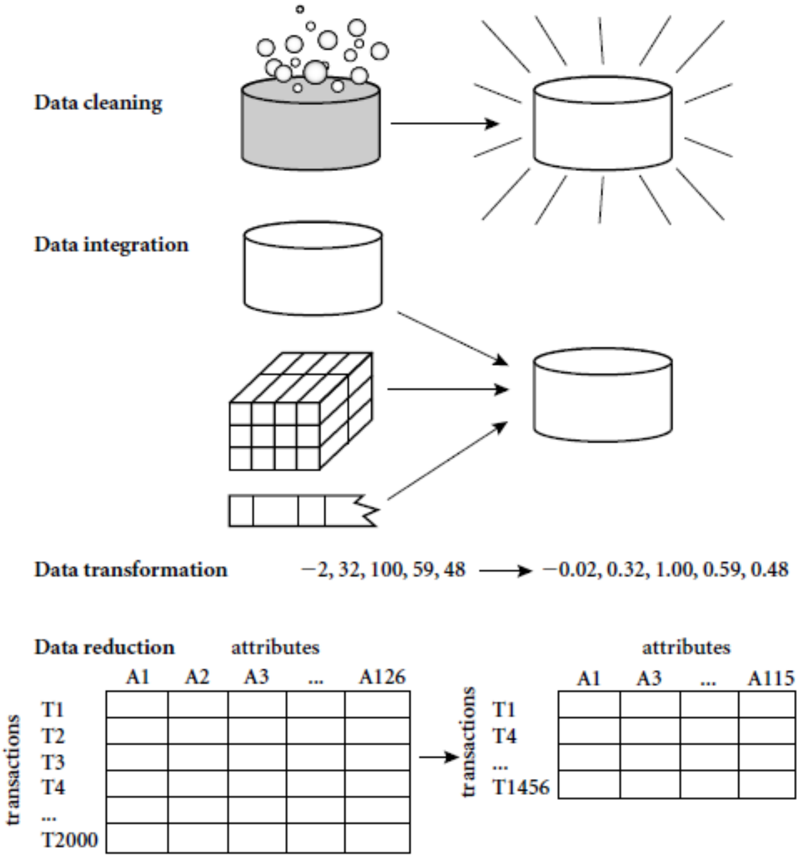
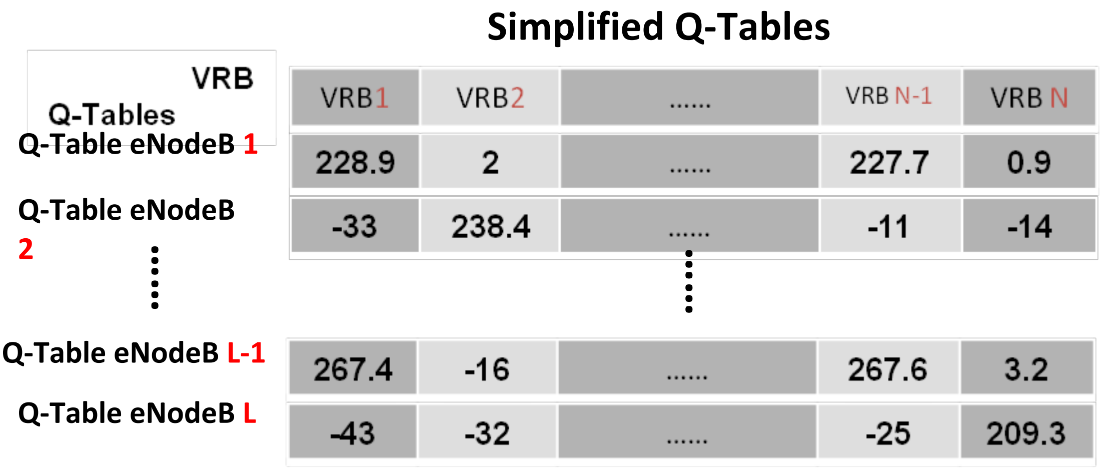


Figure 2.2 Forms of data pre-processing, directly reproduced from [5]

Data pre-processing in cognitive spectrum management

The purpose of data pre-processing in ABSOLUTE's cognitive spectrum management is to transform the raw data received from adjacent eNBs into an understandable, simplified format. We consider a target eNB received Q-Tables from *L* adjacent eNBs, which is shown in Figure 2.3. It is assumed in principal that all eNBs have the access to all *N* available VRBs (frequency bands). $\bigcup_{j \in S} Q_S^j$ are the Q-values of source tasks and Q_T^i are the Q-values of the target task *i*. The received signal strength of eNBs from source eNBs to target eeNB *i* are denoted as $p_r^{j,i} = (p_r^{1,i}, \ldots, p_r^{L,i})$, where *j* is the j^{th} source eNBs and $j \in [1, L]$.



Simplified Q-Tables

VRB Q-Tables	VRB1	VRB2	......	VRB N-1	VRB N
Q-Table eNodeB 1	228.9	2	......	227.7	0.9
Q-Table eNodeB 2	-33	238.4	......	-11	-14
⋮			⋮		
Q-Table eNodeB L-1	267.4	-16	......	267.6	3.2
Q-Table eNodeB L	-43	-32	......	-25	209.3

Figure 2.3 Q-Tables received from adjacent eNBs

The details of data pre-processing in cognitive spectrum management can be shown as follows:

1) Data cleaning (remove noise)

For the Q-tables received from neighbouring eNBs, we are interested in the VRBs with high Q-values (frequently used VRBs). Here, the VRBs with low Q-values mean the signal qualities of these VRBs are not good or these VRBs did not have a chance to be sensed. They are considered as the 'noise', which impact negatively on the processing of data mining. Thus, we set the low Q-values to zero if the Q-values are smaller than a threshold:

$$Q_S^{j,t} = \begin{cases} 0, & Q_S^{j,t} < \alpha \cdot \max(Q_S^j) \\ Q_S^{j,t}, & Q_S^{j,t} \geq \alpha \cdot \max(Q_S^j) \end{cases}$$

where $Q_S^{j,t}$ is the Q-value of t^{th} VRB of j^{th} eNB, α is the factor of removing Q-values. In this report, it is assumed $\alpha = 0.4$.

For the received signal strength from neighbouring eNBs, we assume that the eNBs with low signal strengths have very limited impacts on the target eNB. In other words, these eNBs are far away from target eNB. Thus, the information received from these eNBs can be ignored. Thus, we have:

$$p_r^{j,i} = \begin{cases} p_{r,thres}, & p_r^{j,i} \leq p_{r,thres} \\ p_r^{j,i}, & p_r^{j,i} > p_{r,thres} \end{cases}$$

where $p_{r,thres}$ is the threshold to ignore the information from adjacent eNBs, this value will be equal to 0 after normalization.

2) Data transformation

Here, Min-Max Normalization is selected to normalize both Q-values and the received signal strength into a specific range $B \in [0,15]$. Thus, the data requirement associated with a specific Q-value is 4 bits.

For the Q-tables, the normalized Q-values are:

$$Q_{S,new}^{j,t} = \frac{Q_S^{j,t} - \min(Q_S^j)}{\max(Q_S^j) - \min(Q_S^j)} \cdot (\max(B) - \min(B)) + \min(B)$$

where Q_S^j is the set of Q-values of j^{th} eNB, *B* is the set of the normalized data.

For the received signal strength, the normalized received signal strengths are:

$$p_{r,new}^{j,i} = \frac{p_r^{j,i} - p_{r,thres}}{\max(p_r^{j,i}) - p_{r,thres}} \cdot (\max(B) - \min(B)) + \min(B)$$

where $p_{r,thres}$ is the threshold to ignore the information from adjacent eNBs, B is the set of normalized data.

3) Data integration and reduction

The VRBs with high Q-values, and the eNBs with high received signal strength may provide more interference than the others. Thus, the converted Q-value on a specific VRB based on the information received from adjacent eNBs is shown as:

$$Q_P^t = \sum_{j=1}^{L} (Q_{S,new}^{j,t} \cdot p_{r,new}^{j,i})$$

where $Q_{S,new}^{j,t}$ is the normalized Q-table of j^{th} eNB in t^{th} VRB, $p_{r,new}^{j,i}$ is the normalized received signal strength of j^{th} eNB. A high value of Q_P^t means t^{th} VRB is frequently used by neighbouring eNBs.

2.3.1.2 K-means Clustering Algorithm in ABSOLUTE CSM

K-means clustering is an exclusive clustering algorithm, which is straight forward to implement (computationally faster than hierarchical clustering) and can be applied to even large data sets. This algorithm is widely used in market segmentation, computer vision, geostatistics, astronomy and agriculture [6]. In this deliverable, K-means clustering can be used in our scheme to find the VRBs in which we are interested. Each object is assigned to precisely one of a set of clusters. For this method of clustering it starts by deciding how many clusters (*k* clusters) we would like to form the data. Next, *k* points are selected as the centroids of *k* clusters. Then, each of the points in the data is assigned one-by-one to the cluster which has the nearest centroid, which is shown as:

$$\underset{\mathbf{S}}{\arg\min} \sum_{i=1}^{k} \sum_{\mathbf{x}_j \in S_i} \|\mathbf{x}_j - \boldsymbol{\mu}_i\|^2$$

We recalculate the centroids of the *k* clusters based on the assigned clusters, and then repeat the previous steps until the centroids no longer move. The entire process of the algorithm is summarised in as Table 2.1 [4]:

Table 2.1 K-means clustering algorithm [4]

K-means algorithm:
1. Decide how many clusters (k) we would like to form from our data.
2. Select k objects, use them as the initial set of k centroids.
3. Assign each of the objects to the cluster for which is nearest the centroid.
4. Recalculate the centroids of the k clusters.
5. Repeat steps 3 and 4 until the centroids no longer move.

In Figure 2.4, we consider a set of pre-processed data of the Q-values (Converted Q-Table) in every VRB (channel, 30 VRBs/channels in total). Here, we try to divide the VRBs into three groups (High Q-values, middle Q-values and low Q-values). The initial cluster heads (centroids) of these three groups are set as: the maximum Q-value of the converted Q-Table, the mean Q-value of the converted Q-Table and the minimum Q-value of the converted Q-Table.

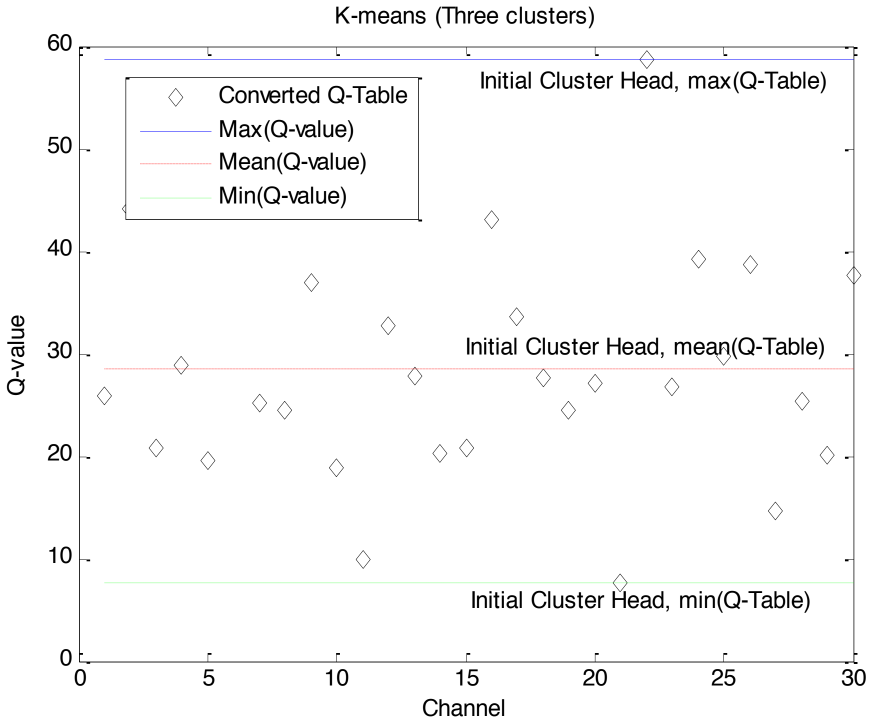


Figure 2.4 K-means algorithm example 1

In Figure 2.5, The Q-values are divided into three clusters based on K-means algorithm. The low Q-value cluster includes the VRBs which are seldom used by neighbouring eNBs. In other words, they are the potential available VRBs of the target eNB.

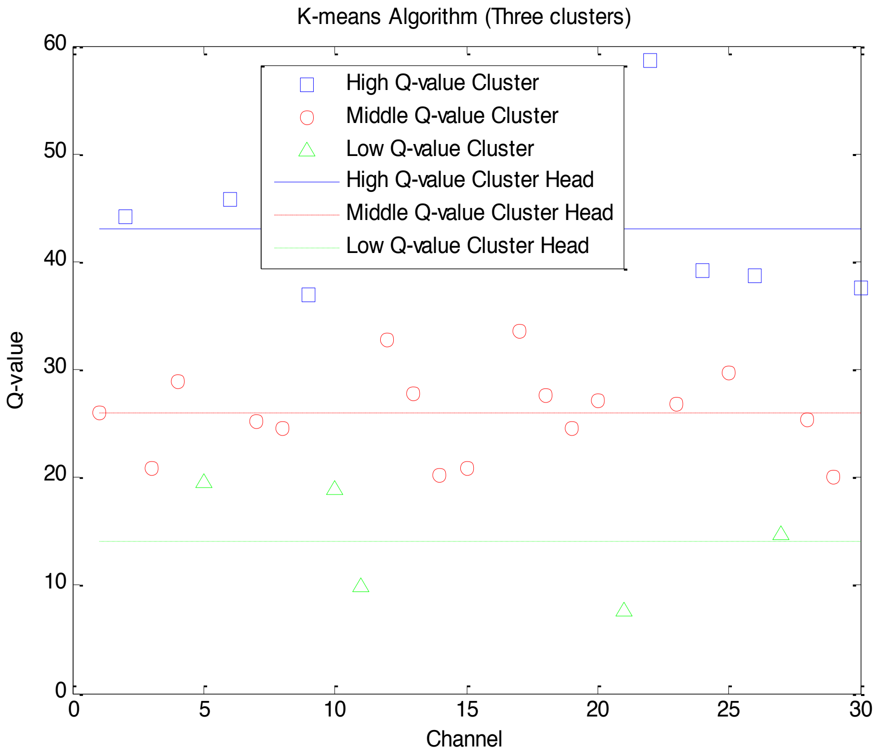


Figure 2.5 K-means algorithm example 2

2.3.2 Knowledge Reward

In this report, knowledge reward is related to how to make use of the information received from neighbouring eNBs in order to improve the system performance. The following conditions are considered in the ABSOLUTE systems:

- Reward when new eNB is deployed
- Reward when an existing eNB tries to explore additional potential available VRBs
- Reward when an existing eNB ties to assign VRBs for new arrivals

The principles of how to reward the above three scenarios are presented in the upcoming sub-sections. The details of these transfer learning algorithms will be introduced in Section 2.4.

2.3.2.1 Reward when new eNB is deployed

When a new eNB is deployed in ABSOLUTE system, it has no historical knowledge about local radio environment. In other words, its Q-Table is empty. That means it may take a long time for this eNB to develop its own 'habit' on spectrum assignment. However, it is possible to generate a converted Q-Table based on the information received from its adjacent eNBs. In Figure 2.6, the newly deployed TeNB1 could receive Q-Tables and received signal strength from its neighbours (TeNB2-7). The less frequently used spectrum bands by neighbours can be considered as the potential good spectrum for TeNB1.

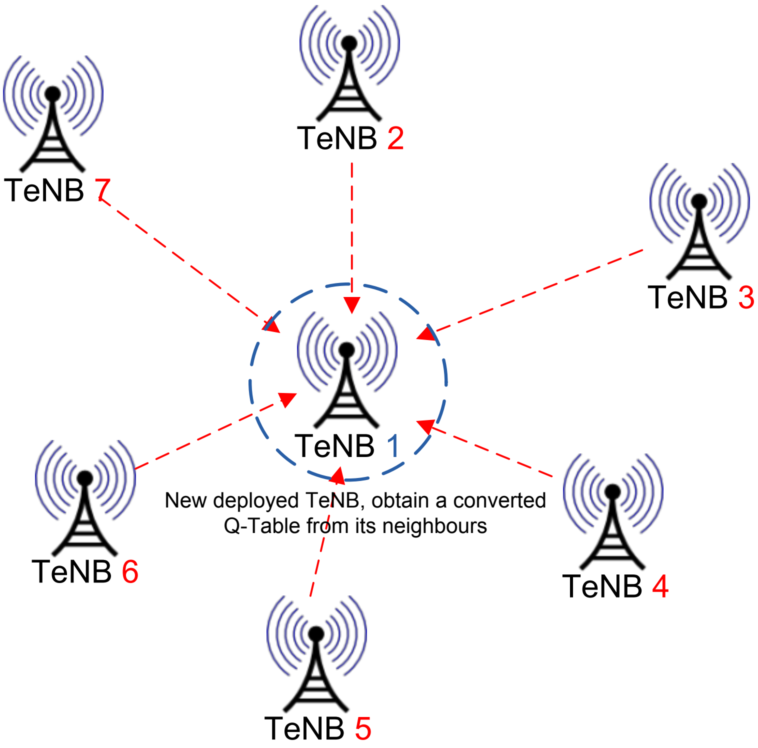

TeNB 2
TeNB 7
TeNB 3
TeNB 1
TeNB 6
New deployed TeNB, obtain a converted Q-Table from its neighbours
TeNB 4
TeNB 5

Figure 2.6 Reward a newly deployed eNB

2.3.2.2 Reward when exploring potential available VRBs

In some situations, the group of most frequently used VRBs of the target eNB may be fully occupied. Thus, the target eNB is required to find other potential available VRBs. The typical solution of learning based algorithm is to sense the high Q-value VRBs in remaining available spectrum bands from local Q-Table. Transfer Learning algorithm can be used here to improve the performance of such process. A joint consideration of spectrum 'habits' from neighbouring eNBs and target eNB may reduce the number of VBRs to be sensed and increase the convergence speed of system.

2.3.2.3 Reward when assigning VRBs for new arrivals

Figure 2.7 is an example of new arriving UE, where TeNB1 can provide the best link. It is assumed that TeNB1-4 are located close to the UE and TeNB5-7 are far away from it, even though TeNB2-7 are the adjacent TeNBs of TeNB1. It can be clearly noticed that the 'habits' of TeNB2-4 play more important roles than TeNB5-7. Thus, the knowledge of TeNB2-4 is considered as useful information when TeNB1 tries to allocate VRBs for this UE. This is because TeNB1 should avoid the frequently used VRBs in TeNB2-4 in order to limit interference at the boundary of two adjacent eNBs. The 'near/far' information from TeNBs to UEs can be obtained when newly arriving UEs try to access the system, where the link environments between UE and potential available TeNBs are measured. The details of how it works in the ABSOLUTE system will be explained in Section 2.4.

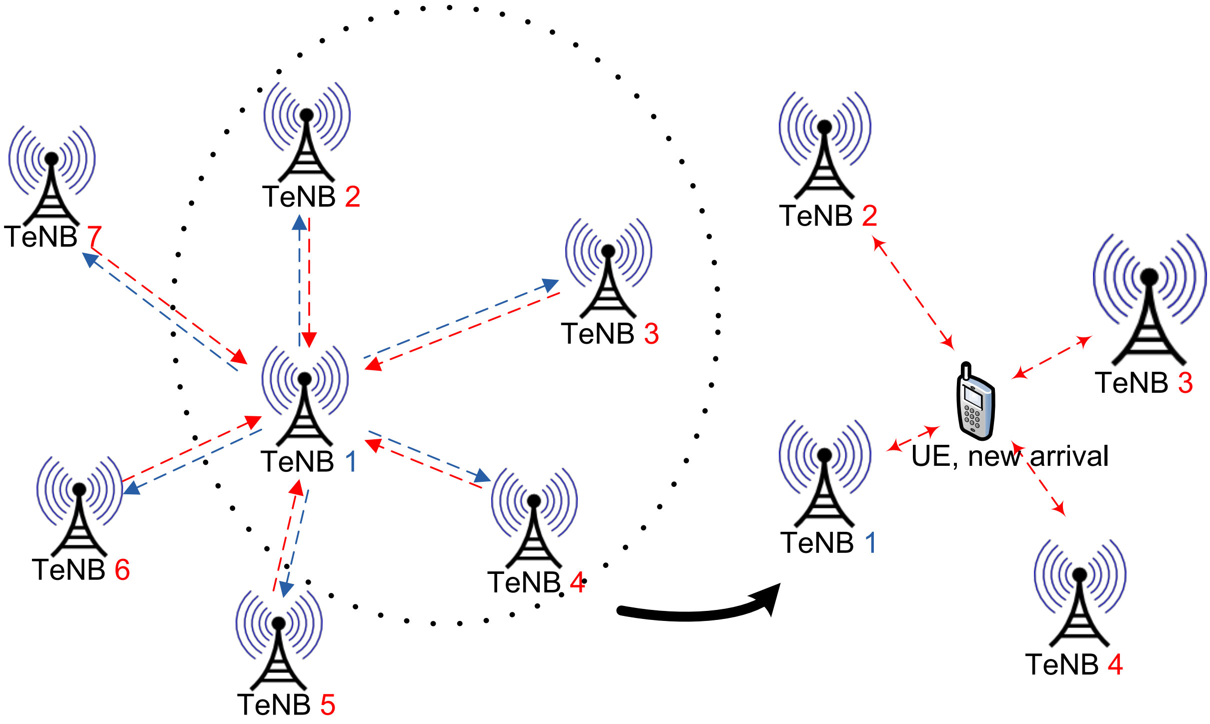


Figure 2.7 New Arrival (UE)

Here, we present a small scale scenario with 7 TeNBs, similar to that shown in Figure 2.7. It is assumed that the number of UEs being served by centre TeNB1 is twice that of other TeNBs'. The offered traffic is set at high level, and the UEs which are located in the middle of two or more TeNBs may provide/suffer significant interference. The number of VRBs (channels) is 10. Figure 2.8 and Figure 2.9 show the number of channels being used by each TeNB after 500, 1k, 3k and 10k events. During the first 500 events, the best SINR scheme is used only to develop eNBs' spectrum selection habits. The transfer learning algorithm works from 501 events. It can be found out that there is no specific spectrum selection 'habit' in the first 500 event diagram. This is because the spectrum assignment scheme is the best SINR scheme before the learning algorithm starts. Based on the performance of TeNB1 from 1k events to 10k events, spectrum is efficient used with low interference (blocking probability = 3%) here. It is noticed that all spectrum bands are frequently used, but some are more frequently. This is because new arrivals only need to avoid the frequently used channels of their own adjacent TeNBs, rather than all the neighbours of target TeNB1.

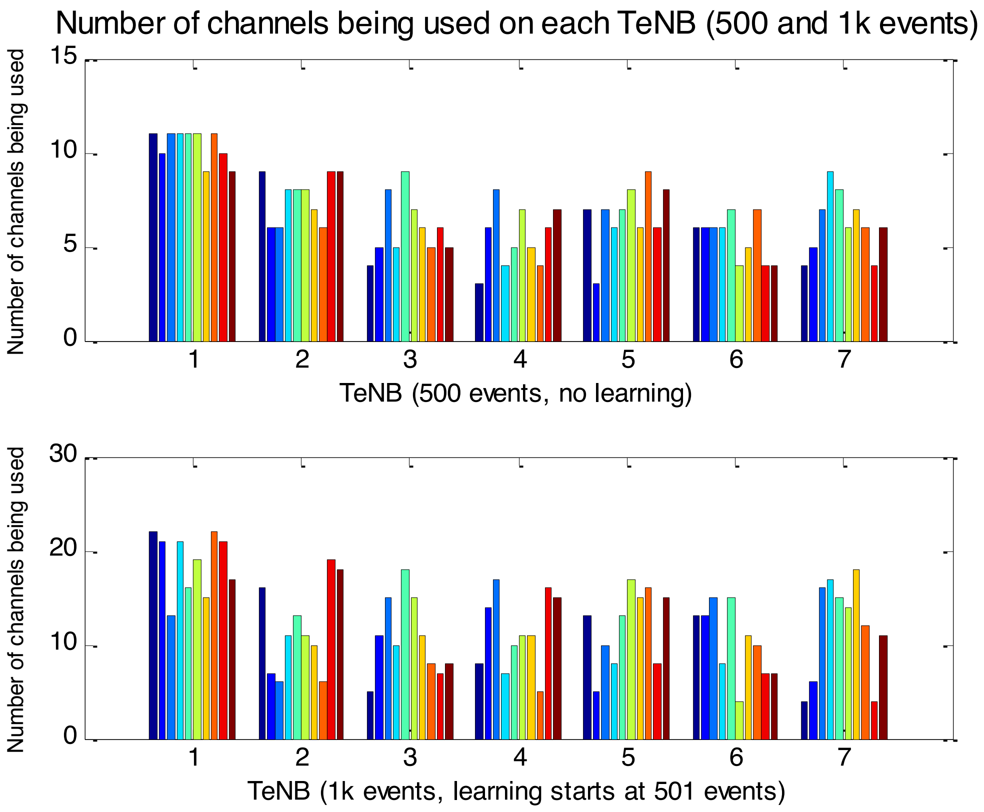


Figure 2.8 Number of channels being used on each TeNB (500 and 1k events)

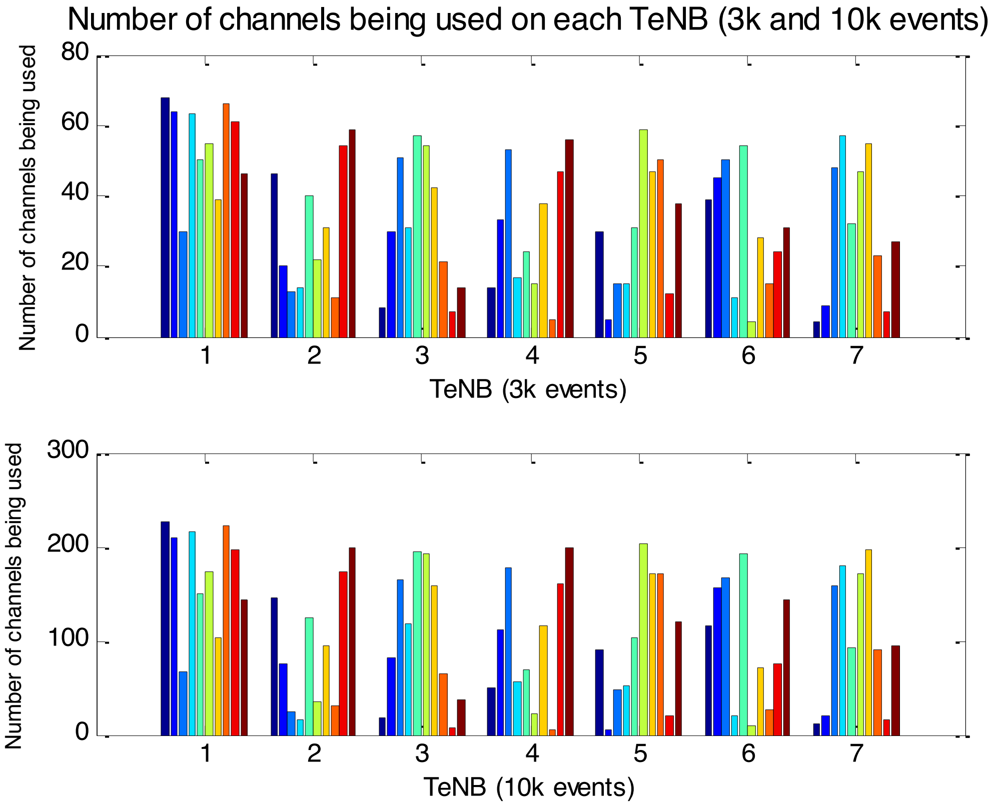


Figure 2.9 Number of channels being used on each TeNB (3k and 10k events)

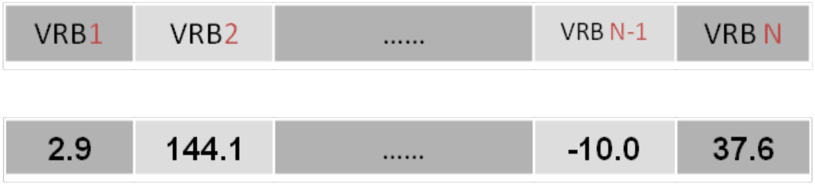
2.3.3 Approach for Utilising Transfer Learning in 4G LTE ABSOLUTE Networks

The main difficulty with applying transfer learning to LTE-A networks is that extra information is transferred from adjacent eNBs. In the current LTE-A standard, such mechanisms for transfer learning have not been designed. However, it is possible for this to be achieved by exchanging information between adjacent eNBs over the X2 interface in LTE network [7]. Here, X2 is defined as a “logical” interface between only neighbouring eNBs and can be switched over the existing transport network [3]. Currently, the X2 interface is mainly used for handover, load management, Co-ordinated Multi-Point transmission or reception (CoMP), Network Optimisation, eNB configuration update, mobility optimisation and general management. Next, it is necessary to optimize the information exchanged (limit data size delivered per second) over X2 interface in order to avoid too much resource being taken by transfer learning mechanism. The total volume of data transferring from one eNB to another can be calculated as:

$$L_{TL-data} = L_{Q-Table} \cdot f_s$$

where $L_{Q-Table}$ is the data length of a typical Q-Table, f_s is the transferring frequency (per second) between adjacent eNBs .

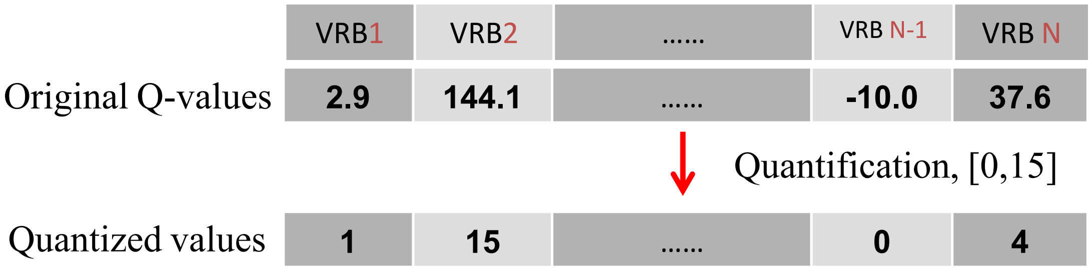
Thus, there are two possible ways to reduce the negative impact on the system capacity: reducing the data size of Q-Tables or increasing the transferring period. Figure 2.10 gives an example of a typical Q-Table, at least 11bits (or 12) is required to deliver a specific Q-value in binary.



VRB1	VRB2	......	VRB N-1	VRB N
2.9	144.1	......	-10.0	37.6

Figure 2.10 Example of a typical Q-Table

However, it is possible to deliver quantized Q-Table instead of the original one. Assuming there are 16 quantized levels Figure 2.11, the volume of a specific Q-value is 4 bits, only one third the volume of data is required to transmit compared to scheme forwarding the original Q-Table to neighbouring eNBs directly.



	VRB1	VRB2	......	VRB N-1	VRB N
Original Q-values	2.9	144.1	......	-10.0	37.6
Quantized values	1	15	......	0	4

Figure 2.11 Quantized Q-Table

2.4 Advanced Transfer Learning CSM with K-means Clustering in details

In this section, the details of transfer learning algorithm in 4G LTE ABSOLUTE cognitive spectrum management are introduced. The learning model on a specific eNB (Single State Q-Learning) is present firstly, and then how transfer learning is applied into the system for different scenarios is discussed.

Single-State Q-Learning is originally proposed to solve Single-state games in Computer Science. A reformulation of the standard Q-learning algorithm is carried out that the Q values of actions are effectively the estimation of the usefulness of the actions in the next step of the learning process. By maintaining a Q value for each action, the agent is able to select the action based entirely on it Q value and the Q value of the selected action will be updated by receiving a reward. The update function is defined as:

$$Q(a) \leftarrow Q(a) + \gamma(r(a) - Q(a))$$

γ is the learning rate ($0 \leq \gamma < 1$) and $r(a)$ is the immediate reward of choosing action a.

Thus, instead of pursuing the optimal policy π^*, the objective of each agent i is to find the action with the highest estimated Q-value Q*. This greatly reduces the complexity of the learning model as the definition of the states and the state transition probability are no longer required. However, in this case the reward $r(a)$ needs to be properly defined so that the feedback of taking an action reflects the successfulness correctly. Particularly in wireless communication systems, the reward is more useful when associated with the physical measurements of the system in order to facilitate the learning process. It is proposed at the early stage that the *SINR* measurements and/or link capacity measurements are taken as rewards. If we assume that an eNB select the spectrum band *r* with action *a* then the reward function can then be defined as follows:

$$r(a) = C_r^i$$

Or

$$r(a) = \gamma_r^i$$

The role of transfer learning is to exchange the Q-values of the learning agents, and then the exchanged information is processed locally to extract useful information. Some aspects of transfer learning based topology management were introduced in D4.1.4 [9]. The flow chart of transfer learning is shown in Figure 2.12. The process includes two parts: knowledge transfer and initial processing, knowledge reward. In other words, they are 'what information is received' and 'how to make use of the information'. The first part has been explained in the previous sections. The data pre-processing techniques are used to process the raw data received from neighbouring eNBs, and then the K-means clustering algorithm is carried out to find the potential available VRBs for the target eNB.

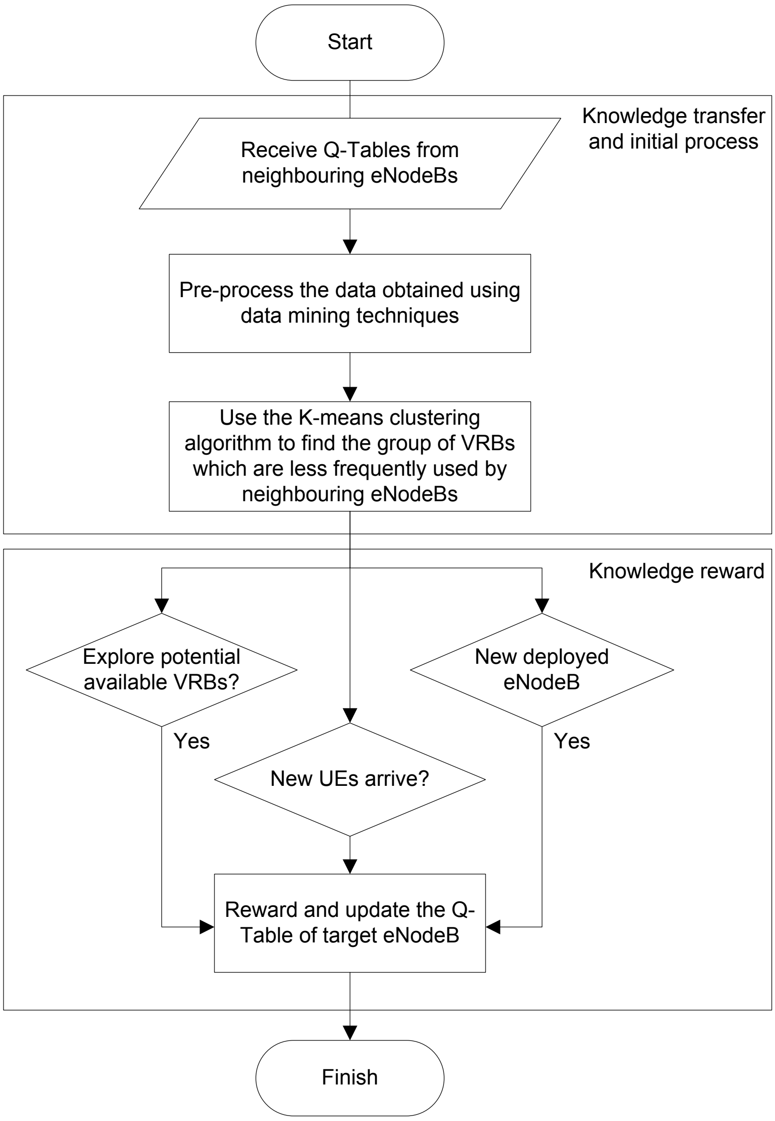

Start
Knowledge transfer and initial process
Receive Q-Tables from neighbouring eNodeBs
Pre-process the data obtained using data mining techniques
Use the K-means clustering algorithm to find the group of VRBs which are less frequently used by neighbouring eNodeBs
Knowledge reward
Explore potential available VRBs?
New deployed eNodeB
Yes
New UEs arrive?
Yes
Reward and update the Q-Table of target eNodeB
Finish

Figure 2.12 Flow chart of transfer learning

In the second part,

$$Q_T^i \leftarrow f_T(\bigcup_{j \in S} Q_S^j, Q_T^i)$$

where f_T is the target predictive function. $\bigcup_{j \in S} Q_S^j$ are the Q values of source tasks and Q_T^i are the Q values of the target task *i*. The focus of the cognitive spectrum assignment task is to discover the most suitable target predictive function f_T in the context of ABSOLUTE. The transfer learning is assumed to provide positive impact on the target eNB for three different scenarios in this deliverable.

2.4.1 Transfer Learning algorithm for Newly Deployed eNBs (TL-ND algorithm)

Transfer Learning algorithm for Newly Deployed eNBs (TL-ND) aims to help the newly deployed eNB learn the spectrum allocation 'habits' from its adjacent eNBs. The less frequently used VRBs will be positive rewarded in the Q-Table of new eNB. In other words, these VRBs are expected to obtain higher priority being selected compared to the remaining VRBs, in order to avoid interference from nearby.

Here, the group of less frequently used VRBs is denoted as VRB_{low}based on the work in Section 2.3. The converted Q-values on all spectrum band is denoted as Q_P^t. Thus, the converted Q-values on the group of less frequently used VRBs are $Q_P^{VRB_{low}}$. The VRB with least converted Q-values means it is the top choice for the newly deployed eNB. The reward is considered as:

$$r(a) = \frac{\max(Q_P^{VRB_{low}}) - Q_P^{VRB_{low},q}}{\max(Q_P^{VRB_{low}}) - \min(Q_P^{VRB_{low}})} \cdot (\max(B) - \min(B)) + \min n(B)$$

where $Q_P^{VRB_{low},q}$ is the converted Q-value on q^{th}VRB, *B* is the set of the normalized data, $B \in [0,10]$.

If the Q-Table of newly deployed eNB is empty, we have:

$$Q_T^i = r(a)$$

2.4.2 Transfer Learning algorithm for Exploring extra available Spectrum Bands (TL-ESB algorithm)

TL-ESB algorithm provides a solution of exploring extra available spectrum bands for existing eNBs. If the number of VRBs being sensed when an UE tries to access the network is higher than a threshold, and this has happened continuously in the past few transmissions, we assume that the convergence of system is not at stable stage. The target eNB needs to find extra potential VRBs. For example, it is assumed that the threshold is equal to two, and we only look back the past three transmissions. Here, a small number of past transmissions help the system start exploration fast. It means that if the VRBs with the top two Q-values in the Q-Table are not suitable for access, over the past three transmissions, extra exploration is required, so that the target eNB needs to process the radio environment information received from adjacent eNBs, in order to find new available VRBs. Here, the reward depends on the difference between the maximum Q-value in the group of low Q-values and the Q-value of q^{th}VRB. The details are presented as:

$$r(a) = (\frac{\max(Q_P^{VRB_{low}}) - Q_P^{VRB_{low},q}}{\max(Q_P^{VRB_{low}}) - \min(Q_P^{VRB_{low}})} \cdot (\max(C) - \min(C)) + \min(C)) \cdot (\max(Q_P^{VRB_{low}}) - Q_P^{VRB_{low},q})$$

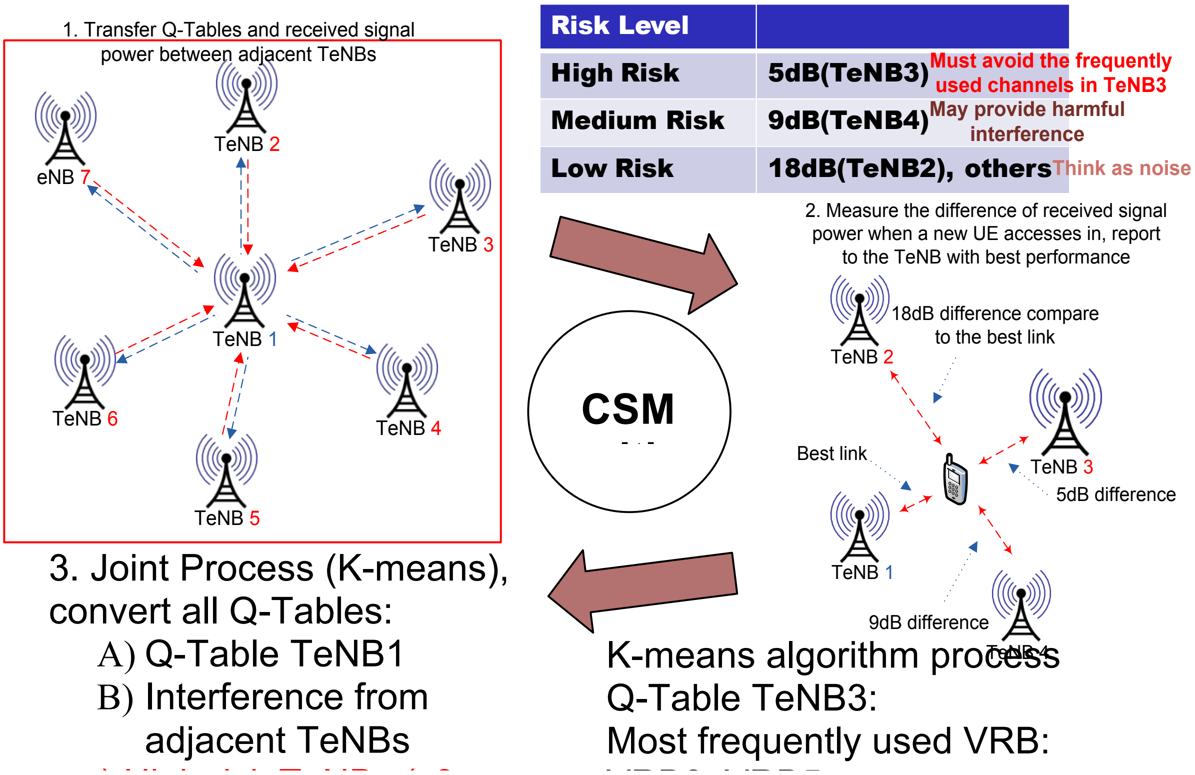
where $Q_P^{VRB_{low},q}$ is the converted Q-value on q^{th} VRB, $\max(Q_P^{VRB_{low}}) / \min(Q_P^{VRB_{low}})$ is the maximum/minimum element in the set of $Q_P^{VRB_{low}}$, *C* is the set of the normalized data, $C \in [0.05,0.3]$.

Then, we update the Q-value vector:

$$Q(a) \leftarrow Q(a) + \gamma(r(a) - Q(a))$$

2.4.3 Transfer Learning algorithm for Assigning Spectrum Bands (TL-ASB algorithm)

TL-ASB algorithm is used when new UE arrives. The basic process is shown in Figure 2.13. The target TeNB receives information from its neighbours firstly. Next, UEs are required to measure channel quality when accessing into LTE networks, and then a TeNB is chosen with the best link. In this example, TeNB1 is selected as the best link. Meanwhile, the UE reports the difference of received signal power to TeNB1 (best performance). A small value of difference means the UE is located in the middle of two TeNBs (TeNB1 and TeNB3), which means the UE may provide/suffer significant interference to/from TeNB3. We then rank TeNB3 as a high risk TeNB for the UE. Thus, the VRBs being assinged to the new arriving UE must avoid the frequently used VRBs in TeNB3. Similarly, we rank the TeNBs (TeNB4) with medium value of difference as medium risk TeNBs for the UE. It may provide/suffer some inteference to/from TeNB4. The TeNBs with a high value of difference are ignored and considered as noise. The final step is a joint process of K-means algorithm and knowledge received. The purpose is to find the potential high interference spectrum bands from the adjacent TeNBs of new arriving UEs depending on their risk levels, and avoid to use these spectrum bands during spectrum assignment. Here, only high and medium risk TeNBs are considered in the process. Meanwhile, we reward high risk TeNBs with a factor of 3.



Risk Level	
High Risk	5dB(TeNB3)
Medium Risk	9dB(TeNB4)
Low Risk	18dB(TeNB2), others

1. Transfer Q-Tables and received signal power between adjacent TeNBs
eNB 7
TeNB 2
TeNB 3
TeNB 1
TeNB 6
TeNB 4
TeNB 5
Must avoid the frequently used channels in TeNB3
May provide harmful interference
Think as noise
2. Measure the difference of received signal power when a new UE accesses in, report to the TeNB with best performance
18dB difference compare to the best link
TeNB 2
CSM
Best link
TeNB 3
5dB difference
TeNB 1
9dB difference
TeNB 4
3. Joint Process (K-means), convert all Q-Tables:
A) Q-Table TeNB1
B) Interference from adjacent TeNBs
K-means algorithm process Q-Table TeNB3:
Most frequently used VRB:

Figure 2.13 Process of TL-ASB algorithm

The converted Q-Table is presented as:

$$Q_P = \sum (Q_i^{\Pr e-P} p_{i,UE}^{\Pr e-P} RF_{i,UE})$$

where $Q_i^{\Pr e-P}$ is the pre-processed Q-Table of TeNB i, $p_i^{\Pr e-P}$ is the pre-processed received signal strength from TeNB *i* to new arriving UE, $RF_{i,UE}$ is the risk factors of TeNB *i* to new arriving UE ($high\ risk = 3; medium\ risk = 1; low\ risk = 0$).

The potential available VRBs with high Q-values and less interference for new arriving UE are shown as:

$$ch_{Q_{t\arg et}t}^{VRB_{ava}} = ch_{Q_{t\arg et}}^{VRB_{high}} \setminus ch_{Q_P}^{VRB_{high}}$$

Where $ch_{Q_{t\arg et}}^{VRB_{high}}$ is the group VRBs wtih high Q-values on target TeNB, $ch_{Q_P}^{VRB_{high}}$ is the group of frequently used VRBs based on the result of converted Q-Table.

2.5 Performance Evaluation

This section examines the performance of cognitive spectrum management with transfer learning in a large scale simulation, with comparisons to conventional Q-Learning algorithm. The results of three TL algorithms will be presented as well.

2.5.1 ABSOLUTE's Large Scale Deployment

Here, we present a large scale ABSOLUTE architecture used in D3.3.2. The PPDR (Callania) scenario is used here for simulation and comparison. The use cases defined based on the work in D2.1 include ABS.UC.01, ABS.UC.02, ABS.UC.03, ABS.UC.04, ABS.UC.05, ABS.UC.17 and ABS.UC.18 [10]. As shown in Figure 2.14, two Aerial eNBs are deployed to provide macro cell services for a 40 km by 20 km rural area. The potential coverage radius of Aerial eNB (AeNB) is more than 10 km. Several terrestrial eNBs are deployed on the north east of the service area aiming to enhance the capacity for local hotspot areas (e.g. little town or village). The hotspot area occupies a 5 km by 4 km area. There are 30 possible deployment locations for Terrestrial eNBs (TeNBs) in the hotspot area, 25 of them have been deployed before the simulation starting. The remaining 5 will be switched on during simulation.

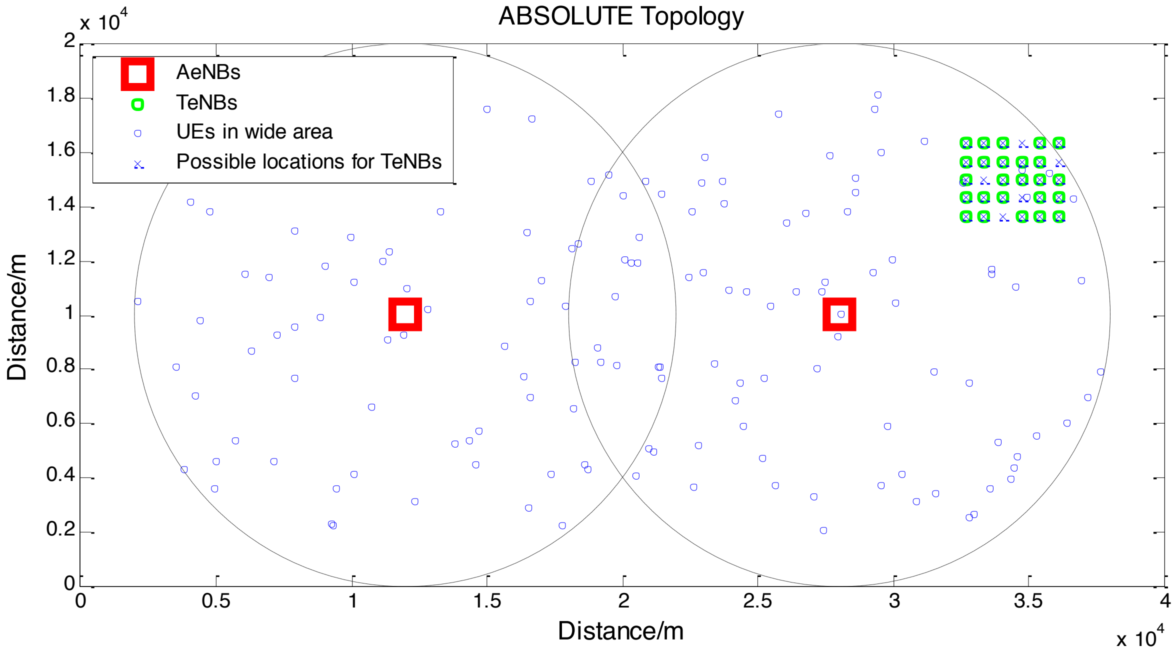


Figure 2.14 ABSOLUTE large scale architecture

The system level simulation uses the parameters proposed on D2.5.2 and D2.6.1. The details are shown in Table 2.2.

Table 2.2 System Parameters

Parameters		Values
Number of AeNBs		2
Number of TeNBs		25+5 (switched on later)
Number of UEs		1500
Number of Events		150000
Probability of UEs in wide area		10%
Probability of UEs in hotspot area		90%
Transmit Power	AeNB	25 dBm
	TeNB	25 dBm
	UE	23 dBm
Thermal Noise		-174 dBm/Hz
File Transfer Traffic Model	Inter-arrival time	Pareto distribution
	Mean file size	1 MB
Antennas		Omni-directional
Bandwidth	AeNB (uplink)	5 MHz; 25 RBs (180kHz/RB)
	TeNB (uplink)	5 MHz; 25 RBs (180kHz/RB)
Carrier Frequency	AeNB	700 MHz
	TeNB	2.6 GHz
Propagation	Aerial	Free space path loss + Log-normal
	Terrestrial	WINNER II D1 (LOS 400-500 m)
Antenna height	AeNB	300m
	TeNB	15m
	UE	1.5m
Back off Time (ms)		LTE Back off Parameter B
Link Selection		Best signal
Transfer Learning Period (ms)		200 ms

2.5.2 Simulation Results

A) Transfer Learning algorithm for Newly Deployed eNBs (TL-ND)

In the first set of results, the TL-ND algorithm is examined to show how transfer learning based cognitive spectrum management schemes reward newly deployed TeNBs. A total number of 25 TeNBs are activated at the beginning of the simulation. 5 more TeNBs will be switched on during the simulation. The average offered traffic per TeNB is kept constant. Convergence is a crucial issue in the previously proposed Q-learning based cognitive radio network, where the speed of achieving a stable state is very slow. Transfer learning with the received information from neighbouring eNBs provides a preliminary knowledge of radio environment. The newly deployed TeNBs can obtain information, like the VRBs that are less frequently used by neighbouring eNBs, and the potentially available VRBs themselves.

Figure 2.15 shows the convergence efficiency of the transfer learning algorithm and a traditional Q-learning algorithm. Here, convergence efficiency is defined as the probability of an eNB being in a stable state. It is assumed that the stable state occurs when the number of tries of sensing VRBs in order to assign a VRB to UE is less than a threshold. A small threshold value of three is selected in this report in order to achieve a fast exploration. The performance is estimated every 2000 data file transmissions. The measurement of convergence efficiency starts from the 5 extra eNBs being switched on.

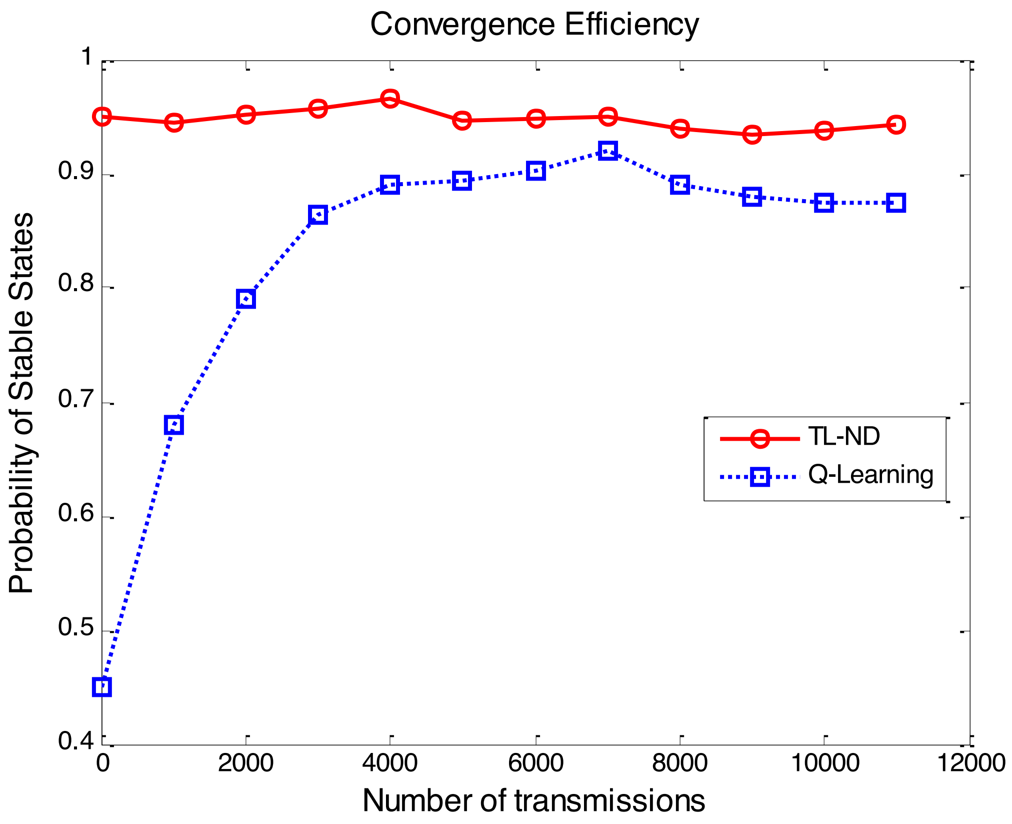


Figure 2.15 Convergence Efficiency

It can be noticed the system with transfer learning algorithm there is almost a constant probability of a stable state while the number of transmissions increasing. This is because an initial Q-Table was generated based on newly deployed TeNBs using the information received from adjacent TeNBs. In contrast, the Q-learning algorithm has very poor performance at the beginning, which is due to the Q-Tables of newly deployed TeNBs being empty. The convergence speed is very slow for the newly deployed TeNBs learning radio environment from scratch. When the Q-learning scheme has converged, the network with transfer learning still has better performance compared to the results of Q-learning, an average improvement of 3%-10% probability of stable states. Thus, we conclude that transfer learning significantly improves convergence on traditional Q-learning.

B) Transfer Learning algorithm for Exploring available Spectrum Bands (TL-ESB)

Secondly, TL-ESB algorithm is presented to show how transfer learning based CSM reward existing eNBs when exploring extra available spectrum bands. The CDF of the number of VRBs being sensed in order to allocate a VRB when an UE tries to access the system is observed in Figure 2.16. It is demonstrated that the 95% of VRBs being sensed in order to assign a VRB are equal or less than two in the Transfer Learning scheme, this value drops to 87% in the Q-Learning scheme. Moreover, the Q-Learning scheme converges to a set of poor VRBs, which causes a significantly number of extra VRBs to be sensed. Thus, the transfer learning algorithm contributes to both good decisions and fast convergence.

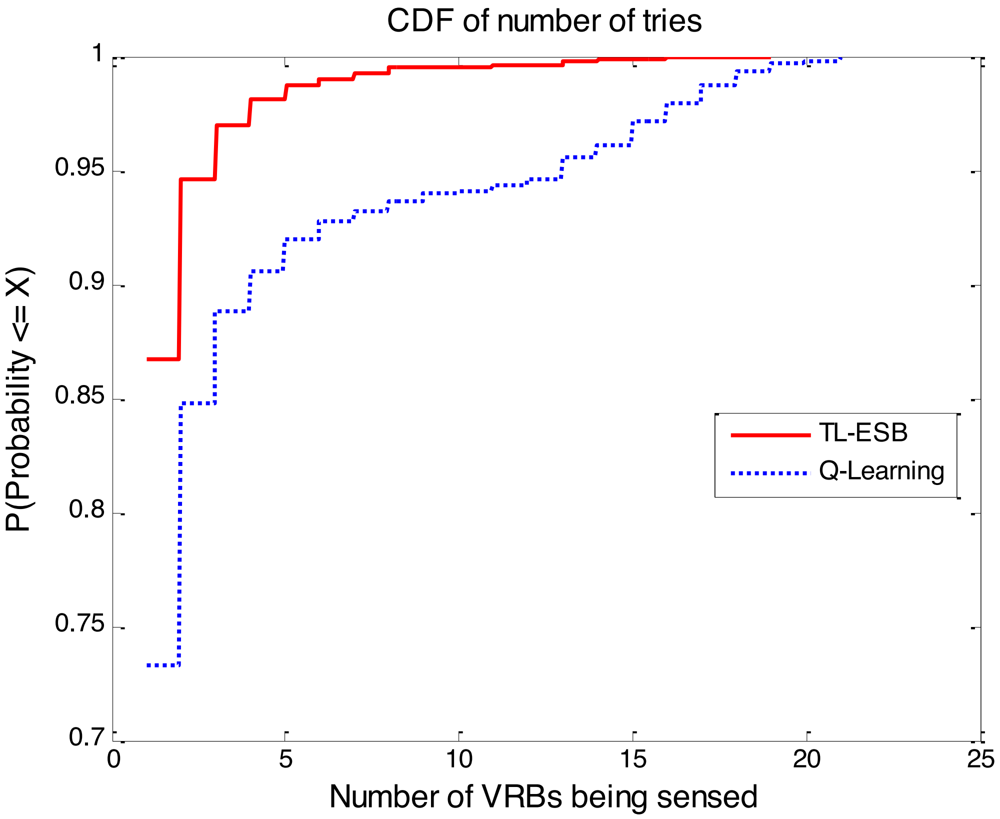


Figure 2.16 CDF of numbers of tries

C) Transfer Learning algorithm for Assigning Spectrum Bands (TL-ASB)

Figure 2.17, Figure 2.18 and Figure 2.19 compare the system performance of the network in terms of the system throughput, retransmission probability and average delay, using typical Q-Learning, the TL for Exploring extra Spectrum Bands (TL-WSB) and the TL for Assign Spectrum Bands (TL-ASB) schemes. The performance of all three schemes at low traffic level has almost no difference. Meanwhile, there is a similar result for system throughput when applying both transfer learning based algorithms and Q-learning algorithm. However, a significant improvement is achieved by transfer learning based algorithms when it comes to the retransmission probabilities and average delay. It is noticed that up to 10%(TL-ESB)/15%(TL-ASB) improvement is obtained when the system offered traffic is at middle/high level for retransmission probabilities, showing the effectiveness of transfer learning techniques in improving QoS in ABSOLUTE systems. This is explained that the transfer learning schemes converge to a set of good VRBs (spectrum pools), which causes less interference compared with pure Q-Learning scheme in simulation. The average delays of transfer learning schemes decrease while the retransmission probabilities reduce. This is because these schemes with better retransmission probability have reduced back off times. Here, the transfer learning based algorithm for assigning spectrum bands (TL-ASB) has the best performance among the other two schemes.

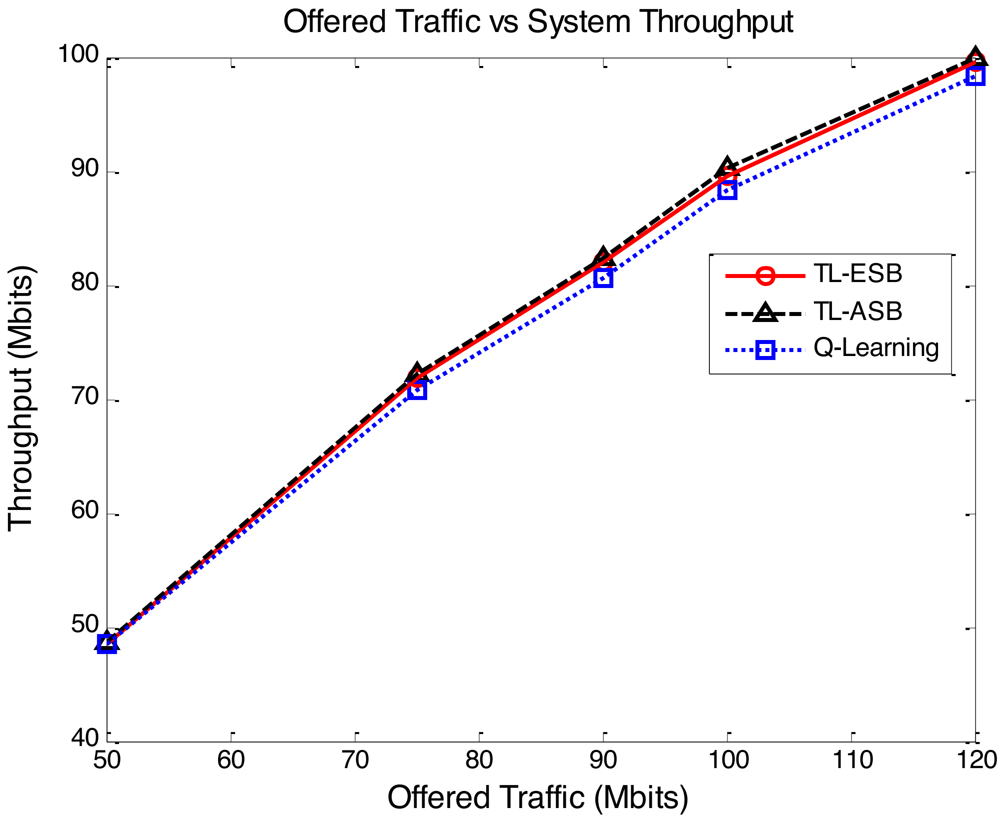


Figure 2.17 System throughput

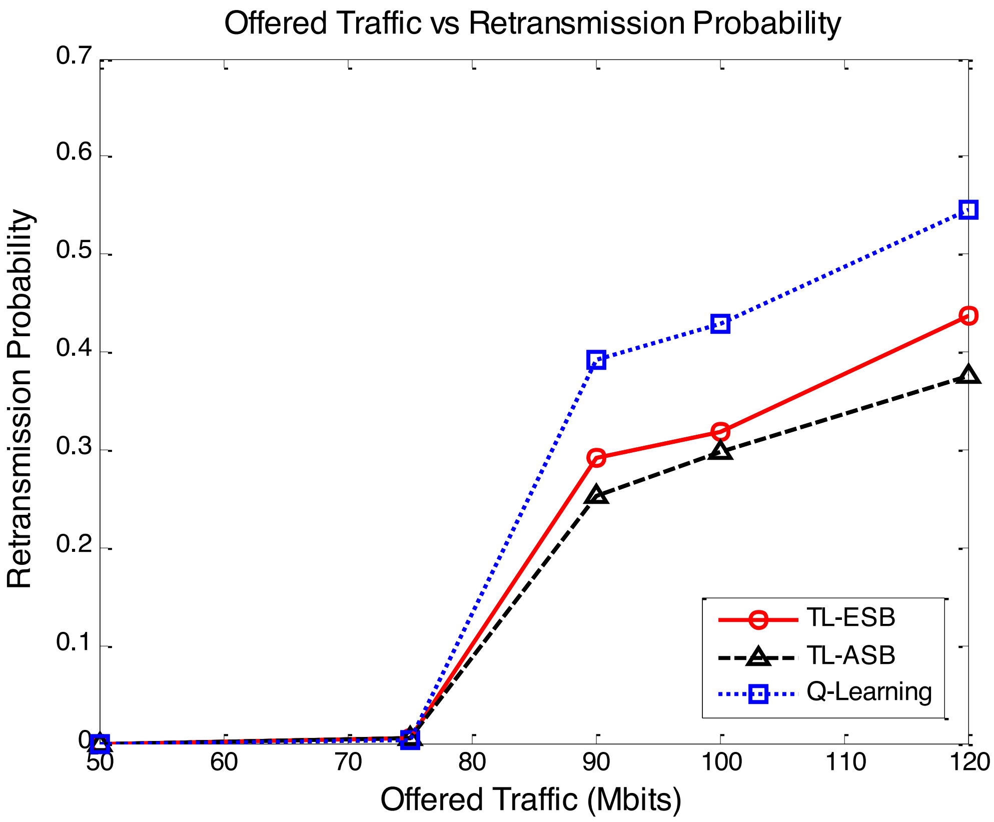


Figure 2.18 Retransmission probability

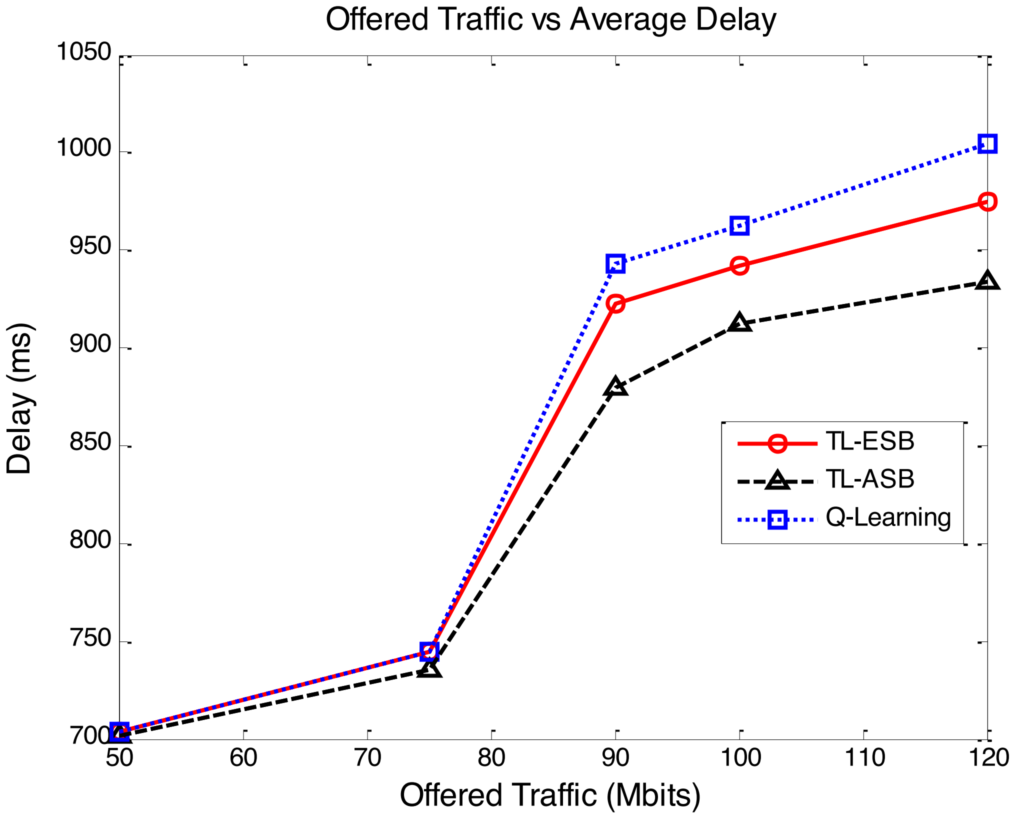


Figure 2.19 Average delay

2.6 Conclusions

In this chapter we investigated the cognitive spectrum management for Callania Public Protection and Disaster Relief scenario. The use cases discussed include ABS.UC.01, ABS.UC.02, ABS.UC.03, ABS.UC.04 ABS.UC.05, ABS.UC.17 and ABS.UC.18. Three transfer learning schemes with K-means clustering to deliver cognitive spectrum management for different conditions within 4G LTE-A ABSOLUTE system. The system model and the major issues of applying transfer learning in an LTE-A network were proposed. The novel transfer learning algorithms have been developed to optimize the system QoS, reduce the convergence delay and number of times that VRBs to be sensed when performing spectrum allocation in comparison to the Q-Learning schemes.

Here, the process of transfer learning is designed as two major functions: knowledge transfer and processing, and knowledge reward. The knowledge transfer and processing function achieves the spectrum information received from neighbouring eNBs. Data mining techniques are used to find the hidden structure of data in this function, including data pre-processing and K-means clustering algorithm. Specifically, the hidden information we are interested is the more/less frequently used spectrum bands by adjacent eNBs. The knowledge reward function determines how to make use of the converted information processed in the previous function for three conditions: newly eNBs deployed, existing eNB looking for extra spectrum bands and existing eNB assigning spectrum bands for new arriving UEs. With the aid of transfer learning, the severe negative impact of topology changes on radio environment can be minimized, and the system can effectively handle the dynamics of user traffic with reduced interference.

A large scale public safety event proposed in D3.3.2 has been used for system level simulation. The simulation results show that the transfer learning algorithms contribute to both good decisions of selecting potential VRBs and fast convergence, where the transfer learning algorithm for assigning spectrum bands (TL-ASB) has the best system performance.

3 Case based Cognitive Cellular Systems for Temporary Events

The purpose of this chapter is to propose a way of improving the stability of reinforcement learning (RL) based dynamic spectrum access (DSA) algorithms for temporary event networks with dynamic topologies that use secondary LTE spectrum sharing. The technique investigated for solving this problem is case-based RL, 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 [11]. In case-based RL these solutions are learned by using an RL algorithm. The combination of these two techniques has been successfully applied to various decision problems, e.g. dynamic inventory control [12] and RoboCup Soccer [13]. However, the only example of applying this methodology in the wireless communications domain is described by us in Subsection 3.2.3 of D3.3.1, where a DSA scheme is designed for a small 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 chapter is organised as follows: Section 3.1 introduces the temporary event scenario. Section 3.2 explains distributed Q-learning based DSA. In Section 3.3 we introduce the concept of case-based RL and propose a case-based Q-learning scheme for dynamic secondary spectrum sharing. Simulation results are discussed in Section 3.4, and the conclusions are given in Section 3.5.

3.1 Temporary Event Scenario

The spectrum sharing problem investigated in this chapter is designed for a temporary event (Bastian, as seen in D2.1) scenario and related to the ABSOLUTE user cases of ABS.UC.01, ABS.UC.02, ABS.UC.19 and ABS.UC.20. It 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 organizers involved in a temporary event, e.g. a football match or a concert, as described in Section 4.2 of D2.1. This scenario is depicted in Figure 3.1, where a small cell network is deployed inside the stadium to provide ultra-high capacity density to the event attendees, and an eNB on an aerial platform is deployed above the stadium to provide wide area coverage.

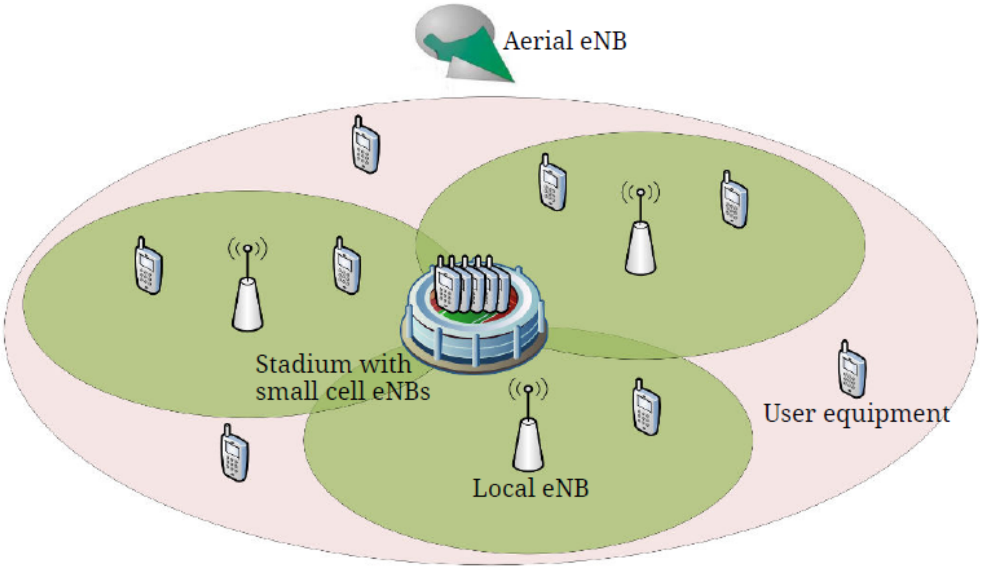


Figure 3.1. Stadium temporary event scenario

The cognitive small cells and the AeNB have secondary access to a 20 MHz LTE channel, also used by a network of 3 local primary eNBs (PeNBs). The goal of the small cell network and the AeNB is to use distributed machine intelligence methods to form a self-organizing heterogeneous cellular system which reuses the LTE spectrum of the local primary LTE network. Furthermore, in this chapter we consider a dynamic topology case, where the AeNB can be switched on and off several times throughout the duration of the event. For example, it can be switched on for providing the event organizers 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 cellular system is to adapt to these sudden changes in their radio environment, while not affecting the QoS in the local primary system.

3.2 Distributed Q-learning Based Dynamic Spectrum Access

In distributed reinforcement learning (RL) based DSA the task of every eNB is to learn to prioritise among the available subchannels only through trial-and-error, with no frequency planning involved, and with no information exchange with other eNBs. In this way, frequency reuse patterns emerge autonomously using distributed artificial intelligence with no requirement for any prior knowledge of a given wireless environment.

3.2.1 Distributed Stateless Q-learning

One of the most successful and widely used RL algorithms is Q-learning, introduced in [14]. Since the distributed DSA learning problem does not require a state representation, a simple stateless variant of this algorithm, as formulated in [15], is used. Figure 3.2 shows a flowchart of the distributed Q-learning based DSA algorithm for one file transmission.

Each eNB maintains a Q-table *Q(a)* such that every RBG a has an expected reward or Q-value associated with it. The Q-value represents the desirability of assigning a particular RBG to a file transmission. Upon each file arrival, the eNodeB either assigns an available RBG to its transmission or blocks it if no RBGs are available. It decides which RBG to assign based on the current Q-table and the greedy action selection strategy described by the following equation:

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

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

The Q-table is updated by the corresponding eNodeB each time it attempts to assign a RBG to a file transmission in the form of a positive or a negative reinforcement. The update equation for stateless Q-learning, as defined in [15], is given below:

$$Q'(a) = Q(a) + \alpha\big(r - Q(a)\big)$$

where *Q(a)* and *Q'(a)* represent the Q-value of the RBG a, before and after the update respectively, 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.

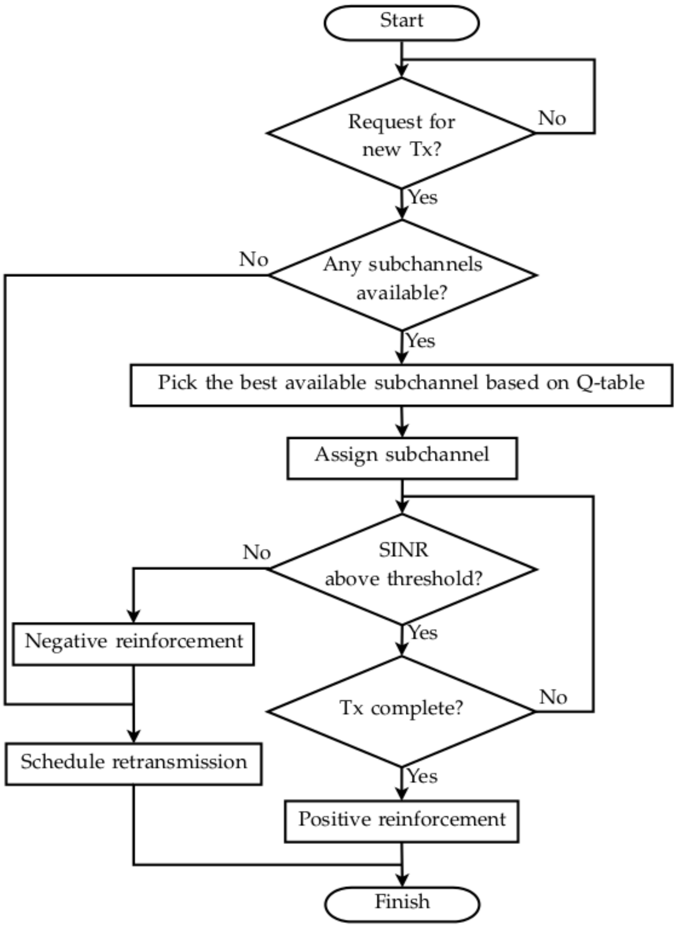


Figure 3.2. Flowchart of the distributed stateless Q-learning algorithm

The reward function returns two discrete values:

- *r=-1 (negative reinforcement)*, if the file is scheduled for retransmission due to SINR being lower than the minimum transmission threshold (1.8 dB) on the selected RBG, or if it is later interrupted for the same reason.
- *r=1 (positive reinforcement)*, if the file is successfully transmitted using the RBG chosen by the eNB, i.e. if SINR is higher than 1.8 dB throughout the whole transmission.

The choice of the learning rate value for this type of distributed Q-learning based DSA problems is thoroughly investigated by us in [16]. The best performance is achieved by using the Win-or-Learn-Fast (WoLF) variable learning rate principle described by the equation below:

$$\alpha = \begin{cases} 0.01 & r = 1 \\ 0.05 & r = -1 \end{cases}$$

There, a lower value of α is used for successful trials (when r=1), and a higher value of α is used for failed trials (r=-1). In this way, the learning agents are learning faster when ``losing" and more slowly when ``winning".

The values in the Q-tables are initialised to zero, so all eNBs start learning with equal choice among all available RBGs.

3.3 Case-Based Reinforcement Learning

Figure 3.3 shows a flow diagram of the processes involved in case-based RL. It also demonstrates that it is an extension of classical RL, i.e. classical RL can be viewed as a special case of case-based RL.

In Figure 3.3, the unfilled blocks and solid lines constitute a flow diagram of a regular RL algorithm. There is an outer output-state-action loop, where outputs of the environment are observed and processed to yield the environment state information, and then the best action is chosen for the current state based on the policy of the learning agent. In the context of our DSA problem, the output of interest is whether or not the last file transmission got blocked or interrupted, and the action is an RBG 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. A policy is then derived from the estimated Q-table and used for choosing an action in the current environment state.

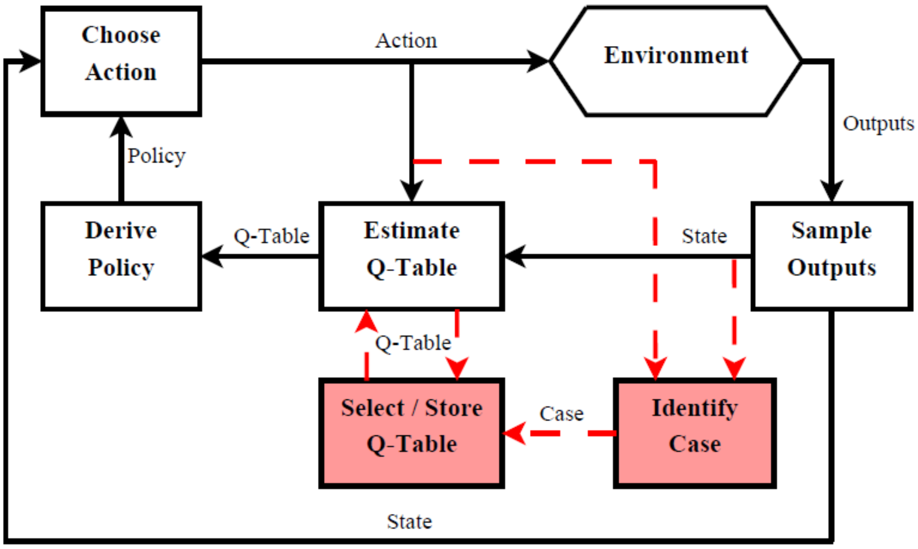
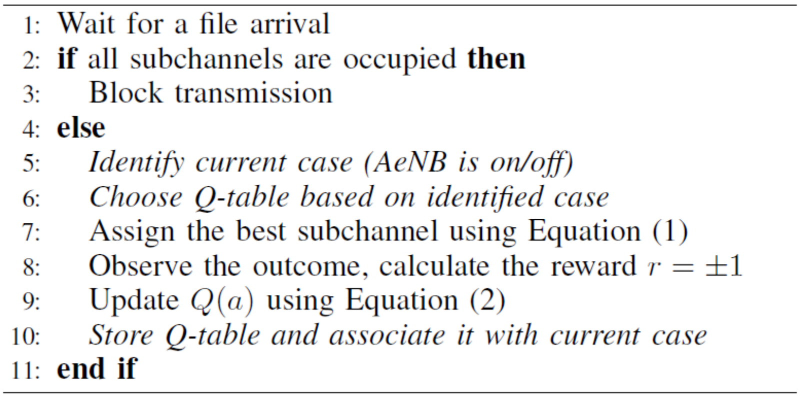


Figure 3.3. Block diagram of processes involved in case-based reinforcement learning

The highlighted blocks and dashed arrows represent additional functionality afforded by case-based reasoning 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. In some circumstances 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 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.

3.3.1 Distributed Case-Based Q-learning

Figure 3.4 shows our proposed adaptation of this case-based RL approach to the dynamic secondary spectrum sharing scenario described in Section 3.1. The functionality afforded by CBR, as an extension to classical RL, is described by steps 5, 6, and 10 of the algorithm in Figure 3.4. Before making an RBG assignment decision based on Q-learning, the given eNB first identifies the current case of the environment, i.e. whether the AeNB is on or off. It then retrieves the Q-table that corresponds to the identified case, or uses its current Q-table, if there are none stored for the required case. Afterwards, when the spectrum assignment decision is made, the outcome is observed and the current Q-table is updated with a positive or a negative reinforcement, the eNB stores the updated Q-table in its case base and associates it with the current case. In this way, all eNBs undergo two learning processes in parallel, depending on whether the AeNB is on or off, with the aim of achieving more stable and reliable spectrum sharing policies.



```
Wait for a file arrival
if all subchannels are occupied then
    Block transmission
else
    Identify current case (AeNB is on/off)
    Choose Q-table based on identified case
    Assign the best subchannel using Equation (1)
    Observe the outcome, calculate the reward $r = \pm 1$
    Update $Q(a)$ using Equation (2)
    Store Q-table and associate it with current case
end if
```

Figure 3.4. RBG Assignment using case-based Q-learning for dynamic secondary spectrum sharing

3.4 Simulation Results and Discussion

The spectrum sharing problem described in Section 3.1 involves an AeNB and a network of small cell eNBs that have to share spectrum among them and with a primary system of local eNBs operating in the area.

The primary system is assumed to employ a dynamic ICIC scheme, where all three eNBs exchange their current spectrum usage as RNTP messages every 20 ms, and exclude the subchannels currently used by the other two eNBs from their available subchannel list [17]. 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 first subchannel to the last. In this way, the primary network makes 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 presented in Figure 3.4 does not assume this and would also work regardless of the spectrum management strategy of the primary system.

The results of implementing the following three schemes in the secondary system are discussed in this section:

- *“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 a distributed Q-learning algorithm described in Subsection 3.2.1.*
- *“Cased-based Q-learning” - all cognitive eNBs run a case-based Q-learning algorithm proposed in Subsection 3.3.1.*

The "dynamic ICIC" approach represents a heuristic baseline DSA scheme, typical for LTE networks [17]*, whereas the ``Q-learning" approach represents a pure RL based approach without the CBR functionality added to it.*

3.4.1 Simulation Setup

The stadium small cell network architecture is shown in Figure 3.5. There, the users are located in a circular spectator area 53.7 - 113.7 m from the centre of the stadium. The spectator area 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. approximately 10,776 user equipments (UEs). 500 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. The offered traffic is 20 Mb/s outside of the stadium and 1 Gb/s inside. The other parameters and assumptions of the simulation model are listed in Table 3.1.

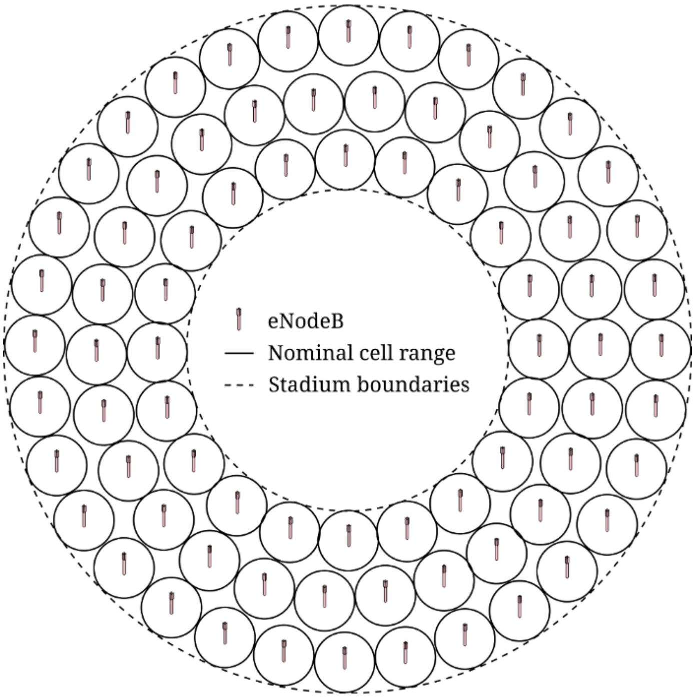


Figure 3.5. Stadium small cell network architecture

The cognitive small cell network and the AeNB which is located above the stadium centre point at 300 m altitude have secondary access to a 20 MHz LTE channel also used by the primary network. The latter consists of 3 primary eNBs (PeNBs) whose coordinates, with respect to the centre point of the stadium, are (-600, -750), (100, 750) and (750, -800) m. Therefore, the goal of the cognitive small cell eNBs and the aerial eNBs is to efficiently utilize the 20 MHz LTE channel, normally reserved for the PeNBs, whilst avoiding interference with them.

Table 3.1. Network model parameters and assumptions

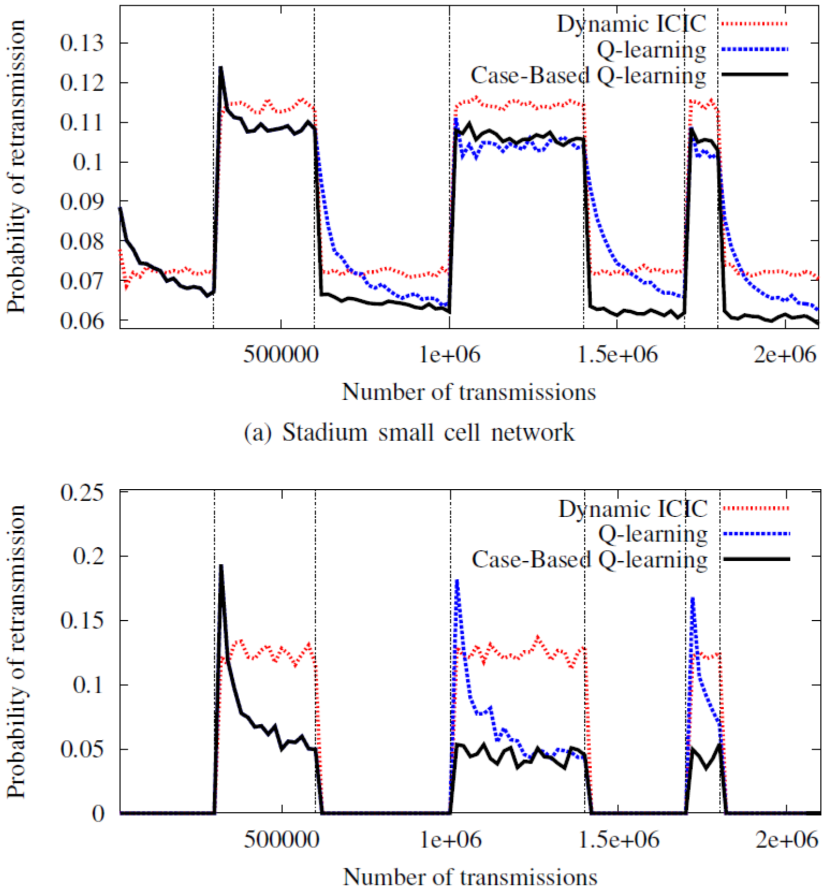
Parameter	**Value**
Channel bandwidth	20 MHz: 100 LTE virtual resource blocks (VRBs)
Subchannel (RBG) bandwidth	4 VRBs: 4 x 180 kHz [18]
Frequency band	2.6 GHz
UE receiver noise floor	94 dBm (290K temperature, 20 MHz bandwidth, 7dB noise figure)
Stadium propagation model	WINNER II B3 [19]
Outdoor propagation model	WINNER II C1 [19]
Propagation model between stadium and outdoors	Combined WINNER II C4 with C1 term [19]
Propagation model between aerial platform and ground	Free space + 8dB log-normal shadowing
Traffic model	3GPP FTP Traffic Model 1 [20], file size – 4.2 Mb (≈ 0.5 MB)
Retransmission scheduling	Uniform random back-off between 0and 960 ms
Link model	3GPP Truncated Shannon Bound Model [21]
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 RBG (4 VRBs) is allocated to every data transmission.	

3.4.2 Temporal Performance

Figure 3.6 shows the average temporal performance of the secondary network in terms of its probability of retransmission (*P(retransmission)*). 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:

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

where N_r is the number of retransmissions and N_s is the number of successfully completed transmissions during a given sampling period. All simulations start with the AeNB being switched off. The vertical dash-dot lines in Figure 3.6 represent the times at which the AeNB is switched on and back off again. The time dimension in the graphs is expressed in terms of the total number of transmissions performed in the simulations, most of which take place inside the densely populated stadium. On average the simulation length is equivalent to ≈2h20'.



(a) Stadium small cell network

(b) Aerial eNB

Figure 3.6. Probability of retransmission in the secondary cognitive network

Figure 3.6a shows how well the stadium small cell network adapts to the sudden irregular changes in its environment caused by the AeNB being switched on/off. It compares the performance of three DSA strategies – "dynamic ICIC", "Q-learning" and the "case-based Q-learning" approach proposed in this paper. At the start of the simulation both learning based schemes start at a poorer QoS level than "dynamic ICIC", but then gradually improve on it until the AeNB is switched on for the first time. At this point, the stadium network starts receiving interference from the AeNB in addition to the primary system, which causes its *P(retransmission)* to significantly increase using all DSA schemes. Afterwards the same pattern of gradual improvement of the RL algorithms compared to "dynamic ICIC" is observed. When the AeNB is switched off again, it takes time for a regular RL algorithm to adapt to the sudden change in the environment. In contrast, the "case-based Q-learning" scheme is able to retrieve the solution to the DSA problem with the AeNB switched off and rapidly improve the network-wide QoS to the previously achieved level. The performance gap between "case-based Q-learning" and "Q-learning" increases further every time the AeNB is switched off again, due to the ability of the former to seamlessly switch between two learning processes.

The difference in performance between the scheme proposed in this chapter and the two baseline schemes is more substantial in Figure 3.6b, which shows the average *P(retransmission)* temporal response of the AeNB. Firstly, both learning schemes significantly outperform the purely heuristic “dynamic ICIC” approach. Secondly, the novel CBR functionality implemented in all stadium small cell eNBs and in the AeNB results in a 70% reduction in *P(retransmission)* experienced by the AeNB users shortly after the AeNB is switched on for the second time and all subsequent times. This demonstrates that by using the “case-based Q-learning” approach the cognitive AeNB can be repeatedly re-introduced into a spectrum sharing environment with no need to relearn its spectrum management strategy.

3.4.3 Primary User Quality of Service

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 3.2 compares the QoS provided to the users outside of the stadium with and without the presence of the stadium users and the secondary network. In addition to the average probability of retransmission, 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 in [20], is given below:

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

where F is the number of files downloaded by the given UE, S_f is the size of the fth file, and T_f is the time it took to download it.

Table 3.2. Primary user QoS with and without the presence of the secondary network (SN)

QoS metric	**Without SN**	**With SN**
Probability of retransmission	3.0×10^{-3}	3.4×10^{-3}
Mean user throughput (UT), Mb/s	3.04	3.07
95th percentile UT, Mb/s	3.16	3.16
5th percentile UT, Mb/s	2.70	2.89
Mean UT 0-100 m from the stadium, Mb/s	2.96	2.89

Table 3.2 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 lowest UT provided to at least 95% of the users and is an important metric for ensuring fair QoS distribution across the whole network. This is because that 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 3.2 emphatically show that it is possible to develop a temporary heterogeneous cognitive network that is capable of servicing a dramatic increase in the offered traffic (1 Gb/s in addition to the original 20 Mb/s, i.e. by a factor of 51), but with no need for additional spectrum and with no degradation in the primary user QoS.

The further work currently underway on case-based RL for dynamic secondary spectrum sharing investigates the application of these principles to ABSOLUTE temporary event networks with more complex and realistic topology management schemes in place.

3.5 Conclusion

In this chapter, a cognitive spectrum management solution for Bastian temporary event scenario is proposed. The related use cases include ABS.UC.01, ABS.UC.02, ABS.UC.19 and ABS.UC.20. The case-based RL method introduced here is an effective and feasible approach to dynamic secondary spectrum sharing in temporary cognitive cellular systems with dynamic topologies. System level simulations that involve a stadium small cell network, an eNB on an aerial platform and a local primary LTE network show that augmenting RL with the CBR functionality results in increased adaptivity of the cognitive cellular system to sudden changes in its radio environment, caused by the aerial eNB being dynamically switched on and off. For example, it is capable of achieving a 70% reduction in the number of retransmissions of the aerial eNB shortly after being switched on, compared to a classical RL approach. Furthermore, the cognitive cellular system, that employs the proposed DSA scheme with only secondary access to an LTE channel, is shown to accommodate a 51-fold increase in the offered traffic with no need for additional spectrum and with no degradation in the QoS of the primary users. In further work on this topic, more complex and realistic topology management schemes are investigated, which provide a greater number of cases and a more challenging case identification task.

4 Distributed Heuristically Accelerated Reinforcement Learning for Dynamic Secondary Spectrum Sharing

Although RL algorithms such as Q-learning described in Section 3.2 have been shown to be a powerful approach to problem solving, their common disadvantage is the need for many learning iterations to converge on an acceptable solution. One of the more recent promising solutions to this issue, proposed in the artificial intelligence domain, is the heuristically accelerated reinforcement learning (HARL) approach. Its goal is to speed up RL algorithms, particularly in the multi-agent domain, by guiding the exploration process using additional heuristic information [22]. In [23], case-based reasoning is used for heuristic acceleration in a multi-agent RL algorithm to assess similarity between states of the environment and to make a guess at what action needs to be taken in a given state, based on the experience obtained in other similar states. In [22], Bianchi et al. prove the convergence of four multi-agent HARL algorithms and show how they outperform the regular RL algorithms. The only example of the HARL approach being applied in the wireless communications domain is the DSA scheme introduced in Chapter 5 of D3.3.2 and used as an integral part of DSS algorithms developed in this chapter. There is no evidence in the literature of the HARL approach being applied to a problem of spectrum sharing between two or more separate cellular systems.

The purpose of this chapter is to propose a novel HARL based framework, which uses a radio environment map (REM), to mitigate the problem of poor temporal performance of RL algorithms applied to DSS problems. The temporary event (Bastian, as seen in D2.1) scenario is selected here, and the following use cases are considered: ABS.UC.01, ABS.UC.02, ABS.UC.19 and ABS.UC.20. The principles and features of the proposed HARL framework also aim to be generally applicable to a wide range of learning problems beyond the wireless communications domain.

The rest of the chapter is organised as follows: Section 4.1 explains the principles behind RL and HARL based DSA. In Section 4.2 we propose a novel HARL framework and show how it can be applied to the spectrum sharing problem in hand. Section 4.3 evaluates the performance of the proposed schemes by simulating a large scale LTE spectrum sharing scenario.

4.1 Cognitive Dynamic Spectrum Access

In order to discuss secondary spectrum sharing, the DSA mechanism for scheduling resources of the cognitive cellular system alone needs to be introduced first. This section presents the concept of heuristically accelerated RL (HARL), and explains the details of the HARL based cognitive DSA algorithm from Chapter 5 of D3.3.2 designed for the secondary system, initially without considering the presence of a primary system.

4.1.1 Heuristically Accelerated Reinforcement Learning

A common disadvantage of machine learning algorithms, such as distributed Q-learning described in Subsection 3.2.1, is that they are normally used to learn solutions only through trial-and-error with no prior knowledge of the problem in hand. Consequently, it takes a large number of trials for them to learn acceptable solutions. This is undesirable in real-time applications such as DSA in cellular systems. An emerging technique to mitigate this poor initial performance problem is the HARL approach, where additional heuristic information is used to guide the exploration process [22].

Figure 4.1 shows our block diagram representation of the processes involved in HARL. It demonstrates that HARL is an extension of regular RL algorithms. The unfilled blocks and solid lines constitute a block diagram of regular RL, whereas dashed lines and shaded blocks indicate the additional functionality afforded by the heuristic acceleration.

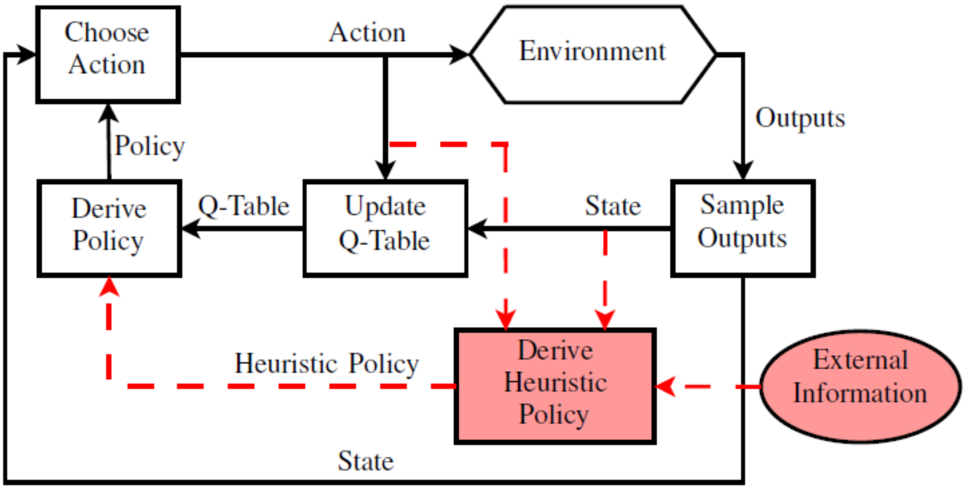


Figure 4.1. Block diagram of heuristically accelerated reinforcement learning

The role of the inner RL loop is to learn a good policy to be used by the learning agent. It is identical to that described in Section 3.3.

The key additional element provided by HARL is the derivation of a heuristic policy. According to Bianchi et al. [22], a heuristic policy is derived from additional knowledge, either external or internal, which is not included in the learning process. Generally, the goal of the heuristic policy $H_t(s, a)$ is to influence the action choices of a learning agent, i.e. to modify its current policy $\pi_t(s)$ in a way which would accelerate the learning process. The format and dimensions of $H_t(s, a)$ should be compliant with the Q-table used by the given learning agent, such that its new combined policy $\pi_t^c(s)$ can be derived using the following equation:

$$\pi_t^c(s) = arg\max_a(Q_t(s,a) + H_t(s,a))$$

where $\pi_t^c(s)$ is the combined policy of the given learning agent for state s at time t based on its Q-table $Q_t(s, a)$ and the heuristic policy $H_t(s, a)$. If $H_t(s, a)$ is always zero, the algorithm becomes a regular Q-learning algorithm with a greedy action selection strategy. In the case of the stateless Q-learning algorithm described in Section 3.2, the heuristic policy does not have a state dimension and can be denoted by $H_t(a)$.

4.1.2 Distributed ICIC Accelerated Q-Learning

The only existing HARL based DSA scheme is known as distributed ICIC accelerated Q-learning (DIAQ), proposed in Chapter 5 of D3.3.2. It uses inter-cell interference coordination (ICIC) signalling in the LTE downlink as heuristic acceleration for a distributed stateless Q-learning algorithm described in Subsection 3.2.1. It achieves dramatic improvements in initial and steady-state quality-of-service (QoS), as well as in learning convergence rate, in a cognitive cellular system with dedicated spectrum.

The format of the messages exchanged between eNBs using ICIC in the LTE downlink is standardized by the 3GPP and referred to as the Relative Narrowband Transmit Power (RNTP) indicator [18]. It contains a bitmap which indicates on which resource blocks an eNB is planning to transmit at high power by setting their corresponding bits to *1*, i.e. on which resource blocks it is likely to cause interference in adjacent cells. For example, in a case where a 20 MHz LTE channel has 25 RBGs, the length of an RNTP message is 100 bits or 25 hexadecimal characters [18]. Since every RBG consists of 4 adjacent resource blocks, every group of 4 bits (i.e. every hexadecimal character) in an RNTP message describes a particular RBG. For example, if an eNB is planning to use high transmit power on a given RBG, its corresponding bits in the RNTP message are *1111* or *0xF*.

The choice of the RNTP threshold used to decide whether a given transmit power is high or low is set to -3 dB with respect to the average transmit power in a cell. To avoid excessive signalling requirements, the time interval between the ICIC message exchanges is assumed to be 20 ms [17].

When a request for a new file transmission is received, the eNodeB starts by aggregating the latest RNTP messages from its neighbours into an ICIC bitmask using a bitwise *OR* operation, as described by the following equation:

$$Mask_{ICIC} = \bigcup_{n=1}^{N} RNTP_n$$

where $Mask_{ICIC}$ is a 25 hexadecimal character string representing the RBGs reserved by *any* of the neighbouring base stations by *0xF*, and representing the “safe-to-use” RBGs by *0x0*, $RNTP_n$ is a 25 hexadecimal character RNTP message of the *n*'th neighbouring eNodeB, and *N* is the total number of neighbouring eNodeBs.

After creating the ICIC mask, the eNodeB creates a heuristic policy $H_{ICIC}(a)$ using the following principle:

$$H_{ICIC}(a) = \begin{cases} h_{ICIC} & Mask_{ICIC}(a) = 0xF \\ 0 & Mask_{ICIC}(a) = 0x0 \end{cases}$$

where $H_{ICIC}(a)$ is the heuristic policy value of RBG *a*, and h_{ICIC} is a fixed negative number with greater amplitude than the difference between the minimum and the maximum possible values in the Q-tables. $H_{ICIC}(a)$ can then be employed to create a temporary masked Q-table $Q_m(a)$ using the following equation:

$$Q_m(a) = Q(a) + H(a)$$

$Q_m(a)$ is then used for heuristically guided decision making, whilst a normal learning process is taking place using the original Q-table *Q(a)*.

By using such a heuristic policy $H_{ICIC}(a)$, the eNodeB is guaranteed to prioritise the RBGs marked as “safe” by the ICIC bitmask before the “unsafe” RBGs by shifting the Q-values of the latter to the bottom of the Q-table, whilst still preserving their respective order in terms of the Q-values (due to the fixed value of h_{ICIC}).

4.2 HARL Based Dynamic Spectrum Sharing

The stadium temporary event spectrum sharing scenario investigated in this chapter is described in Section 3.1. It consists of a network of primary eNBs (PeNBs) operating in a suburban area and a secondary cognitive cellular system that itself consists of two separately operating entities - an aerial eNB (AeNB) for wide area coverage and a small cell network for high capacity density inside the stadium.

A study in [24] has demonstrated that successful dynamic spectrum sharing between a low power stadium small cell system and a relatively high power local PeNB infrastructure can be facilitated using an independent distributed Q-learning algorithm from Subsection 3.2.1 implemented in the former. This is largely because the interference between the two systems is attenuated by the stadium shell. However, the scenario investigated in this paper also involves an AeNB serving line-of-sight (LoS) users both inside and outside the stadium. Therefore, it presents two additional challenges - spectrum sharing between the PeNBs and the AeNB, and spectrum sharing between the AeNB and the stadium small cell network.

Our proposed way of achieving these two spectrum sharing tasks is to use a small scale database, referred to as the radio environment map (REM), to continuously monitor and store the information about spectrum usage of the PeNBs and the AeNB. In this way, the AeNB has a means to avoid interfering with the primary system, and the small cell network can avoid interfering with the AeNB. This type of setup is depicted in Figure 4.2. Secondary spectrum sharing using a spectrum monitoring system and a radio environment map (REM), which is a classical way of achieving coexistence between cognitive radio networks and primary spectrum users, especially in the TV white space context, e.g. [25].

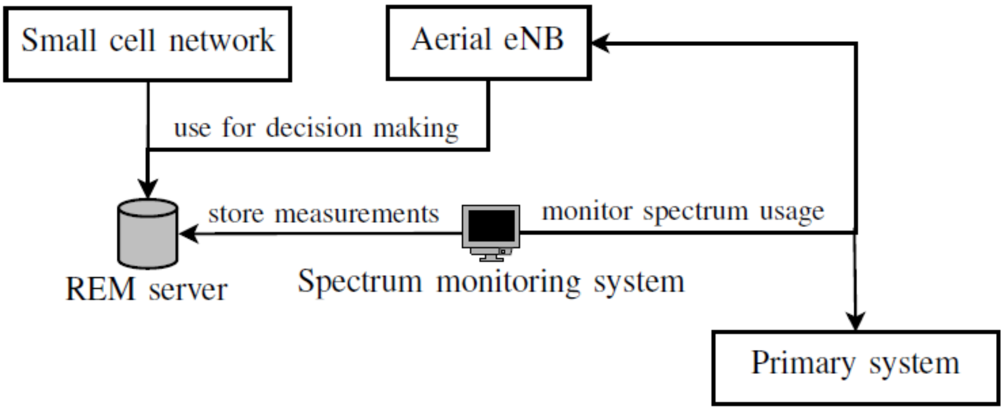


Figure 4.2. Secondary spectrum sharing using a spectrum monitoring system and a radio environment map (REM)

The task of the spectrum monitoring system with a REM database is to detect the occupancy of the spectrum resources used by the PeNBs and the AeNB. It is then possible to estimate the probability of spectrum occupancy at every eNB on every individual RBG that, in turn, can be used to influence the spectrum assignment decisions of the secondary systems.

4.2.1 Spectrum Occupancy Estimation

The spectrum sharing algorithms proposed in this section assume the ABSOLUTE spectrum monitoring system can periodically detect whether or not a particular RBG is being used by a particular AeNB or PeNB. It is designed to return *1* if it is occupied or *0* otherwise.

Given this mechanism for obtaining a stream of binary spectrum occupancy data, it is then important to estimate the probability of RBG occupancy at every observed eNB, i.e. a probability of a particular RBG being occupied at a particular eNB based on the previous observations.

A simple and appropriate way of tracking the mean of a data sequence, whilst simultaneously giving more recent observations higher weight compared to older estimates, is the exponentially weighted moving average (EWMA) method [26]. It can be calculated using the following recursive equation:

$$y \leftarrow (1-\lambda)y + \lambda x$$

where *y* is the mean estimate of the data sequence x, and λ is a factor which controls how quickly the estimated mean adapts to new observations. The role of λ in EWMA estimation is identical to that of the learning rate α in the Q-learning update formula described in Subsection 3.2.1.

We propose adapting the EWMA method to estimate the probability of RBG occupancy *p(occupied)* in the following way:

$$p(occupied) \leftarrow (1-\lambda)p(occupied) + \lambda b, \qquad b \in \{0,1\}$$

where b is a current binary RBG occupancy measurement, i.e. *b=1* if the given RBG is occupied, *b=0* if it is not. In this way, the EWMA equation is used to estimate the mean of a stream of 1's and 0's, representing *p(occupied)*.

4.2.2 REM Based Heuristic Function

A threshold P_{min} to determine whether a particular RBG should be avoided, based on an estimate of *p(occupied)*, can then be defined to obtain the following heuristic function:

$$H_{REM}(a) = \begin{cases} h_{REM} & p_a(occupied) \geq P_{min} \\ 0 & p_a(occupied) < P_{min} \end{cases}$$

where $H_{REM}(a)$ is the value of the REM based heuristic function for RBG a, $p_a(occupied)$ is the EWMA estimate of *p(occupied)* for RBG a, h_{REM} is a fixed negative value which shifts the Q-values of the undesirable RBGs down, such that the other are prioritized before them. This heuristic function follows the same principle of shifting Q-values as the one used in DIAQ (see Subsection 4.1.2).

Such a heuristic function $H_{REM}(a)$ aims to guide the learning process of the cognitive eNBs in a direction desirable for secondary spectrum sharing. The small cell eNBs can coexist with the AeNB by applying this heuristic function to the AeNB spectrum occupancy observations, hereafter referred to as $H_{REM\text{-}AeNB}(a)$. The AeNB in turn can coexist with the PeNBs by applying the same principle to PeNB spectrum occupancy observations. In this case, since the wide area coverage AeNB is going to interfere with all PeNBs in the area of interest, the probability of RBG a being occupied by any PeNB is obtained by calculating the sum of $p_a(occupied)$ values of every individual PeNB:

$$p_{a-PeNBs}(occupied) = \sum_{n=1}^{N} p_{a-n^{th}\,PeNB}(occupied)$$

where N is the total number of PeNBs. The REM based heuristic function $H_{REM\text{-}PeNBs}(a)$ can then be calculated using $p_{a\text{-}PeNBs}(occupied)$.

4.2.3 Superimposed Heuristic Functions

With the introduction of the REM based heuristic function for secondary spectrum sharing, a framework for using several heuristic functions at once is required. For example, in addition to using an ICIC based heuristic function $H_{ICIC}(a)$ introduced in Subsection 4.1.2 for internal dynamic spectrum access, the small cell eNBs are now also required to share spectrum with the AeNB using another heuristic function $H_{REM\text{-}AeNB}(a)$, such that their masked Q-tables $Q_m(a)$ could be constructed using the following principle:

$$Q_m(a) = Q(a) + H_{ICIC}(a) + H_{REM-AeNB}(a)$$

where $Q(a) \in [-1, 1]$ is an original Q-table of a given eNB maintained using the stateless Q-learning algorithm described in Subsection 3.2.1. There, two heuristic functions $H_{ICIC}(a)$ and $H_{REM\text{-}AeNB}(a)$ have to be superimposed to modify a learning eNB's policy, such that it incorporates both ICIC and REM information into its learning process.

We propose a method where every new heuristic function superimposed on the Q-table splits the Q-values into two non-overlapping regions, as shown in Figure 4.3. The normal range of Q-values $Q(a)$ maintained by the stateless Q-learning algorithm from Subsection 3.2.1 is $[-1, 1]$. If the h_{ICIC} parameter of the $H_{ICIC}(a)$ heuristic function is -3, it shifts $Q_m(a)$ values of disapproved RBGs into a non-overlapping region of $(Q(a)-3) \in [-4, -2]$, thus prioritizing them below the RBGs with $Q_m(a) \in [-1, 1]$. If another heuristic function $H_{REM\text{-}AeNB}(a)$ is used and its h_{REM} constant is -7, it will split $Q_m(a)$ into two regions $Q_m(a) \in [-4, 1]$ and $Q_m(a) \in [-11, -6]$ In this way, the RBGs disapproved by $H_{REM\text{-}AeNB}(a)$ are guaranteed to be prioritized below any other RBG. This approach allows an unlimited number of further heuristic functions superimposed on top of each other, as long as their respective importance is known. For example, in this case we prioritize $H_{REM\text{-}AeNB}(a)$ responsible for spectrum sharing above $H_{ICIC}(a)$ responsible for internal stadium network DSA by setting $h_{REM} < h_{ICIC}$.

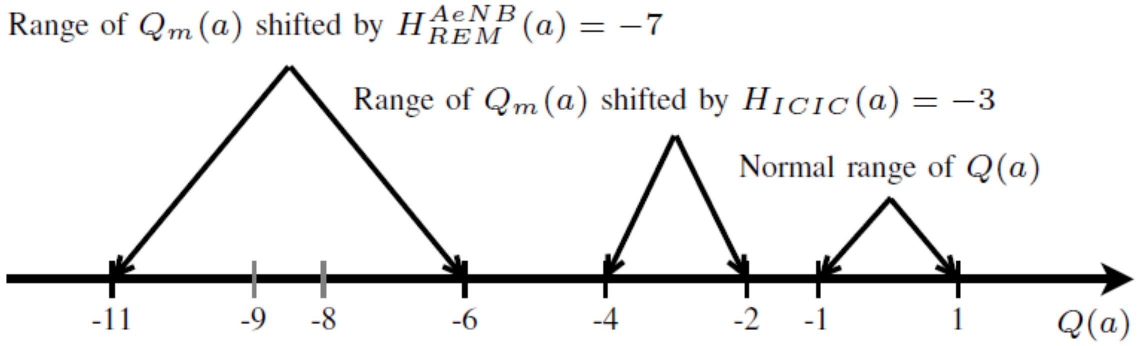

Range of $Q_m(a)$ shifted by $H_{REM}^{AeNB}(a) = -7$
Range of $Q_m(a)$ shifted by $H_{ICIC}(a) = -3$
Normal range of $Q(a)$
-11
-9
-8
-6
-4
-2
-1
1
$Q(a)$

Figure 4.3. An example of the effect of superimposed heuristic functions $H_{ICIC}(a)$ and $H_{REM\text{-}AeNB}(a)$ on the range of masked Q-table values

4.2.4 Q-Value Based Admission Control

The HARL algorithm required for the AeNB to coexist with the primary system only includes one heuristic function $H_{REM\text{-}PeNBs}(a)$, since it is a separately controlled entity with no ICIC-compatible neighbouring base stations. Therefore, it uses the following masked Q-table for guiding its learning process:

$$Q_m(a) = Q(a) + H_{REM-PeNBs}(a)$$

However, another important aspect of secondary spectrum sharing is the primary user protection, i.e. making sure the secondary system, in this case the AeNB, does not produce harmful interference for the primary system, in our case the users connected to the PeNBs. A technique that could be easily and effectively embedded into the HARL framework developed in this paper, i.e. where $H_{REM\text{-}PeNBs}(a)$ shifts part of the Q-values by a fixed negative number $h_{REM\text{-}PeNBs}$, is Q-value based admission control (Q-AC) introduced in [27]. There, a Q-value threshold q_{AC} is defined, such that:

$$A_{allowed} = \{a \mid a \in A' \ \wedge Q(a) \geq q_{AC}\}$$

where A' is the set of currently unoccupied RBGs, i.e. those available for assignment, and $A_{allowed}$ is the set of RBGs allowed for assignment based on the admission threshold q_{AC}. In this way, the RBGs with $Q(a) < q_{AC}$ are never assigned to data transmissions, which are blocked instead.

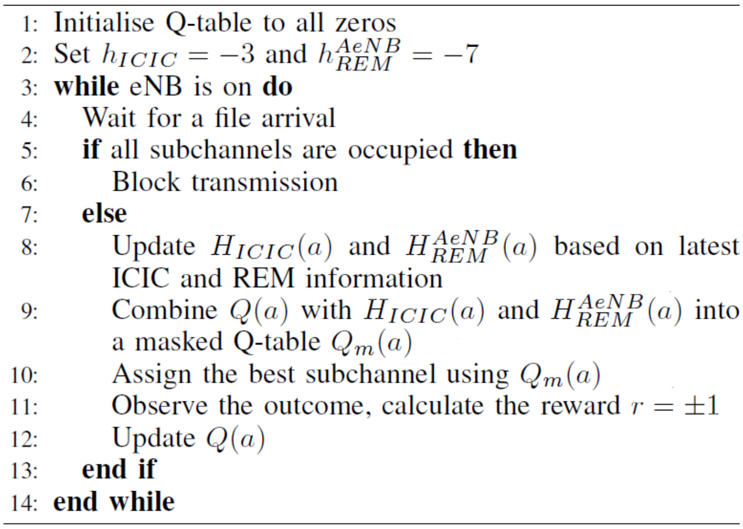
The value of q_{AC} can be chosen such that:

$$q_{max}(a) - h_{REM-PeNBs} < \ q_{AC} < \ q_{min}$$

where q_{min} and q_{max} are the minimum and the maximum possible value of $Q(a)$ respectively. In this way, the RBGs disapproved by the heuristic function $H_{REM\text{-}PeNBs}(a)$ will be forbidden to be assigned at the AeNB, due to their Q-values being shifted below $q_{AC,}$ thus guaranteeing protection of the PeNBs from secondary interference.

4.2.5 HARL Algorithms for Dynamic Spectrum Sharing

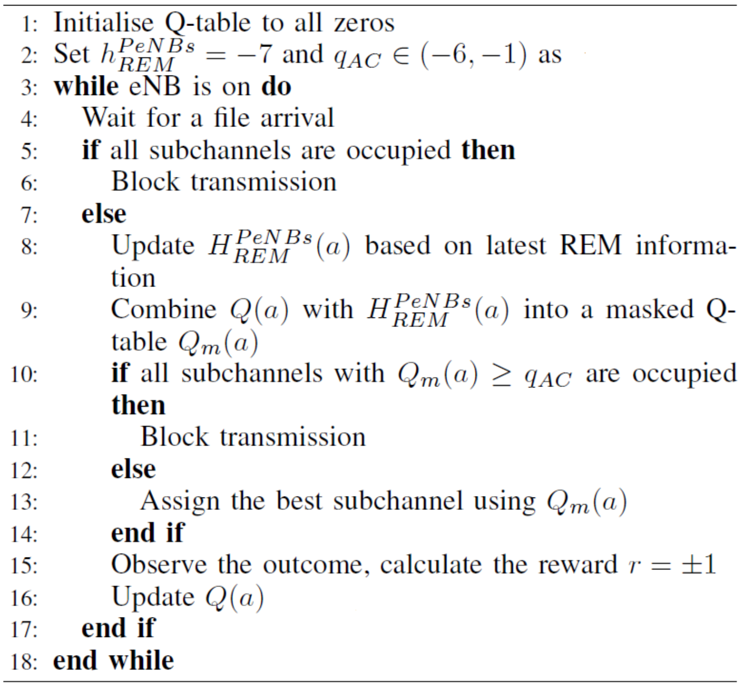
Algorithms in Figure 4.4 and Figure 4.5 summarize the HARL schemes for dynamic secondary spectrum sharing developed in this section. Figure 4.4 shows the sequence of steps in the distributed REM and ICIC accelerated Q-learning (DRIAQ) scheme, designed for stadium small cells to mitigate interference among themselves and the AeNB, using two superimposed heuristic functions. Figure 4.5 shows the REM accelerated Q-learning algorithm with Q-value based admission control (RAQ-AC), designed for the AeNB to share spectrum and avoid interference with the primary system.



```
 1: Initialise Q-table to all zeros
 2: Set $h_{ICIC} = -3$ and $h_{REM}^{AeNB} = -7$
 3: while eNB is on do
 4:     Wait for a file arrival
 5:     if all subchannels are occupied then
 6:         Block transmission
 7:     else
 8:         Update $H_{ICIC}(a)$ and $H_{REM}^{AeNB}(a)$ based on latest
            ICIC and REM information
 9:         Combine $Q(a)$ with $H_{ICIC}(a)$ and $H_{REM}^{AeNB}(a)$ into
            a masked Q-table $Q_m(a)$
10:         Assign the best subchannel using $Q_m(a)$
11:         Observe the outcome, calculate the reward $r = \pm 1$
12:         Update $Q(a)$
13:     end if
14: end while
```

Figure 4.4. Distributed REM and ICIC accelerated Q-learning (DRIAQ) for stadium small cells

Lines {2, 8, 9} of the algorithm in Figure 4.4 and lines {2, 8-12, 14} of the algorithm in Figure 4.5 are specific to the novel HARL schemes developed in this section. If they are removed and $Q_m(a)$ is substituted by $Q(a)$, the algorithms are simplified down to classical stateless Q-learning introduced in Subsection 3.2.1.



```
Initialise Q-table to all zeros
Set $h_{REM}^{PeNBs} = -7$ and $q_{AC} \in (-6, -1)$ as
while eNB is on do
    Wait for a file arrival
    if all subchannels are occupied then
        Block transmission
    else
        Update $H_{REM}^{PeNBs}(a)$ based on latest REM information
        Combine $Q(a)$ with $H_{REM}^{PeNBs}(a)$ into a masked Q-table $Q_m(a)$
        if all subchannels with $Q_m(a) \geq q_{AC}$ are occupied then
            Block transmission
        else
            Assign the best subchannel using $Q_m(a)$
        end if
        Observe the outcome, calculate the reward $r = \pm 1$
        Update $Q(a)$
    end if
end while
```

Figure 4.5. REM accelerated Q-learning with Q-value based admission control (RAQ-AC) for the AeNB

4.2.6 Choice of Parameters

The final details required to complete the design of the REM and the REM based heuristic functions are the values of the EWMA algorithm parameter λ and the probability of RBG occupancy threshold P_{min} for computing the heuristic functions $H_{REM\text{-}AeNB}(a)$ and $H_{REM\text{-}PeNBs}(a)$. We propose using $P_{min} = \lambda$ and $\lambda = 0.008$, while the REM is updated every 200 ms, which is frequent enough to capture the traffic variations of the PeNBs and the AeNB, yet not too frequent to introduce a large overhead of additional REM information that has to be broadcast to all cognitive eNBs.

The value of λ is chosen based on the rate of decay of a $p_a(occupied)$ estimate, e.g. the time it would take for a once heavily used RBG to be assumed unused, if the eNB of interest stopped using it. For example, if $p_a(occupied)=0.99$ and afterwards RBG a is not used for 600 consecutive REM updates, i.e. 2 minutes, the new $p_a(occupied)$ estimate, based on the EWMA equation proposed in Subsection 4.2.1, is the following:

$$p_a(occupied) = 0.99 \times (1 - \lambda)^{600} = 0.00799$$

which is just below $P_{min} = \lambda = 0.008$. Therefore RBG a would no longer be undesirable for secondary reuse, based on the heuristic function $H_{REM}(a)$. This value of λ is high enough to be applicable in dynamic environments where the monitored spectrum usage patterns change over time, yet not high enough to dismiss valuable historical spectrum usage information too quickly. This trade-off between the speed and accuracy of the EWMA algorithm, controlled by the λ parameter, is essential and must be carefully considered, e.g. using numerical examples such as the one described above.

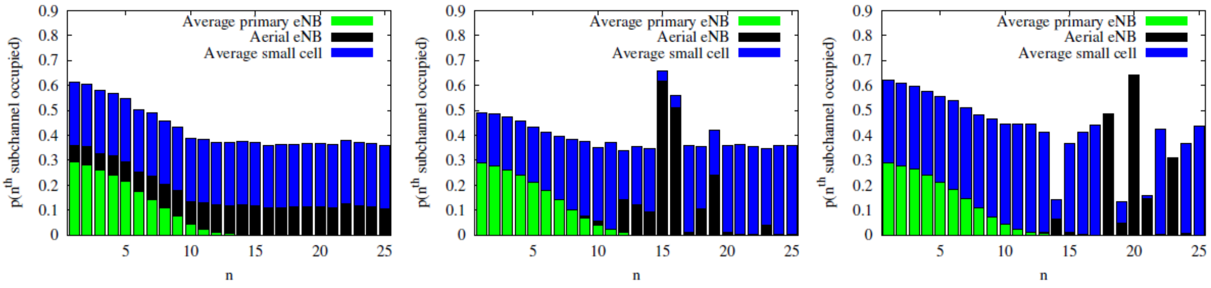
4.3 Simulation Results

The simulation experiments described in this section use the stadium temporary event spectrum sharing scenario described in Section 3.1 with the same simulation model parameters and assumptions as those used in Section 3.4, but with the AeNB being permanently switched on. The results of implementing the following three schemes in the secondary cognitive system are discussed in this section:

- “Isolated ICIC” - the AeNB and the stadium network are working independently, without any learning and without considering coexistence between each other or with the primary system. The stadium network independently employs a dynamic ICIC scheme, such as the one used in the primary system. Every eNB chooses randomly among the subchannels deemed ``safe" by the RNTP messages from its neighbours with the RNTP threshold of -3 dB. The AeNB assigns spectrum randomly, since it operates as an independent one cell network.
- “DIAQ + Q-learning” - all networks are also working completely independently. However, the stadium network employs the DIAQ scheme introduced in Subsection 4.1.2, and the AeNB is using stateless Q-learning from Subsection 3.2.1. This scheme represents a state-of-the-art distributed RL based solution to the spectrum sharing problem.
- “DRIAQ + RAQ-AC” - the combination of novel HARL based schemes developed in Section 4.2 and summarized in Figure 4.4 and Figure 4.5.

4.3.1 Spectrum Occupancy Analysis

Figure 4.6 shows the spectrum occupancy distribution of the PeNBs, the AeNB, and the small cell eNBs using three different spectrum sharing strategies described in the beginning of this section. The distributions were calculated by measuring the amount of time every eNB spent occupying every subchannel and dividing it by the total simulation time.



(a) Isolated ICIC (b) DIAQ + Q-learning (c) DRIAQ + RAQ-AC

Figure 4.6. RBG (subchannel) occupancy of of primary eNBs, aerial eNB and small cells using different dynamic spectrum sharing schemes

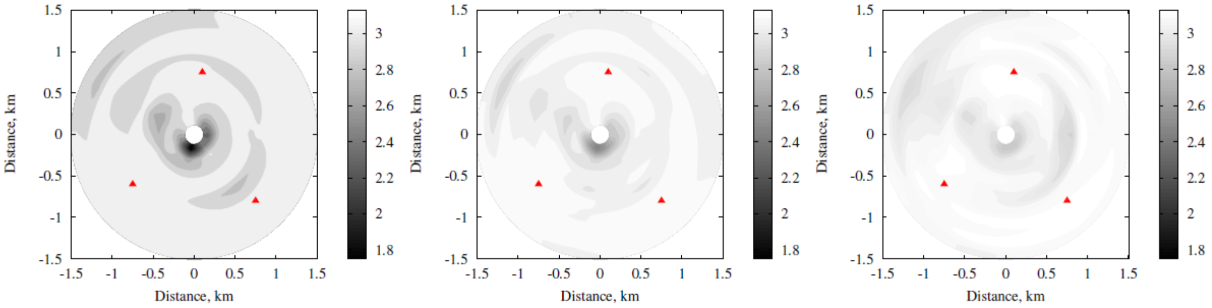
Figure 4.6a shows that, in the case of “isolated ICIC” implemented in all systems with no learning or coexistence control, both the AeNB and the small cell network use the whole spectrum approximately uniformly. Figure 4.6b demonstrates the difference made by introducing distributed Q-learning into the DSA process. The two challenging spectrum sharing relationships associated with this scenario tend to be addressed through distributed machine intelligence:

- the AeNB learns to avoid using the same spectrum as the PeNBs,
- the small cell eNBs tend to learn to use the RBGs preferred by the AeNB less than the others, i.e. they learn to avoid interfering with the AeNB, since it often results in blocked and interrupted file transmissions.

However, Figure 4.6c shows how the novel heuristically accelerated approach further improves the autonomously emerging spectrum sharing pattern by strictly guiding the learning process of the AeNB to avoid interfering with the PeNBs, and discouraging the small cell eNBs from exploring and assigning the RBGs frequently used by the AeNB. Firstly, there is no overlap in the spectrum used by the AeNB and the PeNBs. Secondly, the AeNB uses fewer RBGs, since the small cells successfully avoid using a number of the AeNB's top RBGs. This in turn positively reinforces the use of the same RBGs by the AeNB through the Q-learning algorithm.

4.3.2 Spatial Distribution of User Throughput

Figure 4.7 shows the spatial distribution of user throughput (UT) across the area outside of the stadium, covered by the PeNBs and the AeNB.



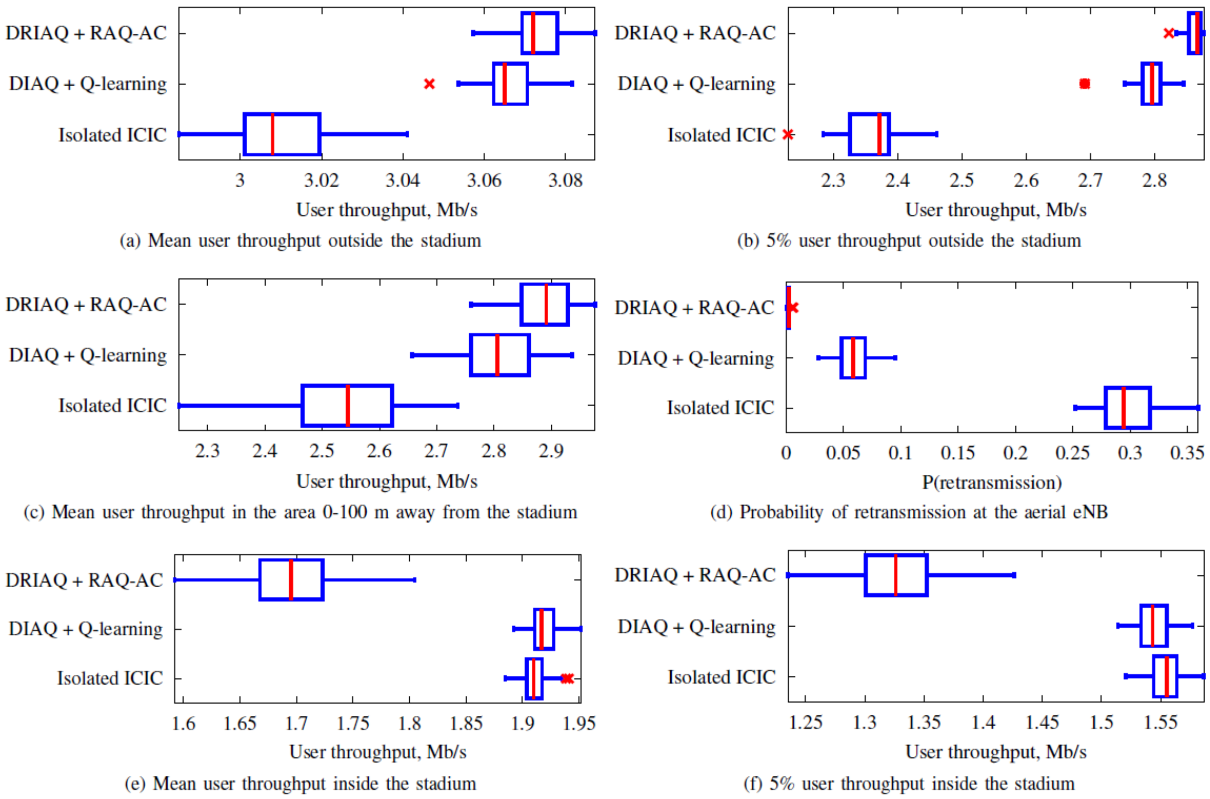
(a) Isolated ICIC (b) DIAQ + Q-learning (c) DRIAQ + RAQ-AC

Figure 4.7. Spatial distribution of user throughput (Mb/s) outside of the stadium (the triangles represent the primary eNB locations)

The contour plots indicate that the area most susceptible to harmful interference is that in the vicinity of the stadium, where the UEs are connected to the AeNB as well as the PeNBs. There is also interference radiating from the ultra-dense stadium small cell network. Figure 4.7a shows that the "isolated ICIC" approach, with approximately uniform spectrum occupancy distribution seen in Figure 4.6a, performs poorly and results in a significant decrease in UT in the vicinity of the stadium. Such performance degradation of the UEs located outside of the stadium is unacceptable from the viewpoint of secondary spectrum sharing. A significant improvement in the spatial UT distribution is achieved by using the learning based "DIAQ + Q-learning" approach. The performance is further improved by using the novel "DRIAQ + RAQ-AC" approach proposed in Section 4.2 due to its ability to autonomously achieve the significantly better spectrum partitioning pattern seen in Figure 4.6c.

4.3.3 Statistical Analysis

So far the results shown in Figure 4.6 and Figure 4.7 have assessed the performance of the networks in one specific simulation scenario, i.e. with the same UE locations, same path losses between each UE and each eNB, and consequently the same UEs connected to the AeNB, each PeNB and each small cell. The results in Figure 4.8 verify the statistical significance in performance improvements gained by using the HARL based "DRIAQ + RAQ-AC" scheme proposed in Section 4.2. It shows the results from 50 different simulation setups, i.e. with different random seeds, in the form of boxplots [28], a compact way of depicting key features of probability distributions. The box boundaries represent the first and third quartile of the distribution, the line between them marks the median result, and the whiskers show the minimum and the maximum point within *1.5×IQR* distance from the box boundaries. IQR is the inter-quartile range, the difference between the 3rd and the 1st quartile (the width of the box). Any results further than *1.5×IQR* away from the box are considered as the outliers and are plotted as individual data points.



(a) Mean user throughput outside the stadium

(b) 5% user throughput outside the stadium

(c) Mean user throughput in the area 0-100 m away from the stadium

(d) Probability of retransmission at the aerial eNB

(e) Mean user throughput inside the stadium

(f) 5% user throughput inside the stadium

Figure 4.8. Boxplots of the primary and secondary system performance from 50 different simulations

Figure 4.8a shows that the variation in mean UT of the users outside the stadium is negligibly small, when comparing different DSA and spectrum sharing strategies. However, the box plots of 5% UT outside of the stadium in Figure 4.8b reveal a more significant difference in their performance. 5% UT for a single simulation is obtained by calculating the 5th percentile of the UT values of 500 users outside the stadium. It is a more important metric than the mean UT, since it represents a minimum quality of service (QoS) guaranteed to 95% of the users, and thus shows how fair the spatial QoS distribution is. Introducing learning algorithms into the spectrum sharing strategies ("DIAQ + Q-learning") results in an 18% increase in the median 5% UT outside the stadium, whereas the "DRIAQ + RAQ-AC" scheme improves it by a further 3%. These improvements are statistically significant since there is no overlap between the boxes in the plot. The same improvement pattern is observed in Figure 4.8c which shows the mean UT of the users located in the vicinity of the stadium (0-100m from the boundary).

Figure 4.8d demonstrates the most notable performance improvement achieved by "DRIAQ + RAQ-AC". It almost entirely eliminates the retransmissions, i.e. the blocked and interrupted file transmissions, at the AeNB. It results in a 99% decrease in the probability of retransmission *P(re-tx)* compared to "Isolated ICIC" and a 96% decrease compared to a significantly better "DIAQ + Q-learning" scheme. *P(re-tx)* is defined as the ratio between the number of retransmissions and the total number of transmissions. This improvement is achieved due to high controllability provided by the REM based heuristic functions designed in Section 4.2. They successfully steer the learning process of the AeNB such that it avoids interfering with the PeNBs, whereas the small cell eNBs are continuously discouraged from occupying the resources preferred by the AeNB.

Figure 4.8e and Figure 4.8f show that the improvements in QoS, provided by the "DRIAQ + RAQ-AC" scheme to the PeNB and AeNB users, come at the cost of a ≈12% decrease in mean UT and a

≈15% decrease in 5% UT provided to the small cell users. However, this concession made by the stadium small cell network is relatively insignificant and essential in the context of dynamic secondary spectrum sharing. It results in the increased feasibility of secondary LTE spectrum reuse by a temporarily deployed eNB on an aerial platform and an ultra-high capacity density stadium small cell network. Furthermore, the "DRIAQ + RAQ-AC" scheme achieves remarkable reliability of AeNB communications (due to the lack of retransmissions). For example, this could be extremely useful in the temporary event scenario for providing robust wireless communication links to event organizers both inside and outside the stadium.

4.3.4 Temporal Performance

Figure 4.9 shows the temporal performance of the two learning based schemes, "DIAQ + Q-learning" and "DRIAQ + RAQ-AC", in terms of the probability of retransmission at the AeNB. All data points are spaced 1 minute apart and were obtained by averaging over 50 different simulations. The time response of "DIAQ + Q-learning" demonstrates that it behaves as a classical RL algorithm, i.e. starts at a poor performance level and gradually improves over time, while the AeNB and the small cell eNBs are learning appropriate spectrum sharing patterns. In contrast, the "DRIAQ + RAQ-AC" time response is a great demonstration of the temporal performance improvements achieved by introducing heuristic acceleration into the learning process. It starts at a superior *P(re-tx)* level and maintains it throughout the whole simulation.

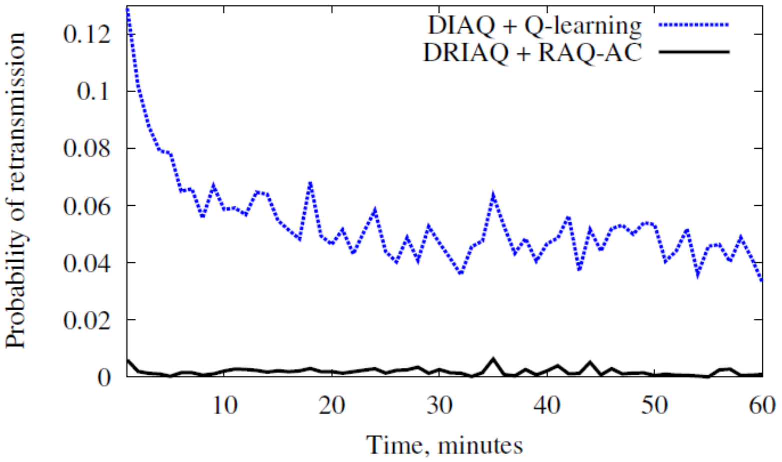


Figure 4.9. Probability of retransmission time response at the aerial AeNB

4.4 Conclusion

The HARL based framework proposed in this chapter utilizes a radio environment map (REM) as external information for guiding the learning process of cognitive cellular systems, which are thus able to reuse the LTE spectrum owned by another cellular network. The performance of the DSS and DSA schemes investigated in this chapter is assessed using system level simulations of a stadium temporary event scenario. This involves an eNB on an aerial platform, a small cell stadium network and a local primary LTE network. Two novel dynamic secondary spectrum sharing schemes are described in detail - distributed REM and ICIC accelerated Q-learning (DRIAQ) used by the small cell network, and REM accelerated Q-learning with Q-value based admission control (RAQ-AC) used by the aerial eNodeB. These schemes are shown to achieve high controllability of spectrum sharing patterns in a fully autonomous way. They also result in a significant decrease in primary system QoS degradation due to the interference from the secondary cognitive systems, compared to a state-of-the-art RL solution and a purely heuristic LTE solution that does not attempt to co-ordinate the cellular networks involved. The spectrum sharing patterns that emerge by using the proposed schemes also result in remarkable reliability of the cognitive eNodeB on the aerial platform due to a 96% decrease in the probability of retransmission compared to a classical RL approach.

Furthermore, the novel principle of superimposed heuristic functions proposed in the context of HARL, as well as the general Q-table mask structure of these functions, are not specific to the investigated spectrum sharing scenario, and are generally applicable to a wide range of self-organization problems.

5 Conclusions and Recommendations

5.1 Conclusions

This deliverable has investigated the performance evaluation of cognitive dynamic spectrum management within the ABSOLUTE project. The cognitive spectrum management schemes developed in this document contributed towards two scenarios: the Callania Public Protection and Disaster Relief (PPDR, as seen in D2.1) scenario and the temporary event (as seen in D2.1 Bastian) scenario. In the PPDR scenario, the related use cases included ABS.UC.01, ABS.UC.02, ABS.UC.03, ABS.UC.04, ABS.UC.05, ABS.UC.17 and ABS.UC.18. Based on the requirement of PPDR in D2.1 (rapid deployable and flexible network solution with broadband service over a large scale area), three transfer learning based algorithms have been designed to provide flexible solutions for different conditions in order to improve the system performance, which focus on the scenario with more unpredictable and random topologies deployment. In the Bastian temporary event scenario, the main goal is to provide network capacity enhancement in the urban areas. Thus, several algorithms have been developed to fit the scenario with more predictable and repeatable patterns of spectrum usage and network topologies. The following use cases have been investigated: ABS.UC.01, ABS.UC.02, ABS.UC.19 and ABS.UC.20.

Chapter 2 investigated the transfer learning scheme with K-means clustering in ABSOLUTE cognitive spectrum management for the PPDR scenario. The system model and the major issues of applying transfer learning in an LTE-A system were introduced. The novel transfer learning algorithms have been developed to optimize the system QoS and reduce the convergence speed comparison to the traditional Q-Learning schemes. In this report, the major process of transfer learning was designed as two functions: knowledge transfer and processing, and knowledge reward. Data mining techniques are used to find the hidden structure of data in this function, including data pre-processing and K-means clustering algorithm. The knowledge reward function determines how to make use of the converted information processed in the previous function for three conditions. With the aid of transfer learning, the severe negative impact of topology changes on radio environment can be minimized, and the system can effectively handle the dynamics of user traffic with reduced interference.

In Chapter 3, a case-based RL method was introduced as an effective and feasible approach to dynamic secondary spectrum sharing in temporary cognitive cellular systems with dynamic topologies. The simulation result show that the augment RL with the CBR increases the adaptivity of the cognitive cellular system to sudden changes in its radio environment, caused by the AeNB being dynamically switched on/off.

Chapter 4 proposed a HARL based framework to utilize a REM as external information for guiding the learning process of cognitive cellular systems, aiming to reuse the LTE spectrum owned by another cellular network for temporary scenario. Two novel dynamic secondary spectrum sharing schemes were described in detail for stadium temporary event scenario: distributed REM and ICIC accelerated Q-learning (DRIAQ) used by the small cell network, and REM accelerated Q-learning with Q-value based admission control (RAQ-AC) used by the aerial eNB. They achieve high controllability of spectrum sharing patterns in a fully autonomous way and decrease the interference from the secondary cognitive systems to the primary system. Furthermore, the novel principle of superimposed heuristic functions proposed in the context of HARL, as well as the general Q-table mask structure of these functions, are not specific to the investigated spectrum sharing scenario, and are generally applicable to a wide range of self-organization problems.

Chapter 5 presented a conclusion of this document and a recommendation of how to make use of all the schemes introduced in ABSOLUTE systems.

5.2 Recommendations

This deliverable has provided a description of how cognitive spectrum management can be applied to Public Protection and Disaster Relief and temporary event networks.

1) The learning techniques (both reinforcement learning and transfer learning) have been tailored to ABSOLUTE to improve the system performance, and in general is becoming more and more applicable to the increasing demand for efficient spectrum reuse. Several learning based algorithms have been developed to meet the special requirements of the two scenarios in ABSOLUTE project. For the Callania PPDR scenario, the transfer learning based schemes are designed to ensure the network is better able to cope with more unpredictable and random topologies. For the Bastian temporary event scenario, the solutions are more applicable in a network with a large number of base stations, where more predictable and repeatable patterns of spectrum usage and network topologies. The schemes developed are in general compatible with existing user equipment, with changes restricted to the network side, or an additional app download on the UE. Future capabilities of ABSOLUTE like network equipment could be readily achieved.
2) The transfer learning with K-means clustering algorithms provide a good solution of the cognitive spectrum management for PPDR scenario in ABSOLUTE, where the PPDR focuses on a rapidly deployable and flexible network solution with broadband service in large scale areas. Three algorithms are designed to meet the requirements of PPDR scenarios: newly deployed eNBs, existing eNBs looking for extra spectrum bands and existing eNBs assigning spectrum bands for new arrivals. These transfer learning based algorithms improve the system performance and reduce the convergence speed. This approach is suitable to exploit in PPDR networks like Callania scenario, where are more unpredictable, flexible and random topologies.
3) Data mining techniques, like K-means clustering, have been shown efficiently dealing with the knowledge process in the transfer learning based algorithms. It helps to find out the hidden data structure behind information exchanged, and should possibly to be used within a broader concept of spectrum management and topology management in the ABSOLUTE network.
4) The intelligent reinforcement learning based approach to spectrum management provides a good method of using spectrum more efficiently. Such intelligent algorithms are especially important for temporarily deployable high capacity networks, where the efficient reuse of spectrum could result in significant capacity enhancements, vital in a PPDR scenario and temporary event scenario.
5) The case-based reinforcement learning approach improves the stability of RL based DSA algorithms in dynamically changing environments with a number of potentially reoccurring network topology patterns. It is computationally inexpensive and does not require access to any additional information, except for the topology updates in the network, e.g. which would be readily available in the REM. This approach should be exploited in temporary event networks, where more predictable and repeatable patterns of spectrum usage and network topology changes

take place. It will have limited effects in the networks like PPDR, which are more incremental in nature.

6) The heuristically accelerated RL methods proposed significantly improve the initial performance and the convergence properties of RL based DSA algorithms in such rapidly deployable cognitive cellular systems as those investigated in the ABSOLUTE project. They make excellent use of the heuristic information contained in the ICIC signals of the secondary systems, as well as the REM of the wider radio environment. These techniques bring significant primary and secondary user QoS benefits to any future RL based DSA network, where such heuristic spectrum management information is available, e.g. ICIC signals or the REM, and thus should be considered in any future network design based on shared spectrum LTE-A scenarios, including LTE-A in the unlicensed band.

References

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

[2] "Description of Work: Aerial Base Station with Opportunistic Links for Unexpected & Temporary Events (ABSOLUTE)," 2012.

[3] Q. Zhao and D. Grace, "Agent transfer learning for cognitive resource management on multi-hop backhaul networks," in *Future Network and Mobile Summit (FutureNetworkSummit)*, Lisboa, 2013.

[4] "Data Mining," [Online]. Available: http://en.wikipedia.org/wiki/Data_mining.

[5] M. Bramer, Principles of Data Mining, Second ed, Springer, 2013.

[6] J. Han, Data Mining: Concepts and Techniques, Second ed, Morgan Kaufmann Publisher, 2006.

[7] "K-means clustering," [Online]. Available: http://en.wikipedia.org/wiki/K-means_clustering.

[8] D. Kimura and H. Seki, "Inter-Cell Interference Coordination (ICIC) Technology," *Fujitsu Scientific & Technical Journal (FSTJ) LTE,* vol. 48, no. 1, pp. 89-94, 2012.

[9] C. B. N. Limited, "Backhauling X2," Limited, Cambridge Broadband Networks, [Online]. Available: http://cbnl.com/sites/all/files/userfiles/files/Backhauling-X2.pdf.

[10] Q. Zhao, S. Rehan, D. Grace, A. Vilhar, A. Svigelj, K. Alic, K. Gomez, T. Rasheed, S. Chandrasekharan, K. Sithamparanathan, M. Thakur, A. Munari and L. Reynaud, "FP7-ICT-2011-8-318632-ABSOLUTE/D4.1.4 Aerial Base Statioins with Opportunistic Links for Unexpected & Temporary Events," 2015.

[11] I. Watson, "Case-based reasoning is a methodology not a technology," *Knowledge-Based Systems,* vol. 12, no. 56, pp. 303-308, 1999.

[12] C. Jiang and Z. Sheng, "Case-based reinforcement learning for dynamic inventory control in a multi-agent supply-chain system," *Expert Systems and Applications,* vol. 36, no. 3, pp. 6520-6526, 2009.

[13] L. Celiberto, J. Matsuura, R. Lopez de Mantaras and R. Bianchi, "Reinforcement learning with case-based heuristics for robocup soccer keepaway," in *2012 Brazilian Robotics Symposium and Latin American Robotics Symposium (SBR-LARS)*, 2012.

[14] C. Watkins, "Learning From Delayed Rewards," PhD Thesis, University of Cambridge, England, 1989.

[15] C. Claus and C. Boutilier, "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*, 1998.

[16] N. Morozs, T. Clarke, D. Grace and Q. Zhao, "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)*, 2014.

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

[18] 3GPP, "LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures (3GPP TS 36.213 version 11.5.0 Release 11)," 2013.

[19] P. Kyosti, J. Meinila, L. Hentila, X. Zhao, T. Jamsa, C. Schneider, M. Narandzic, M. Milojevic, A. Hong, J. Ylitalo, V. Holappa, M. Alatossava, R. Bultitude, Y. de Jong and T. Rautiainen, "IST-4-027756 WINNER II Deliverable D1.1.2: WINNER II channel models," 2008.

[20] 3GPP, "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)," 2010.

[21] 3GPP, "LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Frequency (RF) system scenarios (3GPP TR 36.952 version 11.0.0 Release 11)," 2012.

[22] R. Bianchi, M. Martins, C. Ribeiro and A. Costa, "Heuristically-Accelerated Multiagent Reinforcement Learning," *Cybernetics, IEEE Transaction on,* pp. 252-265, 2014.

[23] R. Bianchi and R. Lopez de Mantaras, "Case-Based Multiagent Reinforcement Learning: Cases as Heuristics for Selection of Actions," in *European Conference on Artificial Intelligence (ECAI)*, 2010.

[24] N. Morozs, D. Grace and T. Clarke, "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)*, Sydney, 2014.

[25] C. Ghosh, S. Roy and D. Cavalcanti, "Coexistence challenges for heterogeneous cognitive wireless networks in TV white spaces," *Wireless Communications, IEEE,* vol. 18, pp. 22-31, 2011.

[26] G. Ross, N. Adams, D. Tasoulis and D. Hand, "Exponentially weighted moving average charts for detecting concept drift," *Pattern Recognition Letters,* vol. 33, pp. 191-198, 2012.

[27] N. Morozs, T. Clarke and D. Grace, "A novel adaptive call admission control scheme for distributed reinforcement learning based dynamic spectrum access in cellular networks," in *International Symposium on Wireless Communication Systems (ISWCS)*, Ilmenau, 2013.

[28] R. McGill, J. Tukey and W. Larsen, "Variations of box plots," *The American Statistician,* vol. 32, pp. 12-16, 1978.

[29] ETSI, *TS 102 721: Satellite Earth Stations and Systems; Air Interface for S-band Mobile Interactive Multimedia (S-MIM),* 2012.

[30] M. Simsek, M. Bennis and A. Czylwik, "Dynamic Inter-Cell Interference Coordination in HetNets: a Reinforcement Learning Approach," in *IEEE Global Communications Conference (GLOBECOM)*, 2012.

[31] M. Dirani and Z. Altman, "A Cooperative Reinforcement Learning Approach for Inter-Cell Interference Coordination in OFDMA Cellular Networks," in *International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt)*, 2010.

[32] P. Vlacheas, E. Thomatos, K. Tsagkaris and P. Demestichas, "Autonomic Downlink Inter-Cell Interference Coordination in LTE Self-Organizing Networks," in *International Conference on Network and Services Management (CNSM)*, 2011.

[33] K. Gomez, T. Rasheed, R. Hermenier, T. Jiang, S. Rehan, D. Grace, L. Reynaud, T. Javornik, I. Ozimek and L. Le Garrec, "FP7-ICT-2011-8-318632-ABSOLUTE/D2.6.1 System-wide Simulations Planning Document," 2013.

[34] T. Jiang, P. Li, C. Liu, N. Khan, D. Grace, A. Burr and C. Oestges, "EU FP7 INFSO-ICT-248267 BuNGee Deliverable D4.1.2: Simulation Tool(s) and Simulation Results," 2012.

Acknowledgement

This document has been produced in the context of the ABSOLUTE project. ABSOLUTE consortium would like to acknowledge that the research leading to these results has received funding from the European Commission's Seventh Framework Programme (FP7-2011-8) under the Grant Agreement FP7-ICT-318632.