Intelligent Radio Resource Management for Mobile Broadband Networks

Qiyang Zhao

A thesis submitted in partial fulfilment
of the requirements for the degree of

Doctor of Philosophy (Ph.D.)

University of York

Department of Electronics

September 2013

Abstract

This thesis studies intelligent spectrum and topology management through transfer learning in mobile broadband networks, to improve the capacity density and Quality of Service (QoS) as well as to reduce the cooperation overhead and energy consumption. The dense deployment of small cell base stations (BSs) is an effective approach to provide high capacity density access. In the meantime, multi-hop wireless backhaul networks enable highly flexible deployment and self-organization of small cell BSs. A heterogeneous small cell access and multi-hop backhaul network is studied in this thesis as mobile broadband system architecture.

Transfer learning is applied to Radio Resource Management (RRM) as an intelligent algorithm to improve the performance of conventional reinforcement learning. In transfer learning, a BS trains its knowledge base relying on knowledge transferred from other related BSs, who are selected using an interference coordination strategy. In a network with static topology, cooperation management is developed to identify the maturity of the knowledge base and control the coordination overhead. It is demonstrated in a multi-hop backhaul network that transfer learning delivers a QoS level that is as high as achieved by a fully coordinated algorithm, but with a very low level of information exchange which is close to a fully distributed algorithm.

Transfer learning is also studied in rapidly changeable network architectures to provide reliable communication. It is carried out during the changes of network topology, through mapping the learner's knowledge base to a prioritized action space with Pareto efficiency. This process assists the BSs to quickly identify and adapt to environment changes, and makes effective decisions. Results show that transfer learning significantly reduces QoS fluctuation during traffic variation and topology changes in a highly dynamic network. Furthermore, a dynamic topology management algorithm is developed to intelligently control the working modes of BSs, based on traffic load and capacity in multiple cells. Topology management is demonstrated to reduce the number of activated BSs with adequate QoS performance provided. Dynamic capacity provision between multiple cells is achieved from transfer learning, which significantly improves QoS and reduces energy consumption.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Dr David Grace. He has offered me great freedom on selecting research directions and provided me with various ideas on up to date research topics. This thesis would not be possible without his guidance and instructions.

I would also like to thank Dr Paul Mitchell and Mr Tim Clarke for their valuable suggestions, assistance and criticisms on my research work.

My thanks go to all colleagues in the Communication Research Group and visitors from other institutes as well, who have created a friendly environment for both academic discussions and social life.

This thesis is dedicated to my parents, for their continuous and enormous support on my PhD study.

Declaration

Some of the work presented in this thesis has been published at or submitted to academic conferences or journals, which are listed at the end of this thesis.

To the best knowledge of the author, all the work claimed as original in this thesis is so. References and acknowledgements to other researchers have been given as appropriate.

Chapter 1. Introduction

1.1 Overview

Traffic density in wireless communication systems has been growing significantly in recent decades. Future mobile broadband systems are targeted at delivering ultra high capacity density networks [1], which will support an increasing number of mobile subscribers and a growing demanding of high speed data rate services. Network capacity will be heavily constrained by spectrum availability in the near future, because of the high level throughput and Quality of Service (QoS) requirement from a growing number of users [2]. An effective approach to high capacity provision under limited spectrum resources is to densely deploy small cellular base stations [3]. By splitting a conventional macro/micro network into small cells, effective spectrum reuse can be carried out with improved link budgets. The Shannon model indicates that the data rate on individual links can then be significantly improved due to wider bandwidth availability and better received signal level [4].

The small cell network architecture has a number of technical challenges. The major issues with respect to Radio Resource Management (RRM) can be categorized as follows:

- Network Complexity

The small cell architecture brings significantly more Base Stations (BSs) into a wireless network. A major issue in this type of network is the backhaul architecture. Traditionally wired fibre or microwave links connect Macro/Micro BSs to the Core Network (CN). However, this approach would incur substantial deployment costs in a small cell network. Wireless backhaul architecture is an effective solution for flexible deployment and cost reduction [5]. Under this approach, a large number of small BSs can be deployed in the locations that have capacity enhancement demands. The wireless backhaul and access network constructs a heterogeneous architecture, which is the baseline of this work.

Network management is another challenge in a small cell architecture. On the one hand, the heterogeneous architecture requires complex algorithms to control various network entities. On the other hand, the control information brings heavy overheads, which reduce network performance. A decentralized self-organizing network management strategy is potentially an effective approach to mitigate these challenges [6], which is an objective of this work.

- Spectrum Management

Spectrum bands are conventionally allocated to BSs and MSs by a centralized Radio Network Controller (RNC) [7]. In a small cell network this approach requires complex algorithms and architecture. Dynamic Spectrum Access (DSA) is a promising approach to simplify spectrum management and to improve spectrum utilization [8]. This paradigm has been widely explored by applying distributed intelligent algorithms [9]. However, a serious drawback of distributed algorithms is that a number of immature decisions should be carried out prior to achieving an improved solution, which cannot guarantee steady and reliable QoS. Moreover, such algorithms become increasingly ineffective in rapidly changeable dynamic environments. The issue of delivering reliable QoS as well as reducing cooperation overhead in a dynamic radio network is the main research topic of this thesis.

- Energy Efficiency

Green communication is becoming vitally important in the future wireless networks. Analysis of energy consumption in typical cellular systems shows that the BSs consume most of the energy [10, 11] in a wireless network. It can be anticipated that the energy issue will be even more serious in small cell networks, because a large number of BSs are densely deployed. An important approach to overcome this issue is to intelligently control the number of activated BSs based on the dynamics of user traffic, as well as maintaining adequate QoS and capacity.

1.2 Hypothesis

The hypothesis of this thesis is that transfer learning can improve system QoS and throughput performance, and reduce cooperation overhead and energy consumption.

Dense capacity wireless networks are proposed to have low level inter-entity control information exchange with sufficient and reliable QoS provision as well as low energy consumption. Conventional radio resource management mechanisms are designed in either a distributed or coordinated manner. The distributed learning strategies require a number of heuristic decisions to learn the radio environment. On the other hand, the inter-entity coordination algorithm provides reliable QoS but requires massive information exchange.

Transfer learning introduces effective multi-agent cooperation into distributed reinforcement learning, which can significantly improve QoS and throughput as well as providing reliable and steady performance in both static and dynamic networks. Distributed cognitive agents can make effective decisions based on the knowledge base trained by information from other agents. A cooperation management algorithm can minimize the coordination overhead and while maintaining high level QoS. In the networks with dynamic traffic and topologies, transfer learning can mitigate environmental impact and provide a steady level of QoS. Furthermore, topology management can reduce energy consumption by effectively controlling the number of activated base stations. Transfer learning with topology management achieves dynamic capacity provision in cellular networks, which significantly improves energy efficiency and QoS.

1.3 Outline

This thesis is organized as follows.

Chapter 2 provides a literature review on the background and established work related to this thesis. Overviews of beyond next generation mobile broadband networks are given, together with self-organization requirements for radio resource management. Particularly, we focus on the dense small cell access networks and wireless backhaul networks as the dense capacity architecture in future networks. Conventional RRM algorithms are reviewed, including those carried out on GSM, WCDMA, LTE and WiFi systems. A comparison between centralized, distributed, coordinated and independent RRM algorithms is provided. Cognitive radio and dynamic spectrum access mechanisms are reviewed, with concentration on learning algorithms for the development of intelligent spectrum management. Furthermore, recent research work on energy efficient wireless cellular networks is given, with an introduction to dynamic network management.

Chapter 3 introduces the wireless network architecture used in this work, including a multi-hop backhaul network and a small cell access network. Various simulation tools are discussed. The detailed modelling methodology is presented, followed by parameters used for performance evaluation. Furthermore, an introduction to Markov analysis used for a theoretical proof later in this thesis is given.

The main contributions of this work are illustrated in Chapter 4 to Chapter 7. In accordance with the architectures of a high capacity density network [12], these chapters are categorized into two parts as illustrated in Figure 1.1.

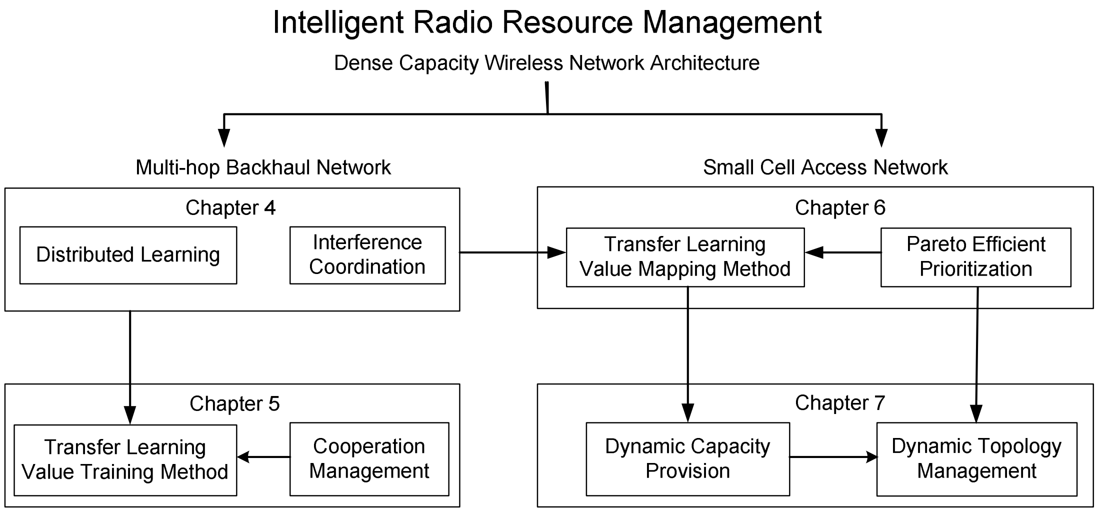

Figure 1.1. Thesis Structure

The first part consists of Chapter 4 and Chapter 5, which focuses on the multi-hop backhaul network. The purpose of this part is to provide high level QoS on backhaul links, as well as reducing cooperation overheads between distributed base stations.

Chapter 4 presents the early work of spectrum management strategies developed for a multi-hop backhaul network. A novel space-division interference coordination strategy for this architecture is presented, by employing channel information exchange. Comparison and analysis of linear reinforcement learning and Q learning algorithms are given, together with improved strategies on a multi-hop backhaul network.

Chapter 5 proposes a newly developed transfer learning paradigm. It is designed as an integration of distributed reinforcement learning and interference coordination, which benefits from distributed decision making as well as high QoS provision. More importantly, cooperation management algorithms are designed to control the amount of information exchanged in transfer learning. Transfer learning with cooperation management is aimed at delivering a high QoS together with a low level coordination overhead.

The second part consists of Chapter 6 and Chapter 7, which focus on the small cell access network. The purpose of this part is to provide reliable communication in the network with dynamic topologies, and reduce energy consumption from dynamic network topology management.

Chapter 6 introduces the concept of a flexible network architecture, which enables the base stations to switch between active and sleep modes. This operation is applied in the scenarios of opportunistic deployments and energy efficient networks. Transfer learning is designed to enhance the knowledge base with topology information during the transition of network architecture. The algorithm prioritizes the action space and maps it with corresponding knowledge base. The target is to provide a steady and reliable QoS level under highly dynamic user traffic and network topology.

Chapter 7 analyses the dynamic capacity provision achieved from transfer learning, with comparisons to a conventional frequency band allocation strategy. A dynamic Topology Management (TM) algorithm is developed to intelligently change the operation mode of base stations, based on user traffic and network capacity. The objective is to effectively reduce energy consumption and provide adequate QoS. Furthermore, transfer leaning is applied to improve QoS and energy efficiency through dynamic capacity provision in multiple cells.

Chapter 8 presents possible future work based on this thesis. Chapter 9 concludes this work and summarizes original contributions.

Chapter 2. Literature Review

2.1 Introduction

This chapter gives a review on published research work related to intelligent radio resource management for the next generation mobile broadband systems. High capacity density network is vitally important to provide broadband wireless access service in densely populated urban areas. It demands a self-organized and flexible architecture design. In this context, a dense small cell access network has been proposed to enhance system capacity in a distributed manner. In addition to this, wireless backhaul is a promising technique to support flexible deployment of small cell base stations (BSs). Radio Resource Management (RRM) is essential in these heterogeneous networks to make the spectrum resources effectively shared by high density wireless users. Multiple Access Schemes are used in RRM to allow multiple users to connect to the same network and share its capacity. Spectrum Management is vital in RRM to effectively divide a common spectrum into resource blocks and assign them to users. Frequency band Allocation (FA) schemes have been widely used to mitigate interference between wireless entities. Dynamic Spectrum Access (DSA) is a newly proposed technique to improve spectrum utilization. Cognitive radio has been studied in wireless network to intelligently control interference in DSA, which is primarily supported by various machine learning techniques. On the other hand, energy efficiency is a crucial aspect in future networks alongside capacity provision. Power components in BSs consume most of the energy in a cellular network, thus dynamic network planning is vital for energy saving.

An overview of high capacity density wireless systems is firstly given in Section 2.2, followed by a review of dense small cell access and wireless backhaul networks. In Section 2.3, a comprehensive review of Radio Resource Management (RRM) techniques in wireless networks is given, which includes multiple access schemes, conventional frequency band allocation strategies and novel dynamic spectrum access mechanisms. Cognitive radio techniques are presented in Section 2.4, which allows wireless nodes to learn the radio environment and make decisions for data transmission. This is followed by various machine learning algorithms. Section 2.5 provides an overview of energy efficiency studies in wireless networks.

2.2 Next Generation Mobile Broadband Networks

2.2.1 *High Capacity Density Wireless Networks*

The capacity demands of mobile communication systems have grown significantly over the past decades, because of the increasing data traffic from mobile subscribers. Mobile broadband access is becoming vitally important in many aspects of our society and people’s daily life. In current and future wireless network, there are a number of different mobile devices (i.e. smartphones, tablets, laptops) transmitting various types of data traffic (i.e. video, data, voice) [13]. It has been reported that mobile networks connect three times more users than wired networks. In addition, cloud networks have been widely investigated in recent years, aimed at connecting mobile devices to a data centre anytime, anywhere [14]. In this context, mobile broadband access is essential to deliver high data rate services in extensive coverage areas.

Mobile user and traffic density vary significantly between rural and urban areas. The densely populated urban areas have ultra high capacity demands, which cannot be effectively supported by current cellular systems [12]. Wireless traffic and user density in metropolitan area have been growing significantly in recent years, because mobile internet changes lifestyle and plays a key role in business. A high capacity density wireless network is thus essential to support the increasing traffic and user density.

The next generation wireless systems are designed to significantly enhance network capacity and wireless link data rate. The IMT-Advanced standard specifies a nominal data rate of 100 Mb/s for high mobility users and 1 Gb/s for stationary users [15]. 3GPP LTE-Advanced and IEEE WiMAX II are mainstream standards on the road to these targets. Furthermore, the FP7 BuNGee project proposes a 1 Gb/s/km^2 high capacity density network for the deployment of beyond 4G wireless system in typical European cities [16].

Spectrum availability is one of the main constraints in a high capacity density network. The Shannon equation [4] indicates that link data rate is limited by channel bandwidth and Signal-to-Interference plus Noise Ratio (SINR). Ultra high user density in a network largely reduces the bandwidth of each channel assigned and increases interference between wireless links. In the physical layer, there are various techniques under research to improve spectrum efficiency of wireless channels, including OFDM, MIMO, Adaptive Modulation and Coding (AMC), Cooperative Communications, etc. [13] The main purpose is to effectively enhance capacity on an individual channel. However, system capacity is also determined by resource utilization and interference management, which is a major research area in RRM.

Spectrum reuse is an essential method to enhance system capacity under limited resources. Spectral efficiency determines the number of users and the volume of traffic that a network can support in a given spectrum band. A maximized spectrum reuse strategy can accommodate more users and higher traffic in the system. However, this could result in excessive interference between multiple links, which in turn reduces channel capacity and system throughput. An effective spectrum reuse strategy should trade off reuse efficiency and co-channel interference.

Channel capacity is also affected by signal strength other than bandwidth and interference, according to the Shannon model. Received signal strength is determined by transmit power, antenna gain and path loss. Power control and power management have been extensively studied for improving received signal gain on multiple users in a network [17], from a conventional water-filling algorithm [18] to intelligent algorithms such as reinforcement learning [19, 20] and game theory [21, 22]. Power allocation is shown to effectively improve system capacity in a given network architecture and propagation environment.

Heterogeneous Networks (HetNet) represent a potentially highly effective method to enhance network capacity, which emerges from conventional cell splitting used in many cellular systems [4]. HetNet significantly improves spectrum reuse by increasing the number of cells. Moreover, signal strength can be largely enhanced with the same transmit power, because HetNet reduces path loss and improves the propagation environment between BSs and MSs.

2.2.2 *Dense Small Cell Access Networks*

A Small Cell Network (SCN) is an effective approach to deliver a dense capacity wireless system, because the smaller cells can afford more subscribers per unit area or higher data rate [3]. BSs in SCN comprise light weight equipments with low transmit power and small antennas. Moreover, they are cost effective in installation, operation and management. The location of SCN BSs can be highly flexible, i.e. on the walls, street lights, and trees. SCN is widely used in next generation mobile broadband systems, in conjunction with conventional Macro cells to construct a HetNet. In the 3GPP LTE standard, SCN is constructed by outdoor Pico cells or indoor Femto cells [23]. Micro cells are used in 2G and 3G systems to enhance system capacity by splitting Macro cells. However, overlapped coverage delivered by SCN provides better spectrum utilization and link selection for MSs. Furthermore, SCN BSs can be deployed in hotspot areas with smaller coverage, similar to a WiFi AP.

Self-organization is a crucial requirement in SCNs. HetNet is a complex architecture that causes significant operational challenges. Frequency Planning (FP) in a SCN could be complicated because of its highly dynamic topology [6]. It is difficult to predict and control the interference between SCNs and Macro cellular network. Moreover, FP algorithms cannot provide effective spectrum reuse and utilization because the user traffic in SCN can be highly dynamic. Last but not least, FP is operated through centralized RRM, which requires BSs to communicate with a RNC for admission control, channel allocation, handover, load balancing, etc. This incurs excessive control information overhead in a HetNet architecture.

Self-organization can reduce the cost of SCNs. This includes but not limited to the cost of BS equipment, network deployment, land rentals and power supply. It is thus important to maximize the capacity provision in a given network topology prior to deploying new BSs. However, user traffic is highly dynamic in different cell coverage areas and different hours of a day. Self-organized RRM is essential to monitor traffic load variation in both time and spatial domains and improve network planning for capacity provision. Furthermore, RRM should dynamically adapt spectrum patterns to the dynamic network environment.

There are three different deployment methodologies proposed for small cell networks [23], including multicarrier, carrier aggregation and co-channel deployment. Multicarrier deployment assigns separated spectrum bands to macro and small cells for interference avoidance. This solution requires an improved load balancing algorithm to transfer the user traffic between these two networks for congestion control. It could be highly inefficient as it creates undesirable bandwidth segmentation. Carrier aggregation is a solution that provides flexibility of spectrum sharing between macro and small cells. In this scheme, one carrier frequency is used for macrocell coverage and another is shared between macrocell and small cells. The interference between overlapped macro and small cells can be avoided as they operate on different spectrum bands. Moreover, the UEs out of small cell coverage can use the spectrum band assigned for small cell BSs, which improves spectrum utilization. Co-channel deployment is one of the most attractive and challenging solutions. In this scenario, all macro and small cell BSs are deployed in the same spectrum band, which avoids bandwidth segmentation and maximizes spectrum utilization. However, interference between adjacent and overlapped cells could be excessively high unless it can be effectively controlled by intelligent RRM algorithms.

2.2.3 *Multi-hop Backhaul Networks*

The backhaul network is a major challenge in a SCN architecture. Conventional fibre or microwave backhaul used in macro or micro cellular networks could be very expensive for connecting a large number of small cell BSs [24]. Moreover, the deployment of fibre backhaul is limited by the geographical environment, and the implementation of microwave backhaul is constrained by propagation environment. Small cell BSs are designed for flexible deployment anywhere anytime [3], thus the fibre and microwave backhaul are not economical and realistic solutions.

A LTE-A relay architecture has been proposed by 3GPP as a candidate HetNet solution for improving coverage and cell edge performance [25]. Relay eNBs are connected to Macro eNBs through wireless backhaul links [23]. The wireless backhaul network is implemented with directional antennas at the transceivers, in order to mitigate interference in the access network and enhance capacity. The topology of the backhaul network could be single or multi hop. The single-hop backhaul architecture has been proposed in [24-26] for the relay network and in [12] for the small cell network. In this architecture, the macrocell BS connects directly to the relay BSs. It has been illustrated in [25] that the role of the relay eNB in LTE-A is mainly to enhance throughput and extend coverage in each sector of a macro eNB. In this case, the backhaul network can be constructed as a simple star topology. In the meantime, [12] uses a similar architecture to construct the backhaul network for the access BSs. The advantage of the single-hop network is that the link has no relay burden, which reduces the amount of radio resources required for the relay traffic. Moreover, the simple topology makes it easy to carry out effective interference management, routing, congestion control, end-to-end reliable connection, etc. However, it becomes inefficient in the scenario where the small cell BSs are densely deployed. The interference at the backhaul hub is excessively high because of the link density. Furthermore, transmit power should be high enough to connect the relay eNBs that are deployed at the edge of macrocell.

The multi-hop backhaul network is proposed in [27-30], which allows BSs acting as wireless relay nodes to forward traffic from other cells. This architecture provides highly flexible deployment of the SCN BSs, and significantly reduces the complexity of the ad hoc network. In this scenario, the locations of BSs are fixed, which do not need complicated distributed routing algorithms. The connection between a MS and a BS remains single-hop, which significantly reduces inter UE interference. Furthermore, the interference at the backhaul hub can be mitigated by reducing the link density. The multi-hop backhaul network also diminishes path loss, because the point to point communication between neighbor BSs reduces transmit power and enhances signal power. This network is particularly reliable on the highways and railways, because the highly fluctuating user traffic between multiple cells can be backhauled via a stable end-to-end connection.

The spectrum management strategy for a wireless backhaul network can be classified as: in-band backhaul and out-of-band backhaul. The in-band backhaul shares the same spectrum pool allocated to access network. In theory, optimal resource utilization can be achieved through this scheme. However, interference between access and backhaul links is excessively high because they are located in the same area. The spectrum allocation algorithm could be very complex [12].

The purpose of spectrum sharing is to overcome the issue of spectrum holes incurred by the dynamics of user traffic. The user traffic in a backhaul link is directly determined by that in the access network. In this context, spectrum sharing between the backhaul and access networks is not vital for the improvement of spectrum utilization. The out-of-band backhaul is potentially an effective strategy to provide sufficient bandwidth and avoid inter network interference. There are several solutions to achieve out-of-band backhaul. The operator could use part of the allocated spectrum, such as 800 MHz, 2.6 GHz (LTE) or 3.5 GHz (WiMAX), as dedicated for the backhaul network [30]. On the other hand, the unlicensed spectrum, such as 5 GHz band, can be freely used for backhaul links. There is also research on 60 GHz mm-wave band for backhaul capacity [12]. In this scenario, significant capacity enhancement can be achieved because of the characteristic natural directivity and considerable propagation loss in mm-wave band. The interference from co-channel links can be reduced because signal attenuation is very high. However, a strict line-of-sight propagation environment is necessary to guarantee sufficient received signal power. In this case, the transceiver antennas of backhaul links should be deployed over a rooftop level.

2.2.4 *Mobile Ad-hoc Networks*

Mobile Ad-hoc Network (MANET) is another promising architecture for the dense capacity wireless system. The network topology in MANET is highly flexible that the nodes are moving and connect with each other. It can handle many-to-many connections and is capable of dynamically updating and optimizing these connections [31]. Routing protocol is especially vital in this network to deliver effective end-to-end QoS.

MANET is highly efficient in providing communications anytime anywhere. It significantly reduces the number of hops that a data file has to be delivered from a source to a destination. Such simplified network architecture reduces the relay burden and thus requires less radio resources to provide capacity. There are many use cases that employ MANET as an effective solution. For example, it allows the User Equipment (UE), such as mobile phone, tablet, camera, laptop, to directly share videos, photos, files with each other in parties, sports games, tourist attractions, shopping malls, etc. Furthermore, P2P services can be effectively implemented on MANET, which allows a UE to obtain data files directly from others in vicinity without causing additional traffic load on the cellular base stations. In some low traffic energy efficient wireless systems such as sensor networks or smart grids, fixed or mobile ad hoc network significantly improves the flexibility in deployment and reduces power consumption. MANET is also popular in future Vehicular Ad hoc Networks (VANETs), where the vehicular are allowed to connected to the traffic control centre for road information, weather, news, etc. Connectivity in VANET is has severe challenge due to the fast moving ad hoc nodes caused by the dynamics of road conditions, which has been studied in [32].

Another example of partial MANET is the multi-hop cellular network proposed in [27], where the network connects a MS with others in the vicinity through multiple hops to the BS. In this scenario, the transmit power could be largely reduced and the cell coverage can be well extended. Such network can be supported by direct mode LTE communication (LTE D2D) [33] that connects multiple UEs in which there is no eNB coverage. The FP7 ABSOLUTE proposes UE clustering techniques, which allows a cluster head UE connect an Aerial eNB with several adjacent UEs [34]. The UE clustering architecture effectively reduces power consumption in the access network and extends the UEs' battery life.

Radio Resource Management in the mobile ad-hoc network has severe challenges. The UEs are usually implemented with omni-directional antennas, which cause excessive interference to each other and largely constraints the data rate. Moreover, the mobility of UEs results in a constantly changing interference environment, which makes the network highly unstable. In Chapter 4, RRM on a multi-hop backhaul network with "tree" topology has been investigated, which can be easily extended to a fully ad hoc/mesh network. This will be discussed in Section 8.3.

2.3 Radio Resource Management

Radio resource management (RRM) is the system level control of co-channel interference and other radio transmission characteristics in wireless communication systems [35]. The objective is to effectively utilize the available spectrum resources for data transmissions. There are various components in RRM, such as handover, channel allocation, power control, etc. This thesis particularly focuses on the aspect of allocating radio resources (time slots, frequency channels, etc.) to wireless users as well as providing system capacity.

RRM has been widely investigated for decades, with a number of algorithms developed [36]. A major target is to maximize the number of users and the volume of data traffic that a system can support. Multiple access techniques represent the baseline of RRM that allows multiple users to connect to the same network. Spectrum management is then carried out to divide and assign a shared spectrum band to multiple users for data transmission. In the following sections, the RRM algorithms are categorized as Frequency band Allocation (FP) and Dynamic Spectrum Access (DSA).

2.3.1 *Multiple Access Techniques*

Multiple access techniques allow multiple users to share the capacity provided by a spectrum band. It is based on a multiple access protocol and control mechanism, namely media access control (MAC). In this section, we categorize various multiple access techniques into channelization schemes and random access schemes.

2.3.1.1 Channelization Schemes

Channelization has been widely applied in wireless cellular systems. In this scheme, the entire spectrum pool is divided into multiple channels in various forms. Channels are assigned not only to multiple users but also to multiple links between transceivers (multiplexing). In a data packet network, a wireless link may have multiple channels assigned simultaneously, because there could be multiple data packets and relayed traffic in transmission simultaneously.

There are four fundamental channelization techniques developed for multiple access or multiplexing [4]:

- Frequency Division: the available spectrum band is divided into several distinct frequency ranges to provide multiple channels. The users are allocated with several frequency channels for transmission. In this scheme, adjacent channel interference exists between neighbour sub-bands, where a guard band is used for channel separation.
- Time Division: the resource pool is divided into several time slots for multiple access. A user can utilize a wider frequency band while time slots should be well synchronized between users. However, transmission delay can cause serious inter-symbol interference. Guard bands are thus necessary between time slots to protect neighbour symbols.
- Code Division: spreading codes are employed to divide signals for multiple users. In this scheme, a spread spectrum technique is used to allow any user to utilize the entire radio spectrum in both time and frequency domains. However, an effective power control algorithm is vital to provide sufficient SINR for users in different locations.
- Space Division: directional antennas are employed to connect users in different locations. Interference is controlled by negative gains on antenna sidelobes. However, narrow beam antennas result in an increased antenna size, which is difficult to implement on MSs and small cell BSs. This scheme is thus usually applied on backhaul links between BSs.

Many current and future communication systems use a mixture of these techniques. OFDMA is widely applied or proposed in 4G and WiFi systems, which defines Resource Blocks in both time and frequency domains [37]. The FP7 BuNGee project also implements directional antennas on the backhaul network, which use both space division and OFDMA techniques [12]. A comprehensive study of the space division scheme on the multi-hop backhaul network will be provided in Chapter 4.

Channelization schemes deliver a contention-free system, where interference is a major issue on multiple users reusing the same channel in the domains of time, frequency, etc. Furthermore, spectrum utilization is a big issue when a user does not require continuous data transmission at all time but only need to use channels occasionally.

2.3.1.2 Random Access Schemes

Random access schemes are widely used in many wireless networks to provide distributed multiple access and flexibility of utilizing the resource pool. ALOHA is a basic random access scheme that allows multiple users to transmit on a common channel. Collisions occur when users contend for the same time slot and random back off is carried out for retransmission. The simplicity of ALOHA makes it a promising technique for the networks that require minimum implementation overheads to save energy, such as Wireless Sensor Networks (WSNs). The intelligent ALOHA protocols with reinforcement learning have been studied in [38, 39] as effective techniques for WSN. This enables a distributed node to learn to avoid collisions with others on the same slot, which thus improves QoS.

Carrier sense multiple access (CSMA) is a more reliable random access scheme that introduces carrier detection before transmitting data packets [40]. The IEEE 802.11 standard implements CSMA using RTS/CTS (Request to Send/Clear to Send) mechanism. A node wising to transmit data will firstly broadcast a RTS frame to the nodes in the vicinity. The destination node replies with a CTS frame. Any other nodes receiving RTS or CTS frames avoid sending data for a given time. The transmitter then starts to send data packets. The receiver replies with an ACK (Acknowledgement) frame when packets are delivered. A packet without an ACK reply in a given time will be considered a lost packet. Various retransmission schemes have been developed for resending the lost packets, including 1-persistent: the transmitter continuously detects the channel and sends data once it is free; P-persistent: the transmitter send data on idle channels with a probability of p; and non-persistent: the transmitter back off the lost packet and wait for a random time to resend [41]. The 1-persistent technique is effective at low traffic loads but may cause excessive collisions at high traffic load, where non-persistent is applied instead.

The random access schemes provide effective ways for multiple access in a distributed manner. However, ALOHA and CSMA are both contention-based system which cannot guarantee reliable QoS. ALOHA schemes are constrained by the random arriving behaviour of data packets. The hidden and exposed node problems can hardly be detected in CSMA [40]. This thesis will mainly focus on cellular communication systems, where channelization schemes are generally applied. In the next two sections we will describe the major methodologies of spectrum assignment.

2.3.2 *Frequency band Allocation*

The Frequency band Allocation (FA) mechanism is widely applied for spectrum management in most of the current cellular communication systems [42]. FA divides the radio spectrum into several distinct frequency bands. A base station is allocated with a frequency band that contains a set of channels to be assigned to radio links on the access or backhaul network. The FA strategy can be carried out in a centralized, coordinated, or distributed manner. However, the common feature of them is that a BS has a fixed size spectrum pool. In this scenario, the network capacity is more constrained by the bandwidth of allocated frequency bands rather than interference.

In this section, we will firstly review different FA strategies categorized by frequency patterns, including cell, zone and antenna based schemes. The operating mode of FA is then presented. Finally the spectrum utilization issue is stated followed by a channel borrowing scheme.

Frequency Planning and Cell Clustering

Frequency planning (FP) is used in most of the FA strategies to mitigate inter-band interference. A typical FP strategy is the clustering algorithm used in the GSM system [42]. This algorithm defines a cluster as a set of adjacent cells that includes all frequencies. The cluster members (BSs) are allocated with different frequency bands in order to avoid inter-cell interference. The same frequency pattern is applied to all clusters in the network. Two cells in a neighbour cluster share the same band [4]. The shape of a cluster could be hexagonal, straight line, square, etc., depending on the location and coverage of the BSs.

Cluster size, namely the number of BSs in a cluster, is a crucial parameter that determines the spectrum efficiency. A smaller cluster size means a larger number of clusters exist in the system. In this case, the bandwidth in each cell is wider because of better frequency reuse capability. However, interference in a small cluster network could be very high due to the short distance between multiple cells sharing the same band. The Shannon equation [4] indicates that link capacity is determined by bandwidth and interference in a given transmitter power and propagation loss. As a result, the cluster size should be carefully designed to increase frequency reuse and to reduce interference.

Fractional Frequency Reuse

FA can also be carried out on fractional zones within a cell. 3GPP LTE proposes the enhanced Fractional Frequency Reuse (FFR) scheme for inter-cell interference coordination in a OFDMA HetNet [43]. It is designed as omni-directional and sectorized schemes based on antenna patterns of eNBs, which is shown in Figure 2.1.

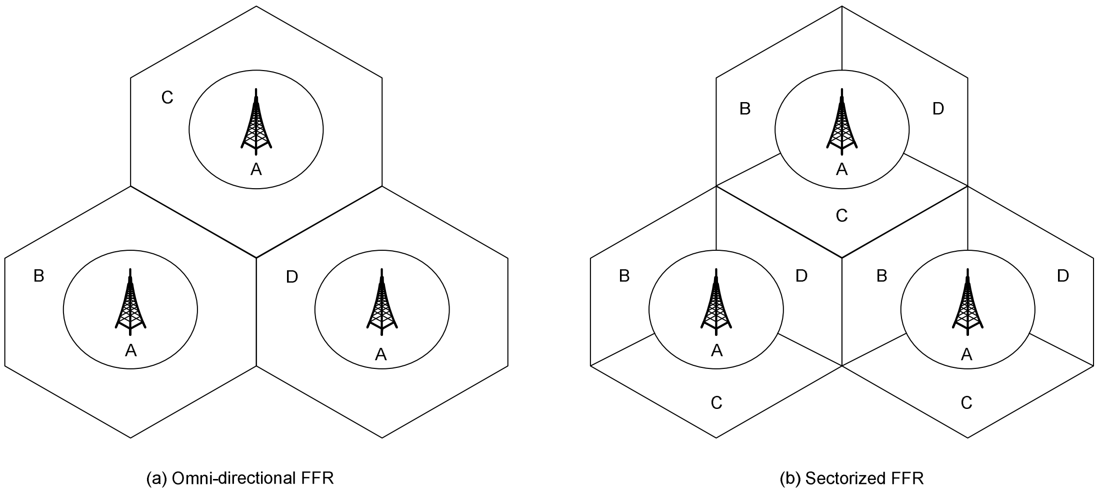

Figure 2.1. Fractional Frequency Reuse

In a small cell with omni-directional antennas, FFR divides a cell into inner and outer zones with different frequency bands allocated to each. The users in the inner zones of adjacent cells can reuse the frequency band. The cell edge users in the outer zones receive interference from neighbour cells, where cluster based FA is applied. This scheme achieves higher system capacity than conventional cluster based FA, because the inner zone has fewer constraints from spectrum division. In LTE the omni-directional FFR scheme is proposed for Pico or Femto cells.

A sectorized FFR scheme has been designed to include all frequencies in a cellular area. In this scheme, three sectorized antennas further divide the outer zone into three sector zones, with separate bands allocated to each. The inner zone can use all frequencies as proposed in omni-directional FFR. The total bandwidth of a cell is maximized to the size of the spectrum pool. The sectorized FFR allows a macrocell to use the entire spectrum pool with mitigated inter-cell interference.

Multi-beam Frequency Planning

We have so far illustrated possible FA strategies on cells and zones. However, some new network architectures have been implemented with directional antennas to establish wireless links. A typical example is the heterogeneous mobile broadband network proposed in the FP7 BuNGee project [12]. The access network is constructed by four directional antennas on ABSs, covering different street areas. The backhaul network is constructed by multiple directional antennas on a HBS, connecting a number of ABSs in a square area. A special FA strategy has been designed based on antenna beams. In this scenario, four frequency bands are allocated to different antenna beams on ABS and HBS. In the access network, an ABS assigns four different frequency bands to four antenna beams covering different streets. Neighbour ABSs are coordinated to avoid the same band used by antennas covering the same street. In the backhaul network, four adjacent antennas are categorized in a group with different bands allocated to each.

The cell-based, zone-based, and antenna-based FA strategies follow the same principle in which frequency bands are used to divide the spectrum. It is highly effective to avoid interference in scenarios where the network architecture is fixed and the FP scheme is carefully designed. For the dynamic network architecture scenarios introduced in Chapter 6 and Chapter 7, adaptive allocation of frequency bands is desired to control interference in different network topologies. A novel dynamic FA strategy is introduced in Chapter 7 to handle this problem.

Protocol Architecture

The architecture of a FA strategy can be centralized, coordinated and distributed. Conventional 2G systems use centralized FA, where the RNC is responsible for planning and allocating frequency bands to various cells through the S1 interface [36]. In LTE systems, the Inter-Cell Interference Coordination (ICIC) strategy is introduced [13], where an X2 interface is employed to exchange control information between eNBs [44]. The frequency band information can be exchanged through X2 links to achieve band separation between neighbouring fractional zones. The degree of information exchanged on X2 links in ICIC is less than that on S1 link in the centralized strategy. The coordination overhead issue occurs only when the network topology is rapidly changeable, where dynamic FA is essential.

Spectrum Utilization and Channel Borrowing

The main issue of the FA strategy is its efficiency of handling traffic dynamics in both time and spatial domains [36]. The number of channels provided by FA to a cell, zone or antenna is fixed to the band size, which constrains the maximum number of users that can be supported. Traffic density is highly dynamic in different time and locations [11]. The uniformly assigned frequency bands are not able to accommodate the dynamics of traffic, which causes users to be blocked in the system according to queuing theory [45]. Spectrum bands cannot be fully utilized in the whole network, where more cells are required to keep adequate QoS. This causes a significant waste of spectrum resources and energy.

A channel borrowing scheme has been proposed in FA to accommodate a non-uniform number of users in different cells. In this context, one cell can borrow free channels from neighbour or adjacent cells when its allocated band is fully utilized. There are two types of borrowing schemes: one is that all the channels in a band can be borrowed for temporary use in other cells; the other is that some channels in a band will be locked for use only in their allocated cell, and the rest of them can be lent out [42]. The channel borrowing schemes can reduce blocking probability in FA by dynamically scheduling radio resources to some extent. However, this may cause overlap between bands which destroys the original FP. It is thus difficult to handle interference in the channel borrowing scheme.

2.3.3 *Dynamic Spectrum Access*

Dynamic Spectrum Access (DSA) is a promising technique under research in the recent years. It is usually applied for RRM in a Cognitive Radio (CR) network [8]. There is a common belief that radio spectrum is suffering a high level of scarcity in recent years [46]. The high speed data rate systems and a dramatic growth in the number of users require significantly greater spectrum than before. Conventionally radio spectrum is assigned or auctioned to operators by regulatory authorities. However, the free unoccupied spectrum is insufficient for mobile broadband networks. A typical example is the deployment of LTE network in many countries. The 800 MHz spectrum band provides comprehensive coverage in LTE, because the lower frequency has better resistance to propagation loss [47]. However, 800 MHz band is the UHF band allocated to analogue and digital TV transmission in many countries. In order to transfer this band to LTE system, Ofcom in the UK has to clear this band and reallocate other spectrum for digital TV stations [48]. Similar actions are carried out in other countries. Furthermore, LTE in 2.6 GHz band suffers serious adjacent channel interference from neighbouring bands, where additional guard bands are needed to provide sufficient separation. These examples indicate that frequency band allocation mechanism is inefficient in supporting high speed communication systems.

Despite the scarcity of frequency bands, spectrum utilization is extremely low in current wireless systems. A study of spatial and temporal spectrum usage in [49] indicates that the spectrum is not used all the time and that the usage depends on location. The main reason for this is the capacity of frequency bands is highly inflexible, which cannot support dynamic user traffic. As a result, the concept of dynamic spectrum access is proposed to assign channels for opportunistic and occasional access. In this manner, the spectrum is expected to be well utilized and the system capacity can be maximized.

Dynamic Spectrum Access Scenarios

DSA is conventionally designed to allow opportunistic “secondary users” (SU) to access the licensed spectrum occupied by “primary users” (PU), namely Opportunistic Spectrum Access (OSA) [9]. The PUs are guaranteed to have reliable QoS and have priority in using the spectrum. The SUs can identify the spectrum holes that are not currently occupied by the PUs and transmit on related channels. Moreover, the SUs should release channels when requested by PUs. This mechanism requires few changes to existing wireless devices in licensed bands. However, it is unrealistic at the current stage for operators to release their licensed spectrum for other uses, because spectrum is one of the most important resources in attracting users and also in some countries there is tremendous cost in purchasing spectrum bands. Furthermore, the PUs are concerned with potential interference and greedy usage from SUs. On the other hand, reliable QoS cannot be guaranteed for SUs, which could be less attractive in the market.

A more realistic DSA can be carried out between multiple wireless entities in a single or heterogeneous network within the same spectrum band held by one operator. Channel allocation can be carried out by either BSs or MSs, though BS based DSA means few changes to existing MSs. In DSA, a common spectrum pool is opened to all BSs or MSs in the network, meaning that all channels can be dynamically assigned to links when traffic arrives and released when transmission finished [50]. The network capacity is constrained by co-channel interference rather than bandwidth, because the system is lacking preliminary FP.

The following part of this section introduces conventional implementation methods of DSA, including Radio Environment Map, Spectrum Sensing and Inter-entity Coordination.

Radio Environment Map

A Radio Environment Map (REM) employs a dynamic database for spectrum management, which contains the information of BS locations and spectrum usage [51]. A BS wishing to assign channels will firstly search the database for empty channels, and update the database after occupying or releasing the selected channel. The database is maintained at a central server but updated dynamically by distributed BSs. This scheme provides up-to-date information of spectrum occupancy and effectively controls interference. However, control information exchanged between distributed BSs and database could be excessively high. Moreover, the database could be very large and difficult to manage, when there is a large number of users or a high volume of data traffic. Storage of such large databases remains an issue.

Spectrum information in REM can also be updated through spectrum awareness. The FP7 FARAMIR project [52] has done comprehensive research in sensing technologies, database storage, resource management and system architectures. The information overhead of REM has been analysed in [51]. Further research will be carried out in the FP7 ABSOLUTE project [33] by using transfer learning technology to reduce cooperation overhead in the REM architecture. The REM with a spectrum database is a standardized technology in IEEE 802.22 Wireless Regional

Area Network (WRAN) [53] and ETSI draft [54] for TV White Space wireless access. Related consultations and research have been carried out by Ofcom in the UK.

Spectrum Sensing

Spectrum sensing has been widely investigated for DSA in cognitive radio networks [55]. The aim of spectrum sensing is to provide wireless users with information of unoccupied frequencies. Interference measurement is the fundamental technique used to evaluate channel quality prior to data transmission. The spectrum sensing module scans the frequency band by gathering the interference power level on each channel using energy detectors. An interference threshold is then set at the sensing entity, to decide whether a channel has sufficient SINR for data transmission [56]. Spectrum sensing can be carried out on either transmitters or receivers, though hidden node problem may occur in both schemes. Transmitter based sensing may not be able to identify potential interference node near the receiver, while receiver based sensing has the issue of selecting channels occupied by the users near the transmitter. Interruptions may occur on either local or neighbor links. Sensing a large number of channels causes long sensing delay and high power consumption. In Chapter 4 reinforcement learning algorithms are developed to improve the system performance with a minimum level of spectrum sensing.

Channel based Interference Coordination

Distributed channel based interference coordination is another potential approach to DSA. This allows BSs to exchange information of channels rather than frequency bands used in conventional ICIC. In this scheme, a BS selects a channel based on the channel usage information from neighbors. Interference can be avoided between coordinated entities. This scheme significantly improves spectrum utilization compared to conventional ICIC. However, it may cause a high level of coordination overhead. Channel usage information should be exchanged over X2 link prior to every data transmission, whereas in ICIC scheme such process is only required when operating initial frequency planning. Interference coordination will be investigated in Chapter 4 together with spatial channel reuse on a multi-hop backhaul network. Furthermore, learning techniques will be developed to reduce coordination overhead.

2.4 Cognitive Radio Techniques

2.4.1 *Cognitive Networking and Cognitive Radio*

Cognitive Radio (CR) is a paradigm for wireless communication, which emerges from Software Define Radio (SDR), Dynamic Spectrum Access (DSA) and Distributed Artificial Intelligence (DAI). CR techniques enable a wireless node to intelligently change its radio parameters to adapt to the dynamic radio environment [57]. SDR is the baseline component in CR that supports dynamic adjustment of radio parameters, including transmit power, channel, AMC scheme, FEC scheme, etc. DSA is the target of CR that gives flexible utilization of radio spectrum. Furthermore, DAI is the most important part in CR, which provides decision making capability on CR agents. CR is not a technique in a specific protocol layer but a group of technologies that constructs an intelligent radio system. A CR agent, either BS or MS, has the capability of dynamically changing radio parameters (making decisions) and implementing them through SDR (taking actions).

The terminologies of DSA, SDR and CR are synonyms in some definitions [58]. However, the key philosophy that differs CR from others is that a CR agent can observe the outside world, learn the decisions and obtain feedback from actions taken. The ability for intelligent decision making based on previous and current actions is the core research area in cognitive radio, which is not supported in other radio systems [59].

Cognitive radio can be extended to the system level scope as a cognitive network [58, 60], which introduce the intelligent decision capability to cross-layer designs. In the physical layer, cognitive network can be employed for effective spectrum sensing [55]. The MAC layer uses cognitive techniques to provide multiple access, interference management, etc, which will be the major research topic in this thesis. In the network layer, cognitive routing has been studied to provide reliable end-to-end connection especially in ad-hoc and mesh networks [61], which will be discussed in Chapter 8. Transport layer protocols for cognitive network have been investigated in [62], which provides end-to-end QoS and throughput via effective utilization of link capacity. Cognitive techniques can be applied to multiple OSI layers, which establish an intelligent radio system. In this thesis, we focus on the Radio Resource Management aspect of cognitive network.

The target of RRM in CR is to achieve effective DSA, which is supported by DAI algorithms implemented in the learning engine. SDR is responsible for taking actions made by CR, which enable the transceivers to operate in different parts of the radio spectrums and communication protocols [57]. A typical cognitive radio follows the cognitive cycle shown in Figure 2.2 with four engines: observation, decision making, action taking and learning.

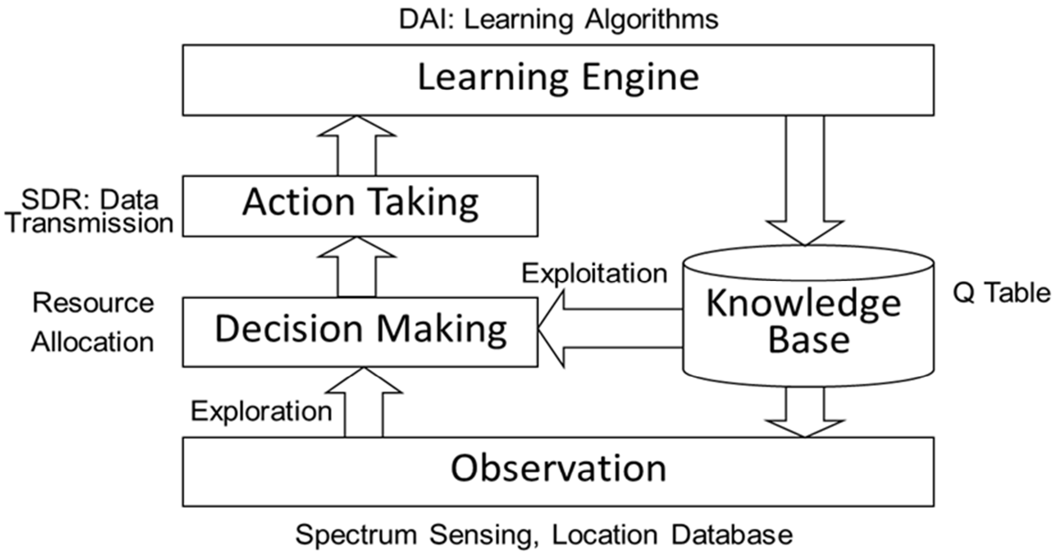

Figure 2.2. Cognitive Cycle

Observation is operated to obtain spectrum information from surrounding area, which takes place before data transmission. Typical techniques in observation are spectrum sensing, radio environment map, random exploration, etc. Observation provides instantaneous knowledge of the scenario, but is not essential in every cognitive cycle. In practical systems it is usually carried out occasionally to assist with the decision making process, because a continuous observation may destroy convergence. Moreover, observation techniques may cause large overheads, including delay and energy cost in spectrum sensing, coordination and database cost in REM. On the other hand, random exploration may cause harmful decisions. The learning engine in cognitive radio is aimed at reducing the level of observation. The impact of spectrum sensing and random exploration will be examined in Section 4.4.

In the *decision making* process, a cognitive agent selects channels for data transmission, namely channel selection. It is illustrated in Figure 2.2 that there are two ways to obtain information for decision making: from the *knowledge base* and from *observation*. A cognitive agent can either *exploit* historical learnt information from the knowledge base, or *explore* instantaneous information from observation [63]. Exploitation provides generalized, averaged and long term knowledge on the channel quality from previous actions. However, environment changes may not be identified quickly from exploitation, because it takes several iterations to train the knowledge base. In this case, exploration is designed to provide external knowledge to decision making and reinforce the knowledge base. A two stage cognitive cycle is proposed in [64], which starts with exploration that acquaints the agent with the radio environment. Exploitation is then operated on the second stage. This strategy is effective in a static scenario. However, the dynamics of user traffic and network topology require observations carried out during environment changes. The ε-Greedy method is developed to periodically operate exploration at a defined probability ε, in order to investigate potential environment changes [65]. Multi-agent cooperation is also a promising technique that assists with decision making, which will be discussed in later chapters.

The *action taking* process refers to data transmission on a selected channel. The SDR module sets up the radio parameters based on information from decision making. The objective is to make the decisions converged on a fixed set of actions. The feedback of an action, including success or failed transmissions, will be transferred to the learning engine.

The *learning engine* is the core module of cognitive radio that acts as the "brain" of a radio system [59]. The role of learning is to train the *knowledge base* that stores experiences of decision making, which is supported by DAI algorithms. There are a large number of learning strategies developed in the computer science society, which can be categorized as single and multi agent learning. A wireless network is a multi-agent environment. The target of learning is to partition frequencies to multiple agents in a distributed manner. This topic will be extensively reviewed in the next few sections.

2.4.2 ***Reinforcement Learning***

Reinforcement learning is learning that maximizes a numerical reward signal. The methodology is to discover which actions yield the most reward by trying them. The characteristics of reinforcement learning are trail-and-error and delayed reward [66]. The implementation scenario of reinforcement learning is the Markov Decision Process (MDP), where a learning agent interacts with its environment to achieve a goal. The agent should observe the state of environment and take actions that affect that state. Moreover, a goal must be introduced relating to the state of the environment.

Reinforcement learning is well suited to cognitive radio, where the action of data transmission interacts with the radio environment and the goal is spectrum separation. There are thus a number of studies on applying reinforcement learning to intelligent spectrum management [20, 64, 67, 68]. A reinforcement learning model has a set of possible states S, a set of actions A, and a set of numerical rewards R. The learning cycle is a state-action-reward process. On a learning iteration t, an agent takes an action $a \in A$ that interacts with the environment. The agent goes into a state $s(t) \in S$ and receives a reward $r(s(t)) \in R$. The objective is to select actions a at each state s, based on maximized reward r. Given a selection policy π, this process is denoted as $a = \pi(s)$.

In the action-value function approach of reinforcement learning, a Q table is setup in every state with elements representing each action. The Q value determines the priority of an action to be selected. Under the policy π, the action-value of a state-action pair (s, a) is defined by

$$\begin{aligned} Q_\pi(s,a) &= E_\pi\{R_t \mid s_t = s, a_t = a\} \\ &= E_\pi\{\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \mid s_t = s, a_t = a\} \end{aligned} \tag{2.1}$$

where R stands for random return which is associated with first taking action a in state s following policy π. The returned value is discounted by γ on each state. A Monte-carlo method is generally involved in averaging values over many random samples of actual returns. In a communication system, this could be a large number of data packet transmissions. The accuracy of Q in a static environment depends on the number of iterations taken in pair (s, a).

The goal of solving a reinforcement learning task is to find a policy that achieves a high reward over the long run. The action-value function (2.1) defines a partial ordering over policies. For finite MDPs, the improved policy can be defined based on high order Q values:

$$Q^{*}(s,a) = \max_{\pi} Q^{\pi}(s,a) \tag{2.2}$$

In a channel assignment scenario, a channel with the highest Q value that is not currently occupied will be selected.

There are many implementation algorithms of reinforcement learning. A typical example is Q learning developed to find an improved action-selection policy for finite MDPs. Initially Q returns arbitrary values $Q(s_0, a_0)$ chosen by the designer. Then each time the agent selects an action and receives a reward in a new state, the Q table is updated based on rewards from the previous state and the selected action. The action-value function is defined as [66]

$$Q_{t+1}(s_t,a_t) = (1-\alpha_t(s_t,a_t))Q_t(s_t,a_t) + \alpha_t(s_t,a_t)[R_{t+1} + \gamma \max_{a} Q_t(s_{t+1},a)] \tag{2.2}$$

where $\gamma \in [0,1]$ is the discount factor that trades off the importance of current and previous states. $\alpha \in [0,1]$ is the learning rate that balances the proportion of historical and instantaneous information, namely the speed of convergence.

Q learning based dynamic channel assignment for a cellular network has been studied in [68] in a centralized manner. Distributed Q learning for interference avoidance has been investigated in [20], in the scenario of a self-organized femtocell network. [19] applies Q learning to power allocation in a wireless mesh network, which is aimed at reducing transmit power consumption. Furthermore, [69] uses Q learning to improve SINR through improving power allocation on distributed cognitive BSs.

Conventional Q learning algorithms have multiple states defined in the system. However, it may be difficult to find states in some scenarios of wireless systems.

This issue occurs particularly in the distributed DSA scenario. Research in [70] proposes the concept of single state Q learning to be applied in the single state scenarios. It uses the same action-value function of multi-state Q learning (2.2) by setting the discount factor γ to 0. The Q value is then updated solely on a single state. Convergence of single state Q learning will be proven in Chapter 4.

2.4.3 *Multi-agent Cooperation*

Distributed reinforcement learning relies on trial-and-error and delayed reward to establish the knowledge base. In the situation where an agent has limited knowledge of the radio environment, arbitrary decisions will be taken and cause harmful impact to others, which in turn reduces QoS. Meanwhile, the delayed reward feature of reinforcement learning results in a large number of iterations to find improved decisions, especially for complicated learning problems.

Multi-agent cooperation has been proposed in many papers as an effective approach to improve distributed reinforcement learning, mainly in two aspects: improving selection policy and speeding up convergence [70]. This is because in a multi-agent environment, the actions taken by one agent interact with others. There are multiple nodes in a wireless network sharing a common set of radio resources. Competition between these nodes can be described through game models [21]. For example, power allocation can be modelled via game theory, because a signal radiated from one user interacts with others in a common interference environment [71]. The target of each individual user is to increase SINR through either selecting high quality channels or increasing transmit power. However, these two actions both increase the interference seen by other users in a finite resource pool, which reduces their SINR. The expected resource utilization policy is that every user has the highest SINR gain in an interactive environment [21, 71].

Multi-agent cooperation is designed to allow a distributed agent to learn behaviours from other agents [70]. The main idea is to exchange information between multiple agents, including location axis, radio parameters, knowledge base, etc. There are various forms of multi-agent cooperation, such as swarm reinforcement learning [72], cooperative game theory [21] and docition [73-77]. A study on the trade-off between independent and cooperative agents can be found in [78]. This work shares the observation results with learnt policies. Analysis and experimental results show that information exchange is beneficial if it can be used effectively, which in turn speeds up learning at the cost of coordination overheads.

Docition is an emerging paradigm in a cognitive radio network employing multi-agent cooperation [73-77]. The philosophy of docition is to enable naïve agents to achieve expert knowledge from mature agents [74]. Conventional reinforcement learning is operated on individual agents. The knowledge base (Q table) is transferred from agents with better performance to those with worse performance. It is proposed in [74] that start-up docition is used to assist newly activated agents and adaptive docition is applied for further performance improvement. Docition has been developed for power allocation on a PU-SU based WRAN network.

Docition is designed to transfer knowledge from mature to naïve agents. However, a practical network could be more complex than this. In the scenario where a new network is initially deployed, the agents may not be able to find anyone with mature knowledge. Moreover, the activation and deactivation of any agents in a network may have serious impact on others in vicinity. The knowledge transfer process should not be limited to teaching new agents.

2.4.4 *Transfer Learning*

Transfer learning is a machine learning technique that focuses on applying knowledge learnt from one problem to a different but related problem [79]. Many machine learning algorithms assume that the agents always stay in the same domain of interest and learn a single training task. However, there are many different tasks in a practical scenario. For example, a cognitive agent could have multiple objectives when moving between different networks or radio environments. When the task changes, most distributed learning algorithms need to rebuild the knowledge base from scratch using newly collected rewards from a trail-and-error process. In a rapidly changeable network, it is difficult for a distributed learning algorithm to quickly train the knowledge base and adapt to each specific environment. The agent has to makes random decisions that could be harmful to QoS.

Transfer learning is developed to improve learning in the target task by transferring knowledge from related source tasks [79]. A learning agent firstly finds some source tasks that have potential impact to the target task. They could be the tasks learnt in the past or on other agents. The knowledge base from multiple source tasks is then transferred to the learner's target task. Finally the agent trains knowledge base with appropriate algorithms. Figure 2.3 [80] compares the process of traditional machine learning and transfer learning.

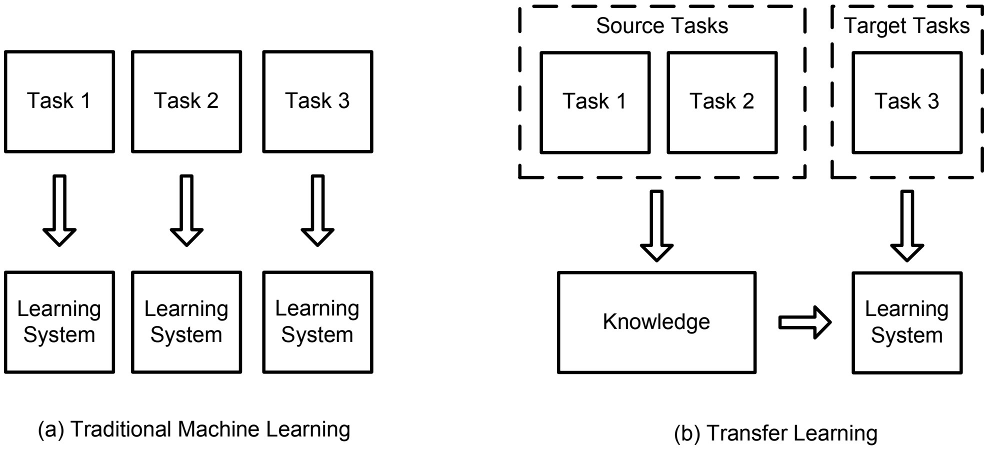

Figure 2.3. Learning process illustration [80]

It is illustrated that transfer learning is not designed to replace traditional learning algorithms, but acts as a supplement to the learning systems on different tasks. The application of transfer learning to reinforcement learning is investigated in [81]. Experiment results show that with transfer learning the agents learn significantly faster. It takes fewer episodes for transfer learning to achieve stable states than reinforcement learning.

The idea of transfer learning is perfectly suitable for resource management in a wireless network. Knowledge transfer between tasks on multiple agents is studied in Chapter 5. Furthermore, the learning task is modelled at a network level in Chapter 6, where knowledge transfer is applied to network changes.

2.5 Energy Efficient Wireless Networks

Energy consumption of wireless networks has become an important research topic in recent years, as CO_2 emissions are a serious environment issue, which may constrain economic development in future. Wireless networks require electricity to operate. However, current design of wireless BSs is particularly poor in terms of energy efficiency. An increasing number of BSs in a high capacity density network could cause significantly more energy consumption [82].

It has been reported that power amplifiers and air conditioners consumes two thirds of the total energy in a wireless network, whereas the data transmission unit consumes only less than 15% [82]. There are a number of research papers focusing on transmit power reduction for energy efficiency. However, this obviously does not reduce the main energy consumption of a wireless network.

Deployment of low power base stations is thus very important in the future networks. The SCN has great advantages in reducing energy consumption as well as enhancing system capacity [83]. The small cells are supported by very light weight base stations that are constructed by low power components. Moreover, flexible and dense deployments of small cell BSs significantly reduce the transmission distance between BS and MS. In addition, the antenna height of the small cell BSs is very low, which significantly reduces shadowing effect on access links. The reduced transmission distance and path loss make low power transmitter possible for high capacity provision.

Network capacity and energy are contradictions in wireless communication systems [84]. Despite BSs in SCNs are implemented with low power components, [82] shows that a large proportion of energy is used to keep the BSs active, such as the cooling systems and power amplifier [84]. There are some recent research papers on dynamic network planning based on traffic patterns [85-87]. The purpose of deploying a BS is to provide adequate capacity in its coverage area. However, user traffic in a cellular network is usually inconsistent and non-uniform. It fluctuates in both time and spatial domains. Figure 2.4 shows the traffic profile of 5 cells in a metropolitan area obtained from Ofcom [88].

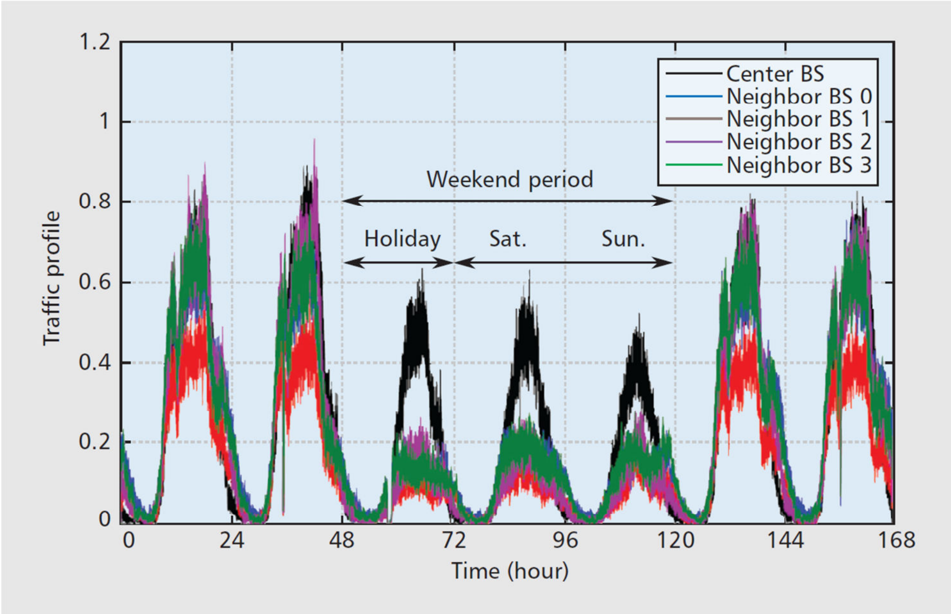

Figure 2.4. Traffic dynamics in time and spatial domains [11, 89]

It can be observed that the average user traffic trends between daytime and evening time, or between weekdays and weekends, vary greatly. Furthermore, the traffic also varies to a large extent in different hours and cells. For example, the peak traffic levels are 50% to 90% higher than low traffic levels. The centre BS has twice the traffic load than others in the weekend.

Dynamic network planning aims to effectively control the number of active base stations according to traffic variations. An excessive amount of energy can be saved by only activating a minimum number of base stations that provides sufficient system capacity. Figure 2.5 illustrates a dynamic network planning paradigm based on system traffic intensity.

A key issue in energy efficient network planning is the time and energy required for switching on or off the base stations. However, there are not many statistical data in this area, mainly because this technique has not been widely applied in practical systems. On the other hand, the definition of "sleep mode" varies with different networks and operators' requirement. Energy models of various types of LTE base stations have been studied by FP7 EARTH project in [90], where the time and power consumption of switching a BS to sleep mode is a further research. The design of dynamic network planning thus should consider this effect, by preventing a BS from switching too frequently between active and sleep modes. This can be achieved by dynamic load management mechanism proposed in Chapter 8.

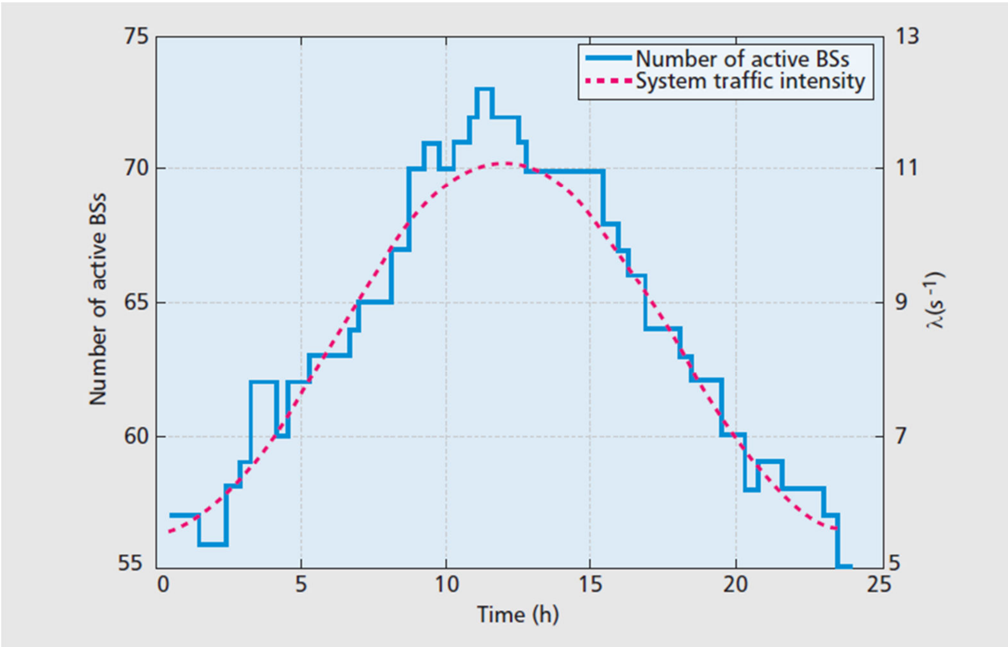

Figure 2.5. Traffic Aware Network Planning [11]

Effective radio resource management is an important aspect for traffic aware network planning. System capacity is determined by spectrum utilization in a finite spectrum pool as discussed before. Better capacity provision can be achieved through improved spectrum utilization without activating new BSs, which in turn saves energy. Capacity enhancement is thus a predominant aspect in network topology management, which will be investigated in Chapter 7.

2.6 Conclusion

This chapter has reviewed background information related to the topic of intelligent radio resource management in high capacity density wireless networks. Research work on dense small cell networks has been given, as a promising architecture to provide broadband wireless access. In this network, system capacity can be enhanced through improved signal power and capacity provision. Wireless backhaul network have been reviewed followed by discussions on single and multi-hop topologies, which provide flexible deployment capability in small cell base stations.

Spectrum management strategies of wireless cellular network have been extensively reviewed, from conventional frequency band allocation to novel dynamic spectrum access. Various typical multiple access and channel allocation schemes have been discussed from the perspectives of capacity provision, spectrum utilization, complexity, operating modes, control information overheads, etc. The RRM requirements in a high capacity density networks have been discussed.

Cognitive radio technology has been illustrated as observation, decision making, action taking, learning engine and knowledge base. It has been proposed as an effective technique to achieve dynamic spectrum access. Conventional reinforcement learning models and algorithms have been reviewed with their application to resource management including channel and power allocation. Multi-agent learning algorithms have been proposed to improve both QoS performance and convergence in distributed reinforcement learning. Finally, research on transfer learning has been reviewed, which has been shown as an effective approach to balance the QoS performance and multi-agent cooperation, and to improve network reliability in dynamic radio environment.

Furthermore, energy efficiency of wireless network has been discussed. The electricity components of base stations consume most of the energy in a network. Related work on dynamic network planning based on traffic patterns has been reviewed as a solution to this problem. The benefit of capacity enhancement on dynamic network planning has been discussed, which could be supported by effective resource management.

Chapter 3. System Modelling and Verification Methodologies

3.1 Introduction

The purpose of this chapter is to introduce simulation and modelling methods for the dense capacity wireless networks, which are major techniques used for performance evaluation in this work. The model needs to be accurate enough to capture the relevant detail of a representative scenario. This chapter will present the assumptions and parameters used in the model, as well as the method of simulation.

The dense capacity wireless network considered in this thesis is an integration of a Multi-hop Backhaul Network and a Dense Small Cell Access Network. These two networks are designed to operate in different spectrum bands, which will be studied separately. Section 3.2 introduces the entire network architecture, and illustrates the models for wireless backhaul and access network, respectively. The simulation tool selected in this thesis is presented in Section 3.3. Detailed models for various aspects of the wireless system are stated in Section 3.4. In Section 3.5, the output parameters used for performance validation are presented. Verification methods for theoretical analysis are demonstrated in Section 3.6.

3.2 Network Architecture

In this thesis, the dense capacity wireless network is considered to be a construction of two networks: a Multi-hop Backhaul Network and a Dense Small Cell Access Network. These two networks are operated on individual spectrum bands to prevent interference. As a result, they will be investigated, modelled and analysed separately.

The system is derived from the architecture proposed by the FP7 BuNGee project [12], which was aimed at providing 1 Gbit/s/km^2 capacity density in an urban area. It has been suggested in this project that such capacity density requirements can be provided by a dense deployment of small cell Access Base Stations (ABSs) providing an access network to high density mobile users, as shown in Figure 3.1. The ABSs are designed as portable, light-weight devices, which can be densely deployed and easily managed. The major role of ABSs is to provide extremely high data rate to Mobile Stations (MSs) on a street level (where indoor services are not considered). In this case, a below rooftop deployment of ABSs (e.g. on street lamps) is proposed in [12] to mitigate interference between streets, by using the shadowing effect from buildings.

To achieve such a network cost-effectively, one option is to backhaul the offered traffic via multi-hop links connecting a line of ABSs to a Hub Base Station (HBS). Thus the role of the HBS is to provide backhaul connections to an operator's core network, rather than connecting with mobile users directly. An important feature of this backhaul network is that the multi-hop links are provided by directional antennas on each ABS and HBS, which substantially reduces interference, and improves the link budget. Moreover, spatial resource reuse on directional antennas can significantly improve the network capacity, especially for a multi-hop network.

Figure 3.1 illustrates the system architecture of the high capacity density network. The system is proposed to be highly self-organized, where the ABSs are entitled to establish and manage both the access and backhaul networks. The HBS here acts only as a backhaul hub without any management functions to the ABSs.

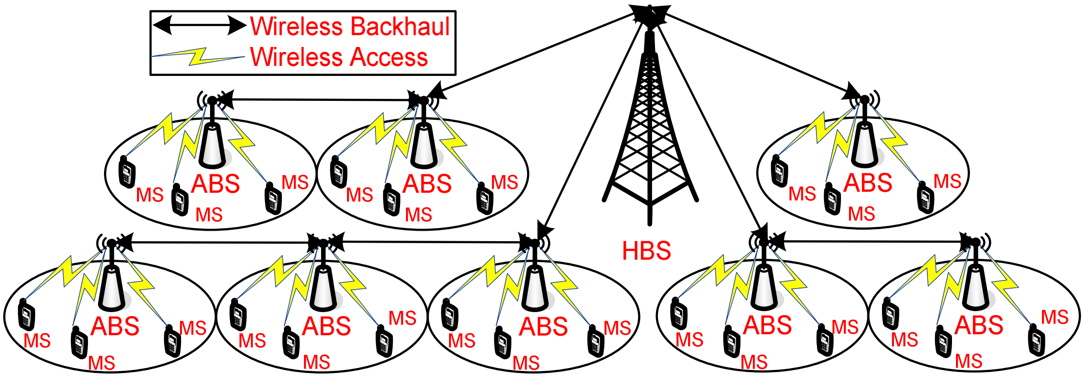

Figure 3.1. High Capacity Density Network Architecture

3.2.1 *Multi-hop Wireless Backhaul Network*

The multi-hop backhaul network is proposed in [67, 91, 92] as a potential solution to the BuNGee backhaul architecture. Multi-hop networks have a number of advantages in providing backhaul services compared to the single-hop network in BuNGee, which has been characterized in Section 2.2.2. The general network model is illustrated in Figure 3.2 below, which consists of a HBS in the centre and several ABSs around it. It can be observed that a HBS serves a set of x branches. On each branch there are a set of y hop ABSs, with elements numbered outwards from the central HBS. Each link has access to a common pool of frequency channels.

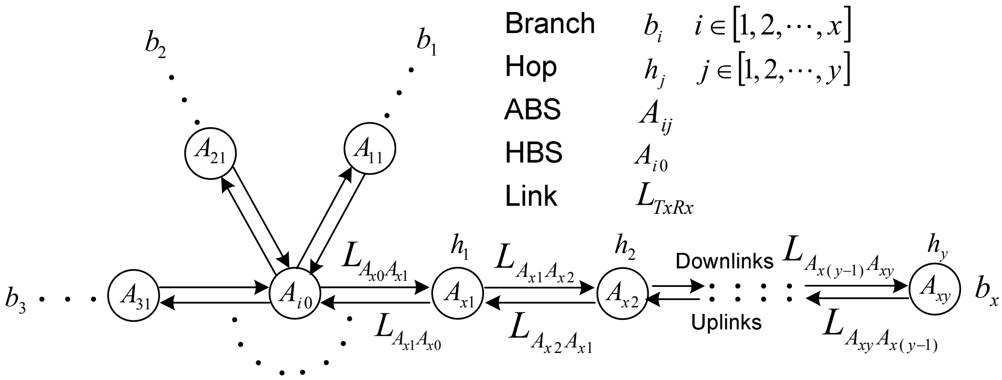

Figure 3.2. Multi-hop Wireless Backhaul Network Model

The traffic flow is generated from a source ABS and transmitted to the HBS on the uplinks, or from the HBS to a destination ABS on the downlinks, through an end-to-end route established by multiple links. The link budget on an end-to-end route is constrained by an individual link with lowest link quality, namely the bottleneck.

Data transmission could be constrained by this bottleneck regardless of performance on other links.

3.2.2 *Flexible Small Cell Access Network*

The access network of BuNGee is constructed by a dense deployment of small cell ABSs at a below rooftop level of urban streets. In this thesis, the ABSs are deployed on the high streets around building area, which is dedicated to provide augmented capacity for high street users as shown in Figure 3.3. The coverage radius of each ABS is 90m to a maximum, with omni-directional antennas implemented.

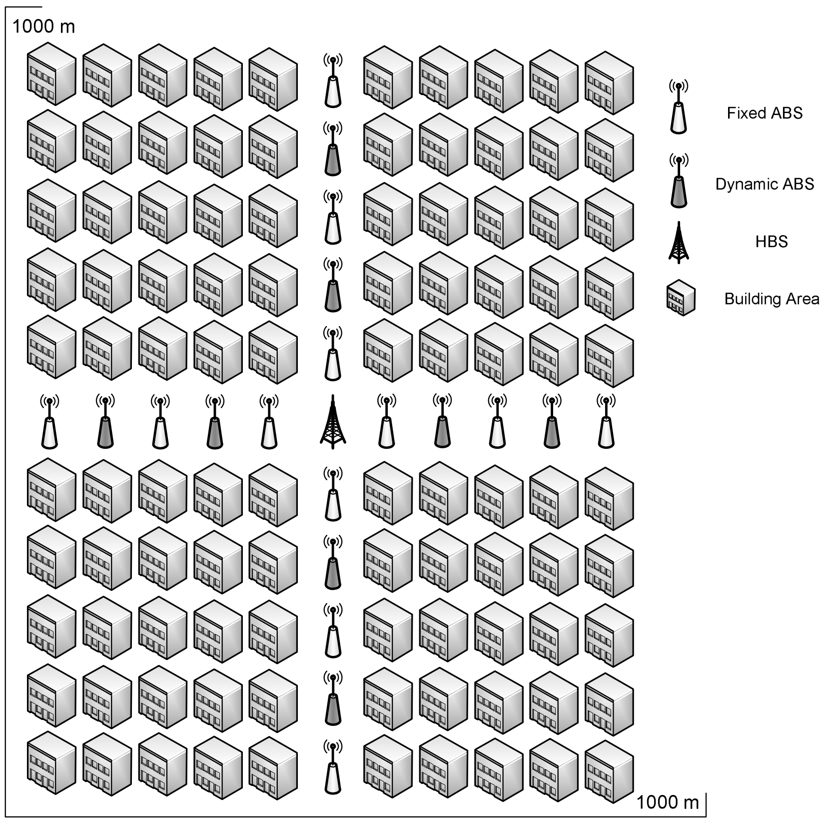

Figure 3.3. Flexible Small Cell Access Network Model

The role of ABSs is to enhance the network capacity where the traditional macrocell BS (co-located with HBS) has no sufficient resource to support. In this case, the number of ABSs required largely depends on the traffic density in a specific area during a specific time. In the future applications of wireless communication, an ultra high capacity requirement could happen occasionally when a number of users gather on the streets. However, a dense deployment of ABSs could incur a significantly high amount of energy consumption if it is not effectively managed.

For the purpose of achieving a balance between capacity and energy issue, a flexible small cell access network is introduced in Figure 3.3. This includes two types of ABSs proposed: the fixed ABSs are always active; the dynamic ABSs are activated only when the user traffic cannot be supported by the fixed ABSs, and in which case the traffic will be transferred accordingly. The flexible network architecture will be controlled by the topology management strategies developed in Chapter 7.

3.3 Simulation Techniques

There are a wide range of simulation tools available to model wireless communication systems. However, different protocol layers have their preferred tools and modelling methodologies.

Programming languages such as C and C++ can be directly used for modelling the wireless systems. It has been traditionally used for software simulations, especially during the early days where advanced simulation tools such as OPNET and NS were not available. C/C++ is one of the most commonly used programming languages, which is especially effective in compiling and executing. In simulation, it has a great advantage in iterative computations since the source code is compiled to binary code in advance and can be re-executed repeatedly, rather than using run-time interpretation in some other languages. Its flexibility in memory management can also avoid overflow when a large number of stochastic simulations are required. Moreover, object oriented programming with C++ can significantly reduce the complexity of code when the same protocols and algorithms are operated on a large number of nodes. Last but not least, C and C++ are standard languages for many practical implementations of communication systems, for example on many DSP and USRP broads and most of the TCP/IP protocols. The use of C and C++ in software simulation makes implementation easier.

C/C++ in software simulation also has some disadvantages. The absence of GUI component makes it difficult for developers to debug the codes, or obtain temporal results. Experiences show that a normal C/C++ debugging error could have multiple reasons other than the code itself, such as the debugger, memory and operating system, which will increase the developing time.

Matlab developed by MATHWORKS Inc provides powerful matrix calculation and graphing routines as well as a number of mathematical and professional functions [93]. It provides convenience in programming and debugging, which makes the work easier and more visible. As a type of interpreted language Matlab programs can be debugged step by step without the requirement of compiling, which provides an easy way to find errors. Matlab is a preferred tool to build up the architecture of this work because a large number of matrixes are used for evaluating various parameters for multiple nodes. Moreover, Matlab provides effective ways to produce graphical results for performance evaluation. Furthermore, this work will consider a number of dynamic network behaviors, such as traffic, channel usage, network topology, etc. Matlab can significantly reduce the time for code development. In recent years, Matlab is commonly used in both academe and industry. A system level simulator in Matlab provides transportable codes for some other researchers.

There are other network simulators that could potentially be used for this type of simulations, such as OPNET, NS2, NS3, etc. However, they are originally designed to model the detail of protocols and their interactions. In this work, system performance is one of the most crucial aspects considered whereas the protocol behavior is less important. As a result, Matlab is selected in this work as the major simulator, which provides the capability of modeling complex system architecture with visible performance validation.

3.4 Wireless Network Modelling

3.4.1 *System Level Simulation*

System level simulation is widely used in following chapters to analyze performance and validate designed approaches. It is developed to model practical network architectures and capture related performance. This work mainly focuses on the RRM aspect of wireless cellular and backhaul networks, which requires link level and data traffic modeling.

The simulator is developed to be applicable for different types of scenarios, which consists of multiple modules (functions) modeling different aspects of a network. The structure of this simulator is illustrated in Figure 3.4.

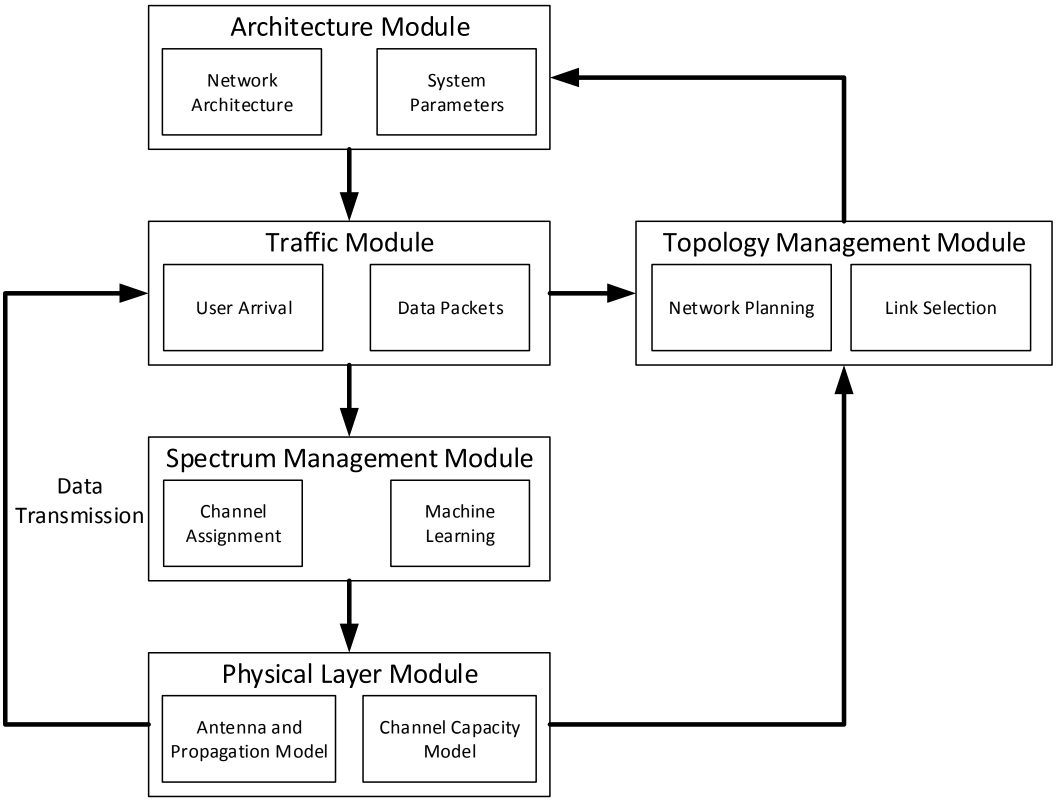

Figure 3.4. Simulator Structure

The architecture module is firstly developed to include the network topology (location of elements and possible connections) and system parameters (power, frequency, bandwidth, noise, etc.). This module provides most of the constant setups throughout the entire simulation process, which is designed as a black box with interfaces connected to other parts of the simulator. The network topology and system parameters depend on the scenarios, which will be detailed in the following related chapters.

The physical layer module is developed to model the wireless link between transmitters and receivers. This includes the antenna model at the transceivers, the propagation model defining transmission loss, as well as the channel capacity model representing the modulation and coding schemes. The output of this module is the received signal power, interference level and the link data rate. Detailed models will be described in Section 3.4.2.

The traffic module describes the traffic characteristic of the network, including the number of users and their related arrival and departure time. In traffic engineering, these characteristics are classified as events. Monte-Carlo method [94] is widely used in event based simulation. The output of Monte-Carlo simulation is a long term averaged result of a large number of repeatedly sampled values from random distributions, which removes the interim fluctuation and achieve a statistical reliability. The Monte-Carlo method and event based simulation are widely used in the traffic module of this work, to obtain stabilized results and to capture temporal performance.

The spectrum management module is mainly responsible for assigning channels to the network elements in different locations, to establish wireless link using the physical layer module. The assignment behaviour is controlled by algorithms designed, including learning strategies demonstrated in the following chapters. This module is one of the most important parts in this work, which is aimed at improving the network capacity through effective spectrum utilization.

The topology management module is designed to dynamically control the network topology according to traffic level. This module manages the location of network elements and the connections between them. In a multi-hop backhaul network, it provides the routing table on each ABS for end-to-end connections. In a small cell access network, a novel dynamic topology management algorithm is proposed. The network energy consumption can be reduced by controlling the number of active base stations and related traffic. The detailed algorithm of this module will be presented in Chapter 7.

In a system level simulation, these modules are connected with each other via related inputs and outputs. The Monte-Carlo events generated from the traffic model determine when these modules are used. The results could be obtained from multiple modules. In order to have reliable validation, the number of events should be high enough to remove interim randomness, and the result should be evaluated on a steady state within a finite number of iterations.

3.4.2 *Physical Layer Models*

The physical layer models are used to capture the physical characteristics of wireless links. Performance evaluation in RRM is mainly at a network level rather than on an individual link between transceivers. Thus these models simplify the simulation of physical layers but also provide essential characteristics of the practical systems.

3.4.2.1 Antenna and Propagation Model

The antenna and propagation models provide the antenna gain and transmission loss between the transceivers, respectively [95]. In this thesis, various models have been used in different networks and scenarios.

The network architecture presented in Section 3.2 indicates that directional antennas are implemented in the multi-hop backhaul network and omni-directional antennas are deployed in the small cell access network. The directional antenna model demonstrated here is thus for the backhaul links.

In a backhaul network, the location of BSs is fixed and the architecture is static. The antenna mainlobe can be implemented in the direction of links. Multiple antennas are implemented on a base station, to transmit or receive signals in different directions. There are two models used in this thesis. In Chapter 4 a simplified aperture antenna model is used, which is originally designed in [96]. The model defines the mainlobe curve of antenna gain pattern, which describes antenna sidelobe as a fixed value of relative power, normally -30dB. The designed valid antenna beamwidth is less than 90° and the full range of radiation angle is within ±90°. In order to adapt this model to our scenario, we have extended the sidelobe to ±180° by using the same sidelobe gain. The antenna pattern is demonstrated in Figure 3.5. It can be observed that the effective beamwidth of this antenna is less than 90° where the peak gain is positive.

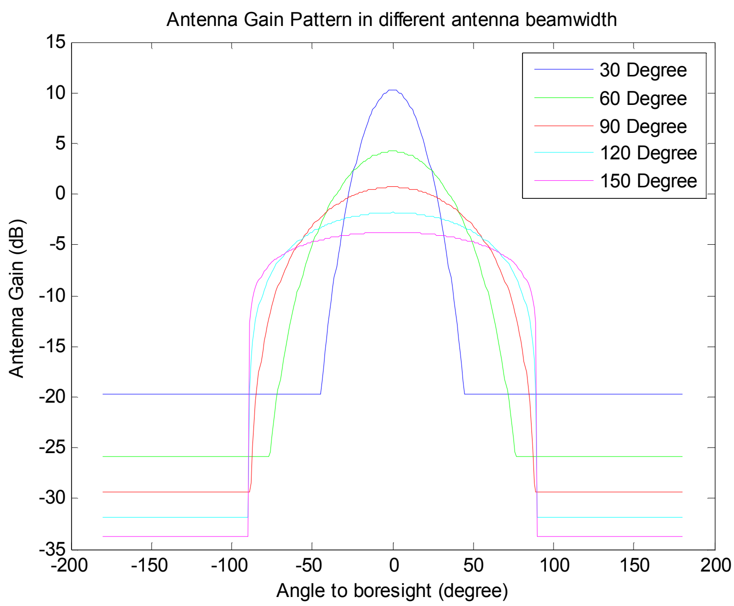

Figure 3.5. Aperture Antenna Model

The antenna gain is the intensity of an antenna at a given direction compared to the ideal hypothetical antenna, which is defined as [95]

$$G(\theta, \emptyset) = eD(\theta, \emptyset) \tag{3.1}$$

where D is antenna directivity and e is antenna efficiency. In this model, only the horizontal polar with degree factor θ is considered. The directivity is calculated from

$$D = \cos(\theta)^n \frac{32\log 2}{2(2\arccos(\sqrt[n]{\frac{1}{2}}))^2} \tag{3.2}$$

where n is a power factor defined as

$$\left(cos\frac{\theta_{3dB}}{2}\right)^n = 0.5 \tag{3.3}$$

θ_{3dB} is the 3dB beamwidth where the radiation power drops down to half of its peak value, which is an important factor for changing the shape of antenna mainlobe. The characteristic of mainlobe will change the interference environment, which is one of the main considerations in designing spatial reuse strategies.

Another newly designed antenna model proposed in [12] is used for other parts of this work. It is obtained from a practical product developed by Cobham. The antenna pattern is demonstrated in Figure 3.6.

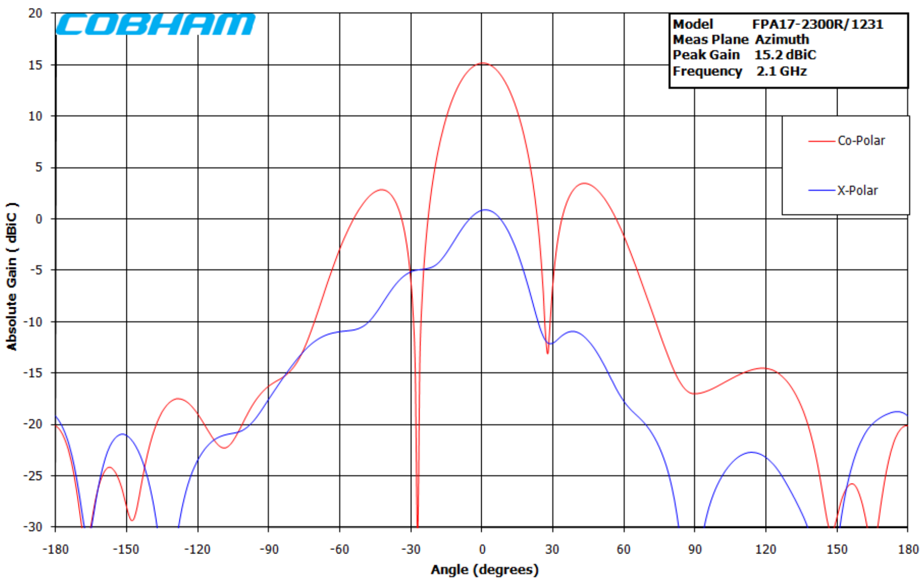

Figure 3.6. Directional antenna developed by Cobham [12]

It can be observed that its first sidelobe has a significant high power level, which rolls off smoothly to sidelobe power less than -15dB. The multi-hop backhaul topology significantly reduces the number of directional antennas required on an HBS (Figure 3.1), which in turn reduces interference level between adjacent links.

The propagation model provides the path loss between transceivers on a wireless link. There are two propagation models used in this thesis. The COST-HATA model is used in Chapter 4, which is designed in [97] as a frequency range extension for Okuma-HATA model. The path loss is calculated by

$$\mathrm{L[dB]} = 46.3 + 33.9\mathrm{log}f_c - 13.82\mathrm{log}h_{te} - a + (44.9 - 6.55\mathrm{log}h_{te})\mathrm{log}d[m] + C_M$$

$$C_M = \begin{cases} 0\mathrm{dB}, & \text{for Rural and Suburban Area} \\ 3\mathrm{dB}, & \text{for Urban Area} \end{cases} \tag{3.4}$$

where f_c is the carrier frequency, h_{te} is antenna height and d is distance between transceivers.

The other parts of this thesis use the WINNER II channel model B5a proposed in [98], which is designed for the small cell scenarios in metropolitan areas. The path loss between transceivers is calculated by

$$\mathrm{L[dB]} = 23.5\log_{10} d[m] + 42.5 + 20\log_{10}\frac{f_c[GHz]}{5} + X \tag{3.5}$$

where X is the log-normal shadow fading, with standard deviation $\sigma = 4$.

3.4.2.2 Channel Capacity Model

Channel capacity is the rate of bits that can be delivered over a communication channel. According to the Shannon-Hartley theorem [4], channel capacity on a wireless link is determined by Signal-to-Noise Ratio (SNR) and channel bandwidth B, which can be obtained from

$$\mathrm{C} = \mathrm{B}\log_2\left(1 + \frac{\mathrm{S}}{\mathrm{N}}\right) \tag{3.6}$$

This indicates the maximum data rate that can be achieved on a wireless link. However, in a practical system and channel capacity could be constrained in physical layer, including the modulation and coding schemes. A Truncated Shannon model has been developed in [99], which is a representative of rates that can be achieved in practice given an Adaptive Modulation and Coding (AMC) codeset [100]. The achievable data rate for a specific user on a channel can be expressed as

$$C = \begin{cases} 0 & \gamma < \gamma_{\mathrm{MIN}} \\ \alpha B \log_2(1+\gamma) & , \gamma_{\mathrm{MIN}} \leq \gamma \leq \gamma_{\mathrm{MAX}} \\ \alpha B \log_2(1+\gamma_{\mathrm{MAX}}) & \gamma > \gamma_{\mathrm{MAX}} \end{cases} \tag{3.7}$$

where $\alpha \in [0,1]$ is an attenuation factor representing the implementation loss compared to Shannon bound (3.6). γ is the Signal-to-Interference plus Noise Ratio (SINR) achieved at the receiver, which can be obtained from

$$\gamma = \frac{g_c(L) g_t g_r p}{n + \sum g_c(L) g_t' g_r' p} \tag{3.8}$$

where $g_c(L)$ denotes the channel gain, g_t and g_r are antenna gains at the transmitter and receiver base stations, and those of g' are gains on interfering transceivers using the same channel. *n* is the thermal noise power and *p* is the transmit power.

In (3.7), γ_{MIN} and γ_{MAX} are introduced to represent the effective SINR range that can be used for the employed AMC codebook in a practical system. In order to adapt the 3GPP parameters presented in [100], the Truncated Shannon model has been defined in [99] as: the minimum SINR for maintaining a communication link: $\gamma_{\mathrm{MIN}} = 1.8\mathrm{dB}$; the SINR where a maximum capacity can be achieved in AMC codeset: $\gamma_{\mathrm{MAX}} = 21\mathrm{dB}$; and the implementation loss: $\alpha = 0.65$. By applying these parameters in (3.7), the data rate curve matches the AMC codebook defined by 3GPP [100].

3.4.3 *File Transfer Traffic Model*

The traffic model is designed to model the behaviour of data traffic across wireless network. Future wireless communication systems are designed to be fully packet-switched. The channel bandwidth is shared by multiple users rather than persistently allocated to a dedicated user in a traditional circuit-switched telephone network [41]. This approach delivers more reliable end-to-end connections because failed data can be retransmitted rather than dropped.

OSI and TCP/IP are typical conceptual models that characterize the functions of communication systems [41], by partitioning it into abstraction layers. Each layer has logical links connected with the same layer on other nodes when physical link is established. The definition of data units varies on different layers. For example, bits on physical layer, frames on data link layer, packets on network layer, etc. Traditional network simulators are designed to model the above data units delivered in a network, which is essential to capture and analyse the performance of each layer. However, these data unit models could be very complex and inefficient in simulation when modelling a large scale network architecture with a huge amount of data traffic in transmission.

The purpose of network simulation in this work is to investigate and validate the radio resource and network topology management methodologies, where the system performance is of most interest. These types of characteristics can be obtained when a sufficient amount of packets/frames are delivered in a network. For the purpose of capturing the system performance in a heterogeneous architecture as well as reducing the simulation complexity, a file transfer traffic model has been developed for this work to model data traffic behaviours.

In the file transfer model, a file is defined as an entity grouping data payload in the application layer. It could be a succession of packets, frames or bits in lower layers [101]. Characteristics of files delivered in a practical UMTS network have been reviewed in [102]. Compared with a conventional packet based traffic model, the file transfer model investigates the data packet transmission at a larger time scale, whereas an individual packet has minor impact on the system performance. In addition, the network protocols are assumed to be well established in this model.

The simulator randomly generates the file inter-arrival time (arrival rate) and the file size, which follow a defined statistical distribution. The offered traffic can be controlled by varying the inter-arrival time. The transmission time of a file is determined by the channel capacity (3.7) and the SINR level (3.8) in a given file size (defined in bits). A file can be backed off and retransmitted from an interruption point, because it contains a succession of data units.

The long-tailed distribution is suitable for modelling the inter-arrival time of files delivered in a practical network [102]. In this model, the large files mainly contribute to the network burden, even though the probability of large files being generated is fairly low compared to the small files. For example, web browsing is the major application on the Internet, which has a large number of short file transmissions. However, streaming media, where a long session is delivered, occupies the majority of bandwidth resource.

The Pareto distribution is a typical implementation of long-tailed distributed network traffic, which has its CDF function defined as [103]

$$F(x) = 1 - \left(\frac{x_{min}}{x}\right)^{\alpha}, x > x_{min} \tag{3.9}$$

Here α is the shape parameter and x_{min} is the scale parameter. The number sequence in (3.9) should always be no smaller than x_{min}. The mean value of the Pareto distribution is

$$E(X) = \frac{\alpha x_{min}}{\alpha - 1}, \alpha > 1 \tag{3.10}$$

So α can be derived from (3.10) as

$$\alpha = \frac{E(x)}{E(x) - x_{min}} \tag{3.11}$$

The variance of the Pareto distribution is

$$VAR(X) = \frac{{x_{min}}^2 \alpha}{(\alpha-1)^2(\alpha-2)}, \alpha > 2 \tag{3.12}$$

In summary, the shape parameter α can be obtained from (3.11). However, it should follow the condition of $\alpha > 2$, in order to obtain a valid distribution.

The Inverse Transform Sampling is an effective method for generating pseudo-random number of any distribution based on CDF [103]. The principle is that if *X* is a continuous random variable with CDF $F(X)$, then the random variable $Y = F(X)$ has a uniform distribution on [0,1]. Following this, the sequence can be generated from the inverse function of CDF. The Pareto distributed inter-arrival time *t* can be generated from

$$t = \frac{t_{min}}{x^{\frac{1}{\alpha}}} \tag{3.13}$$

where *x* is a uniform distributed random sequence following $x \in [0,1]$.

The mean value in (3.10) is effectively $\bar{t}$. To satisfy $\alpha > 2$, t_{min} should follow

$$t_{min} > \frac{1}{2}\bar{t} \tag{3.14}$$

According to the Little's Law [41] and the relationship between the mean arrival rate λ and inter-arrival time $\bar{t}$, the offered traffic is defined as

$$G = \lambda T = \frac{T}{\bar{t}} \tag{3.15}$$

where T is transmission time.

In the network simulation, file size is randomly generated following a uniform distribution with a defined average value. The back off time for retransmission is uniformly distributed with mean value of $\bar{t}$. A file can be consistently retransmitted until successfully delivered.

3.5 Performance Evaluation Techniques

System performance is measured by a number of parameters looking at different aspects. Conventional QoS is used to evaluate the network performance, which includes network throughput, delay, retransmissions, etc. Several parameters are defined to capture the learning behaviour. In addition, cooperation overhead and energy efficiency are measured from a percentile perspective. The results are produced in both average and temporal formats, in order to evaluate performance from different perspectives.

3.5.1 *Quality of Service*

Quality of Service is widely used in evaluating the performance of contemporary communication networks. This includes a number of parameters measuring the system in different aspects. Some of them are selected in this work to produce interested results in radio resource management.

The number of blocked and dropped calls are conventional parameters used for measuring QoS on a call based network, which represent a call is prevented from accessing the network or interrupted during transmission [104]. A telephone user usually has more tolerance on blocked calls than dropped calls. These two parameters are used in Section 4.3 for a link level performance evaluation.

In a packet based network, retransmission is carried out on a file that is either blocked initially or interrupted during transmission. The probability of blocked and interrupted files have little difference in user experience, because they are applied with the same retransmission scheme. Moreover, for an interrupted file, the user only needs to retransmit the remaining part of the file that has not been delivered. In this context, the probability of retransmissions is used for QoS evaluation, which is defined at time *t* can as follows

$$P_r(t) = \frac{N_B(t)+N_I(t)}{N_T(t)} \tag{3.16}$$

where $N_T(t)$ is the number of total transmissions, $N_B(t)$ and $N_I(t)$ are the number of blocked and interrupted files.

Throughput in a wireless network is defined as the average rate of successful data delivery. In system level research, throughput can be affected by both transmission and back off delay. The average throughput of the entire network is measured from

$$S(t) = \frac{N_{bit}(t)}{t} \tag{3.17}$$

$N_{bit}(t)$ is the number of bits delivered within time *t*, which is contributed by the files delivered by all the users in the network, including those still in transmission.

The delay of a file consists of transmission delay and back off delay. Transmission delay is the time required to push all the bits of a file into the wireless link, which mainly depends on the channel capacity (3.7). The back off delay is the time consumed by a file waiting for retransmissions. Moreover, the propagation and signal processing delays are not considered in this work because they are relatively small compared to others. The queuing delay is assumed to be effectively handled by well defined transport layer protocols [41].

In summary, the average delay of a file is calculated by

$$\overline{D} = \frac{1}{N_{File}} \sum_{i=1}^{N_{File}} \left(\frac{N_{bit}(i)}{C} + \sum_{j=1}^{N_B(i)+N_I(i)} D_r(j) \right) \tag{3.18}$$

This equation includes the time consumed to deliver files with N_{bit} bits, as well as the time spent to back off N_B blocked files and N_I interrupted files.

Cumulative Distribution Function (CDF) is used in this work to provide the statistical behavior of a large amount of results measured by a Monte-Carlo simulation. The sampled results in simulation are a set of discrete random variables *X*, thus the CDF of *x* can be calculated from

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} P(X = x_i) \tag{3.19}$$

Error Bar

Error bar is a graphical representation of the variability of data, which can be used to evaluate the accuracy of Monte-Carlo simulation results. It identifies the probability (confidence level) that a given set of results will be within a specified range (confidence interval) [105]. The longer a simulation run, the smaller confidence interval is for a specific confidence level, and vice versa.

A technique used for obtaining confidence limits in this thesis is shown in (3.20) below [106].

$$e = \mu \pm z_c \frac{\sigma}{\sqrt{N}} \tag{3.20}$$

where μ is the sample mean, σ is the standard deviation, N is the number of trials, and z_c relates to the chosen confidence level.

The values of z_c are given in Table 3.1 for several common confidence levels, which is valid given the results fit a normal distribution.

Table 3.1. Confidence levels and corresponding z_c

Confidence Level	90%	95%	99%	99.9%
z_c	1.645	1.96	2.58	3.29

The resulting confidence limits can be plotted as error bar, which is demonstrated in Figure 7.13 in Section 7.5. It shows the system average delay, which is evaluated during a period of simulation on a large file transfer events, as specified in Table 6.3. The simulation configurations are consistent in this thesis, thus the confidence level of all the results can be represented by Figure 7.13.

3.5.2 *Learning Efficiency*

The application of machine learning to wireless communications is one of the major topics and original contributions in this thesis. It is thus important to directly investigate the efficiency of learning algorithms in a wireless environment.

A cognition cycle has the steps of decision making, action taking, and learning [58]. A learning iteration is associated with a decision on which channel to select, as well as an action to establish link between transceivers on a selected channel. The outcome of decisions, either success or failure, indicates the quality of learning. It also shows the traffic level that the learning process leads to instability and ineffective configurations.

The probability of failed decisions is used as a parameter to measure learning efficiency, which is obtained from

$$P_{fail}(t) = \frac{N_{fail}(t)}{N_{it}(t)} \tag{3.21}$$

where $N_{fail}(t)$ is the number of failed decisions and $N_{it}(t)$ is the number of iterations during time *t*. It should be noted that $N_{it}(t)$ includes the number of transmissions $N_T(t)$, plus the iterations where the transceivers try to establish a link on selected channels but fail. The relationship between these parameters follows

$$\begin{gathered} N_{it}(t) \geq N_T(t) \geq N_F(t) \\ N_{fail}(t) \geq (N_B(t) + N_I(t)) \end{gathered} \tag{3.22}$$

The learning performance of distributed reinforcement learning and transfer learning is demonstrated in Figure 5.6 in Section 5.7.

Convergence is another important target of machine learning. An effective learning algorithm is not only to find a stable and reliable set of decisions but also to achieve this stable state quickly, which is evaluated by convergence rate. Conventional measurement of convergence by the computer science community normally uses the number of episodes over the number of iterations. However, this method is based on a repeated simulation over a known target. In the wireless network scenario, the target of learning cannot be discovered before taking a sufficient number of actions. Moreover, the dynamics of the environment could continuously change the learning target. In this case, the learning efficiency cannot be measured by the iterations taken to achieve a targeted state.

The stable state is defined as a learning agent staying on a fixed set of actions, which can be used to measure the convergence rate of learning. A high probability of stable states indicates that the learning agent achieves a targeted solution. In this work, a novel stable state evaluation method is developed and applied not only to measure the performance of learning but also to determine the time for information exchange in knowledge transfer. The detail of this method will be illustrated in Section 5.5, with examples demonstrated.

The overall target of spectrum management is to partition the shared spectrum to individual users. Following this, the spectrum usage probability has been defined and used to evaluate such partitioning behaviour [64]. Effective spectrum usage corresponds to some channels being used at a significant higher probability than others. Detailed equations will be presented in Section 4.4.

3.5.3 *Cooperation Overhead and Energy Efficiency*

Cooperation Management and Topology Management are two major original contributions in this work other than Transfer Learning, which reduce coordination overhead and energy consumption in distributed wireless networks, respectively. Related parameters are defined to validate these strategies.

The probability of information exchange is used to evaluate the amount of control information (i.e. channel usage indication) transmitted between multiple agents. In Chapter 4 this parameter is used to investigate the performance of cooperation management strategies. It is calculated from the number of information exchange $N_e(t)$ over the number of iterations taken in time *t*:

$$P_E(t) = \frac{N_e(t)}{N_{it}(t)} \tag{3.23}$$

Energy efficiency is a crucial parameter used to measure the performance of topology management strategies. An effective way to inspect energy consumption is to use a practical energy model for the entire system including the energy used for radio transmission, power amplifier, cooling system, etc. However, it is difficult to obtain a generalized model representing energy consumption in different systems. Instead of producing the result in actual energy units, a proportional energy consumption parameter will be used to indicate the energy saving from topology management. In the flexible small cell architecture presented in Figure 3.3, a set of dynamic ABSs is introduced as capacity enhancement to the baseline fixed ABSs. The activation of dynamic ABSs has direct impact on energy consumption of the overall architecture. In this scenario, a parameter of energy consumption ratio is defined for measuring the amount of extra energy required over the baseline architecture with fixed ABSs only, which is defined by

$$E(t) = \frac{N_{ON}(t) - N_{Base}}{N_{Base}} \tag{3.24}$$

N_{Base} is a baseline energy level calculated from the energy used by all fixed ABSs. $N_{ON}(t)$ is the energy consumed by all activated ABSs during time *t* in the network, including the dynamic ABSs.

3.6 Verification Methodologies

Verification is used to analyse the system performance through mathematical models. Theoretical results are produced to validate the designed strategies. In this thesis, queuing theory is used to analyze the small cell access network with dynamic architectures, to validate the resource and topology management strategies.

Queuing Theory and Markov Models

Queuing theory is an effective tool to analyse the QoS and capacity of wireless communication systems, which has been extensively studied in [45]. The traffic behaviours of users in a network can be modelled as a queuing system, including file generation, transmission and interruption. Queuing theory is a tool to investigate user behaviour in a limited amount of resources. In the related theoretical models, a resource block is normally assumed to be assigned for a dedicated user, unless a reuse pattern should be defined to permanently fix the overall network capacity.

A big challenge of modelling a cognitive radio network as a queuing system is the dynamics of spectrum reuse in the system. In such scenarios, the decision of spectrum selection is made by distributed users and varies from time to time, which frequently changes the network capacity. It is thus difficult to directly model the system during the learning process. However, the improved and converged solution achieved by learning can be modelled with analytical tools.

In this thesis, a queuing system is used to model the dynamic small cell network. The dynamic spectrum access system is modelled in a stable state such that, a cluster of adjacent base stations assign different channels to their users without interference. The overall network capacity can then be calculated from the spectrum size. This is the theoretical optimized state of distributed learning. However, transfer learning has the potential to achieve such state effectively following the Pareto efficient prioritization algorithms. The detailed models and algorithms will be illustrated in Chapter 7.

The classical M/G/k/1 queue is used to model the system, which follows the conditions that user arrives at a Poisson process; the service time has general distribution; k servers (channels) are provided in the system; the queue length is 1 that the blocked users are considered to be lost in the system. The blocking probability indicates system throughput.

The Markov model is an essential mathematical tool to analyze queuing systems. A Markov chain models all user behaviors (arrivals, departures, blockings) in different system states. In a communication network, a state is defined as the number of channels occupied in the system. The state transition probability is determined by the arrival and departure rates. Conventional Markov analyses are carried out in one dimension, which only models a single system. However, practical wireless networks are constructed from multiple base stations, with a number of channels (servers) allocated to each or shared by all. In this work, a novel Multi-dimensional

Markov model has been designed and used to analyse multi-cell performance. An example of a two dimensional queuing system is illustrated in Figure 3.7.

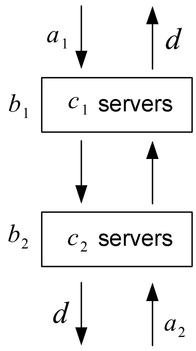

Figure 3.7. Multi-dimensional Queuing System

The system has two base stations b_1 and b_2 with channel set c_1 and c_2 assigned, respectively. By defining arrival rate of users in these cells as a_1, a_2 and the departure rate as d, the probability of the system having j_1 and j_2 channels occupied in each cell is

$$(a_1 + a_2 + (j_1 + j_2)d)P(j_1, j_2) = a_1 P(j_1 - 1, j_2) + a_2 P(j_1, j_2 - 1) + (j_1 + 1)dP(j_1 + 1, j_2) + (j_2 + 1)dP(j_1, j_2 + 1) \quad (j_1 < c_1, j_2 < c_2) \tag{3.25}$$

The Markov chain and equilibrium equation varies for different systems, which will be discussed in Chapter 7.

3.7 Conclusion

This chapter has described the method of modelling, simulation and analysis methods used in this thesis. The generalized network models of multi-hop backhaul and flexible access network have been demonstrated. Matlab is selected as the software tool for network simulation. The simulator is constructed by several modules to cover different aspects of the system, including the architecture, physical layer, traffic, spectrum and topology management modules. Selected antenna, propagation and traffic models have been discussed. The output parameters including QoS, learning efficiency, cooperation and energy evaluation are presented to validate the developed strategies. Furthermore, queuing theory is demonstrated to analyse the flexible access network, with newly designed multi-dimensional Markov models.

Chapter 4. Distributed Learning and Interference Coordination

4.1 Introduction

The objective of cognitive resource management is to enable the cognitive agents (base stations or mobile stations) to dynamically select resource blocks (channels) to set up a communication link. It is designed to carry out resource management in a distributed manner without a centralized Radio Network Controller (RNC), which significantly reduces the complexity of the network architecture. However, the co-channel interference becomes a serious issue in a cognitive radio network, because the cognitive agents could be randomly located and may have dynamic access to a common spectrum pool. In the cognitive radio scenarios where centralized planning and scheduling algorithms are not available, the cognitive agents are required to coordinate and learn the radio environment to avoid interference.

For the purpose of operating distributed resource management, a cognitive agent is required to either communicate with others in vicinity or learn the surrounding environment. In this context, two strategies are proposed and investigated in this chapter: an Interference Coordination strategy and a Distributed Learning strategy. The Interference Coordination strategy is designed to directly exchange spectrum usage information between adjacent cognitive agents, in order to avoid the same channels being used simultaneously. On the contrary, the Distributed Learning strategy enables the cognitive agents to learn the spectrum usage and user activity.

This operation is designed to be fully independent, and not require information exchange. Under the exploration and exploitation of learning algorithms, the cognitive agent can converge to a preferred spectrum pool after a number of learning iterations.

This chapter presents the earlier work on fully distributed and fully coordinated cognitive resource management approaches for the multi-hop backhaul network of the high capacity density network architecture. The radio environment of the multi-hop network is firstly investigated, including the hidden/exposed terminal problems and the issue of bottlenecked traffic. In Section 4.3, a Space-division Interference Coordination strategy is proposed, which is based on the spatial resource reuse between antenna beams to provide fair resource utilization across multiple hops.

The second part of this chapter investigates a distributed learning algorithm that is applicable to the multi-hop backhaul network. This includes a Linear Reinforcement Learning algorithm and a Single State Q-learning algorithm. Theoretical convergence is evaluated as a performance comparison of the two algorithms in different scenarios. Furthermore, improved decision making schemes with physical parameters are investigated, to improve the spectrum sensing efficiency.

The purpose of this chapter is to investigate further improvements to the conventional distributed resource management approaches, which motivates the design of transfer learning in further chapters.

4.2 Radio environment of multi-hop networks

The architecture of a multi-hop backhaul network has been illustrated in Figure 3.2, which consists of a HBS in the centre and several ABSs around it. The HBS serves a set of x branches, where a set of y hop ABSs are connected on each. On a backhaul network, the downlink traffic is transmitted from the HBS to an ABS while the uplink traffic is vice versa. The traffic flow can be delivered only if an end-to-end link has been established, which contains multiple hops from source to destination.

The role of cognitive resource management is to assign data channels to individual links (between two base stations), in order to establish an end-to-end link for communication. In the scenario of a multi-hop network, multiple channels may be required on an individual link, to deliver both local and relayed traffic. As well as this, the backhaul architecture incurs higher relayed traffic load on the links near the hub, which require more resources to be assigned.

4.2.1 *Interference Issue*

The major target of resource management here is to mitigate interference between backhaul links. Interference is caused by the links using the same channels, which is determined by the location of transceivers, antenna profile, the transmit power, etc. The backhaul network has a more static interference environment than the access network, because the location of base stations is normally fixed and directional antennas are implemented on both ends of the link, as detailed in Section 3.2.1. Conventional Minimum Interference (MI) [56] and Maximum SINR [107] channel assignment schemes have the capability of identifying interfering terminals in the vicinity. However, the hidden terminal problem occurs when a transmitter cannot identify the potential interfering terminals near the receiver, because their interference power could appear low at the transmitter but high at the receiver. This normally happens when a receiver is near another transmitter. For instance, when an ABS has co-located transmitter (Tx) and receiver (Rx) antennas, the transmitter may choose the same channel of the receiver because the receiver antenna does not radiate signal power towards the new transmitter's antenna. This issue can be illustrated in Figure 4.1 below, where three multi-hop links are connecting four hops of ABSs. The arrows denote the direction of antenna main lobes.

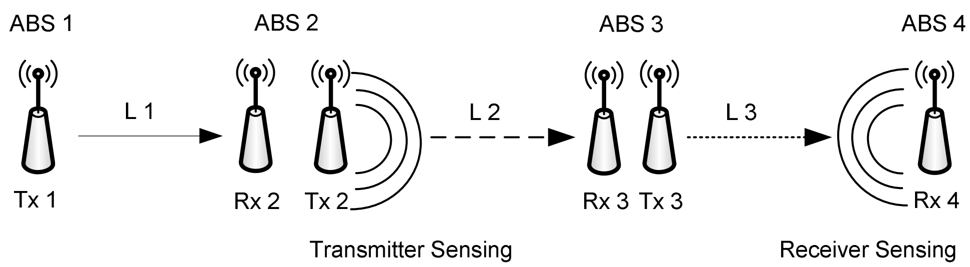

Figure 4.1. Multi-hop network interference environment

It can be seen from Figure 4.1 that in a multi-hop architecture, a relay node (ABS 2/3) may have co-located receivers (Rx) and transmitters (Tx). For the purpose of establishing an end-to-end link, these transmitters and receivers should relay traffic flows. However, while routing information (OSI layer 3) is routinely transferred across multi-hop links, the exchange of channel usage information (OSI layer 2) is much more difficult, and is not practical with many existing and future protocols standardised today. The MAC (in contrast to routing) protocols conventionally work at the individual link level, meaning that it is difficult to implement a centralized assignment solution on a multi-hop network. As a result, the distributed channel assignment strategies, presented in Chapter 2, remain the most practical solution. In this manner, a transmitter can identify a receiver only if the interference level is above a threshold for establishing a communication link.

The interference is radiated from the transmitter, whereas the quality of a link depends on the SINR at the receiver. This causes the hidden terminal problem, where a receiver is out of an interference detection range [40]. In a multi-hop backhaul network, the antenna directionality and the transceiver's location make the interference environment different from an access network. In the example architecture of Figure 4.1, L1 is a link that has already been established. The second hop transmitter Tx2 operates spectrum sensing to establish L2. In this case, it may not detect excessive interference from Tx1 because: 1) its antenna is pointing in the reverse direction of L1; 2) it is located at a distance from Tx1. However, Tx2 may incur interference to Rx2 because they are co-located.

A similar issue occurs when a receiver operates spectrum sensing. It can be illustrated from the same figure when the third hop establishes a link L3. Rx4 may not detect interference from Tx2 because it is out of the signal range. In this case Tx3 incurs interference to Rx3.

In summary, the antenna directionality and spatial location of transceivers causes fully distributed channel assignment schemes to be inefficient in some scenario. The distributed transceivers should either exchange channel usage information or learn the radio environment, to avoid such negative impact.

4.2.2 *Bottleneck Issue*

The multi-hop backhaul architecture, as presented in Figure 3.2, has a HBS connecting a set of ABSs. Traffic generated from a source ABS should pass through multiple hops until arriving at the HBS, and vice versa from the HBS to a destination ABS. The system performance thus relies on the end-to-end QoS, which is constrained by a single hop with lowest QoS. It can be observed from the network architecture Figure 3.2 that a traffic flow can be delivered only if all hops between source and destination are assigned channels. The number of channels required on hop h_i, including link $L_{A_{x(i-1)}A_{xi}}$ and $L_{A_{xi}A_{x(i-1)}}$ follows

$$C(\,h_i) = \max(h) - i + 1 \tag{4.1}$$

This indicates that lower hops require more channels to deliver relayed traffic than higher hops. On the other hand, the backhaul architecture indicates that the HBS suffers from higher interference than the ABSs, because the hub connects all the multi-hop branches.

In summary, a drawback of multi-hop network is that more resources are required to deliver relayed traffic compared to single-hop network. However, with directional antennas implemented, a multi-hop topology significantly reduces interference density on the HBS compared to single-hop topology. A better spatial resource reuse can be carried out to reduce the relay burden.

4.3 Space-division Interference Coordination

4.3.1 *Interference Coordination Mechanism*

The interference coordination resource management mechanism has been applied recently to some distributed networks without a centralized RNC, as reviewed in Chapter 2. One typical application is in an LTE network, where the adjacent eNBs are allowed to exchange channel usage information to avoid the same sub-spectrum being used simultaneously, namely Inter-Cell Interference Coordination (ICIC) [23]. The motivation of interference coordination is to eliminate interference in a defined area through information exchange.

Following the analysis of the interference environment on a multi-hop backhaul network in Section 4.2, a straight-forward coordination strategy is to define the interference range of a link covering neighbouring hops. The interfering links L^I_{TxRx} of a communication link $L_{A_iA_j}$ are defined as

$$\forall L_{A_iA_j}, L^I_{TxRx} = \bigcup(Tx \vee Rx = A_i \vee A_j) \tag{4.2}$$

In the basic interference coordination strategy, the communication link $L_{A_iA_j}$ exchanges channel usage information with the interfering links L^I_{TxRx} before every file transmission. Simulations in later sections prove that such a strategy effectively eliminates most of the dropped links caused by the hidden terminal problem.

4.3.2 *Spatial Reuse Methodology*

Spatial reuse is designed to improve the resource utilization in the spatial dimension. The directional antennas implemented on backhaul links effectively reduce interference radiated in unwanted directions, which potentially provides further resource reuse capability in the spatial domain.

It has been illustrated in Figure 4.1 that neighbouring links along the same direction incur excessive interference. However, interference from those links in the reverse direction may be controlled by directional antennas, even though they are in the interference range.

ABS Spatial Reuse

Figure 4.2 illustrates the designed space-division resource allocation strategy. A two hops ABSs network is presented in the example architecture, with both downlinks and uplinks constructed. The downlink $L_{A_1A_2}$ and uplink $L_{A_3A_2}$ have receiver antennas on the same ABS A2, pointing in opposite directions. In this case interference between them could be fairly low, according to the directional antenna profile. Resources can be reused on these two neighbouring links without interference. Similarly $L_{A_2A_1}$ and $L_{A_2A_3}$ have the same behaviour. In general, spatial reuse can be operated on links L^R_{TxRx} if

$$\forall L_{A_iA_j}, L^R_{TxRx} = \bigcup(Tx = A_i \vee Rx = A_j) \tag{4.3}$$

Compared to (4.2), the number of channels required with spatial reuse is only a half of the original interference coordination strategy, which is contributed by the channel reuse between neighboring uplinks and downlinks.

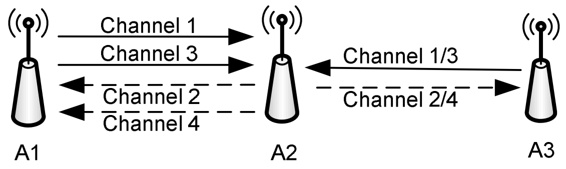

Figure 4.2. Spatial reuse between multiple hops

A crucial objective of spatial resource reuse is to reduce the relaying burden on multi-hop networks. Traditionally increasing a multi-hop link by one hop requires one more channel to be assigned for relayed traffic. However, the relay link could reuse channels selected by a neighbouring reverse link when spatial reuse is introduced. In this manner, no additional channels are required compared to a single-hop topology with the same number of ABSs.

This can be illustrated in the example architecture Figure 4.2, where the second hop fully reuses channels assigned on the first hop. It can be observed that the same number of channels is required when constructing this network using a single hop architecture, where individual links are established to connect each hop with A1. The same methodology applies to other hops, where the lower hop can always reuse channels assigned to the neighbour higher hop. In conclusion, the space-division resource allocation strategy effectively eliminates the relay burden caused by multi-hop architecture.

HBS Spatial Reuse

The space-division interference coordination can be applied not only between neighbouring hops on ABSs but also between different branches on a HBS. Figure 3.2 demonstrates that a HBS connects multiple branches in different directions. Directional antennas are implemented to isolate interference between the links in the same direction, where channels can be reused. In this case, spatial reuse can be operated within downlinks or uplinks.

An important issue for spatial reuse on the HBS is that the directional antenna radiates a lower signal power on the sidelobes as well, which has the potential of interfering with adjacent links on the same direction. This can be illustrated from an example in Figure 4.3, where $L_{A_0A_1}$ and $L_{A_0A_2}$ have a small angle. The signal ranges of these two links overlap according to the antenna profile. As a result, spatial reuse between these links may incur excessive interference. A similar issue occurs between uplinks $L_{A_1A_0}$ and $L_{A_2A_0}$. On the other hand, A_3 is out of the signal range of A_1 and A_2. Thus $L_{A_0A_3}$ can reuse channels on $L_{A_0A_1}$ and $L_{A_0A_2}$, and vice versa on reverse links.

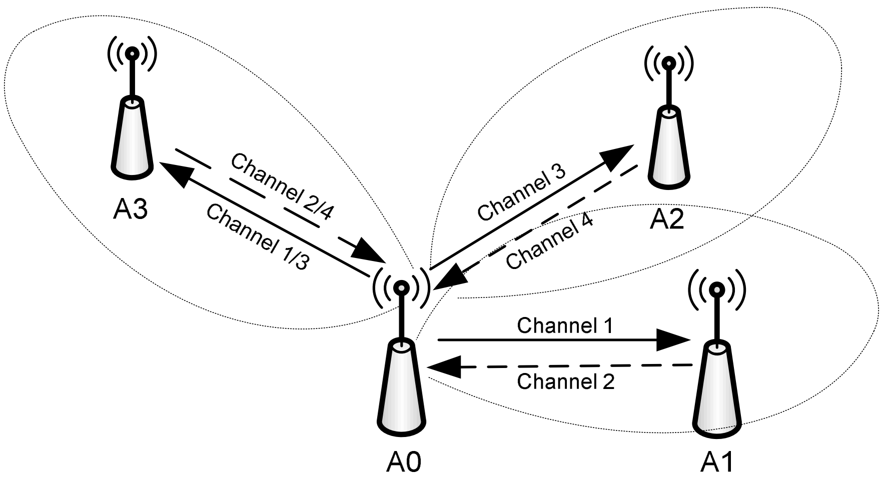

Figure 4.3. Spatial reuse between multiple branches

4.3.3 *Simulation*

In this section, simulation results are presented to validate the Space-division Interference Coordination strategy, which is based on the general architecture presented in Figure 3.2. A network topology with 38 ABSs connected to a central HBS on 8 branches is used, based on the topology shown in Figure 4.4 below.

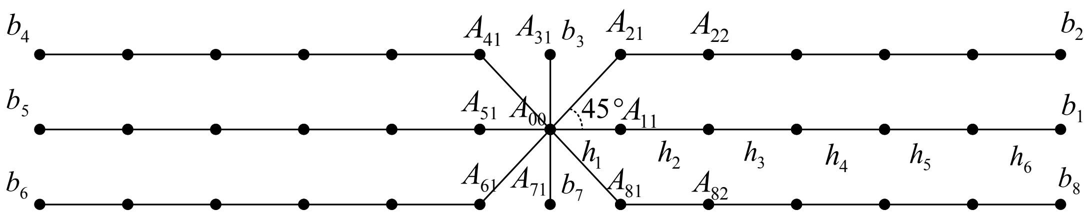

Figure 4.4. Multi-hop backhaul network simulation topology

The simulation parameters for this section are shown in Table 4.1. The Minimum Interference channel selection strategy [56] is used as a baseline comparison, which selects channels with minimum interference level at the transmitter. The interference coordination with no spatial reuse, ABS spatial reuse and HBS spatial reuse are evaluated as described in Section 4.3 before.

Table 4.1. Simulation Parameters

Parameters	Values
Transmission Power	-50 dBm
Antenna Model	Aperture antenna [96]
Antenna beamwidth	30°
Antenna height	5 m
Propagation Model	HATA PCS Extension (Urban) [108]
Channel Bandwidth	12 MHz
Thermal Noise[a]	-174 dBm/Hz
SINR Threshold[b]	9.05 dB
Distance between ABSs	30 m

a. Noise power in a resistor at room temperature [109]

b. QPSK and 7/8 coding rate at 10^{-6} BER [17]

The first part of simulation investigates the link level, by assigning each ABS with 1 Erlang offered traffic. Figure 4.5 shows the number of dropped links with different numbers of available channels.

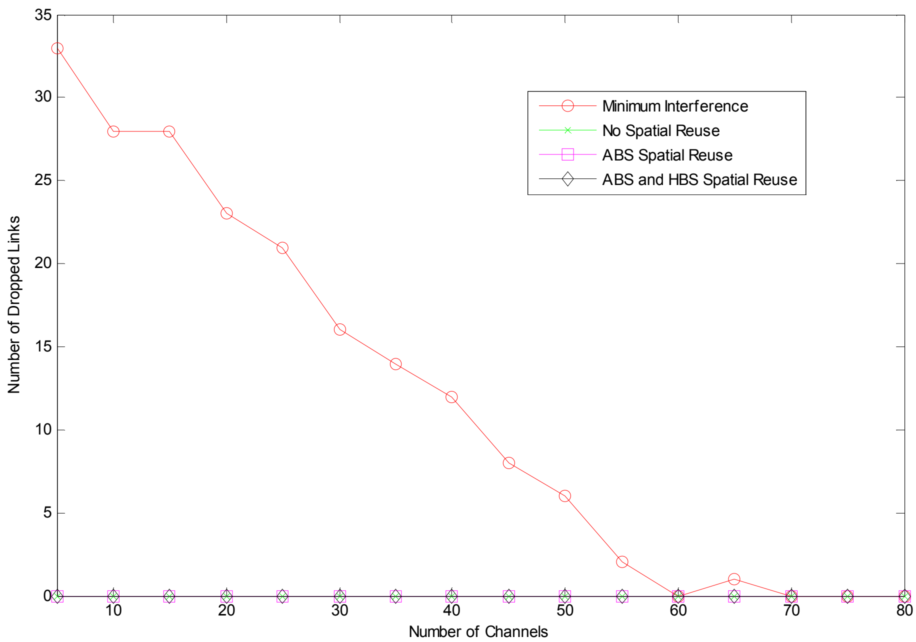

Figure 4.5. Number of Dropped Links

It can be observed that the conventional Minimum Interference scheme starts from a high number of dropped links, because it can hardly avoid the interference from hidden terminals. On the other hand, the three Interference Coordination schemes are shown to effectively control the interference. Moreover, the spatial reuse schemes prevent dropped links as achieved by the no spatial reuse scheme, which validates the methodologies presented in Figure 4.2 and Figure 4.3.

Figure 4.6 presents the number of blocked links, which is caused by insufficient available channels. It can be investigated that by introducing channel reuse between neighbour hops of ABSs in (4.3), the ABS spatial reuse scheme slightly improves the blocking performance compared to the no spatial reuse scheme. On the other hand, the spatial reuse scheme on both ABSs and HBS achieves much lower blocked links than the other two coordination schemes. The result illustrates that a significant traffic bottleneck occurs on the first hop of links connecting a HBS and multiple branches, which largely affects the system performance. The space-division interference coordination strategy is shown to effectively eliminate the bottleneck issue by introducing spatial reuse appropriately between neighbouring uplinks and downlinks. Moreover, the Minimum Interference scheme has no blocked links because it has no constraints in selecting a channel from the entire shared pool. However, a high volume of dropped links demonstrated in Figure 4.5 results in poor overall network performance.

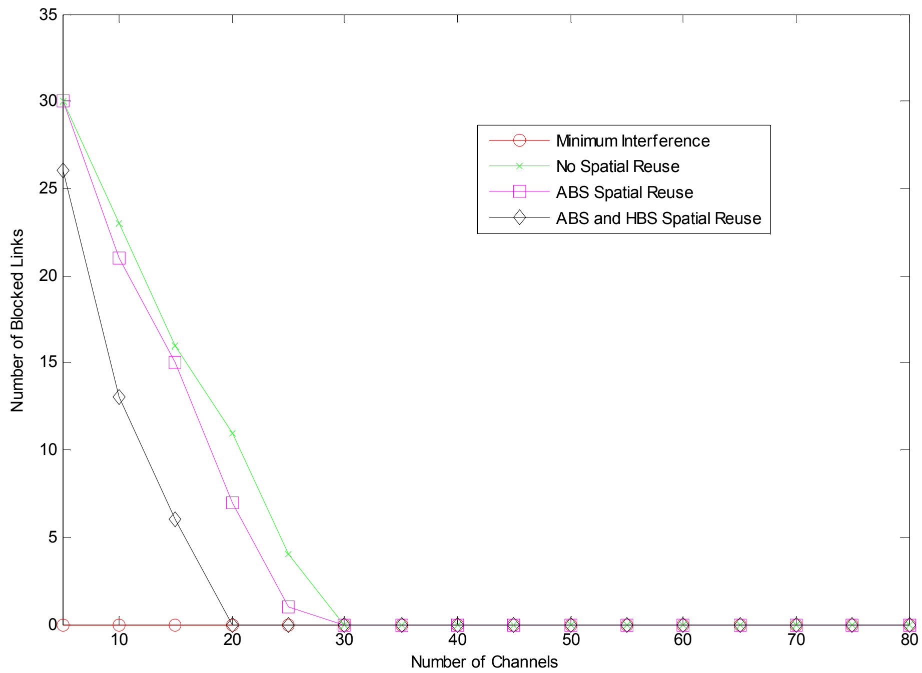

Figure 4.6. Number of Blocked Links

The impact of the antenna beamwidth has the potential to reduce the capability of spatial reuse between different branches, as illustrated in Figure 4.3. This could largely affect the network QoS. Figure 4.7 demonstrates the number of blocked, dropped and overall failed links for a range of different antenna beamwidths when the HBS spatial reuse scheme is applied. By comparing with the network topology in Figure 4.4, it can be seen that blocked and dropped links occur from 40° onwards, because the mainlobe of neighbour HBS antenna beams starts to have overlap with each other. Moreover, the number of dropped links reaches a peak level near 90° and 180°, where the edge of a HBS antenna mainlobe covers the receiving ABSs on adjacent branches. On the other hand, the number of dropped links can be reduced when the beamwidth varies between these peak levels, because interference on the receiving ABSs is high enough to be detected.

It can be concluded here that the designed spatial reuse strategy between different branches (Figure 4.3) could be applied only if the antenna beamwidth is smaller than the mainlobe's angle with adjacent links.

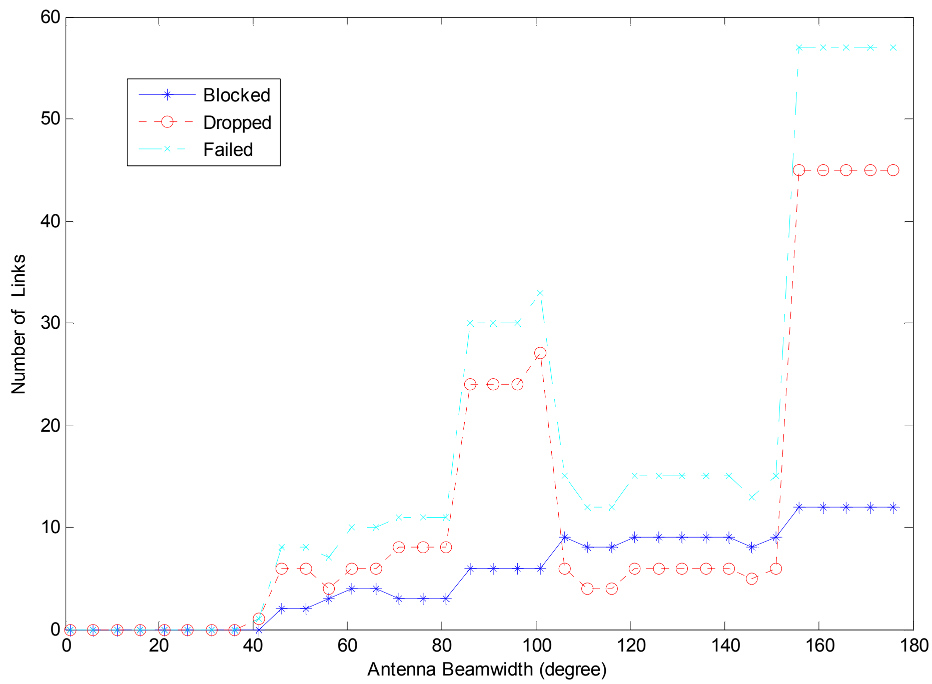

Figure 4.7. HBS Spatial Reuse performance according to antenna beamwidth

In the second part of the simulation, the traffic level performance is investigated. To assist traffic modelling, a file transfer model is introduced, which is capable of representing a succession of packets. The detailed modelling process has been stated in Section 3.4.3. The total number of channels is set to 30.

Figure 4.8 presents the network throughput performance along with offered traffic. It is demonstrated that the network can deliver higher throughput when a more flexible spatial reuse is applied under 30° beamwidth. The ABS and HBS Spatial Reuse scheme can afford a much higher offered traffic than others, because network resources are reused at a maximized level between different hops and branches. Compared with this, the ABS Spatial Reuse scheme has much lower throughput, because it is constrained by the bottleneck relayed traffic on the HBS. The No Spatial Reuse and Minimum Interference schemes achieve similar performance. However, Minimum Interference incurs more dropped links as illustrated in Figure 4.5, which is supposed to be more harmful than blocked links in a network [4].

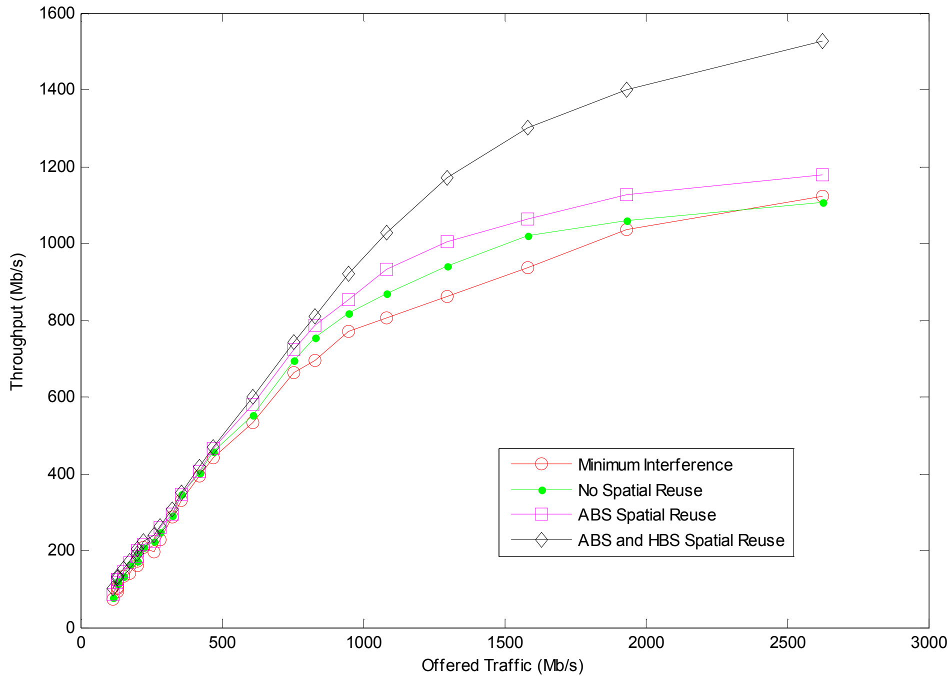

Figure 4.8. Network Throughput

The total network delay is presented in Figure 4.9, which is the accumulated delay of all files delivered. A similar performance of all schemes is shown compared with throughput, where the HBS and ABS Spatial Reuse scheme contribute to a lower network delay before the network is saturated. The throughput and delay performance shows that by applying spatial reuse between the transmitters or the receivers, the network capacity can be enhanced by reducing the number of retransmissions from fewer blocked and dropped links.

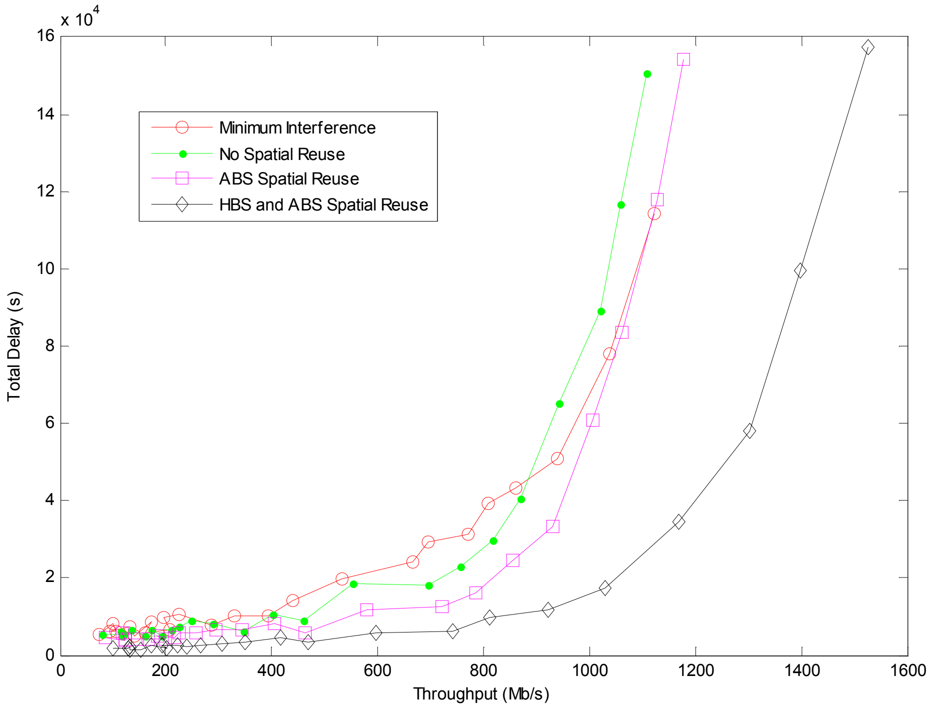

Figure 4.9. Network Delay

4.4 Distributed Reinforcement Learning

This section demonstrates the application of distributed reinforcement learning to channel allocation on a multi-hop backhaul network. The objective of this technique is to operate the network in a fully distributed manner, without multi-agent coordination and information exchange. The application model of reinforcement learning to multi-hop networks will firstly be presented. Two typical distributed reinforcement algorithms will be proposed and discussed, followed by analysis on convergence performance. Moreover, an improved decision making strategy using interference information from spectrum sensing will be demonstrated.

4.4.1 *Cognitive Models for Multi-hop Networks*

The aim of distributed reinforcement learning in channel allocation is to partition channel sets for links in different geographical areas, which in turn reduces the interference between them. A cognitive radio cycle was originally defined in [58] as *observation, making decisions, taking actions, and learning*. In radio resource management, the action space $a = \{a_1, a_2, \dots, a_n\}$ is an available spectrum pool with multiple channels (actions) a_k. In the decision making process, an agent selects a channel according to $a_k = \pi(a)$ [66], where π denotes a decision making policy. This is based upon a knowledge base constructed from a set of actions associated with Q values. The learning strategy updates the knowledge base, following the outcome of the selected action such as when a file is: *(a) successfully delivered, (b) interrupted during transmission or (c) initially blocked*. A successful outcome reinforces the policy by increasing the associated Q value whereas a failed outcome reduces the Q value and hence the probability that the action is employed next time. A successful action will have a higher accumulated Q value in the knowledge base.

The motivation for applying distributed reinforcement learning to radio resource management is to improve QoS including throughput, delay, retransmissions, etc. On a selected channel these parameters are largely affected by the SINR level γ, as defined in (3.8). The reward function is designed to use the Q value to represent the outcome of decisions, which is designed as follows

$$R_{a_k} = \begin{cases} 1 & \gamma \geq \gamma_{\min} \\ -1 & \gamma < \gamma_{\min} \end{cases} \tag{4.4}$$

where R_{a_k} is reward value on a selected channel a_k, $\gamma_{\min}$ is the minimum acceptable SINR threshold for establishing a communication link. The objective of learning is then to maximize Q values on successful transmissions and minimize those on interrupted or blocked transmissions. In the dynamic spectrum access scenario, the expected channel set for a distributed agent varies with the dynamic behaviour of other agents in both time and spatial domains. Therefore, it is difficult to define a target action space for reinforcement learning.

In a multi-hop backhaul network, individual Q tables are created on each link in Figure 4.4, in order to perform learning in a distributed manner. The Q value is updated on a link-by-link basis in the situations where a connection is blocked, interrupted, or released. The learning behaviour on each hop is independent.

4.4.2 *Distributed Reinforcement Learning Algorithms*

The learning algorithm is designed to reinforce the knowledge base for future decisions, by applying the rewards (4.4) on every iteration. Under this operation, both the historical and instantaneous information will be introduced into the knowledge base. The learning function has the responsibility in controlling the proportion of them in different learning stages.

The objective of the learning algorithm is to find an improved action space for distributed cognitive agents, as well as converging speedily to this space. These two processes operate concurrently in a cognitive radio network. However, a targeted action space could be invisible and dynamic to cognitive agents during learning period. The action-value function is designed to intelligently find the improved action space as iterations are taken, and keep the decisions stable once this space is approached.

4.4.2.1 Linear Reinforcement Learning

Linear Reinforcement Learning was initially proposed in [64], which defines an action-value function which updates the knowledge base on every learning iteration:

$$Q(t) = fQ(t-1) + R \tag{4.5}$$

where t is the number of learning iterations conducted, Q is an array of Q values assigned to each possible action, representing the knowledge of decisions made in the past. f is a weighting factor that controls the impact of rewards on Q value, as well as the convergence speed.

The transition function contains both *historical* $Q(t-1)$ and *instantaneous* R information. The weighting factor f determines the proportion of these two parts of information in building up the knowledge base for decision making. The reward function (4.4) represents two possible reward states: success and failure of transmissions. Clearly the character of the Q array is determined by the decision making history of an agent. Analysis of how the rewards affect Q value provides a valuable insight into mechanisms later described that aim to reduce disruption of service due to the learning process.

We begin by considering the dynamics of Q when a protracted sequence of the same rewards occurs. We denote this reward state as S_i. We will then look at the effect on Q of switching to a different reward state, S_{i+1} (e.g. from a sequence of successful actions to one of unsuccessful actions – although the treatment is equally valid in the reversed transition).

So, according to (4.5) the Q value after t iteration in reward state S_i is

$$Q_{S_i}(t) = f^t Q_{S_i}(0) + f^{t-1}R + \cdots + R \tag{4.6.1}$$

$$= \begin{cases} f^t Q_{S_i}(0) + \left(\dfrac{1-f^t}{1-f}\right)R, & f \neq 1, f > 0 \\ Q_{S_i}(0) + tR, & f = 1 \end{cases} \tag{4.6.2}$$

$$= \begin{cases} f^t \left(Q_{S_i}(0) - \dfrac{1}{1-f}\right)R + \dfrac{R}{1-f}, & f \neq 1, f > 0 \\ tR + Q_{S_i}(0), & f = 1 \end{cases} \tag{4.6.3}$$

Here $Q_{S_i}(0)$, f and R are constant factors, only the iteration number t increases. The time derivative of $Q_{S_i}(t)$ is

$$\frac{d}{dt}Q_{S_i}(t) = \begin{cases} f^t \left(Q_{S_i}(0) - \dfrac{1}{1-f}\right)R \ln f, & f \neq 1, f > 0 \\ R, & f = 1 \end{cases} \tag{4.7}$$

$$\lim_{t \to \infty} \frac{d}{dt}Q_{S_i}(t) = \begin{cases} 0, & 0 < f < 1 \\ R, & f = 1 \\ \infty, & f > 1 \end{cases} \tag{4.8}$$

It can be concluded from (4.7) and (4.8) that the time rate of change of $Q(t)$: (a) $f \in (0, 1)$: decreases exponentially; (b) $f = 1$: stays constant; (c) $f \in (1, \infty)$: increases exponentially. The gradient also indicates the proportion of historical and instantaneous information in the Q value.

Now, consider a reward state transition occurring after n iterations: R is now returned as a different value in (4.4). Resetting $t = 0$, we have a reward state transition:

$$Q_{S_{i+1}}(0) = Q_{S_i}(n) \tag{4.9}$$

By defining I(.) as either the proportion of historical learnt or newly acquired reward information in the knowledge base, equation (4.6.2) indicates that $I(Q_{S_i}(n))$ and $I(R)$ in $Q_{S_{i+1}}(t)$ follow the distributions listed at Table I:

Table 4.2. Historical and Instantaneous Information

	$I(Q_{S_i}(n))$	$I(R_{S_{i+1}})$
f∈(0,1)	Exponential Decay	Exponential Growth
f=1	Constant	Linear Growth
f∈(1,∞)	Exponential Growth	Exponential Decay

From Table I, it can be seen that historical information, the Q value contribution from previous reward states, decreases quickly and the acquired reward information increases dramatically when f∈(0,1). The converse behaviour occurs when f∈(1,∞). These ranges of f will result in either historical or newly acquired information being quickly lost.

In [110] we chose f=1 for a strategy which adopts linearly increasing reward information following a reward state transition $S_i \to S_{i+1}$. Since $Q(0)$=0 in this case, the knowledge base can be decomposed into:

$$Q(t) = t_{S_+}R_{S_+} + t_{S_-}R_{S_-} \tag{4.10}$$

where S_+ and S_- are the sets of all actions that incur positive or negative rewards, respectively. The reinforcement learning process naturally partitions S_+ and S_- through the decision making policy:

$$\pi(a) \in \arg\max Q_{a_k}(t) \tag{4.11}$$

The Q table is set up with arbitrary values in the start-up stage when a limited number of actions has been taken. Decisions are thus made on a random basis, which may cause harmful actions.

4.4.2.2 Single State Q-Learning

Single State Q-Learning was originally proposed in [70], as a Q learning solution to systems without defined multiple states. The distributed cognitive radio network, as illustrated in previous sections, is generally a stateless system where a learning target is dynamically changing according to the radio environment. As a result, this single state Q learning is potentially suitable for the DSA scenario.

The action-value function of Single State Q Learning is defined as

$$Q(t) = (1-\alpha)Q(t-1) + \alpha R, \lambda \in (0,1) \tag{4.12}$$

where the convergence speed is controlled by the learning rate α.

Compared with standard multi-state Q learning presented in Section 2.4.2, this algorithm takes the "discount factor" as 0. The component regarding to previous states in the equation is not included.

Compared with linear reinforcement learning, it can be observed that the learning rate α and the weighting factor f on $Q(t-1)$ follow such relation:

$$\alpha = 1 - f, \text{ if } f \in (0,1) \tag{4.13}$$

As a result, the relationship between *historical* $Q(t-1)$ information and the control parameter in Q learning is contrary to that in linear reinforcement learning, as illustrated in Table 4.2. Moreover, the reward component is also controlled by the learning rate.

To better compare this algorithm with linear reinforcement learning, we begin the same analytical process which investigates the dynamics of Q under different reward states. According to (4.12) the Q value after *t* iterations in reward state S_i is

$$Q_{S_i}(t) = (1-\alpha)^t Q_{S_i}(0) + \alpha R\left((1-\alpha)^{t-1} + (1-\alpha)^{t-2} + \cdots + (1-\alpha)^0\right) \tag{4.14.1}$$

$$= (1-\alpha)^t Q_{S_i}(0) + \alpha R\left(\frac{1-(1-\alpha)^t}{1-(1-\alpha)}\right) \tag{4.14.2}$$

$$= (1-\alpha)^t Q_{S_i}(0) + R\left(1-(1-\alpha)^t\right) \tag{4.14.3}$$

Here $Q_{S_i}(0)$, α and R are constant factors, only the iteration number *t* increases. The time derivative of $Q_{S_i}(t)$ is

$$\frac{d}{dt}Q_{S_i}(t) = (1-\alpha)^t \ln(1-\alpha)(Q_{S_i}(0)-1) \tag{4.15}$$

$$\lim_{t->\infty}\frac{d}{dt}Q_{S_i}(t) = 0 \tag{4.16}$$

It can be concluded from (4.16) that $Q(t)$ stays consistent after several of iterations. According to (4.15), the convergence rate of $Q(t)$ depends on α. Moreover, equation (4.14.3) indicates that *historical* information$Q(t-1)$is exponentially decreased and the *instantaneous* information R is exponentially increased. Following (4.14.3), the converged Q value can be obtained from

$$\lim_{t->\infty}Q_{S_i}(t) = R \tag{4.17}$$

This indicates that Q learning converges to the reward value given by (4.4). It can also be deduced that when the same reward state transition function (4.9) is applied, Q learning will converge to the new reward value in that state.

4.4.2.3 Convergence Comparison

The analysis of the dynamic variation of Q values in these two distributed learning algorithms shows that they have significant different temporal behaviours in a cognitive radio scenario. The linear reinforcement learning provides more random exploration during the initial stage, with more steady decisions on a long-term basis. On a contrary, the single state Q learning converges quickly (managed by the learning rate) to the reward value. However, in the cognitive radio scenario, the rewards may be changing very quickly because of a highly dynamic radio environment. In this case, the Q values could be fluctuating very frequently and an expected action space can hardly be found. Furthermore, as the historical information in single state Q learning is exponentially decreased, a cognitive agent may quickly loose learnt information in previous reward states.

The objective of distributed learning in a cognitive radio scenario is to partition the resource set to different agents. Following such motivation, a distributed agent should finally make steady decisions on a converged action space. The highly dynamic radio environment could result in highly fluctuating rewards on the converged action space. However, the learning algorithm should not be affected too much by these rewards, because steady decision is of higher priority. In this case, the linear reinforcement learning keeps more information on the iterations learnt in the past reward states but the Q learning relies heavily on the most recent state. It can be concluded that linear reinforcement learning is more suitable to achieve a stable solution, because it is less sensitive to reward changes.

4.4.3 *Interference Weighted Decision Making Strategy*

Exploration and exploitation are two fundamental stages in the cognitive cycle shown in Figure 2.2, which enables the agent to explore the environment to gain new information, or exploit the information that it has learned [111]. In the start-up stage, exploitation provides fairly limited information because the Q table has highly arbitrary values. Conventional reinforcement learning algorithms make random decisions in the exploration process, and also in the exploitation process when a set of actions have the same Q value. However, such random decision could be very harmful to the learner and surrounding agents, particularly at the start-up stage.

The traditional distributed dynamic channel assignment strategy with spectrum sensing [8] is effective in the DSA scenario. For instance in the Minimum Interference (MI) channel assignment scheme, the transmitter or the receiver sense the instantaneous interference level within the spectrum pool and assign channels with minimum interference in a random order. [8] presents a heuristic interference threshold based MI scheme in a cellular network where the user locations fit a uniform distribution. However, the interference level at a cognitive agent is an approximate determination of their distance to others. Hence it is difficult to set an interference threshold for all the base stations. Moreover, a channel with the minimum interference level at the transmitter may not be the best channel for the receiver, and on the contrary, a good channel at the receiver could interfere with the links near the transmitter.

The physical information from channel sensing can be used as an estimation of the channel quality. Here we have designed an improved decision making policy: the Interference Weighted (IW) strategy. In this strategy, the probability of selecting a channel depends on its interference level. The idea is to give a smaller probability for the channels with higher interference level to be selected, in order to achieve low interference at both ends of the link. In this scheme, channels are weighted by interference as

$$W_{a_i} = \frac{1}{\sqrt{I_{a_i} + n}} \tag{4.18}$$

where I is the interference level, n is the thermal noise. The probability for channel a_i to be selected is

$$P_{a_i} = \frac{W_{a_i}}{\sum_{a_k \in A} W_{a_k}} \tag{4.19}$$

Here A denotes channels that have not been used between two base stations. With (4.18) and (4.19) the lower interference channels have higher probability to be selected, and vice versa.

The ε-greedy method [66] is widely used in many exploration strategies to explore channels with a probability of ε. The problem with the original approach is that a large amount of random selections have been taken initially when the information in the Q table is fairly limited [65]. Some approaches make decisions on a Boltzmann distribution on the Q values from learning [65, 111] to reduce the inaccuracy of the Q values. The interference weighted decision making strategy is based on the instantaneous interference level, which provides more accurate information, thereby delivering more effective decisions especially in the early stages.

4.4.4 *Simulation*

In this section, several simulations are conducted to validate the distributed reinforcement learning algorithm on a multi-hop backhaul network. The simulation parameters are shown in Table 4.3. The ε-Greedy exploration probability is 0.2. Steady state performance is evaluated from 10s onwards.

The random channel selection strategy is used as a baseline comparison, which selects channels based on a uniform distribution. This strategy is also integrated with reinforcement learning based channel selection.

The interference weighted strategy is proposed to improve the decision making process of conventional minimum interference and random strategies. Figure 4.10

and Figure 4.11 demonstrate a comparison between these schemes, where the random selection scheme does not require spectrum sensing.

Table 4.3. Simulation Parameters

Parameters	Values
Number of Branches	6
Number of Hops	3
Carrier Frequency	3.5 GHz
Transmit Power	7 dBW
Bandwidth	40 MHz
Number of Channels	30
Thermal Noise	-174 dBm/Hz
Inter-arrival time	Pareto distribution
Mean File size	5 Mb
Antenna Model	Multi-beam model from [99]
Propagation Model	WINNER II B5a [98]
Simulation Time	60 s

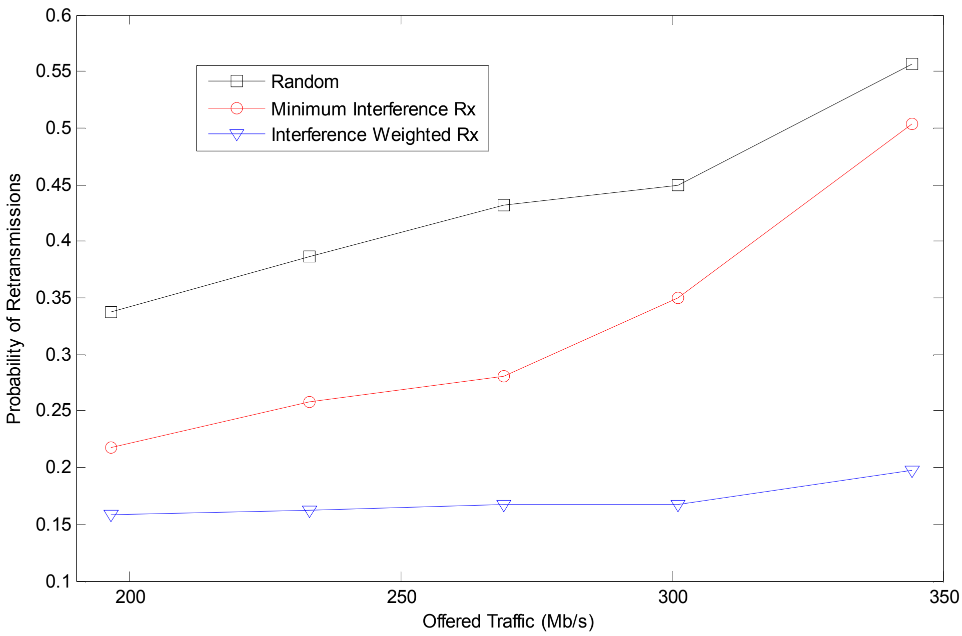

Figure 4.10. Probability of Retransmission (Decision Making strategies)

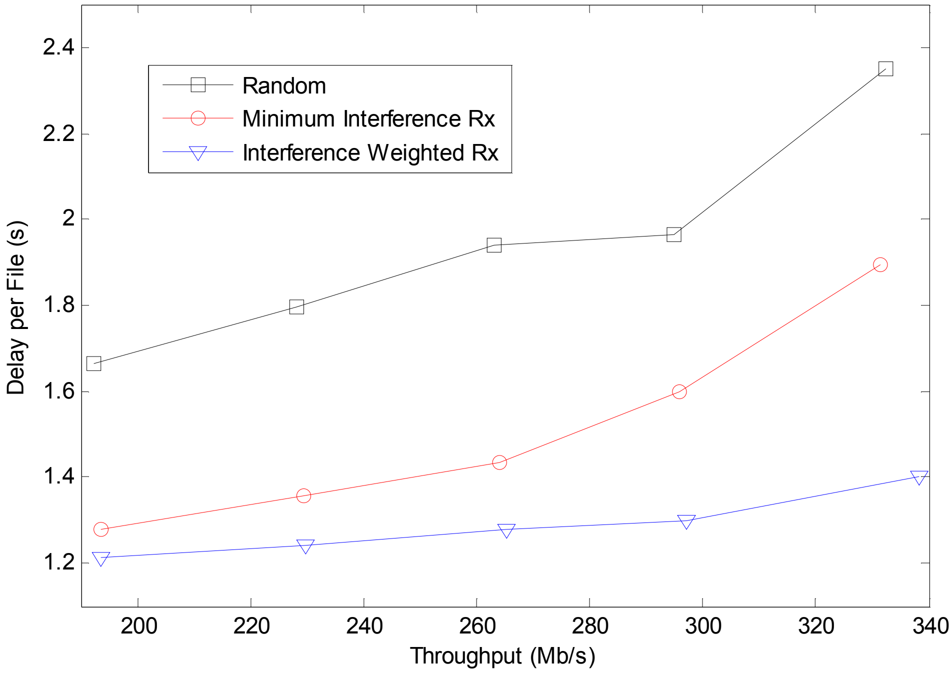

Figure 4.11. Throughput and Delay (Decision Making strategies)

It can be observed that the Interference Weighted strategy achieves lower retransmissions and delay than the Minimum Interference strategy. This demonstrates that channels with the lowest interference levels at a receiver may not be the best selection for communication. This process may cause a high level of interference to other activated receivers near the transmitter. The Interference Weighted scheme allows some probability for channels with higher interference to be selected, which is proven to be effective for the overall QoS. The random selection scheme is shown to achieve much higher retransmissions and delay than the other two, which demonstrates that spectrum sensing based decision making can significantly improve QoS over this range of traffic level.

Spectrum sensing that provides interference information can be operated on either the transmitter or receiver end of a link. Previous research in [56] shows that in a single-hop cellular network, selecting channels by sensing interference at the transmitter can support greater traffic than that at the receiver. However, for a multi-hop backhaul network, the interference at the receiver site from a neighbouring transmitter on the adjacent link may dominate the performance, which is because of the hidden terminal problem illustrated in Figure 4.1.

To compare the decision making strategies and validate the improvement of learning, we have performed simulations for transmitter (Tx), receiver (Rx) based IW and random strategies. Figure 4.12 shows the probability of retransmissions at various offered traffic levels.

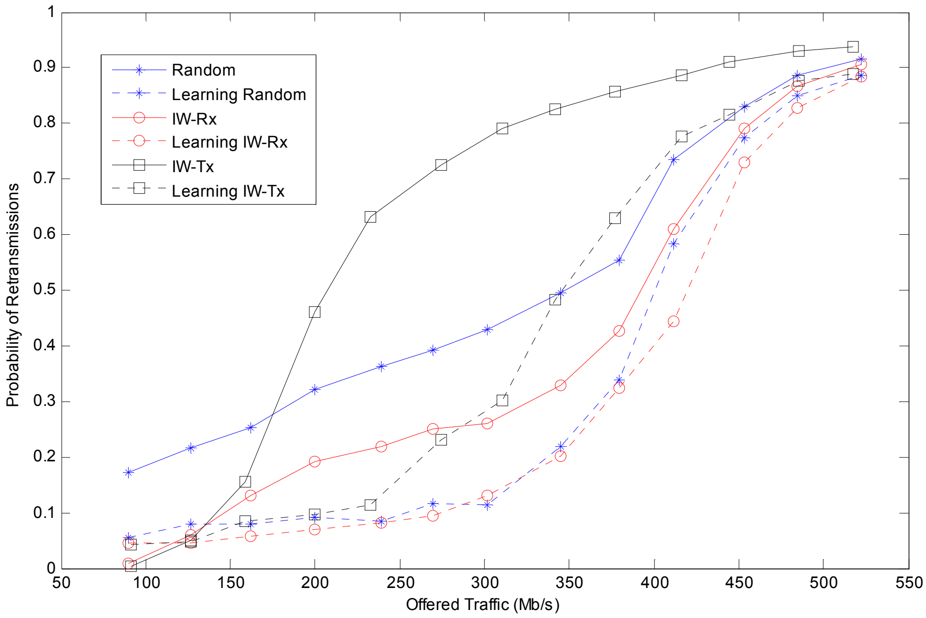

Figure 4.12. Probability of Retransmissions

It can be seen that without learning, the transmitter based interference weighted strategy has retransmissions 15% lower than random strategy when the offered traffic is lower than 150 Mb/s. However, it increases dramatically after and is much higher than the others. On the contrary, the receiver based IW strategy continuously performs with the lowest level of retransmissions, as the interference information at the receiver provides a more accurate estimation of the channel quality than that at the transmitter.

The linear reinforcement learning scheme can effectively improve the overall QoS to a large extent for all the decision making strategies. It is shown in Figure 4.12 that between the offered traffic levels of 150 Mb/s and 450 Mb/s, reinforcement learning improves up to 40% on transmitter based IW, 20% on random and 10% on receiver based IW. This is the most useful range of offered traffic where the network is neither idle nor saturated. It can be concluded that with the improvement of reinforcement learning, spectrum sensing based decision making strategies have less impact on the averaged performance.

Figure 4.13 below shows the overall network performance described by throughput and delay. The transmitter based IW strategy can support only up to 350 Mb/s throughput and with linear reinforcement learning it has higher delay than others. The receiver based IW strategy with learning performs the best with 1.50 s lower delay and 20 Mb/s higher throughput than the random strategy.

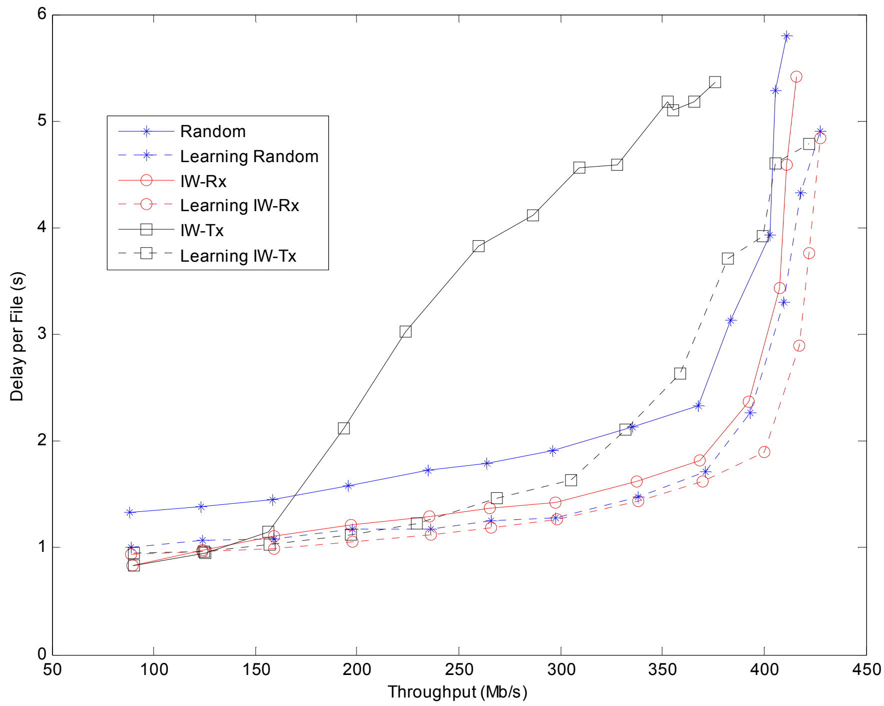

Figure 4.13. Throughput and Delay

Reinforcement learning is a delayed process where the base stations need to learn by taking a number of actions to obtain experience. The convergence rate is a crucial parameter that measures the efficiency of learning, which shows the time taken by the base station to learn the preferred channels. Slow convergence in learning can cause be harmful (in terms of excessive interference) to the base stations during the initial stages. Figure 4.14 presents the temporal performance at an offered traffic of 230 Mb/s (averaged over 10 simulation runs), in order to provide representative performance of the speed of the learning schemes.

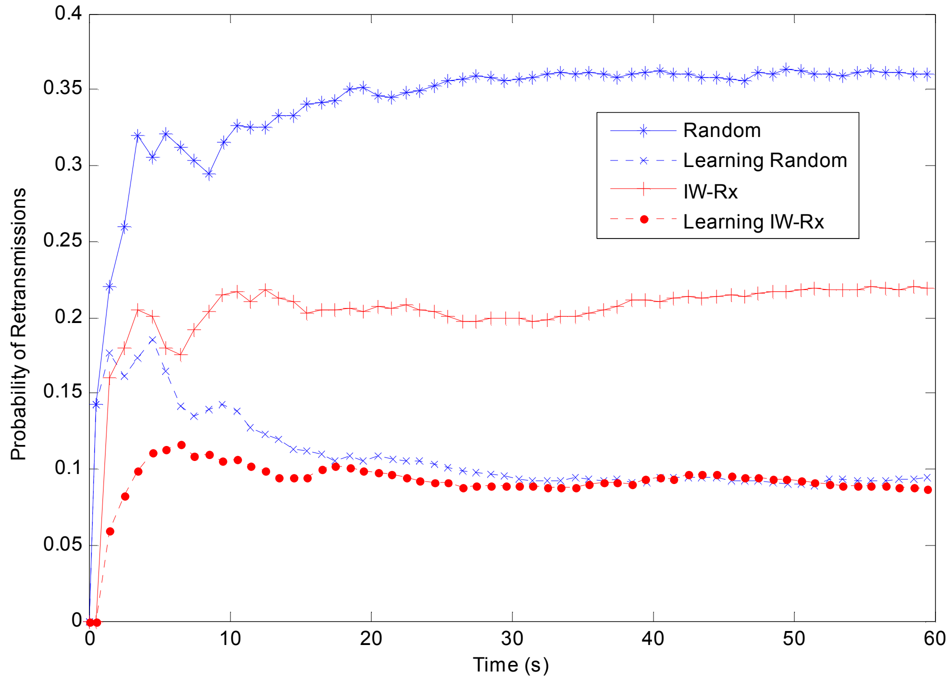

Figure 4.14. Temporal performance of QoS

It can be seen that in the early stage linear reinforcement learning with a random selection has a high level of retransmissions, approaching the level of the non-learning scheme. The linear reinforcement learning with receiver based IW strategy has 10% lower retransmissions at this stage and quickly converges to a stable QoS.

The probability of channel usage in Figure 4.15 shows the proportion of preferred channels selected by all the base stations in the network. By ranking the number of channels used in descending order on link j, we have a ranked channel set U_j, where $(U_{j1} > U_{j2} > \cdots > U_{jn})$. For an m links n channels network, the overall usage probability on channel i is defined as

$$P_i = \sum_{j=1}^{m} U_{ij} \Big/ \sum_{i=1}^{n} \sum_{j=1}^{m} U_{ij} \qquad (4.20)$$

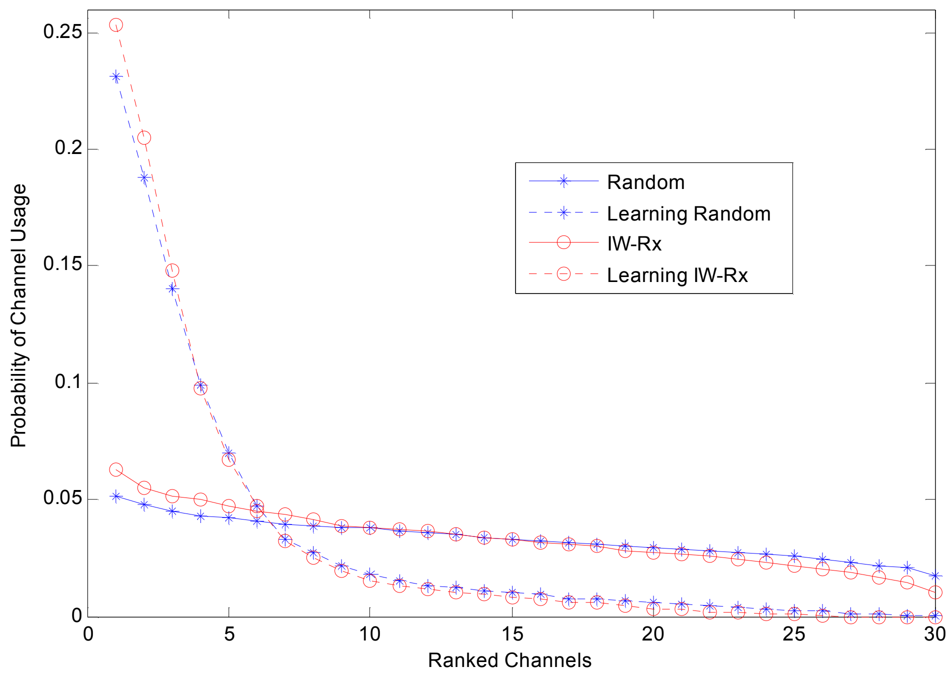

Figure 4.15. Probability of Channel Usage

From Figure 4.15 we can see that without learning the base stations selects different channels with nearly equal probability. In the learning schemes there are around 6 highly preferred channels. Accordingly 6 channels are never used, which is desirable since adjacent links with the transmitter and receiver on the same base station cannot in practice reuse channels. Compared with Figure 4.14, it can be seen that the reinforcement learning improves QoS through an effective channel partitioning.

4.5 Conclusion

This chapter has investigated a fully cooperative interference coordination and a fully distributed reinforcement learning strategy applied to resource management in a multi-hop backhaul network.

The interference coordination strategy is studied through channel usage information exchange before data transmission. A novel spatial reuse scheme has been developed on a multi-hop backhaul network, which allows channel reuse on adjacent transmitter or receiver antennas co-located on the same node, provided that the angle between them is larger than the antenna beamwidth. It is demonstrated in simulation that the hidden terminal problem is effectively controlled by interference coordination. Spatial channel reuse on both HBS and ABSs reduces the number of channels required for relaying, which thus significantly mitigates the bottleneck issue on multi-hop networks. The network throughput and delay is largely improved with fewer retransmissions and blocked/dropped links.

Distributed Reinforcement Learning is studied to allow base stations to learn the radio environment and carry out effective channel selection. The convergence behaviour analysis shows that linear reinforcement learning provides more information from the previous decision to the knowledge base than Q learning, which in turn provides more steady decisions in a dynamic radio environment.

The linear reinforcement learning scheme keeps the base stations on preferred channels as more actions have been taken, which assists with the establishment of stable end-to-end links. It is demonstrated that by effectively partitioning a set of channels to the base stations, the learning scheme achieves up to 30% lower retransmissions and 150 ms lower mean delay than random selection, and delivers similar steady QoS as achieved by the spectrum sensing based schemes. A novel Interference Weighted decision making strategy has been developed, which selects channels based on a probability generated from the interference level. It is shown to provide higher QoS than a conventional minimum interference scheme and it speeds up the convergence for reinforcement learning.

In general, this chapter provides analysis of the multi-hop backhaul network radio environment with novel interference coordination and distributed reinforcement learning strategies developed to deliver effective QoS and throughput. However, a fully coordinated strategy increases the complexity of protocol development while a fully distributed strategy requires long-term investigation to achieve stable performance. As a result, a potential better solution for cognitive network could be a partly distributed/coordinated strategy, which is supposed to achieve a balance between the coordination and QoS requirement.

Chapter 5. Transfer Learning with Cooperation Management

5.1 Introduction

In Chapter 4 two dynamic resource management strategies were presented for a cognitive multi-hop backhaul network: fully distributed learning and full interference coordination. It is demonstrated in Figure 4.12 and Figure 4.13 that distributed reinforcement learning provides effective QoS without spectrum sensing. Moreover, Figure 4.8 and Figure 4.9 illustrates that by applying multi-agent coordination for spatial reuse through channel usage information exchange, the network QoS has been significantly improved from interference mitigation compared to a sensing based Minimum Interference algorithm. However, the exchange of channel usage information between distributed agents incurs an excessive amount of control traffic, which is inefficient for self-organized networks.

The balance between Quality of Service (QoS) and control information overhead across distributed self-organized networks has been a key research issue in recent years. Self-organization of the network architecture becomes a compelling solution for simplified and efficient RRM [8]. A centralized frequency planning strategy is inflexible in supporting fluctuating offered traffic levels in different areas. Distributed interference coordination techniques impose high additional traffic loads on control links that exchange resource block occupancy information [36]. Therefore, a crucial objective of the next generation wireless networks is achieving efficient Quality of Service (QoS) in a distributed manner with low levels of information exchange.

Next generation wireless systems introduce the idea of implementing a flat architecture, to reduce system complexity and entity coordination [13]. On the other hand, interference coordination is demonstrated to significantly improve distributed resource management. For the purpose of providing effective distributed operation, it is possible to integrate a minimum amount of information exchange between distributed agents. The system is then expected to benefit from both distributed learning and interference coordination.

In this chapter, a brand new method for implementing distributed intelligent algorithm is introduced over a network based upon transfer learning [79]: the transfer of learning knowledge between multiple tasks. A learning task is modelled as the learning target on multiple agents. The use of transfer learning is demonstrated in this chapter where the exchange of appropriate information from surrounding agents that have an interference impact on the learner enables the learning process to converge more quickly to a better, more stable state. The intention is that by applying a cooperation management strategy in transfer learning, the information exchange between independent learning agents can be reduced to a minimum level whilst achieving learning performance close to that of a fully coordinated network. When compared to traditional, centralized or coordinated RRM mechanisms, transfer learning enables the degree of coordination to be significantly reduced. So, the target of transfer learning is to achieve an effective balance between cooperation overhead and QoS on distributed networks.

5.2 Transfer Learning: Value Training Method

Reinforcement learning is a delayed reward process [66] where agents usually have relatively limited information to inform policy when they are initially activated.

During this initial stage, and in the case here, channels are selected on an almost random basis as there is no discriminatory information.

In this chapter, a value training method is developed in the context of transfer learning, to enable the transfer of knowledge between multiple agents. It is designed to exploit prior learning by transferring a set of Q tables from related *source agents* to the local *target agent*. From the perspective of the network level, reinforcement learning is operated iteratively throughout multiple agents, depending on the source and destination of offered traffic. A cognitive agent could consequently either act as target agent when it is transmitting or receiving a file; or as a source agent when requested for information exchange from other source agents in vicinity. Knowledge can be transferred between all agents in the same interference environment, but not necessarily from an agent more knowledge to that with less.

Cooperation management is one of the most important modules in transfer learning. It controls the degree of knowledge transferred between multiple agents. The role of cooperation management is to identify and transfer useful information provided by source agents, as well as to stop transfer learning once it has no positive impact for the target agent. With cooperation management, the cognitive agent is expected to achieved significant higher QoS than distributed learning and lower information overhead than interference coordination.

The framework of transfer learning is illustrated in Figure 5.1. The cooperation management algorithm firstly decides whether a transfer learning is necessary. The source agent selection module is then operated to obtain Q tables from related agents in the vicinity. Finally the target agent training algorithm generates a new Q table under the information from multiple sources, for the next iteration.

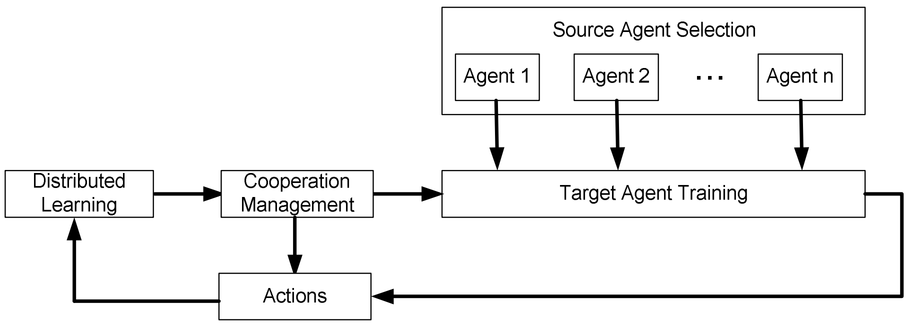

Figure 5.1. Transfer Learning: Value Training Method

The source agent selection, target agent training and cooperation management define where, what and when to transfer, respectively. The motivation of designing these algorithms is to utilize positive information and avoid negative information.

5.3 Source Agent Selection

The source agents in transfer learning are the agents that can provide useful information to the knowledge base on the target agent. In a wireless network scenario, the selection of source agents depends on the interference level incurred by the agents in the vicinity.

The SINR level potentially provides the information of interference impact from a dedicated agent, provided that only one source agent is sharing the same channel with the target agent. Thus an effective approach for selecting source agents is to send a beacon signal on a common control channel; the adjacent active agents reply with the SINR indication back to the target agent, which is then used to evaluate their qualification to be source agents. In the multi-hop network scenario, a cognitive agent has a transmitter and a receiver. The receiver side can measure SINR directly while the transmitter side can only sense the interference level. The source agents should include both transmitters and receivers in the vicinity, to reduce the hidden and expose terminal problems. Table 5.1 illustrates the process of the source agent selection scheme on both transmitter and receiver ends.

The SINR level γ can only be measured at the receiver end. P_{signal} is the received signal power from its corresponding transmitter, and P_{beacon} is the received beacon power from other agents. Transfer learning operated on a target agent is aimed at: 1) reducing the interference to receivers in vicinity; 2) avoiding interference from transmitters in vicinity. Thus the source agent selection strategy is different between transmitter and receiver ends of a target agent. Moreover, γ_{min} denotes the minimum adequate SINR level for a communication link as defined in Section 3.4.2.2, which is an important criterion for selecting source agents.

Table 5.1. Process of Source Agents Selection

Transmitter end of a target agent	**Receiver end of other active agents**
Send a beacon frame calling for Source Agents at the Receivers end	
	Send SINR $\gamma = \frac{P_{signal}}{P_{beacon}}$ to target agent
Qualify source agent if $\gamma < \gamma_{min}$	
Receiver end of a target agent	**Transmitter end of other active agents**
Send a beacon frame calling for Source Agents at the Transmitters end	
	Send a beacon frame back to target agent
Qualify source agent if $\frac{P_{signal}}{P_{beacon}} < \gamma_{min}$	

The radio environment on a multi-hop backhaul network has been analysed in Section 4.2, followed by interference coordination algorithms designed in Section 4.3. Transfer learning is designed as an algorithm integrating distributed learning and coordination, thus the source agent selection strategy is proposed to follow the interference coordination algorithm. Figure 3.2 and Figure 4.2 illustrated that under the spatial division from directional antennas, the receiver end of a target agent (e.g. A_{x_2} of $L_{A_{x1}A_{x2}}$) could be heavily interfered by a neighbour source agent that has its transmitter co-located (e.g. $L_{A_{x2}A_{x3}}$). On the other hand, the transmitter end of a target agent (e.g. A_{x_1} of $L_{A_{x1}A_{x2}}$) could cause high interference to a neighbour source agent that has its receiver co-located (e.g. $L_{A_{x0}A_{x1}}$). However, the other neighbour agents on the reverse link direction (e.g. $L_{A_{x3}A_{x2}}$ and $L_{A_{x1}A_{x0}}$) can in practice share the same channel with the target agent.

For a target agent $L_{TA} = L_{A_{x_1 y_1} A_{x_2 y_2}}$, the source agents L_{SA} are selected according to a *Source Agent Selection* scheme:

$$\forall x_i' y_i',\ L_{SA} = L_{A_{x_1'y_1'}A_{x_1y_1}} \cup L_{A_{x_2y_2}A_{x_2'y_2'}} \tag{5.1}$$

Following this approach, the relationship between potential good actions (a_+) and bad actions (a_-) on source and target agents is

$$\begin{aligned} L_{a_-(\text{Target})} &= L_{a_+(\text{Source})} \\ L_{a_+(\text{Target})} &= L_{a_-(\text{Source})} \end{aligned} \tag{5.2}$$

The source agent selection strategy identifies the agents that incur excessive interference. The knowledge transfer process, as illustrated in Figure 5.1, is operated in a single direction from multiple source agents to a single target agent.

5.4 Target Agent Training

The target agent training module defines an approach of transferring positive information from source agents' knowledge base to the target agent. The knowledge base in distributed reinforcement learning is represented in the form of a Q table. The target agent training scheme is designed to combine the Q tables from multiple source agents, and reinforce the target agent's knowledge base for decision making. The aim of this process is to maximize the positive impact and minimize the negative impact from source agents. There are two issues to be considered:

1. The position of source agents. The interference impact received from multiple source agents depends on their path loss, mainly transmission distance, to the target agent.
2. The action-value function of the distributed learning algorithm. The functions designed in transfer learning should accelerate the convergence process of distributed learning, and assist the distributed agents with identifying effective steady selections.

There are two major approaches to utilize the information provided by source agents. A straight forward method is to apply the spatial constraints introduced in Section 4.3. However, this only provides instantaneous information of the radio environment. Alternatively, the Q table learnt from distributed learning has the ability of assisting transfer learning, provided that a value training function is effectively developed to combine multiple Q tables from different agents.

5.4.1 ***Value training function***

In a linear reinforcement learning algorithm, the good actions have continuously increasing Q values and the bad actions have continuously decreasing Q values. The objective is to maximize positive Q values on good actions and minimize negative Q values on bad actions, because the actions with highest Q values are selected on each iteration. For the purpose of accelerating this process, the value training function is designed to append the Q table with transferred Q values.

According to the source agent selection strategy, the objective is to assist the target agent to avoid channel reuse with the source agents. In this scenario, the source agents' good channels could cause harmful interference to the target agent. On the contrary, the bad channels at source agents could be reused at the target agent. Consequently, the Q table from source agents could be added conversely to the target agent, which provides channel partitioning between them. According to the action-value function (4.10) of linear reinforcement leaning in Section 4.4.2, the objective of transfer learning is to maximize Q value on good actions $Q(a_+)$ and minimize Q value on bad actions $Q(a_-)$:

$$\begin{cases} Q(a_+) = \max\left(t_{S_+}R_{S_+}\right) \wedge \min(t_{S_-}R_{S_-}) \\ Q(a_-) = \min\left(t_{S_+}R_{S_+}\right) \wedge \max(t_{S_-}R_{S_-}) \end{cases} \tag{5.3}$$

Furthermore, the value training function should also balance the information from source and target agents, in order to avoid the Q table from either distributed or transfer learning dominating the knowledge base. The Q tables $Q_{(SA)}$ from source agents L_{SA} is transferred to the target agent $Q_{(TA)}$ as follows:

$$\forall a_k \in a, \quad Q_{a_k(TA)}(t) = Q_{a_k(TA)}(t) - \frac{\sum_{i \in L_{SA}} Q_{a_k(SA_i)}}{\left|L_{SA}\right|} \tag{5.4}$$

where $|L_{SA}|$ is the number of source agents. The value training function virtually exchanges R_{S_+} and R_{S_-} on t_{S_+} and t_{S_-} when transferring $L_{SA}(Q)$ to $L_{TA}(Q)$. Equation (5.5) illustrates this process, which follows the targets expressed in (5.3).

$$L_{TA}(Q(t)) = L_{TA}(t_{S_+}R_{S_+}) + \frac{\sum L_{SA}(t_{S_-}R_{S_+})}{\left|L_{SA}(t_{S_-})\right|} + L_{TA}(t_{S_-}R_{S_-}) + \frac{\sum L_{SA}(t_{S_+}R_{S_-})}{\left|L_{TA}(t_{S_+})\right|} \tag{5.5}$$

This function adds opposite Q values from source agents $Q_{a_k(SA)}$ to the target agent $Q_{a_k(TA)}$, which can improve linear reinforcement learning in two aspects: 1) a new agent can quickly find good channels from transferred positive values; 2) a mature agent can quickly recognize bad channels from transferred negative values.

5.4.2 *Space-division Coordination*

The space-division coordination introduced in Section 4.3 provides a straight forward method in spatial channel partitioning, which can be applied in transfer learning. The spatial constraints partially lock the available action space for decision making, which prevents the channels used on source agents to be selected by the target agent. Following the source agent selection strategy, the available channel set to a target agent C_A provided from spatial constraints is

$$C_A = C - C_T \cup \left(\bigcup_{\forall i \in L_{SA}} C_{S_i} \right) \tag{5.6}$$

where C_T and C_S are channels occupied by target and source agents, respectively.

The space-division coordination provides interim channel usage information to the target agent, which is expected to be more precise than the value training function. However, since transfer learning is expected to reduce the coordination overhead in a distributed network, the target agent may lose all information from source agents once the knowledge transfer is stopped. On the contrary, based on the value training function, the knowledge base has the memory of channel usage information, which provides long-term improvement.

5.5 Stable State Evaluation

Stable state evaluation is designed to identify whether a learning target has been achieved. In transfer learning, it is employed in cooperation management (detailed in the following section) to control information exchange and knowledge transfer.

An objective of learning algorithms is to achieve a stable state. In a cognitive resource management scenario, the decisions made from a cognitive agent on a stable state should converge to a fixed set of channels, which is also referred to as a solution learnt from consecutive iterations. Nevertheless, a stable state does not necessarily indicate that high QoS is achieved, which could potentially cause two extreme case: 1) Positive State: resource partitioning is achieved and adjacent agents converge to different set of channels; 2) Negative State: all the agents converge to the same channel set, which incurs excessive interference. The source agent selection strategy in transfer learning effectively avoids the negative stable state, based on (5.5).

In the computer science community, the stable state of a reinforcement learning scenario is normally well defined as a goal. The intelligent agents find the goal [72] based on the action-value function. However, it is difficult or unrealistic to define a stable state in a cognitive radio scenario, because the radio environment varies with offered traffic, spectrum size, location of adjacent agents, etc. In this case, the expected decisions may be dynamically changing as well. As a result, it is important to estimate various stable states in order to guarantee convergence.

A Q table in reinforcement learning provides two parts of information. 1) The level of the Q values indicate the probability of corresponding actions being selected; 2) The ranking of actions by Q value indicates the priority of each action. Namely, the learning information in a Q table is represented by the relative Q value between actions rather than the absolute Q value on each. A channel usage probability method is proposed in [110] to measure the channel partitioning status in a stabilized scenario. However, this cannot provide a unified measurement, because the number of stable channels required by an agent is a dynamic value depending on the traffic level. Additionally, the channel usage method only provides a protracted snapshot measurement to the system behavior, while evaluation on a specific iteration cannot be provided. As a result, an action ranking method is developed here to evaluate a stable state.

The decision making policy in reinforcement learning is based on the action with a high Q value. In a single action system, the stable state can be defined as the action with highest Q value which remains the same over iterations. However, in a wireless network, an agent may transmit files on multiple channels simultaneously, for either local and relay traffic. As the decision making process is based on the Q values from high to low, the selected action space is supposed to always have high Q values.

Therefore, a stable state is defined to exist when the rankings of the occupied channels C_R (sorted by their related Q value) stays consistent. The strategy is described as:

Table 5.2. Stable State Evaluation

On learning iteration t
1: Record the number of visited actions $N(t)$
2: Set $N(t) \& N(t-1) = \min(N(t), N(t-1))$
3: $C_R' = Sort(a)$, for $\forall\, a_i \in a, i \neq \lvert a \rvert$, $Q_{a_i} \geq Q_{a_{i+1}}$
4: Set C_R, where $C_R \in C_R'$ & $\lvert C_R \rvert = N(t)$
5: If $C_R(t) = C_R(t-1)$
6: Stable state reached
7: End

The convergence of learning is achieved when a stable state exists. This tool can be used to evaluate the convergence of learning in cognitive radio networks. In a stable state, the learning agent is expected to achieve a stable QoS level. Figure 5.2 and Figure 5.3 illustrate the relationship between stable state and QoS with linear reinforcement learning and the Q learning algorithms. The simulation is based on the architecture from Figure 3.2 with parameters listed in Table 4.3.

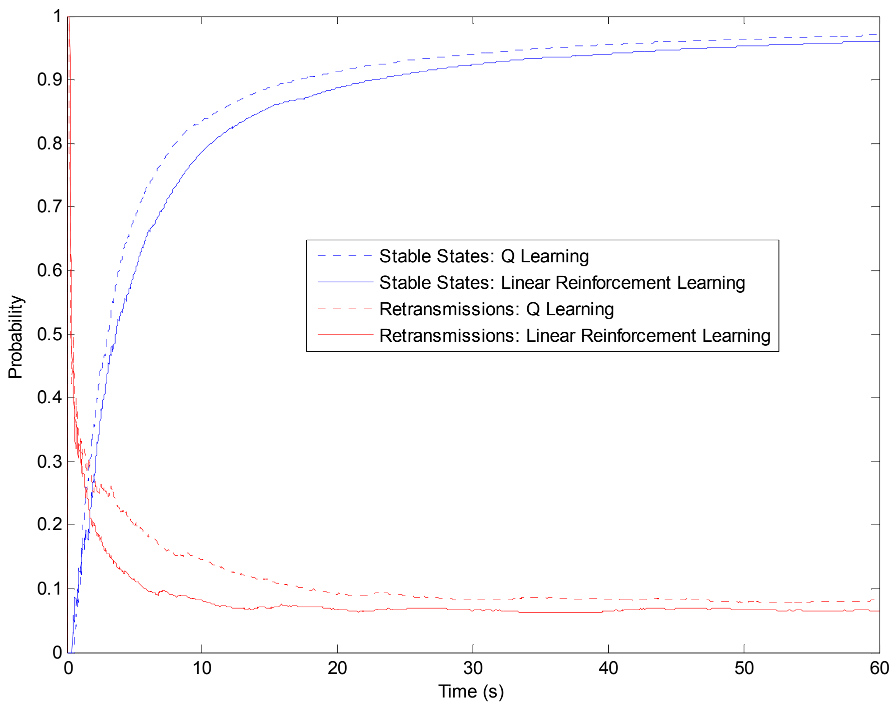

Figure 5.2. Stable States and Retransmissions: Low Traffic Level

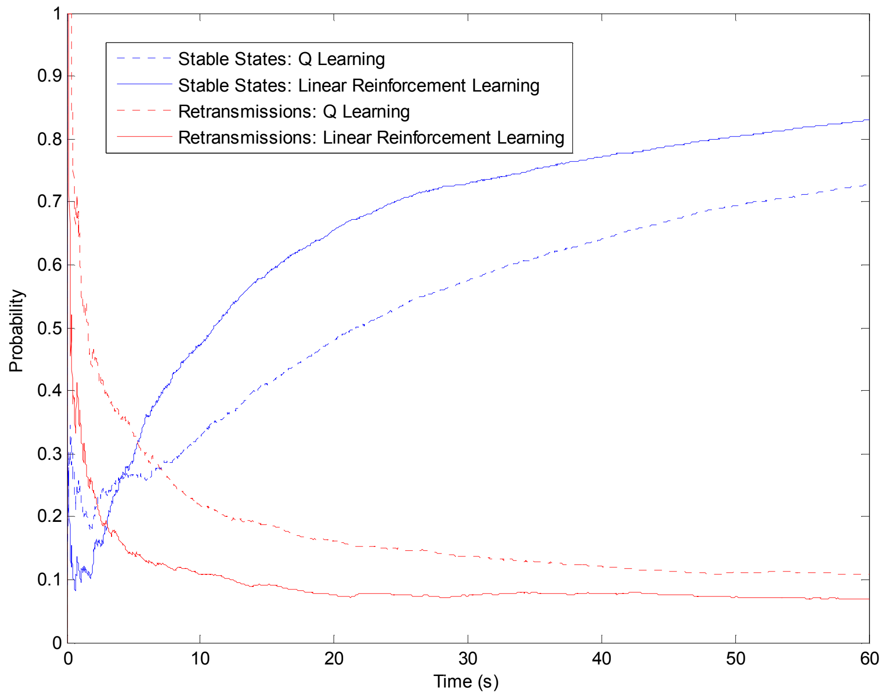

Figure 5.3. Stable States and Retransmissions: High Traffic Level

Figure 5.2 demonstrates that the learning algorithms converge at 20s with 90% stable states achieved. The retransmission probability also converged to 10% at the same time. More stable states results in fewer retransmissions. It can be concluded that the stable state evaluation algorithm provides an effective approach in investigating the convergence of learning.

In Figure 5.3, the linear reinforcement learning achieves a significantly higher stable state probability than Q learning, with a lower retransmission probability. It demonstrates that a higher QoS level can be achieved when distributed learning converges to a fixed channel set. The convergence analysis in Section 4.4.2 indicates that Q learning is more sensitive to the environment changes than linear reinforcement learning. Comparing their stable state probability in Figure 5.3, it can be concluded that Q learning is very ineffective in finding converged channel sets because of a highly dynamic user traffic, which in turn causes a low level of QoS. On a contrary, linear reinforcement learning is more effective in a high traffic scenario. In summary, the stable state evaluation method provides a definition of convergence in learning for radio resource management scenario, which also complies with the stability of temporal QoS.

5.6 Cooperation Management

The Cooperation Management (CM) module is designed to start transfer learning when extra knowledge is needed on the target agents, and stop information exchange when a stable state is achieved from the knowledge base. Based on the target, cooperation management should have the ability to evaluate the information transferred from source agents, and to identify its impact on the target agent.

The Q table in a transfer learning cycle, as demonstrated in Figure 5.1, is updated by the action-value function from linear reinforcement learning (4.5) and value training function from target agent training (5.4). Thus it is possible to carry out a stable state evaluation on the knowledge base learnt from either function.

5.6.1 *CM on Value Training Function*

The motivation of Cooperation Management in transfer learning is to stop the information exchange when the impact on the learning process of transferred knowledge is significantly diminished. One direct method of measuring such impact is to inspect the action ranking C_R changes after applying the value training function (5.4). This method directly measures the impact of transfer learning on the target agent's Q table. The structure of this scheme is described in Figure 5.4 below.

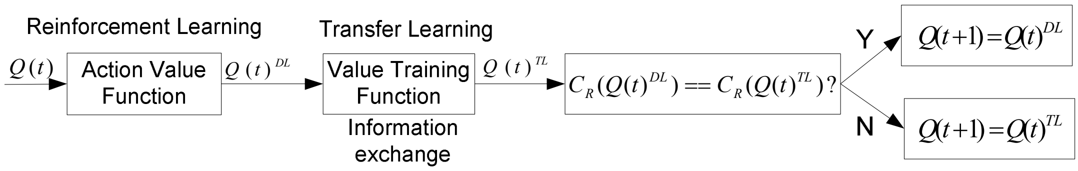

Figure 5.4. CM on Value Training Function

Figure 5.4 indicates that information exchange is carried out before evaluating stable states on knowledge base. However, the target of transfer learning is to reduce multi-agent coordination. Additional control schemes are necessary in this algorithm to stop the evaluation process when stable knowledge base is identified appropriately. Here we develop initial and adaptive control schemes on knowledge transfer, by comparing channel rankings C_R between a distributed reinforcement learning process $\cdot\ (t)^{DL}$ and a transfer learning process $\cdot\ (t)^{TL}$.

5.6.1.1 Initial Evaluation

The motivation of initial evaluation in cooperation management is to improve the convergence performance when cognitive agents are initially activated. The algorithm is designed to assist the naïve agents to quickly build up their knowledge base. Figure 5.3 shows that in distributed learning, the QoS performance is improved as more stable states are achieved. As a result, one target of transfer learning is to speed up the initial convergence process. The stable state can be used as a criterion to terminate the information exchange operation. The algorithm is designed as follows.

This algorithm stops transfer learning once it has no changes to $C_R(t)$, which improves the performance before a stable state is achieved.

Table 5.3. CM on Value Training Function (Initial)

Operate distributed learning, obtain $Q(t)^{DL}$
Evaluate $C_R(t)^{DL}$ over $Q(t)^{DL}$
Operate information exchange with (5.1)
Operate value training function with (5.4), obtain $Q(t)^{TL}$
Evaluate $C_R(t)^{TL}$ over $Q(t)^{TL}$
If $C_R(t)^{TL} = C_R(t)^{DL}$
    Set $Q(t) \rightarrow Q(t)^{DL}$
    Terminate this algorithm
Else
    Set $Q(t) \rightarrow Q(t)^{TL}$
End

5.6.1.2 Adaptive Evaluation

The initial evaluation strategy only exchanges information from the start until a stable state is reached, and then terminates. It will not respond to changing dynamics of the surrounding environment, i.e. the impact of the activation of new agents, the variation of offered traffic, the mobility of agents, etc. The converged selections learnt from previous knowledge transfer may not be a good solution to the new environment. For the purposes of exploring further environment changes after the initial termination of knowledge transfer, we propose a ε-Greedy exploration scheme, as described in Table 5.4, to extend transfer learning:

Table 5.4. CM on Value Training Function (Adaptive)

```
19: Operate 9 to 18 iteratively until 15 reached
20: If rand(.) < ε
21:     Operate 19
22: End
```

This algorithm activates transfer learning through random exploration after its initial termination, and then stops it until another stable state is achieved. The agent can then explore potential environment changes periodically. The accuracy of exploration and the information exchange cost depend on both $C_R(t)$ changes in the training function and an exploration probability ε. The learning agent could monitor the surrounding environment more frequently by setting a higher ε. However, this causes more information exchanges.

Cooperation management on the value training function could be made to operate continuously without initial or adaptive information exchange control, as illustrated in Figure 5.4. It is expected to provide effective performance when operating the network in such a fully coordinated manner, although information exchange cannot be controlled in practice. Nevertheless, this mode of operation will be shown in Section 5.7 to establish a theoretical bound, for the comparison of QoS reduction in the initial and adaptive control schemes.

5.6.2 ***CM on Action-Value Function***

Cooperation management on the value training function has the limitation that $C_R(t)$ is investigated *after* the transfer process, which makes it difficult to control the information exchange effectively. The action-value function also provides stable state information on the basis of action ranking, although it may take more trial-and-error iterations for the changes to be investigated. However, by evaluating the action ranking over the local knowledge base before applying the value training function, the agent has no need to exchange information prior to measuring learning stability. The structure of cooperation management on action-value function is as follows:

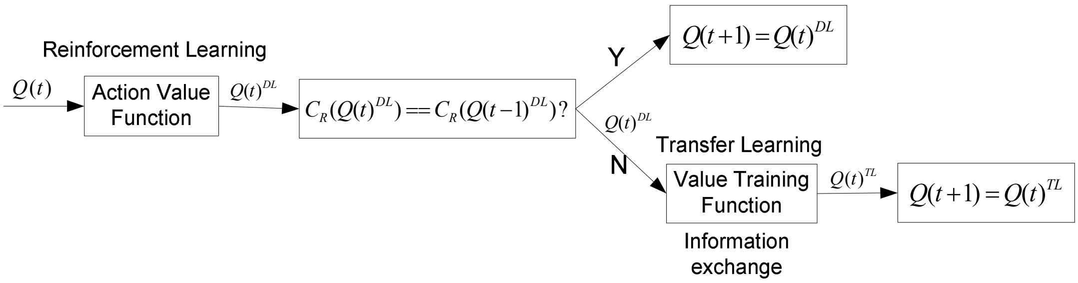

Figure 5.5. CM on Action-Value Function

The action ranking is evaluated on a reinforcement learning iteration only, before information exchange. The reinforcement learning is responsible for identifying environment changes, to activate transfer learning. This algorithm continuously monitors the knowledge base in a fully distributed manner. It is entirely based on temporal learning information rather than heuristic greedy exploration and is, therefore, more efficient.

This algorithm is presented in Table 5.5 below.

Table 5.5. CM on Action-Value Function

Evaluate $C_R(t-1)$ over $Q(t-1)$
Operate distributed reinforcement learning, obtain $Q(t)^{DL}$
Evaluate $C_R(t)^{DL}$ over $Q(t)^{DL}$
If $C_R(t)^{DL} = C_R(t-1)$
    Operate information exchange with (5.1)
    Operate training function with (5.4), obtain $Q(t)^{TL}$
    Set $Q(t) \rightarrow Q(t)^{TL}$
Else
    Set $Q(t) \rightarrow Q(t)^{DL}$
End

It has been illustrated before that the objective of transfer learning is to provide environment information through the knowledge base transferred from source agents to the target agent. Transfer learning is thus more sensitive in identifying environment changes than reinforcement learning. Cooperation management on the action-value function could be less effective than that on value training function when the target agent converges on an action space for a long time, because the historical information from the last reward state dominates the Q value, while the instantaneous information increases slowly.

5.7 Simulation

This section presents simulation results showing the benefits of applying cooperation management strategies to reduce information exchange during transfer learning, whilst maintaining high levels of QoS. The simulation is based on the architecture presented in Figure 3.2, using the key parameters listed in Table 4.3.

The learning process is applied when a file is either successfully delivered or delayed for retransmission. An ε-Greedy selection is introduced to distributed reinforcement learning, for the purpose of providing a low level of random exploration to find potentially better decisions. Similarly, ε-Greedy exploration is also used in CM on value training function with adaptive control, to find potential environment changes. The ε value is set as 10% in distributed learning and 1% in cooperation management, respectively. The algorithms are examined on a long-term averaged basis at different traffic levels, and also on temporal basis at a traffic level of 470 Mb/s. Two conventional resource management schemes are used as performance comparison. In the first, a *fully distributed reinforcement learning* scheme enables agents learn independently of each other. In the second case, a *full transfer learning* scheme exchanges agents' knowledge bases at every learning iteration.

Figure 5.6 demonstrates the probability of failed decisions of the fully distributed reinforcement learning and full transfer learning schemes, which illustrates the degree of unsuccessful channel selections made by learning algorithm, as defined in equation (3.21) in Section 3.5.2. Reinforcement learning has higher failed decision probability than transfer learning throughout, with significant increase from 30% to 90% when offered traffic is higher than 350 Mb/s. On the other hand, transfer learning is shown to make effective decisions when offered traffic is below 450 Mb/s. It can be concluded that the transfer learning algorithm (5.5) significantly improve the decisions made by reinforcement learning (4.5), which supports the network at a higher traffic level.

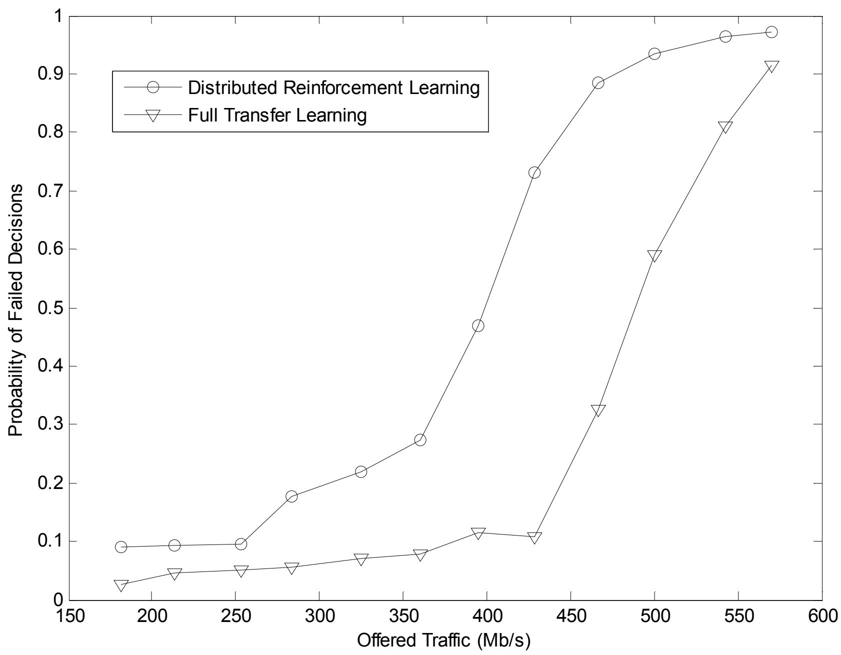

Figure 5.6. Learning Efficiency

Figure 5.7 and Figure 5.8 demonstrate the temporal performance of the different cooperation management algorithms proposed in Section 5.6. The value training function is applied.

The probability of failed decisions is presented in Figure 5.7. Transfer learning achieves 20% to 40% lower failed decision probability than distributed learning. In particular, the performance of the fully coordinated transfer learning demonstrates that the value training method effectively provides expert knowledge to the distributed agents.

In the same figure we see the performance of the three cooperation management algorithms presented earlier. The initial CM on the value training function has failed decisions gradually decreasing from 25% down to 15%, which illustrates that transfer learning at the start-up stage significantly improves decision making. Moreover, the result is shown from 5000 iterations. It can thus be concluded that the initial CM scheme has slow convergence because knowledge transfer is permanently stopped once a stable state is achieved. A cognitive agent in turn has to make decision fully based on reinforcement learning. Adaptive CM on value training function achieves a steady 15% failed decision probability from 5000 iterations onward, which benefits from periodic random explorations to activate transfer learning after the first stable state. CM on action-value function achieves the same 10% failed decision probability as full transfer learning. It becomes slightly worse in latter iterations because the control strategies terminate the information exchange. However, it still achieves lower failed decision probability than adaptive CM on value training function, because the CM decision is made on stable states evaluated from the learning function rather than random exploration after the initial termination of information exchange.

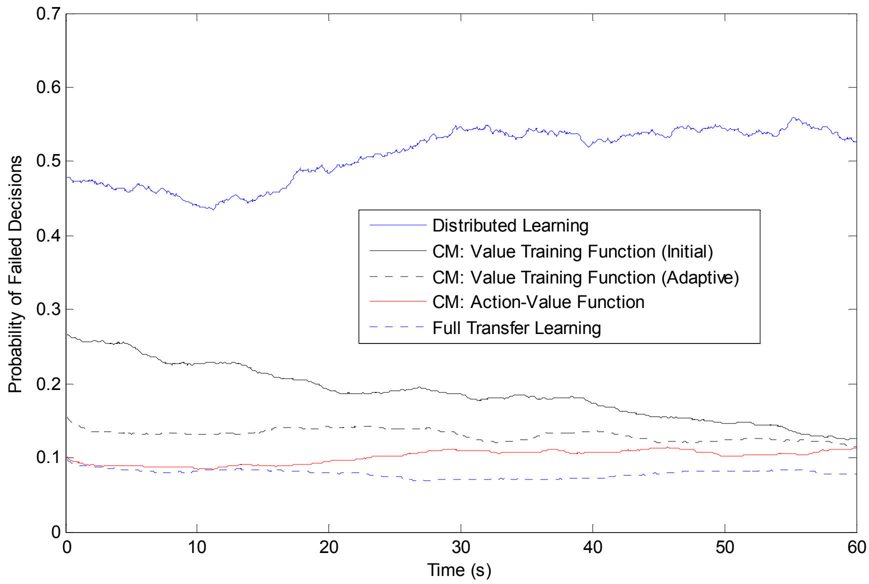

Figure 5.7. Probability of Failed Decisions (Cooperation Management)

Figure 5.8 shows that the cooperation management strategies largely reduce the information exchanged by more than 80% compared to full transfer learning. CM on the value training function achieves 3% information exchange at 5000 iteration. Transfer learning is stopped on all agents at 5500 iteration with initial CM, whilst is maintained at a 1% level with adaptive CM. Moreover, CM on the action-value function has a higher coordination overhead at 20%, with better QoS achieved in Figure 5.7. It can be concluded that cooperation management algorithms effectively control the amount information exchanged during the converging period of learning and achieve high level of QoS.

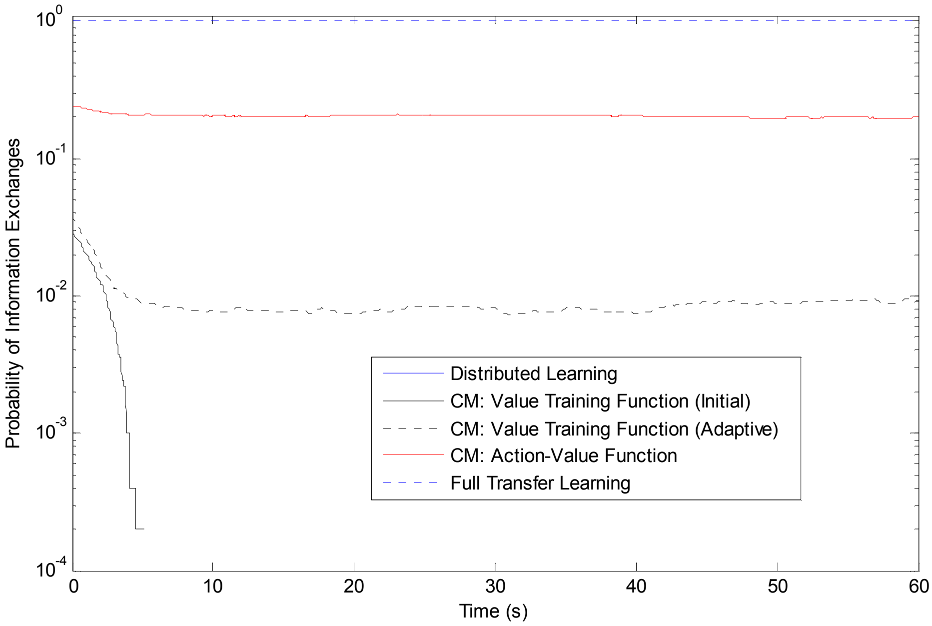

Figure 5.8. Probability of Information Exchanges (Cooperation Management)

A comparison of different Target Agent Training strategies from Section 5.4 is demonstrated in Figure 5.9 and Figure 5.10. We use the cooperation management on the action-value function, because the interference coordination strategy does not have a value training function. The simulation is operated by applying the training strategies of the value training function, interference coordination and an integration of the two.

It is illustrated in Figure 5.9 that three training schemes achieve significantly lower failed decision probability than the fully distributed learning algorithm. However, their convergence behavior has big difference. The value training strategy exhibits on almost constant failed decision probability of 10%. The performance of the interference coordination strategy starts from 6% and gradually increases to 8%. This is because instantaneous channel usage information effectively avoids interference as demonstrated in Chapter 4. However, interference coordination does not exchange learning information based on past experience, thus the failed decision probability increases when coordination is terminated by the cooperation management scheme.

The combination of the value training and the interference coordination strategy provides highly effective convergence performance to a low level of failed decisions, from 10% reducing down to 5%. It can be concluded that transfer learning effectively improves interference coordination after the termination of information exchange. Furthermore, it shows even better performance than the fully coordinated transfer learning, because negative information in knowledge transfer is effectively removed.

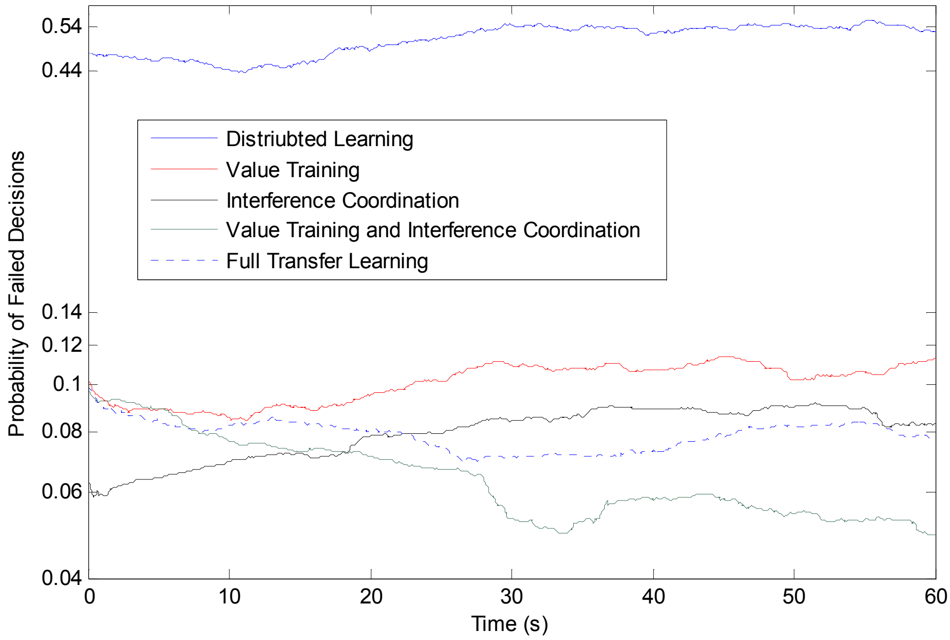

Figure 5.9. Probability of Failed Decisions (Target Agent Training)

Figure 5.10 illustrates the coordination overhead of these schemes. It can be observed that interference coordination incurs a high level of cooperation overhead. The value training function on the other hand significantly reduces the amount of cooperation down to 20% throughout. Moreover, the combined scheme achieves the same level. Compared to Figure 5.9, it can be concluded that the value training function is vital to reduce cooperation overhead whilst keeping adequate QoS.

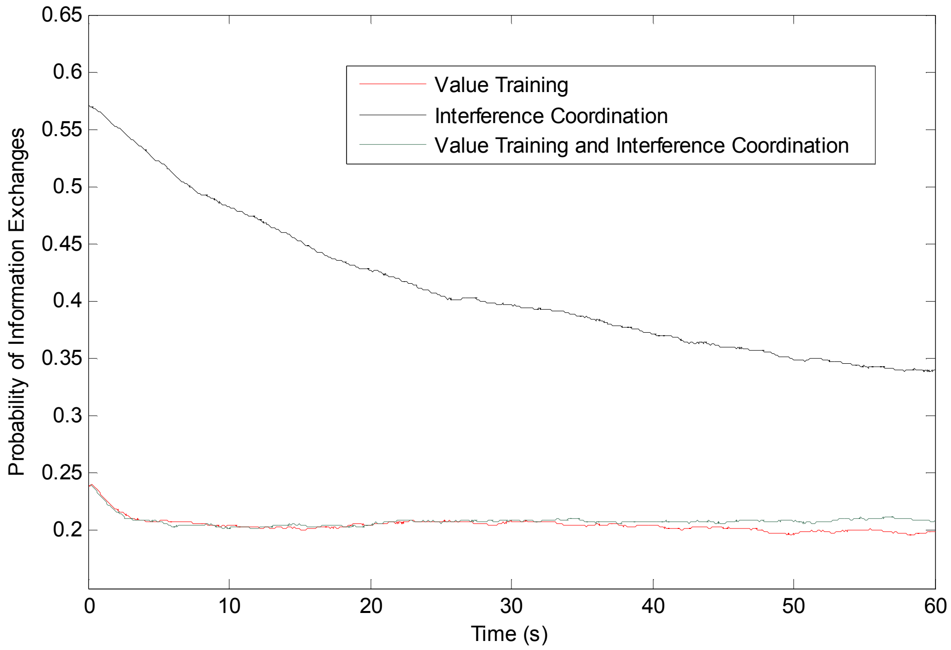

Figure 5.10. Probability of Information Exchanges (Target Agent Training)

In the second simulation, we evaluate a long term average performance on a stable network, and demonstrate QoS under cooperation management over a wider range of traffic loads from idle to saturated. In the following simulations, we introduce a Frequency Planning (FP) scheme derived from BuNGee [12], which divides the spectrum pool into 2 equal size sub-bands and allocates them differently to any neighbouring links. The FP approach is expected to perform efficient interference avoidance. However, the link capacity is highly constrained by the fixed band size, especially on the links closer to the HBS because of the bottleneck caused by relay traffic. The QoS and throughput is shown in Figure 5.11 and Figure 5.12.

It can be observed from Figure 5.11 that the network with the transfer learning benefits from much fewer retransmissions compared with the reinforcement learning and frequency planning scheme. Compared with the full transfer learning, the retransmission probability of the initial CM on value training function becomes higher while the adaptive CM on the value training function stays the same, when the offered traffic grows beyond 450Mb/s. CM on the action-value function has slightly lower retransmissions than full transfer, with far fewer information exchanges as demonstrated before. This is because the Q value information from source agents could have a negative impact, i.e. the agents operate almost random selection during the initial stages. This issue has been addressed as *negative transfer* in [80] and thus an effective stable state evaluation strategy is crucial to decide when to transfer.

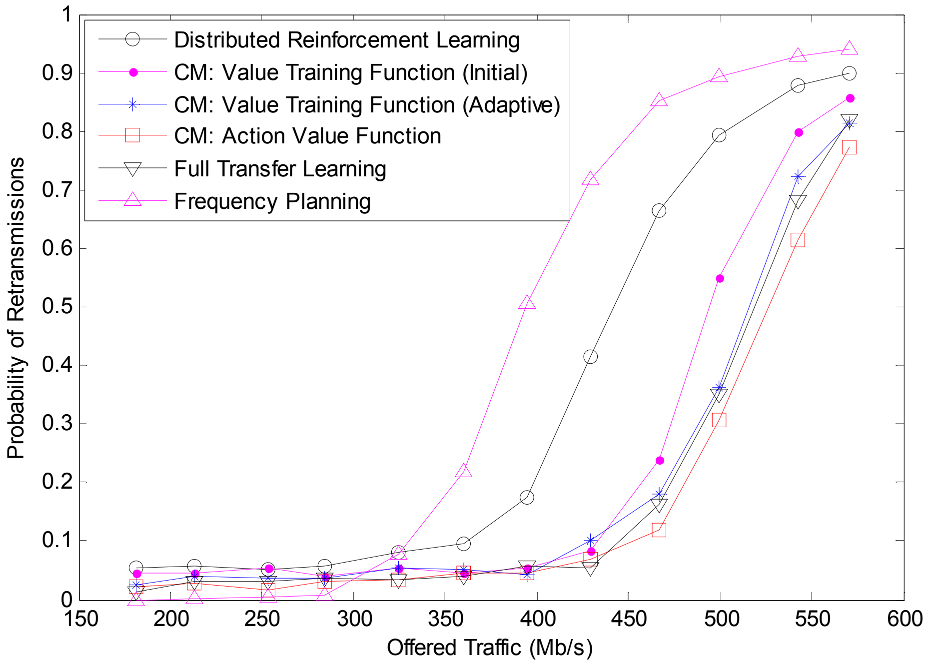

Figure 5.11. Probability of Retransmissions

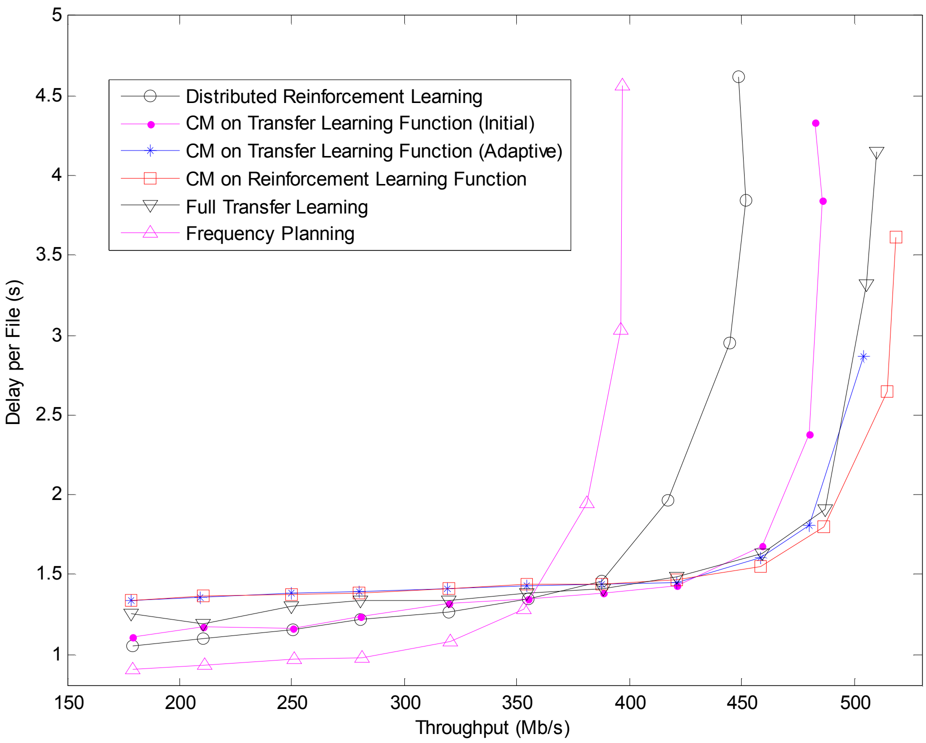

Figure 5.12. Mean Delay per File

Figure 5.12 shows the mean delay and the throughput that the network can support in the offered traffic levels presented in Figure 5.11. In this scenario, the transfer learning strategies can support significantly higher throughput than the frequency planning and distributed reinforcement learning strategy. For the cooperation management strategies, the delay properties, seen in Figure 5.12, follow similar trend to those for the probability of retransmission, as seen in Figure 5.11. Adaptive CM on value training function shows slight improvement over the initial CM on value training function, while CM on action-value function delivers the highest throughput of them all.

In the third simulation, we model a scenario where offered traffic gradually increases from 90Mb/s to 360Mb/s, stepping up by 90Mb/s every 40s, as shown in Figure 5.13. In a practical distributed network, such a traffic increase would cause QoS to reduce quickly because further channel partitioning is required. Here we assess performance using a cumulative window from the start of simulation.

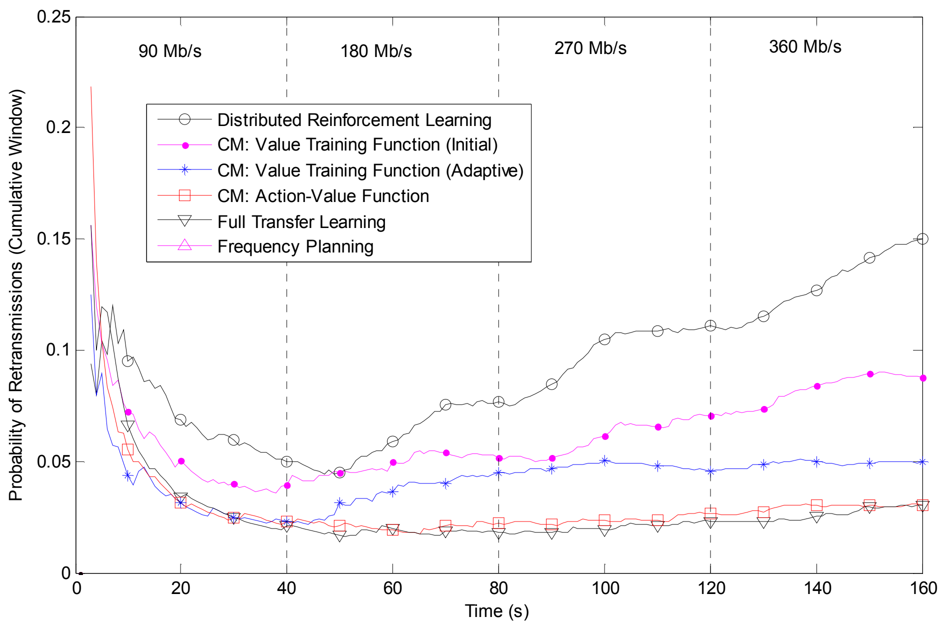

Figure 5.13. Probability of Retransmissions (Dynamic Traffic)

The probability of retransmissions with the distributed reinforcement learning scheme increases with offered traffic. It reaches up to 3 times the original value at the highest traffic level. However, using full transfer learning, the probability of retransmission does not change. The value training function effectively assists the agent to avoid poor decisions by reducing their Q value in the knowledge base.

In the same figure we see the effects of the four different cooperation management schemes presented earlier. First comparing the full transfer learning with CM on action-value function, both strategies achieve the same 2% level of retransmissions. We conclude that for a static network topology, stable state evaluation on the action-value function provides effective information exchange control.

Initial CM on value training function exhibits a lower probability of retransmissions than the distributed reinforcement learning. However it increases gradually with offered traffic. This is because transfer learning terminates when the agent reaches a stable state in the initial 90Mb/s phase, and environment changes cannot subsequently be identified. The ε-Greedy based adaptive CM method provides opportunities for agents to conduct transfer learning after the initial stable state. Therefore, a lower retransmission level is achieved than using the previous method. However, it still increases over the 180Mb/s phase, after the initial termination of transfer learning. These results demonstrate that cooperation management can effectively control information exchange by evaluating the stable states of the knowledge base on the action-value function rather than the value training function.

The probability of information exchange is shown in Figure 5.14. It is measured over a sliding window of 40s. The cooperation management strategies reduce information exchange by more than 95% when compared with the full transfer learning. Adaptive CM on value training function exhibits a 1% probability of information exchange after the initial stable state reached. This is a result of ε-Greedy exploration. The information exchange probability of CM on action-value function fluctuates between 1% and 3%. As Figure 5.13 showed, using this level of exchange, it achieves much better QoS than CM on value training function and a similar QoS level as full transfer learning.

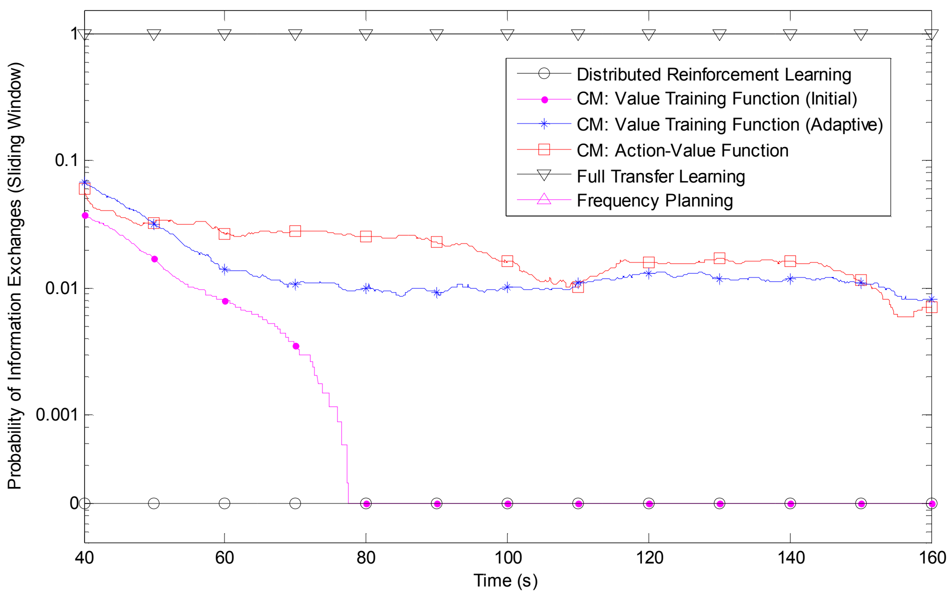

Figure 5.14. Probability of Information Exchanges (Dynamic Traffic)

One of the key motivations for multi-agent cooperation in transfer learning [73] is to enable new, naïve agents to benefit from the experience of mature agents. Therefore, we perform another simulation using two groups of backhaul branches which begin to transmit at different times. Six branches are used in the model, with other parameters as summarized at Table II. In this simulation, the ABSs on branches b_1, b_3, b_5 transmit continuously from the start whilst those on b_2, b_4, b_6 join the network after 30s. This emulates a practical scenario in which ABSs can be automatically switched on and off in order to save energy, by responding to variations in the number of active mobile users in the network. We expected that transfer learning would effectively overcome difficulties with the dynamic topology that tend to adversely affect the performance of reinforcement learning [21].

Figure 5.15 below shows the performance of the two groups of ABSs in accessing the network at different times. The result is captured as the cumulative probability. CM on action-value function is applied, given that it has been demonstrated to be an effective cooperation management scheme in this Transfer Learning model.

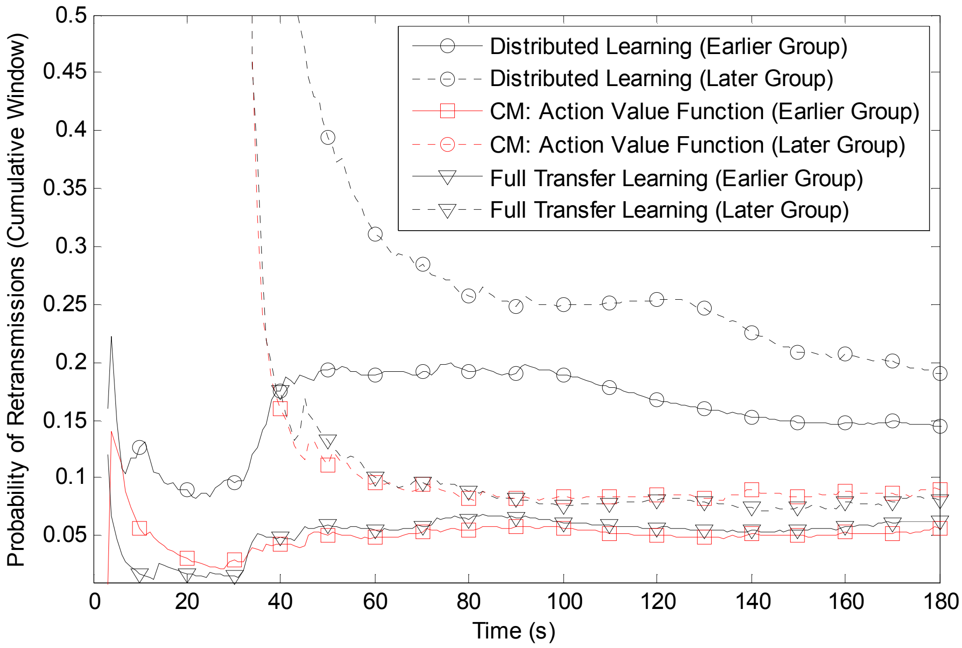

Figure 5.15. Probability of Retransmissions in Dynamic Architecture

It demonstrates that, using a distributed reinforcement learning strategy, the latter group has a harmful impact on the performance of the earlier group, causing retransmissions to increase by 10% between 30s and 40s. However, using transfer learning, the earlier group stays at 5% retransmissions and the later group converges to 8%, which is 8-10% lower than the distributed reinforcement learning scheme and exhibits much faster convergence. We conclude that transfer learning can either protect established agents from being harmed by the newly-activated ones, or support the new ones in converging to the expected decisions attained from the start of the simulation.

5.8 Conclusion

This chapter has introduced a Transfer Learning strategy to improve radio resource management on a multi-hop backhaul network. Transfer learning is proposed to be transfer learnt knowledge from selected source agents, to assist distributed learning on a target agent. Three components are proposed in transfer learning: the Source Agent Selection module identifies the harmful agents that cause excessive interference; the Target Agent Training module transfers learning information between agents and reinforces the knowledge base; the Cooperation Management module controls the level of information exchanged and maintains QoS.

The source agent selection scheme is developed from the spatial channel reuse scheme demonstrated in the previous chapter. Two target agent training strategies are investigated: the value training strategy is shown to provide effective QoS and the interference coordination strategy is demonstrated to provide further improvements.

The Cooperation Management strategy provides an effective solution to balance the Quality of Service (QoS) and the cooperation overhead. A series of cooperation management strategies have been developed and demonstrated, which assess stable states from value training function and action-value function. Cooperation Management (CM) strategies are shown to reduce cooperation overheads between distributed agents by up to 90%. CM on value training function and action-value function effectively control negative transfer and achieve similar performance as theoretical full transfer. Adaptive CM on value training function controls information exchange and QoS more effectively than the initial scheme by using ε-Greedy exploration on environment changes. CM on action-value function can be operated in a fully distributed way without exploration, and is shown to be the most efficient scheme in terms of QoS and information exchange probability. It has also been demonstrated that transfer learning provides efficient convergence in a network with both dynamic topology and offered traffic, reducing the harmful effect of agent activation and traffic increase.

Chapter 6. Transfer Learning for Dynamic Network Architectures

6.1 Introduction

In this chapter, flexible small cell access networks are studied to provide broadband radio access in dense populated urban areas. The scenario described in Section 2.2.2 is used where the low power Access Base Stations (ABSs) are deployed at a below rooftop level of high streets. Omni-directional antennas are implemented on each ABS. The architecture is designed to utilize the building area between streets to reduce interference through the shadowing effect.

The small cell access network with dynamic topology is examined in the following work. Flexible network architectures have a number of application scenarios including femto base stations, energy efficient topology and opportunistic networks, which will be described in Section 6.2. The conventional reinforcement learning algorithm suffers from serious QoS fluctuations with changes of traffic load or network topology. This is because the cognitive agents need sufficient iterations to learn the new radio environment, which will be analysed in Section 6.3. Transfer learning is redesigned in Section 6.4, with a value mapping method used to effectively learn the topology transitions, in order to reduce QoS fluctuations and provide reliable communication. A dynamic frequency reuse clustering scheme is proposed in Section 6.5, which defines clusters for multi-agent coordination. A novel Pareto efficient action space prioritization algorithm is developed in Section 6.6, which is designed to eliminate interference between cells and maximize resource utilization in a cluster. This is followed by an action-value mapping strategy in Section 6.7, which associates the Q value learnt in the previous task with the prioritized action space for the new task. Simulation results and conclusions are discussed in Section 6.8 and Section 6.9, respectively.

6.2 Dynamic Network Environment

The interference environment of the access network is much more complex and dynamic than the backhaul network, because of the omni-directional antennas and the highly random user locations. In this context, the variations of user traffic and network topology have more impact on decision making and convergence in conventional reinforcement learning, which in turn affects network QoS and reliability.

6.2.1 *Dynamic User Traffic*

The offered traffic in a cellular network is typically fluctuating in both the time and spatial domains. It has been shown in Figure 2.4 that the average offered traffic between different hours in a day, different days in a week, or different cells varies significantly.

Conventional reinforcement learning based resource management strategies are usually designed and examined with different static traffic levels, in order to converge quickly to a fixed set of action space [64]. However, the changes of offered traffic in practical networks require a number of “blind” iterations on the intelligent agent to reinforce the knowledge base, which causes a period of low QoS. It has been illustrated in Chapter 4 that when an agent needs to assign multiple channels, multiple reinforcement learning processes need to be carried out independently and simultaneously on an action space. An increasing offered traffic level obviously requires more channels to be learnt on a base station. In this context, the learning agent needs to carry out learning on actions without Q value information, which may cause a number of random explorations.

6.2.2 ***Dynamic Network Topology***

The dynamics of the network topology also has impact on distributed reinforcement learning. There are many practical scenarios where a dynamic topology applies. Here we provide some typical scenarios including femto BSs, energy efficient architectures and opportunistic networks. Challenges of reinforcement learning in these networks are illustrated.

Femto Base Stations

Future wireless networks are designed to be cost and energy efficient. A number of portable and light-weight base stations are expected to share the spectrum with a conventional cellular network, in order to enhance the capacity density. A typical example is the femtocell (HeNB) proposed in 3GPP LTE [100], which is managed by consumers and connected directly to the internet through DSL or a cable line. The purpose of implementing HeNB is to provide extra network capacity in hotspot areas, such as homes, offices, café shops, etc. The location and working time of these HeNBs could be very dynamic and unpredictable, because the consumers can easily switch them on or off according to the user requirements in a local area. Coexistence of HeNBs and macro eNBs in a common spectrum pool is a crucial issue in the LTE network. Conventional reinforcement learning has a big challenge in this scenario. On one hand, the newly activated HeNB requires a number of trial-and-error iterations to learn the surrounding environment. On the other hand, the converged action space on existing HeNBs and macro eNB may be destroyed by new HeNBs.

Energy efficient network architecture

Energy efficiency is an important target in future wireless networks, which has been reviewed in Chapter 2. Traffic aware network management is a hot topic proposed in many recent research papers as an effective paradigm to reduce energy consumption in cellular networks [11, 85, 86, 89]. It is expected that a dominant proportion of energy can be saved by switching the base stations between working and sleeping mode based on the local offered traffic level [112]. Figure 2.5 illustrates that in this paradigm, the number of activated base stations varies with offered traffic in different hours of a day. In order to maintain QoS for the cells in sleep mode, the users in these cells should be covered by their neighbouring cells, which make changes to the network topology.

Figure 6.1 shows a model of dynamic small cell network based on the architecture in Figure 3.3. The MSs are connected to the nearest Fixed or Dynamic ABSs when they are all activated, as marked by dots and stars, respectively. The fixed ABSs can extend their coverage to the holes incurred by the deactivated dynamic ABSs, and take over corresponding user traffic.

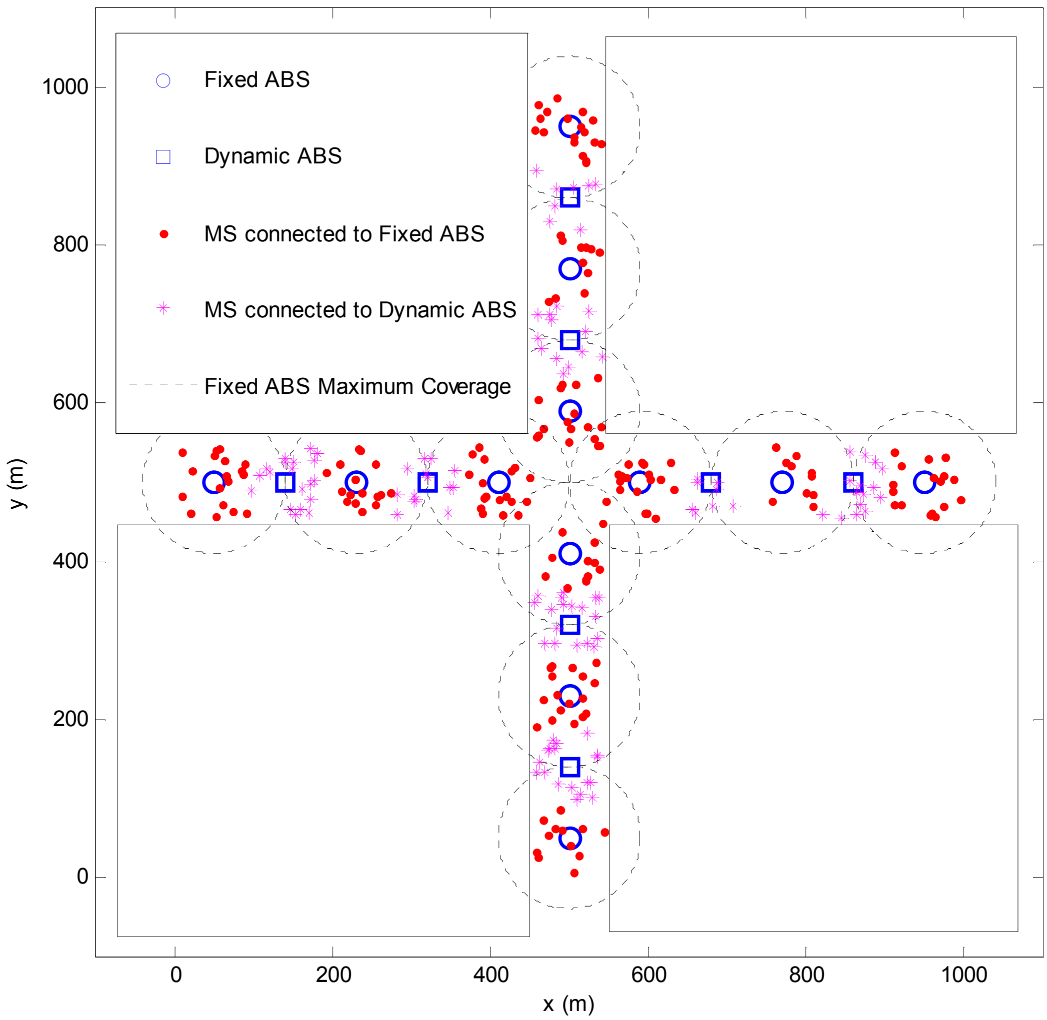

Figure 6.1. Small Cell Network with Dynamic Topologies

The topology transition process significantly changes the radio environment, especially the frequency reuse pattern. It can be seen from Figure 6.1 that when all

ABSs are activated, adjacent Fixed ABSs are separated by a dynamic ABS. In this case, the MSs connected to the same type of ABSs have the potential to reuse the same frequency. However, the fixed ABSs become neighbouring cells when the dynamic ABS is switched off. Conventional reinforcement leanring uses a single Q table for channel assignment. After switching off a dynamic ABS, the MSs transferred to the fixed ABS may suffer from excessive interfernece, because the two neighbouring fixed ABSs initially have the same preferred acton space. It may take a large number of iterations for them to learn a new policy, which causes plenty of failed decisions. On the other hand, it is difficult for a newly activated dynamic ABS to quickly find a preferred action space, because the spectrum may be fully utilized by fixed ABSs. Moreover, converged action space on fixed ABSs may be broken by the dynamic ABSs during their environment adaption process.

Opportunistic network architecture

Opportunistic networks are newly proposed architectures in the FP7 ABSOLUTE project for dealing with unexpected and temporary events [113]. In the unexpected event scenario, it is designed as a reliable communication infrastructure that provides critical services including emergency recovery operations, critical infrastructure restoration, post-disaster surveillance, etc. In the temporary event scenario, it is also used to support high data rate services and enhance network capacity.

Disasters and temporary events require a reliable, rapidly deployable and cost-effective communication architecture, which can be easily rolled out and rolled back at the beginning and the end of the events. The ABSOLUTE project proposes light-weight Aerial eNodeBs (AeNB) and Terrestrial eNodeBs (TeNB) to provide augmented coverage and capacity. The network topology of this architecture can be highly dynamic based on different phases of the event. A fast convergence learning algorithm is desired on eNBs to quickly adapt the radio environments under different topologies.

6.3 Learning in Dynamic Environment

In conventional reinforcement learning, a cognitive agent updates the knowledge base according to the environment feedback from actions. A typical reward function

(4.4) applied in the resource assignment scenario uses positive and negative values to represent the success and failure of actions [110]. The convergence behaviour of linear reinforcement learning and Q learning has been analysed in Chapter 4, where the Q value is based on both historical and instantaneous information. In the dynamic traffic and topology scenarios, a newly activated agent should adapt to the surrounding radio environment. Moreover, the existing agents should identify the environment changes and learn a new policy. The environment adaption and identification process require a number of actions to be taken in reinforcement learning, in order to obtain sufficient rewards.

The environment adaption process is essential when a cognitive agent is initially activated. The knowledge base in such cases contains arbitrary values without learnt information. A transfer learning approach has been proposed in Chapter 5 and [91, 92] on the backhaul network to improve the start up performance. Similar ideas could be applied in the access network based on related spectrum pattern. The base station can then approach faster convergence having been provided preliminary with environmental knowledge.

The environment identification process is more complicated. Reinforcement learning is designed to maximize Q values in the preferred action space. Thus the only situation when an agent could drop out from the converged action space is to reduce their corresponding Q values to a lower level than others in the spectrum pool. However, the only way to reduce Q value in conventional reinforcement learning is to take failed decisions. It has been analysed in Chapter 4 [92] that the speed of this process is determined by the number of iterations taken in the past, the reward values, and the learning rate.

It has been illustrated in (4.10) that in linear reinforcement learning, the Q value is determined by the number of positive t_{S_+} and negative t_{S_-} decisions and their related reward values (R_{S_+} and R_{S_-}). Equation (4.7) indicates that rate of increase or decrease of the Q value is R when $f = 1$. In order to reduce an action's Q value to a lower level than others, the number of negative decisions required is equal to the number of positive decisions taken previously. It can be concluded that learning could cause a number of harmful decisions during environment changes if the agent converges to an action space for a long time.

In single state Q learning, the Q value converges at the reward value R in a stable state (4.17). The rate at which a Q value increases or decreases is a function determined by the learning rate α and the initial Q value in a reward state $Q_S(0)$. The gradient of such rate variation follows exponential growth according to (4.15), which indicates that the new reward information R quickly dominates the Q value. Under this effect, the single state Q learning can quickly drop out from one action space and converge to another, as analysed in Section 4.4.2.3. However, this is not expected when the topology becomes static. The learning rate has no information on the dynamics of network environment.

6.4 Transfer Learning: Value Mapping Method

The motivation of Transfer Learning is to use the network topology information, i.e. base station location, coverage area, to improve the knowledge base at the start of learning in each network topology. This is particularly important in the dynamic networks discussed before. The agent then receives a lower impact from the environment changes, and thus a consistent and reliable QoS level can be provided.

In order to achieve fast environment identification and adaption, a value mapping method is designed in the context of transfer learning to apply network topology information to the knowledge base. An agent can improve decisions by using location information from others. Transfer learning, as originally proposed in the computer science society [79, 80], is aimed at improving learning in the new *Target Task* by leveraging knowledge from the related *Source Task* that has been learnt. In the dynamic network scenario, the source and target tasks are defined as the learning target in the network before and after topology changes, namely when an ABS is switched on or off. Transfer learning is designed to associate the Q values learnt from the *Source Task* with a newly prioritized action space in the *Target Task*, which in turn generate a new Q table that has been adapted for use in the new network topology. The topology changes can be directly identified through control information from the adjacent ABS that is switched on or off.

In the scenario where an agent is newly activated, a start-up value generation function is applied to provide initial knowledge of the network. The structure of value mapping method can be illustrated in Figure 6.2.

$$\xrightarrow{Q(0)} \boxed{S(0)} \xrightarrow{T(Q(0))} \boxed{S(1)} \xrightarrow{T(Q(1))} \boxed{S(2)} \xrightarrow{T(Q(2))} \boxed{S(3)} \ \cdots \xrightarrow{T(Q(t-1))} \boxed{S(t)}$$

Figure 6.2. Transfer Learning: Value Mapping Method

Here $S(t)$ indicates the states of a network topology at time t. Transfer learning algorithm T is carried out on the Q table learnt in each state. $Q(0)$ is the initial setup strategy of the Q table based on the network topology, in order to provide channel priorities when there is insufficient learning knowledge. The algorithm will be illustrated in the following sections.

A key benefit of applying transfer learning to the knowledge base is that only a single Q table is needed throughout the whole process, which minimizes the memory requirement for storing the knowledge base. Moreover, the reinforcement learning algorithm carried out on each network state can further improve the Q table provided from the transfer learning function, and effectively reduce the potential negative transfer.

6.5 Dynamic Frequency Reuse Clustering

The topology information is the fundamental "expert knowledge" that can be utilized to improve the knowledge base in the transfer learning. Frequency reuse clustering is an effective approach to understand the interference environment and select source agents in a cellular network. In transfer learning, a similar idea is introduced but operated in a fully distributed manner. Moreover, inter cell coordination is carried out only when a new topology is established, with the purpose of prioritizing Q values.

Clustering of cells is a classical approach for interference mitigation and frequency planning in most conventional spectrum management strategies [100]. It is used to manage the degree of frequency reuse between cells. A frequency reuse cluster is designed to be the smallest number of cells used to include all frequencies. Co-channel interference can be avoided between any cells in a cluster. The cluster size (number of cluster members) determines the distance between any two co-channel cells in neighbouring clusters, which consequently controls the interference level. A network with small clusters normally benefits from effective spectrum utilization but also suffers from excessive interference, and vice versa. The Shannon equation (3.6) indicates that under fixed transmit power, system capacity is constrained by bandwidth and interference. Thus the cluster size should be carefully designed to obtain a maximum gain from both aspects.

Clustering of adjacent cells is potentially an effective way to apply network topology information for transfer learning in the dynamic small cell networks of Figure 6.1. The construction of clusters varies on the cellular lattice shapes and the size of clusters depends on the signal attenuation between base stations. The BuNGee [12] project defines a fixed transmit power on all ABSs. An ABS suffers interference mainly from those on the same line-of-sight street according to the scenario in Figure 3.3. In this case, clustering the neighbouring cells potentially reduces a large proportion of interference.

A dynamic clustering strategy is essential to apply dynamic topology information to frequency reuse. It can be observed from Figure 6.1 that the activation or deactivation of a dynamic ABS shrinks or extends the cells of neighbouring fixed ABSs. By keeping the number of cluster members fixed, there will still be the same number of cells that separate the nearest two ABSs using the same frequency. Thus the interference between neighbouring clusters can be kept at the same level in different network topologies, though the coverage of a cluster may vary in size. The activation of a dynamic ABS increases the number of clusters in the network, which provides better spectrum reuse and enhances system capacity. The ABSs should reconstruct the cluster when the network topology changes. Capacity enhancement through dynamic clustering will be further analysed in Chapter 7.

The purpose of dynamic frequency reuse clustering in transfer learning is to define the level of multi-agent cooperation for information exchange, which is similar to "source agent selection" defined in Chapter 5. In this context, source agents are selected as other cluster members except from the target agent. The following section will illustrate the methods of information exchange in a cluster.

6.6 Action Space Prioritization

6.6.1 *Pareto Improvement Resource Prioritization*

In the dynamic spectrum access scenario, the cell capacity can be dynamically changed according to the traffic load level, rather than being constrained by an allocated spectrum band. This flexibility is particularly important in a dynamic environment where loading across cells can change rapidly. However, interference between adjacent cells becomes a crucial issue. A frequency reuse cluster is allocated the whole spectrum band. By avoiding interference between cluster members, dynamic spectrum access can provide maximized spectrum utilization in the whole network. The target of transfer learning is to achieve this in a fully distributed manner through channel prioritization.

The *Pareto Improvement* process is a resource allocation strategy that allows any individual in a group to occupy more resources without causing interference with others [114]. The *Pareto Efficiency* is the upper bound of this process that defines the maximum number resources that can be allocated to all individuals. The cluster based dynamic spectrum management scenario can effectively use the Pareto improvement concept to avoid interference. It can be modelled as given a fixed spectrum pool, each BSs in a cluster can assign any number of channels to users without causing interference to others, unless the entire pool is occupied. By defining the shared channel pool set as *C*, a cluster of cells as *K*, the channel set selected in each cell as c_k, the Pareto improvement process can be illustrated as follows:

$$\bigcap_{k=1}^{|K|} c_k = \varnothing \text{ , if } \sum_{k=1}^{|K|} |c_k| \leq |C| \tag{6.1}$$

This indicates that the occupied channel set in any cell c_k is different from others, which prevents blocked or interrupted transmissions in a cluster.

Table 6.1 is an example of Pareto improvement in a three cell cluster model. A total number of 12 channels are shared by 3 BSs in a cluster. The priorities indicate the order of channels selected in each cell. The priority set effectively demonstrates the channel prioritization process. The entire channel pool *C* can be divided into 4 sets when each BS occupies 4 channels. An individual BS with more than 4 channels selected could reduce the number of channels available to others. However, this does not interfere with the channels in use, because the top prioritized channels for one agent are placed to the bottom in a reverse order for others. In this context, the entire cluster could always utilize a total of 16 channels without causing inter-cell interference, which is not affected by the number of channels occupied in each cell.

Table 6.1. Pareto Improvement Priority Table

Priorities		I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
Channels	BS 1	8	7	6	5	9	1	10	2	11	3	12	4
	BS 2	4	3	2	1	9	5	10	6	11	7	12	8
	BS 3	12	11	10	9	5	1	6	2	7	3	8	4
Priority Sets		A	B	C	D	D		C		B		A	

Channel prioritization is essential to achieve Pareto improvement (6.1) in a distributed manner. The reinforcement learning algorithms use the Q value to discriminate the priority of channels. It is thus important to associate the Pareto improvement priority table with the Q table. The cluster members can then operate distributed assignment following Pareto improvement. The cluster capacity could be maintained to the entire spectrum pool regardless of traffic variation in the spatial and time domains.

6.6.2 *Algorithm*

This subsection demonstrates an action space prioritization algorithm designed to achieve Pareto improvement resource allocation. In distributed reinforcement learning algorithms, the Q value determines the priority of channels being assigned. In order to prioritize channels in a Pareto improvement manner, the first step is to obtain the original channel priority from the Q table. A channel ranking table can be obtained by sorting channels in a descending order as

$$[q,p] = Sort(Q)\colon Q_{p(i)} \geq Q_{p(i+1)} \;\&\; q(i) \geq q(i+1), \text{ for } \forall i \in [1, |Q|-1] \tag{6.2}$$

Q and q are the original and sorted Q tables, respectively. p is the sorted channel table, similar to Table 6.1. The cognitive agent effectively assigns channels according to their position in p.

After operating (6.2) throughout all the cluster members, it is then important to change p to a Pareto improvement order. Table 6.1 indicates that the top $\frac{|C|}{|K|}$ channels on one agent should be set to the bottom in a reverse order on other agents, which could avoid interference between cluster members. A source priority set $p^{(s)}$ on an agent k is defined as

$$p_k^{(s)} = \bigcup_{i=1}^{\frac{|C|}{|K|}} p_k\left(|K| - i + 1\right) \tag{6.3}$$

which can be obtained from reversing their original channel order. $p^{(s)}$ effectively assists the other agents to avoid using the same channel priorities.

The knowledge transfer process is then operated to combine $p^{(s)}$ obtained from all other agents in the cluster into a single priority table, namely the target priority set $p^{(t)}$. $p^{(t)}$ will be placed to the bottom of the priority table on the target agent, because they are top prioritized channels on source agents. The target priority set is built following the channel order on each $p_i^{(s)}$:

$$p_k^{(t)} = \bigcup_{j=1}^{\left|p^{(s)}\right|} \bigcup_{\forall i \in K^{(s)}} P_i^{(s)}(j) \tag{6.4}$$

The top priority channels on the target agent are those not included in all the $p^{(s)}$, which can be obtained from extracting the complement set of $p^{(t)}$ from the original p. $p^{(t)}$ is then placed to the bottom in the priority table for the target agent:

$$p_k = (p_k - p_k^{(t)}) \cup p_k^{(t)} \tag{6.5}$$

The priority table here is a strict order set, thus (6.1) to (6.5) effectively change the element positions in set p.

This prioritization process should be carried out iteratively in the cluster until every agent has been trained by all the others. It is particular important that the source agents should be those have been trained by other agents before, which guarantees the effectiveness of transferred knowledge. In this condition, the source agent set

$K^{(s)}$ increases by the number of iterations from 0 to K. The number of iterations required to achieve Pareto improvement priority table under this approach is

$$N_T = 2|K| - 2 \tag{6.6}$$

This is because for the first round of multi-agent coordination in the cluster, a base station obtains information from only part of the entire cluster because the source agent set $K^{(s)}$ is being built up. On the second round, all base stations have source agents $K^{(s)} = K - 1$, and the coordination can be completed. Moreover, on the first round the first agent is acting as source agent only providing information to others and the last agent already has sufficient $K^{(s)}$, thus the coordination can be stopped 2 iterations before the second round is finished.

6.7 Action-Value Mapping

The action-value mapping strategy is designed to map the Q values learnt from reinforcement learning in the previous source task to the action space prioritized by transfer learning. The mapping function associates the sorted Q table in (6.2) with the priority table in (6.5), which effectively use the original Q value on different channels:

$$Q\big(p(i)\big) = q(i) \tag{6.7}$$

The output of Transfer Learning is a Q table that has been prioritized in a Pareto improvement manner. The base stations can then operate distributed assignment in later iterations with information learnt from frequency reuse clustering. Furthermore, the reinforcement learning algorithm can be operated on the Q table that contains output from transfer learning, which is effective for removing potential negative transfers.

Start-up Q value generation

The start-up stage is the time when a cognitive agent initially starts to learn the environment. It can be referred to the “first task” in transfer learning. This is a special case that no previous source task exists. However, the feedback from the

Pareto improvement prioritization still provides valuable start-up knowledge, which can be used to generate an initial knowledge base.

A conventional reinforcement learning algorithm sets arbitrarily numbers (usually 0) to the channels. A "warm-up" strategy has been introduced to speed up the environment adaption process [64]. However, it may take a number of iterations for an agent to find the preferred channel set, because random exploration is conducted when few rewards are obtained from learning.

In transfer learning, the initial Q table is generated with discriminating Q values. The ranking of their corresponding channels is based on the Pareto improvement priority table learnt from (6.2) to (6.5). This method provides Q tables with the same characteristics achieved from (6.7).

We have so far introduced the framework of the proposed value mapping method in transfer learning, including dynamic frequency reuse clustering, action space prioritization and action-value mapping. The entire algorithm is presented as pseudo code in Table 6.2.

Table 6.2. Transfer Learning: Value Mapping Method

On environment state transition $S(t-1) \rightarrow S(t)$
1: Reconstruct cluster member, maintaining $|K(t)| = |K(t-1)|$;
2: $\forall K_i$, set source agent $K_i^{(s)} \rightarrow \emptyset$;
3: Operate repeatedly in K: for a target agent K_i
4: Obtain priority table q and p based on (6.2);
5: Notify source agents $K_i^{(s)}$, transfer $p \rightarrow p^{(s)}$ based on (6.3);
6: Transfer $p^{(s)} \rightarrow p^{(t)}$ from source agents $K_i^{(s)}$ based on (6.4);
7: Transfer $p^{(t)} \rightarrow p$ based on (6.5);
8: Transfer $Q(t-1) \xrightarrow{p} Q(t)$ based on (6.7);
9: Set $K_{i+1}^{(s)} = K_i^{(s)} \cup K_i$;
10: Terminate when $\forall K_i$, $K_i^{(s)} = K - K_i$

6.8 Simulation

In this section, we examine the system in multiple aspects, to investigate and validate the value mapping algorithm of transfer learning in dynamic network environments.

The small cell access network architecture in Figure 6.1 is used. Simulation parameters are listed in Table 6.3 below.

Table 6.3. Simulation Parameters

Parameters	Values
Number of Fixed ABSs	12
Number of Dynamic ABSs	8
Number of MSs	600
Frequency Reuse Clustering	Neighbouring Cells
Transmit Power	-3 dBW
Bandwidth	20 MHz
Number of Channels	20
Thermal Noise	-174 dBm/Hz
Inter-arrival time	Pareto distribution
Mean file size	0.5 Mb
Antenna	Omni-directional
Propagation	WINNER II B5a [98]
Link Selection	Best signal

6.8.1 ***Start-up Performance***

This section examines the network performance when all 20 ABSs are newly activated. During the start-up stage, the ABSs have to build up the knowledge base by learning the environment from a number of trial-and-error actions. Convergence is a crucial issue in traditional reinforcement learning based cognitive radio networks [64, 68, 73], where the speed of an agent achieving a stable channel set is very slow. Transfer learning with start-up Q value generation provides the intelligent agent with preliminary knowledge of the surrounding radio environment. In the high traffic load scenario, initial prioritization is particularly important because the action space should be quickly partitioned.

In this section, a set of temporal performance results will be assessed to demonstrate the convergence efficiency. The system is examined until all the ABSs have an offered traffic level of 270 Mb/s. A transfer learning algorithm is performed on linear reinforcement learning and Q learning, with a comparison to these algorithms operated in a fully distributed manner without transfer learning.

The convergence efficiency is demonstrated in Figure 6.3, which is assessed through the probability of stable states defined in Table 5.2. The performance is evaluated every 1000 learning iterations.

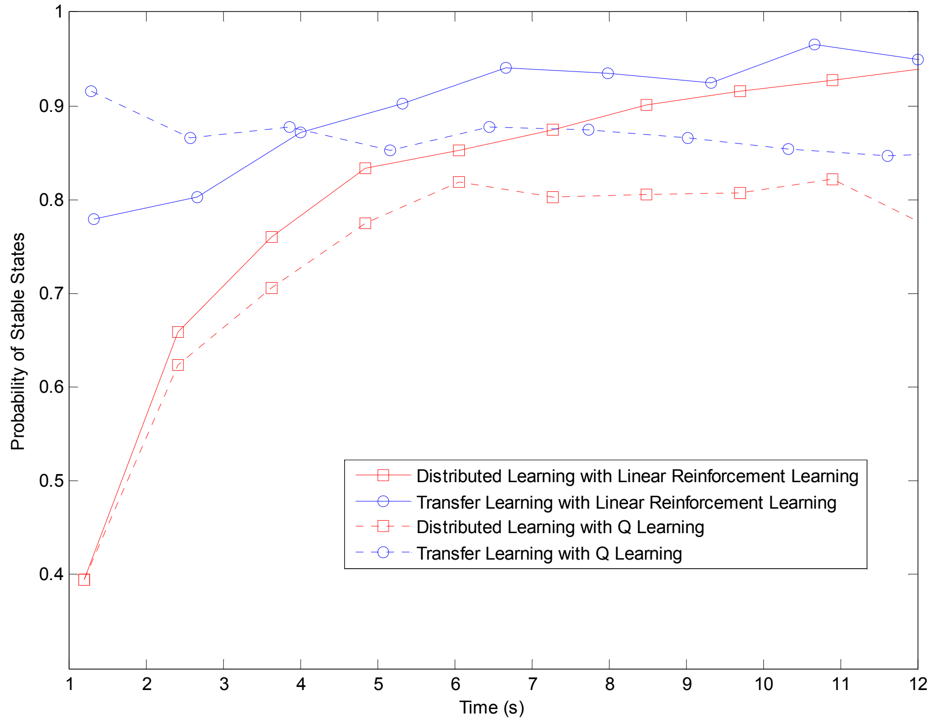

Figure 6.3. Convergence Efficiency (Start-up Performance)

It can be seen that the network with transfer learning benefits 40% to 50% more stable states than the distributed learning from the starting stage. Transfer learning keeps the stable state probability at 80% to 90% throughout. The network with distributed learning algorithms approaches convergence 3000 iterations later than with transfer learning. It then stays at a slightly lower level. We conclude that transfer learning significantly improves convergence on reinforcement leanring. Furthermore, it can be observed that the convergence efficiency of linear reinforcement learning strategy is 10% better than Q learning. This also validates the conclusions produced in Chapter 4.

The QoS is demonstrated as a CDF of the retransmission probability and mean delay per file, respectively. The retransmission performance shown in Figure 6.4

effectively presents the probability of failed decisions in learning, which affects the back off delay of file transmission.

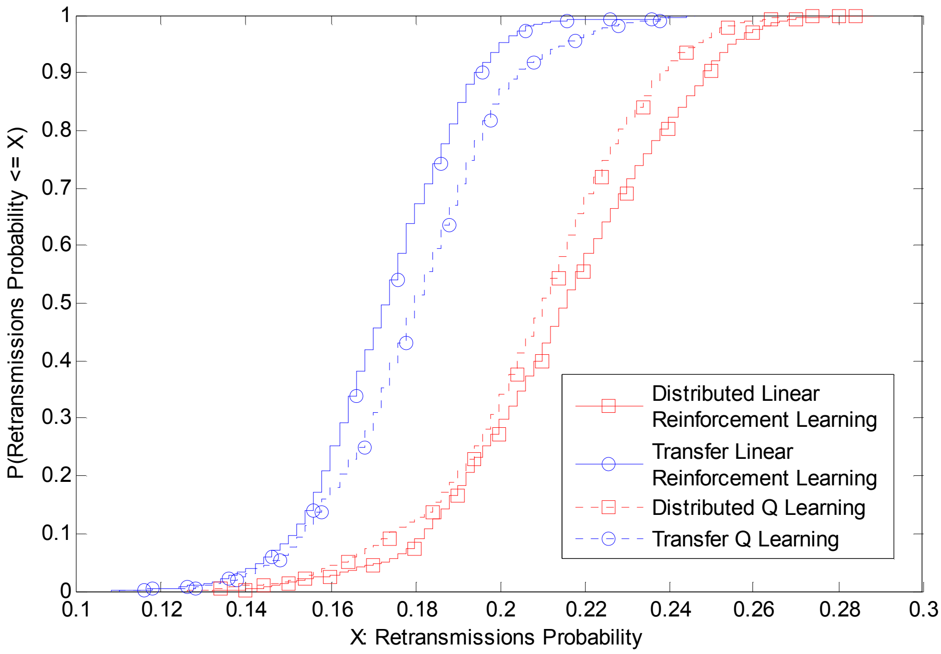

Figure 6.4. Probability of Retransmissions (Start-up Performance)

It can be seen that the network retransmission probability of transfer learning is between 15% and 20%, while the distributed learning algorithms are 4% higher. Compared with Figure 6.3, it can be concluded that transfer learning provides a higher QoS through faster convergence. Moreover, distributed learning converges to a set of poor channels, which causes significantly more retransmissions. Transfer learning algorithms effectively partition the channel set for each agent, which contributes to both good decisions and fast convergence.

The CDF of mean delay per file through all the learning iterations is presented in Figure 6.5. A similar improvement when applying transfer learning is achieved to that of the retransmissions probabilities. It can be seen that 90% of the iterations have a delay lower than 0.6s with transfer learning, whereas with distributed learning only 50% achieve the same delay band. It can be concluded that transfer learning quickly constructs a low latency network in a newly established architecture.

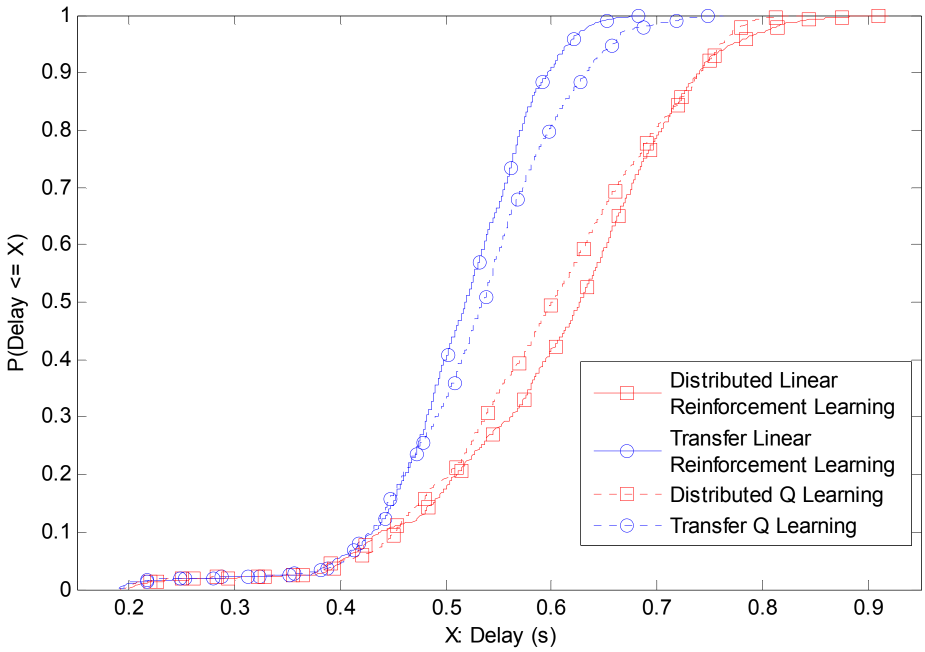

Figure 6.5. Mean Delay per File (Start-up Performance)

6.8.2 *Traffic and Topology Transition*

This section examines transfer learning during changes to traffic load and the network topology. In a flexible network architecture scenario, dynamic base stations are deployed during periods of heavy user traffic loads. A steady QoS level is desired provided that more base stations are activated. However, this is usually hard to achieve in practice, because a distributed cognitive agent require a number of iterations to learn the changes in the traffic and topology environment. The transfer learning policy presented before is designed to solve this problem.

Figure 6.6 presents a typical transition of traffic profile and network topology. A burst of user traffic occurs after 12000 data files generated in the network, lasting until 43000 files generated. The traffic level increases from 150Mb/s to 300Mb/s. The 8 dynamic ABSs are activated to provide capacity enhancement to the other 12 fixed ABSs during this period, and are deactivated after the traffic burst.

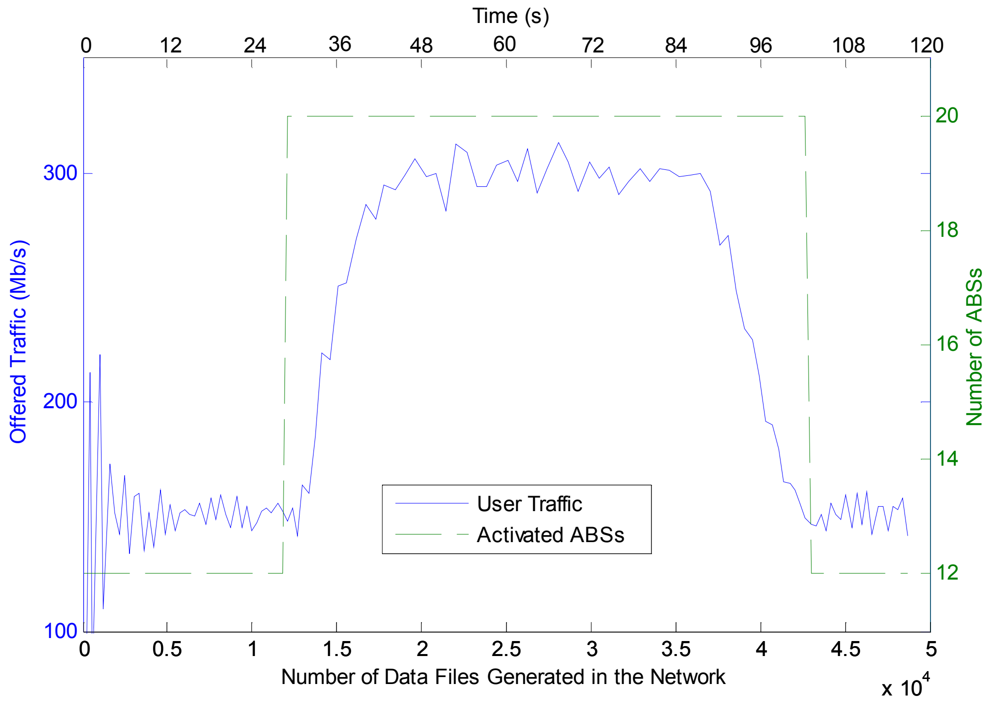

Figure 6.6. Traffic and Topology Transitions

The QoS performance is assessed by considering the retransmission probability and delay on a temporal basis, which clearly shows the network behaviour at different stages of the transition. Figure 6.7 demonstrates that the network with transfer learning achieves a steady and reliable retransmission probability throughout all the phases. The probability of retransmission with either linear reinforcement learning or Q learning is around 0.5% after the initial convergence of 50000 generated files. The network with distributed learning has a step change in retransmission probability up to 3% after the activation of dynamic ABSs, because it takes a number of iterations for them to learn the action space in the new environment. This learning period also interferes with the adjacent ABSs that have already converged to a set of channels.

The distributed reinforcement learning algorithm stays at a high probability level, while the distributed Q learning brings retransmissions down to 1% as more iterations being learnt. This also validates the convergence analysis of reinforcement learning algorithms in Section 4.4.2.3, showing that Q learning is more adaptable to environment changes and linear reinforcement learning and provide more stable decisions.

The network with transfer learning is not affected by the traffic variation and the changes of the dynamic ABSs. A reliable 0.5% retransmission rate is achieved with very small fluctuations during the whole event. It can be concluded that the harmful impact of traffic and topology changes is largely mitigated by transfer learning.

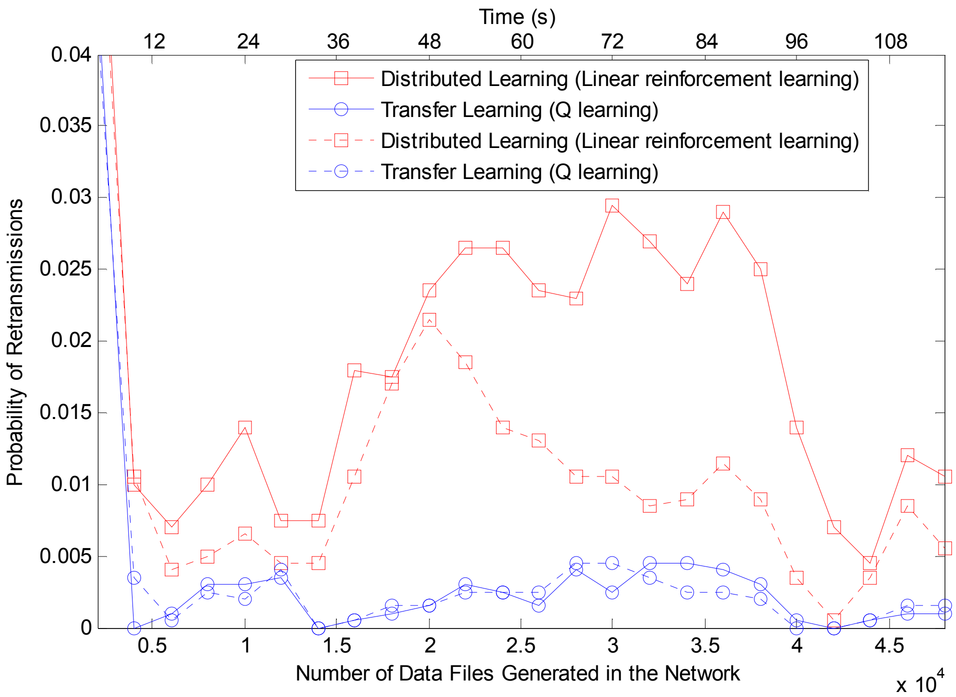

Figure 6.7. Probability of Retransmissions (Single Transition)

The delay performance is presented in Figure 6.8, which has similar behaviour compared to the retransmission probability. A consistent 0.15s delay is achieved with the transfer learning algorithm, regardless of the changes in the architecture and traffic. The distributed learning algorithms experience a significant step change in the delay during the high traffic period, which increases to around 0.25s. Q learning reduces the delay significantly down to 0.18s in the later stages, which behaves the same as the retransmission probability. The retransmitted files largely contribute to the increase in delay, which is also caused by protracted rewards from the environment changes.

It can be concluded that transfer learning reduces the negative impact of user traffic and network topology transitions down to a minimum, and provides a flexible operation of ABSs according to traffic level. A steady and reliable QoS level is provided to users regardless of the increasing number of ABSs and user traffic. The interference between dynamic cells is largely mitigated.

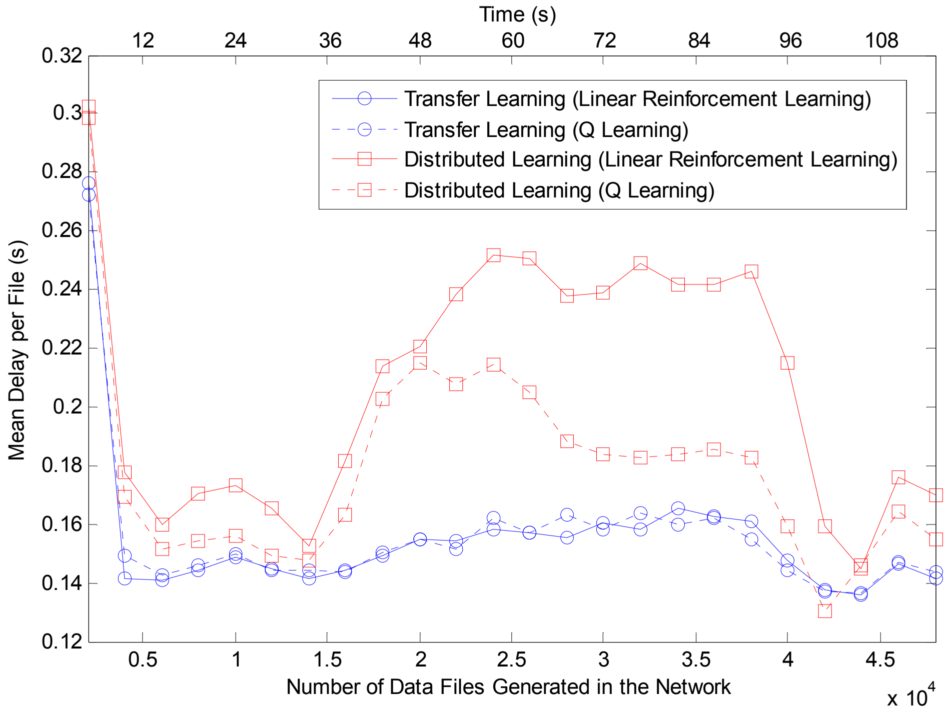

Figure 6.8. Mean Delay per File (Single Transition)

6.8.3 *Dynamic Traffic and Topology Fluctuation*

The traffic profile in an urban area is highly dynamic in both the time and spatial domains throughout different times of a day or a week, as illustrated in Figure 2.4. In this context, the dynamic ABSs can be switched on and off frequently to follow the variation of user traffic. The continuous and rapid changes of the network architecture could incur great challenges in controlling the interference between cells and managing the capacity allocated to each.

In this section, a regular fluctuation of traffic profile and network architecture is examined as presented in Figure 6.9. The traffic load changes periodically between 150Mb/s and 300Mb/s. The low traffic period lasts for 6000 generated files and the peak traffic period lasts for 30000 generated files. The dynamic ABSs are dynamically switched on and off based on the traffic level. This profile presents the same concept illustrated in Figure 2.4 where the regular peaks and low traffic periods represent the daytime and evening time in a week, respectively.

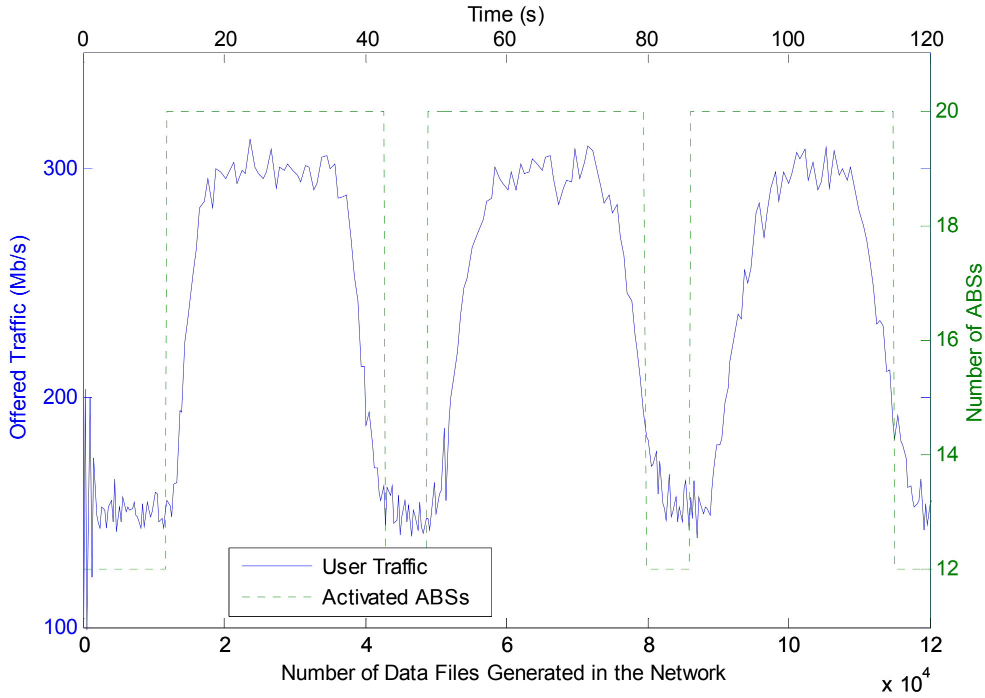

Figure 6.9. Dynamic Traffic and Topology Fluctuations

In this scenario, a cognitive agent (base station) has the memory to store their knowledge base in both active and sleep modes. The Q table learnt from previous phases of traffic/topology is directly applied to the new phase. In reinforcement learning, this may provide past experience to the agents but it may not be applicable to the new environment.

The probability of retransmission is demonstrated in Figure 6.10. It is clearly shown that the distributed learning algorithms are largely affected by the traffic and network variation. Linear reinforcement learning has three peak levels at 2.5% during the high traffic period. Q learning benefits from the previous learning experience in the high traffic period, where the second peak retransmission rate is 0.5% lower than the first. This can be illustrated from (6.1) and (6.2) where Q learning has a faster transition time because the Q value on the previous selected actions can be reduced exponentially. The learning rate assists an agent quickly converging to a different channel set in the new phase, which contributes to better QoS. In linear reinforcement learning, it is difficult for an agent to drop out from a converged action space where the Q value is relatively high.

The transfer learning algorithm is shown to significantly improve linear reinforcement learning and Q learning algorithms. A stable retransmission probability at around 0.3% is achieved, with a small variation of 0.2% during the periodical changes of traffic and network topology. The average retransmission probability is significant lower than that achieved in distributed learning algorithms. It can be concluded that transfer learning effectively supports dynamic topology operation in a rapidly changeable user traffic scenario.

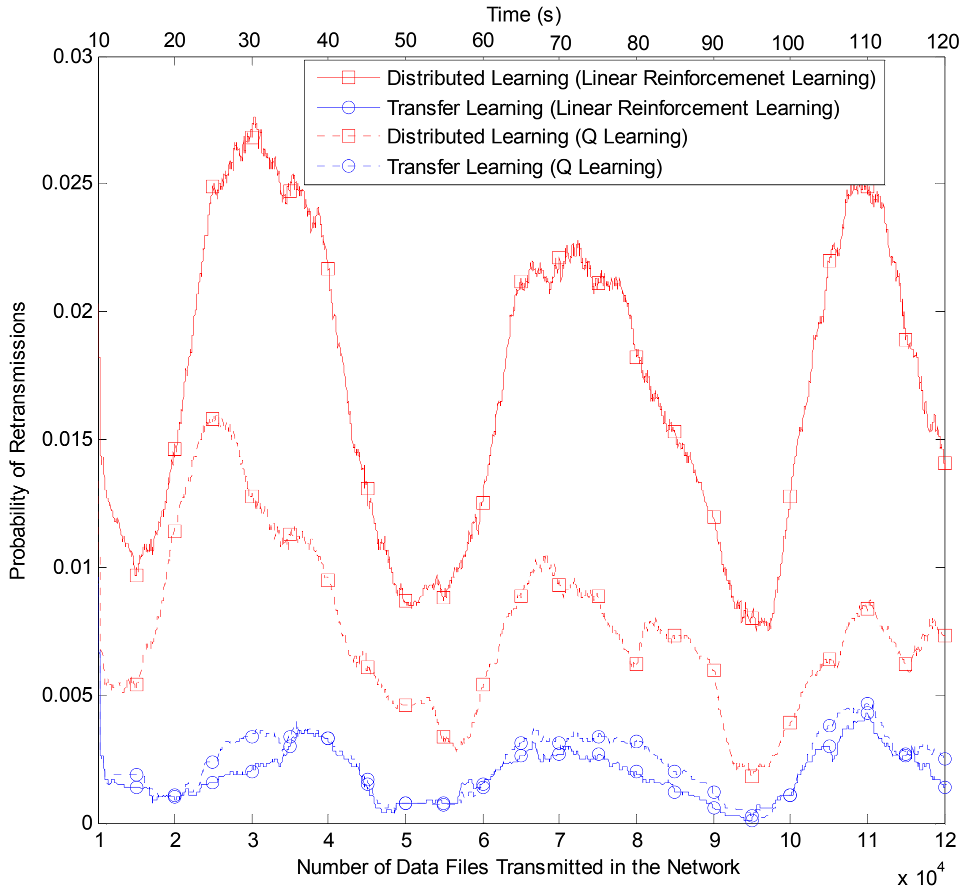

Figure 6.10. Probability of Retransmissions (Dynamic Fluctuations)

The corresponding delay performance is presented in Figure 6.11. The distributed reinforcement learning has a large variation of delay between 0.16s and 0.24s. Q learning achieves lower delay than reinforcement learning, with a continuous improvement in the subsequent high traffic period. The transfer learning algorithm achieves a much smoother variation than distributed learning, reaching 0.02s. It achieves up to 0.09s lower delay in the high traffic period and 0.02s lower delay in the low traffic period. The delay performance of transfer learning shows a higher fluctuation than retransmission probability. This is because channel reuse in a high traffic period reduces the data rate. However, steady and reliable delay performance is achieved through prioritization of Q tables at the start of each traffic and topology phase.

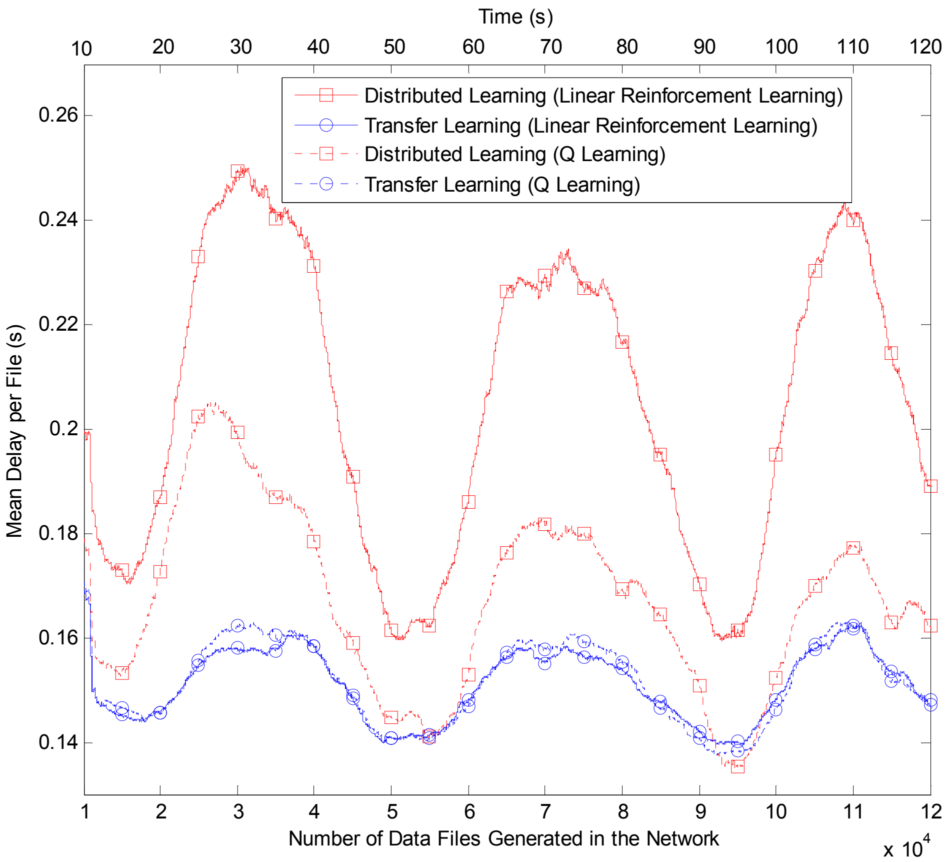

Figure 6.11. Mean Delay per File (Dynamic Fluctuations)

6.9 Conclusion

This chapter proposes a value mapping method in the context of Transfer Learning to improve the distributed reinforcement learning in dynamic radio scenarios. The dynamics of offered traffic and network architecture have been examined, with linkage to three typical scenarios including femto cells, energy efficient architecture and opportunistic networks. The environment identification and adaption efficiency of conventional distributed learning algorithms has been analysed.

A value mapping algorithm is designed under transfer learning, to train the knowledge base during environment state transitions. The dynamic frequency reuse clustering strategy is proposed based on keeping the cluster size fixed, to maintain the same frequency pattern and control the inter cell interference. A Pareto improvement resource prioritization method has been developed, which dynamically share the capacity between cluster members. A action-value mapping strategy is proposed to associate Q values with prioritized action space. This enables individual agents to carry out fully distributed resource management after transfer learning. Furthermore, a Q value generation scheme is designed to provide discriminated information to the Q table at the start-up stage.

Transfer learning is designed as a generic algorithm which is applicable to many reinforcement learning algorithms. This section has examined its application to linear reinforcement learning and single state Q learning in dynamic traffic and topology scenarios. In the start-up stage performance, transfer learning is shown to converge much faster than distributed learning, with better QoS achieved. A steady and reliable QoS level is achieved on transfer learning during the transition to different user traffic levels and network topologies. Furthermore, transfer learning effectively reduces the QoS fluctuations in a highly dynamic network environment.

Chapter 7. Dynamic Capacity Provision and Topology Management

7.1 Introduction

In this chapter, dynamic topology management is developed to autonomously manage the architecture of the small cell access network. Chapter 6 introduces the flexible network architecture scenarios with transfer learning applied to provide reliable communication. However, the network topology was manually controlled by the service provider. This chapter investigates topology management strategies based on the dynamics of the traffic profile in both the time and spatial domains. Decisions of the time, location and number of ABSs deployed/removed can be made in a distributed and self-organized manner.

Capacity provision is the fundamental requirement of resource and topology management in a wireless cellular network. The deployment of a base station enhances the network capacity by providing additional frequencies or better spectrum reuse in a local area. Resource management schemes have a direct impact on capacity provision, especially handling dynamic traffic variations.

Section 7.2 defines topology management in cellular networks and studies its relationship with system capacity. Section 7.3 analyses capacity provision of frequency band allocation and transfer learning strategies through a Markov model.

The dynamic topology management algorithm is proposed in Section 7.4. Simulation results and conclusion are provided in Section 7.5 and Section 7.6.

7.2 Network Topology and Capacity in Cellular Systems

Network topology is a terminology used in a number of wireless scenarios with different applications. It is generally specified as the location of nodes and the connections between them. In a wireless ad-hoc network, topology management generally refers to routing algorithms in the network layer. The target is to establish an end-to-end connection through multi-hop links from a source node to a destination node. Thanks to the high flexibility of the wireless link selection in this type of network, a wide range of topology management techniques can be applied, such as clustering or evolutionary algorithms.

Topology management in cellular networks is much simpler than in ad-hoc networks. In current and near future cellular communication systems, the single hop wireless link are generally used between mobile stations and access base stations [23]. The wireless backhaul network is managed by a hub base station with a highly controlled network topology [12]. This is because cellular communication is designed to provide highly reliable links with steady QoS, throughput, capacity, etc. A multi-hop architecture in the access network requires additional hardware functionality in mobile stations and makes network management excessively complex. Topology management in cellular network is thus generally considered to be the planning of base stations and their connection with mobile users, namely cell planning and access link selection.

Dynamic topology management is vital to support the network in various scenarios. Energy efficient network management is a major scenario that requires topology management. It has been reviewed in Section 2.5 that 60% to 80% of the total energy consumption is contributed by the operation of base stations [115]. Dynamically switching on and off the base stations according to local traffic variations can thus reduce a significant amount of energy consumption. Topology management is designed to autonomously identify traffic intensity and provide a seamless map of base station deployment. The algorithm should guarantee an adequate level of throughput and QoS, which is not reduced by turning base stations into sleep mode.

On the other hand, it is aimed at reducing the number of active base stations down to a minimum in order to save energy.

Opportunistic architectures represent a newly proposed communication network for unexpected and temporary events scenarios the ABSOLUTE [113] project. The network is aimed at providing coverage and capacity where conventional architecture is destroyed or cannot provide adequate QoS, during the period of disaster relief or unplanned events. The roll-out and roll-back of an opportunistic network requires a dynamic traffic aware network planning strategy to deploy and remove the base stations. A dynamic deployment map is desired to provide the number, location and time of different types of base stations in various phases. In this context, an autonomous topology management strategy is essential to deliver fast and adaptable network architecture. It also saves the energy on opportunistic base stations [113] where steady power supply is not always available. Furthermore, the cost of network deployment to the operators can be significantly reduced.

Given the same objective of energy and cost saving in both scenarios, a generic topology management strategy can be designed to control the base station's working modes.

The major role of a cell in a wireless cellular network is to provide capacity. The target of dynamic network planning is to match the level of capacity with user traffic. Capacity provision is determined by different resource management schemes. It can be supplied by either extra spectrum resources or enhanced spectrum reuse. Chapter 2 reviews two major categories of resource management strategies: Frequency band Allocation (FA) and Dynamic Spectrum Access (DSA). Capacity of FA is more constrained by the spectrum band and that of DSA is limited by interference. Transfer learning proposed in Chapter 6 achieves Pareto efficient resource utilization in a cluster of cells, where capacity of a cell can be automatically adjusted by the user behaviour. Interference can be eliminated in a group of cells before the whole spectrum is fully utilized.

7.3 Dynamic Capacity Provision

Dynamic capacity provision between cell is essential to reduce the number of base stations required in a network, because it has the potential to improve resource utilization in a given network topology. This section analyses capacity provision from Frequency band Allocation and Transfer Learning strategies, through the use of Markov models. Dynamic capacity provision is validated on transfer learning under the Pareto efficient action space prioritization. The Erlang B queuing model is considered here, where traffic buffering is not available [4].

Frequency band allocation and transfer learning are operated on a multi-cell environment. A multi-dimension Markov model is thus essential to describe the system behaviour. [116] presents a two dimensional Markov model for the coexistence of two overlaid aerial cells, with fixed frequency bands allocated to each.

The network scenario considered here has multiple neighbouring cells with no overlapped coverage, as the access network applies a shortest path routing strategy, where the users are connected to the nearest activated base station. However, the neighbouring base stations may cause excessive interference with each other. A connection may be interrupted if the same channel is shared between. The Markov model is built on a cluster basis, which is considered to be a group of cells with the entire spectrum pool allocated.

An example three BS model is illustrated in Figure 7.1, where b_3 is a base station that can be switched off in a low traffic period. The size of the entire cluster is fixed. The coverage area of b_1 and b_2 can be extended when b_3 is off. In this context, the cluster $\{b_1, b_2\}$ can be split into $\{b_1, b_3\}$ and $\{b_2, b_3\}$ after b_3 is switched on. Furthermore, the user arrival rate on b_1 and b_2, denoted as λ_1 and λ_2, is partly taken by b_3 as well depending on their locations.

The main purpose of topology management is to provide sufficient capacity through base station deployment in its corresponding coverage area. This energy efficient network architecture is aimed at supporting a high offered traffic with a minimum number of active base stations. The Markov analysis is firstly based on a two cell single cluster model. A three cell two overlapped cluster model is then analysed when b_3 is activated.

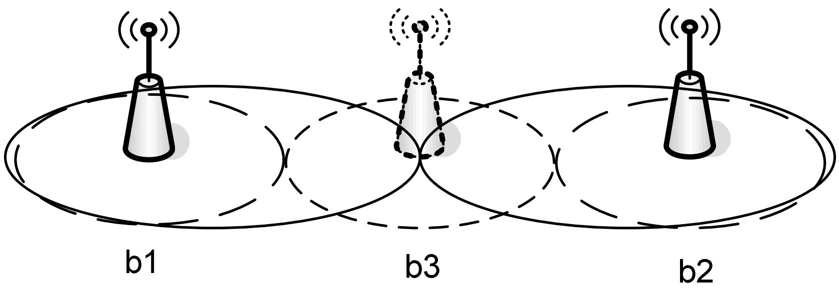

Figure 7.1. Three BS dynamic topology model

7.3.1 *Two Cell Single Cluster Model*

Cluster $\{b_1, b_2\}$ is constructed when b_3 is inactive, following the frequency reuse clustering scheme. A total number of n channels are available to the users in this cluster area, which are not permitted for reuse in order to avoid interference. The user arrival rate in each BS under equal cell sizes is λ_1 and λ_2. The departure rate is μ.

In order to compare the frequency band allocation and the transfer learning scheme, a heterogeneous Markov diagram is developed as shown in Figure 7.2. The whole triangle diagram represents the states under transfer learning, where Pareto efficient resource utilization is achieved in a cluster. The diagonal line denotes the states where all the n channels are occupied by users in the cluster. The system capacity will be full when approaching this line. The rectangular area denotes the states under frequency band allocation, where n_1 and n_2 are the size of frequency bands allocated to each cell. It can be observed that transfer learning provides more flexible states to the system, thus the probability of blocking can be reduced.

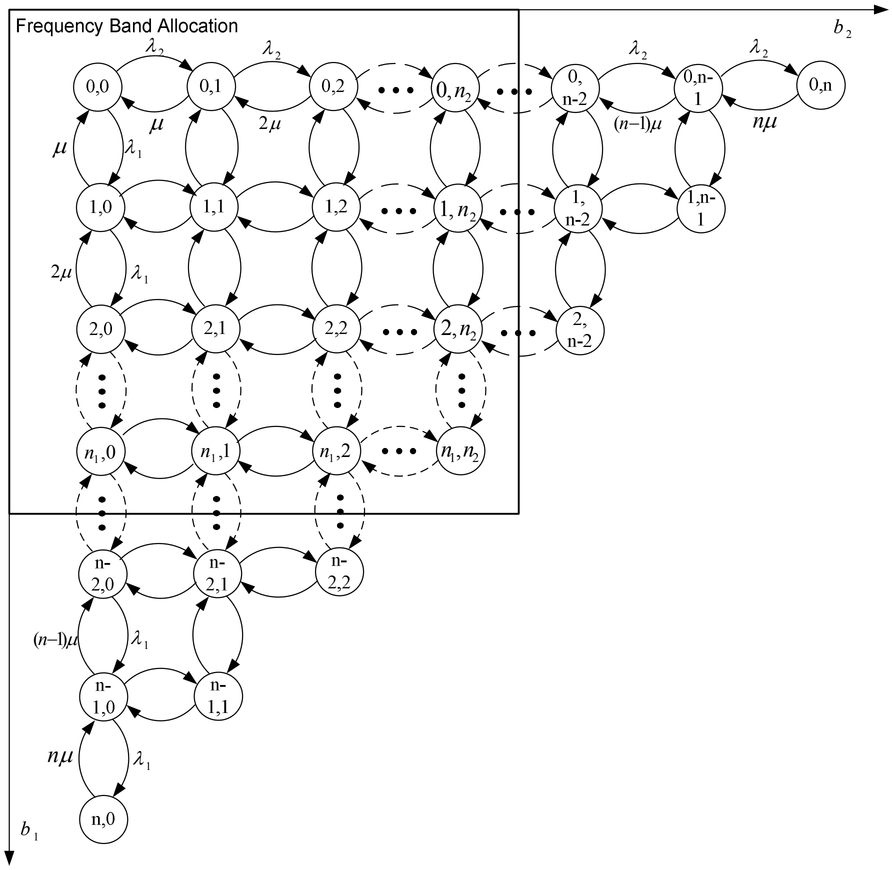

Figure 7.2. Two Cell Single Cluster Markov model

Transfer Learning

Under the Pareto efficient action space prioritization delivered by transfer learning, the global function of the transfer learning scenario can be expressed as a summation of states in the triangle area:

$$\sum_{j_1=0}^{n}\sum_{j_2=0}^{n-j_1} p(j_1, j_2) = 1 \tag{7.1}$$

The transition probability between states can be described in the equilibrium function. For a general state (j_1, j_2), $j_1 \in [0, n]$, $j_2 \in [0, n - j_1]$, we have

$$(\lambda_1 + \lambda_2 + j_1\mu + j_2\mu)p(j_1, j_2)$$

$$= \lambda_1 p(j_1 - 1, j_2) + \lambda_2 p(j_1, j_2 - 1) + \mu p(j_1 + 1, j_2) + \mu p(j_1, j_2 + 1) \tag{7.2}$$

No states exist when $\forall j_1, j_2 < 0$, we have

$$p(j_1, -1) = p(-1, j_2) = 0 \tag{7.3}$$

Similarly, no states exist beyond a boundary state $j_1 + j_2 > n$:

$$p(j_1, n - j_1 + 1) = p(n - j_2 + 1, j_2) = 0 \tag{7.4}$$

The purpose of the Markov analysis is to obtain the system probability at each state. There are multiple ways to solve a multi-dimensional Markov chain. The numerical method provides a straightforward approach to the solution [117]. It defines a transition matrix **P** to include all transition probabilities between states, which can be derived from (7.2) to (7.4). The distribution over the states can be written as a stochastic row vector **x** with the relationship of

$$\mathbf{x}^{(n)} = \mathbf{x}^{(n-1)}\mathbf{P} = \mathbf{x}^{(n-2)}\mathbf{P}^2 = \cdots = \mathbf{x}^{(0)}\mathbf{P}^n \tag{7.5}$$

where $\mathbf{x}^{(0)}$ is an initial probability. **x** is expected to converge to a stable vector after a sufficiently large number of iterations n, which is effectively the system probability at each state. MATLAB simplifies the process of solving such complex matrix operations. It is possible to get the probability over a group of states when the condition is well defined. The blocking probability in this scenario can be written as a summation of states with n channels:

$$P_b = \sum_{j=0}^{n} P(j, n-j) \tag{7.6}$$

Frequency band Allocation

The rectangular part of the diagram in Figure 7.2 with the defined band sizes n_1 and n_2 shows the frequency band allocation scenario. The global function is expressed as

$$\sum_{j_1=0}^{n_1} \sum_{j_2=0}^{n_2} p(j_1, j_2) = 1 \tag{7.7}$$

The equilibrium function for a general state is the same as (7.2) and (7.3). The probability beyond a boundary state that $j_1 > n_1$ and $j_2 > n_2$ is expressed as

$$p(n_1 + 1, j_2) = p(n_1, j_2 + 1) = 0 \tag{7.8}$$

The state probabilities can be obtained based on the global and equilibrium functions, through numerical methods presented in (7.5). The blocking probability is expressed as

$$P_b = \sum_{j_2=0}^{n_2} \frac{\lambda_1}{\lambda} p(n_1, j_2) + \sum_{j_1=0}^{n_1} \frac{\lambda_2}{\lambda} p(j_1, n_2) \tag{7.9}$$

Analytical Results

The objective of this analysis is to investigate how transfer learning provides dynamic capacity to cluster members, as well as delivering adequate QoS at high traffic loads without switching on more base stations.

Equation (7.5) is used to generate the analytical results. A total number of 20 channels are allocated to the whole cluster area. In the first part of analysis, we set up an equal offered traffic to each cell, namely $G_1 = G_2 = 7$ (Erlang). Figure 7.3 presents the system probabilities for each state, where channels are dynamically shared between the two cells. The diagonal line is the boundary of the system capacity where the summation of b_1 and b_2 coordinates is 20. The rectangular boundary indicates the border of the frequency band allocation scheme with 10 channels assigned to each cell. The colour depth denotes the state probabilities.

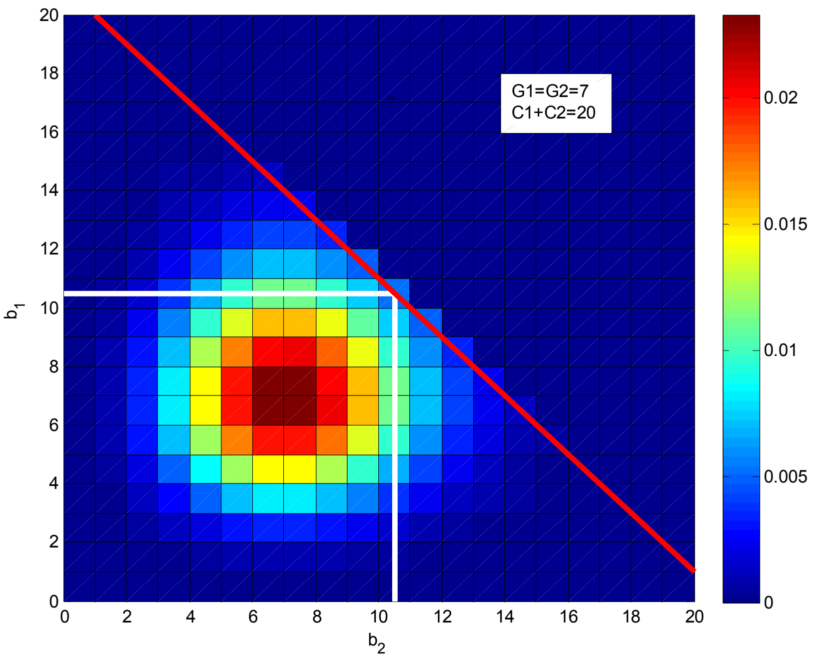

Figure 7.3. State Probabilities of the Two Cell Markov Model

It can be observed that the highest state probability occurs when 7 channels are occupied in each cell, which is due to the 7 Erlangs traffic load. On the other hand, the system may stay in some states beyond the 10 channel bound in each cell, because the user arrives according to a Poisson random distribution. This potentially causes transmissions to be blocked in the frequency band allocation scheme. In the scenario where the user traffic is unequal, the rectangular boundary should follow the state probability pattern, otherwise the system capacity is highly constrained.

The blocking probability under a variation of offered traffic between two cells is presented in Figure 7.4, which is also based on the topology in Figure 7.1. The x axis indicates a variation of traffic load proportion between b_1 and b_2, starting from an even traffic level shared at 1. The frequency band allocation scheme fixes 10 channels to each BS, with blocking probability solved from (7.7) to (7.9). A channel borrowing scheme is modeled which changes the band size according to traffic load, as presented in Section 2.3.2. The Pareto efficient bound, solved from (7.1) to (7.6), indicates the situation where all channels can be utilized regardless of traffic variation, which can be achieved by transfer learning illustrated in Chapter 6.

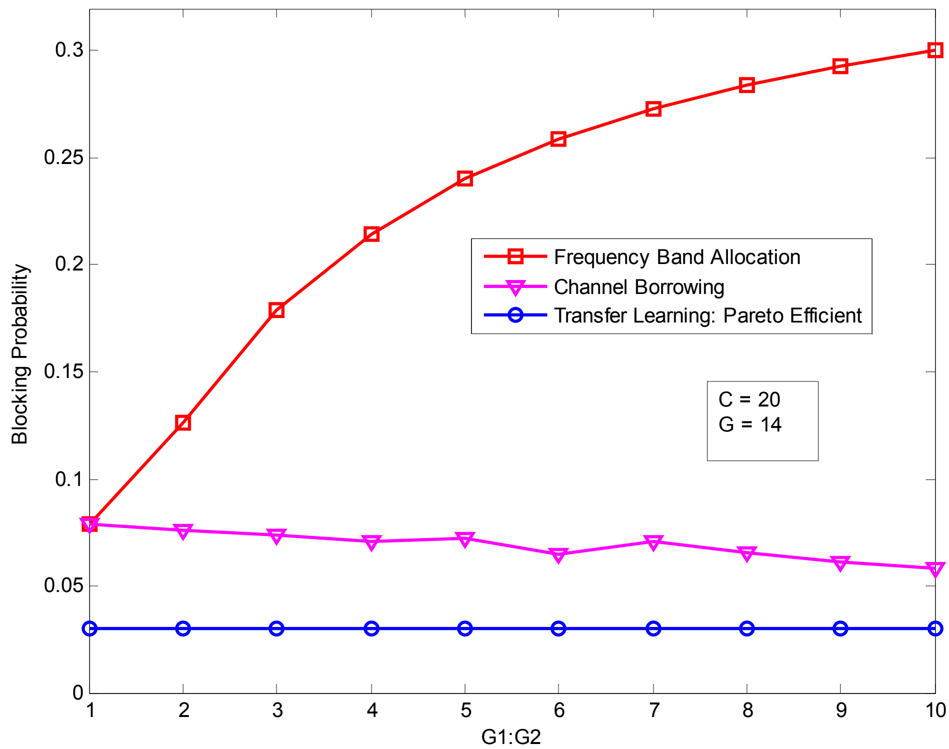

Figure 7.4. Blocking Probability vs Traffic Load Proportion

It can be seen that a consistent 3% blocking probability is achieved under Pareto efficient resource utilization. The channel borrowing scheme has a blocking probability that is twice as the Pareto efficient bound because in Figure 7.3, the states beyond the rectangular boundary may be visited by the system with some fail probability. The frequency band allocation strategy has a dramatic increase in blocking probability because of the fixed band size, particularly when the traffic ratio in two cells G1:G2 < 4. It can be concluded that transfer learning provides a consistent QoS with dynamic capacity provision. An effective channel borrowing scheme can significantly improve QoS under frequency band allocation. However, it is difficult to implement this scheme in a practical network due to interference between overlapped bands, as described in Section 2.3.2.

7.3.2 ***Three Cell model***

7.3.2.1 Two Cluster Model with Frequency Reuse

This section presents a three dimensional Markov model for the three cell scenario when b_3 is switched on. The system has two overlapping clusters: $\{b_1, b_3\}$ and $\{b_2, b_3\}$ and frequencies can then be reused between b_1 and b_2. The Pareto efficient resource allocation may not be achieved in each cluster, because b_1 and b_2 can only assign channels not currently occupied by b_3.

Figure 7.5 demonstrates a heterogeneous system model including both frequency band allocation and the transfer learning strategy. In transfer learning, the $\{b_1, b_3\}$ and $\{b_2, b_3\}$ planes have a limit of n channels as in Figure 7.2. The $\{b_1, b_2\}$ plane has the probability of assigning $2n$ channels, though it depends on the number of channels used in b_3. By defining the boundary state as $P(j_1, j_2, j_3)$, $j_1 + j_3 = n$ applying to plane $P(n, n, 0) - P(0,0, n) - P(0, n, 0)$, and $j_2 + j_3 = n$ applying to plane $P(n, n, 0) - P(0,0, n) - P(n, 0,0)$. Furthermore, the value of j_1 and j_2 varies inversely with j_3. The frequency band allocation strategy assigns n_1, n_2, n_3 channels to b_1, b_2, b_3, respectively. Similarly, $n_1 + n_3 = n$ and $n_2 + n_3 = n$ also applies.

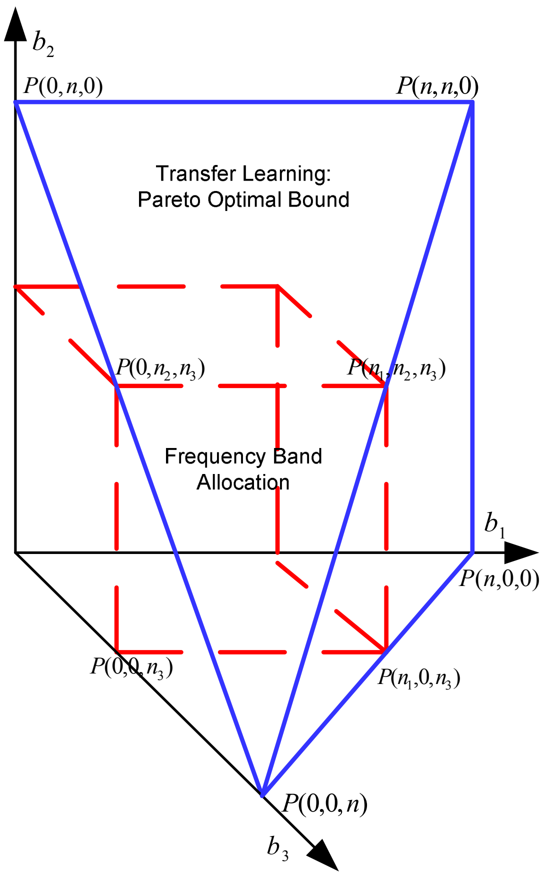

Figure 7.5. Three Cell Two Cluster Markov model

Transfer Learning

The global function of transfer learning in this scenario is a summation of states in the blue polyhedron, which can be written as

$$\sum_{j_3=0}^{n}\sum_{j_1=0}^{n-j_3}\sum_{j_2=0}^{n-j_3} p(j_1, j_2, j_3) = 1 \tag{7.10}$$

For a general state $P(j_1, j_2, j_3)$, $j_1 + j_3 \in [0, n]$ and $j_2 + j_3 \in [0, n]$, we have

$$(\lambda_1 + \lambda_2 + \lambda_3 + j_1\mu + j_2\mu + j_3\mu)p(j_1, j_2, j_3) \tag{7.11}$$

$$= \lambda_1 p(j_1 - 1, j_2, j_3) + \lambda_2 p(j_1, j_2 - 1, j_3) + \lambda_3 p(j_1, j_2, j_3 - 1) +$$

$$(j_1 + 1)\mu p(j_1 + 1, j_2, j_3) + (j_2 + 1)\mu p(j_1, j_2 + 1, j_3) + (j_3 + 1)\mu p(j_1, j_2, j_3 + 1)$$

There are no states exist when $\forall j_1, j_2, j_3 < 0$, we have

$$p(-1, j_2, j_3) = p(j_1, -1, j_3) = p(j_1, j_2, -1) = 0 \tag{7.12}$$

Similarly, there are no states existing beyond a boundary state $j_1 + j_3 > n$ and $j_2 + j_3 > n$, so we have

$$p(j_1, n - j_3 + 1, j_3) = p(n - j_3 + 1, j_2, j_3) = 0 \tag{7.13}$$

The blocking probability is then the states on two triangle border planes

$$\begin{aligned} P_b = \sum_{j_3=0}^{n} \sum_{j_2=0}^{n-j_3-1} \frac{\lambda_1 + \lambda_3}{\lambda} p(n - j_3, j_2, j_3) + \\ \sum_{j_3=0}^{n} \sum_{j_1=0}^{n-j_3-1} \frac{\lambda_2 + \lambda_3}{\lambda} p(j_1, n - j_3, j_3) + \sum_{j_3=0}^{n} p(n - j_3, n - j_3, j_3) \end{aligned} \tag{7.14}$$

Frequency band Allocation

The global function of frequency band allocation in this scenario is a summation of states in the red cube, which can be written as

$$\sum_{j_1=0}^{n_1} \sum_{j_2=0}^{n_2} \sum_{j_3=0}^{n_3} p(j_1, j_2, j_3) = 1 \tag{7.15}$$

The general state expression is the same as (7.11) and (7.12). There are three boundary planes follows n_1, n_2, n_3, where $n_1 + n_3 = n$ and $n_2 + n_3 = n$.

$$p(n_1 + 1, j_2, j_3) = p(j_1, n_2 + 1, j_3) = p(j_1, j_2, n_3 + 1) = 0 \tag{7.16}$$

The blocking probability is a summation of states on three rectangle border planes

$$P_b = \sum_{j_2=0}^{n_2} \sum_{j_3=0}^{n_3} \frac{\lambda_1}{\lambda} p(n_1, j_2, j_3) + \sum_{j_1=0}^{n_1} \sum_{j_3=0}^{n_3} \frac{\lambda_2}{\lambda} p(j_1, n_2, j_3) + \sum_{j_1=0}^{n_1} \sum_{j_2=0}^{n_2} \frac{\lambda_3}{\lambda} p(j_1, j_2, n_3) \tag{7.17}$$

7.3.2.2 Single Cluster Model without Frequency Reuse

The single cluster model is based on placing all the three cells into a cluster. This scenario is used as a comparison to the frequency reuse strategy that reconstructs the cluster when b_3 is switched on. The Markov diagram is illustrated in Figure 7.6, which is shown as an extension to the two cell model in Figure 7.2.

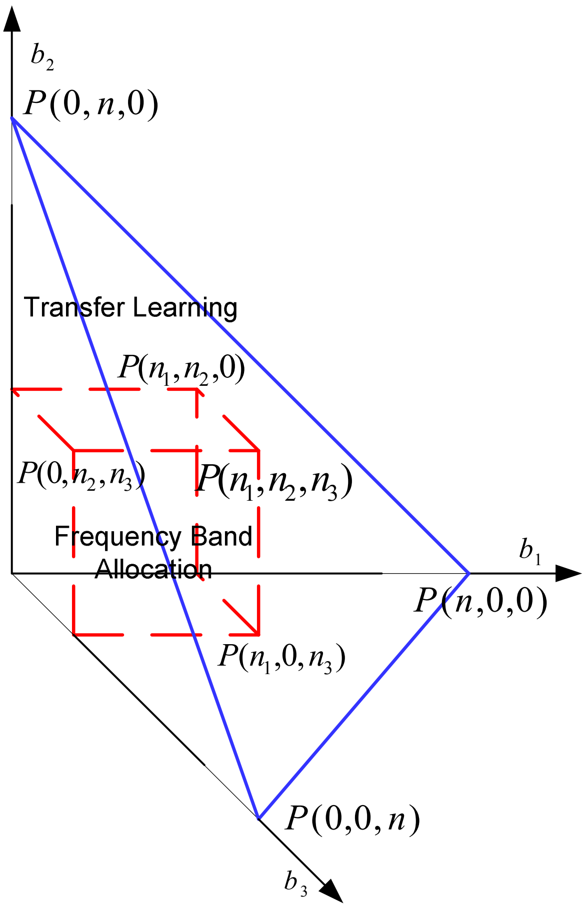

Figure 7.6. Three Cell Single Cluster Markov model

The triangular plane is the only boundary for transfer learning where $j_1 + j_2 + j_3 = n$. Similarly, the band sizes in frequency allocation follows $n_1 + n_2 + n_3 = n$. The global function for transfer learning is expressed as

$$\sum_{j_1=0}^{n}\sum_{j_2=0}^{n-j_1}\sum_{j_3=0}^{n-j_1-j_2} p(j_1, j_2, j_3) = 1 \tag{7.18}$$

The equilibrium function is the same as (7.11) and (7.12). The boundary limit is

$$p(j_1, j_2, n - j_1 - j_2 + 1) = p(j_1, n - j_1 - j_3 + 1, j_3)$$

$$= p(n - j_2 - j_3 + 1, j_2, j_3) = 0 \tag{7.19}$$

The blocking probability under transfer learning is

$$P_b = \sum_{j_1=0}^{n}\sum_{j_2=0}^{n-j_1} p(j_1, j_2, n - j_1 - j_2) \tag{7.20}$$

For the frequency band allocation strategy, the global function is

$$\sum_{j_1=0}^{n_1}\sum_{j_2=0}^{n_2}\sum_{j_3=0}^{n_3} p(j_1, j_2, j_3) = 1 \tag{7.21}$$

The equilibrium function is the same as (7.11) and (7.12). The boundary limit is

$$p(n_1 + 1, j_2, j_3) = p(j_1, n_2 + 1, j_3) = p(j_1, j_2, n_3 + 1) = 0 \tag{7.22}$$

The blocking probability under frequency band allocation is

$$P_b = \sum_{j_2=0}^{n_2}\sum_{j_3=0}^{n_3} \frac{\lambda_1}{\lambda} p(n_1, j_2, j_3) + \sum_{j_1=0}^{n_1}\sum_{j_3=0}^{n_3} \frac{\lambda_2}{\lambda} p(j_1, n_2, j_3) + \sum_{j_1=0}^{n_1}\sum_{j_2=0}^{n_2} \frac{\lambda_3}{\lambda} p(j_1, j_2, n_3) \tag{7.23}$$

7.3.2.3 Analytical Results

The purpose of this analysis is to validate the dynamic capacity provision achieved by transfer learning, and also to examine the capacity enhancement from switching on b_3 and frequency reuse.

In the first part of analysis, the user density is uniformly distributed in the whole area. Namely, the offered traffic in each cell follows $G_1 = G_2 (= G_3)$ in the scenario that b_3 is either on or off. The activated BSs are allocated 6 channels in total. For the frequency band allocation scheme, an equal number of channels are assigned to each cell. Namely each cell is assigned with $n/2$ channels in the frequency reuse scenario and $n/3$ channels in the non-reuse scenario when b_3 is activated.

Figure 7.7 demonstrates the system blocking probability with various traffic levels under a dynamic working/sleeping mode of b_3. It also illustrates a comparison between two cluster frequency reuse and the single cluster no reuse scheme when b_3 is switched on.

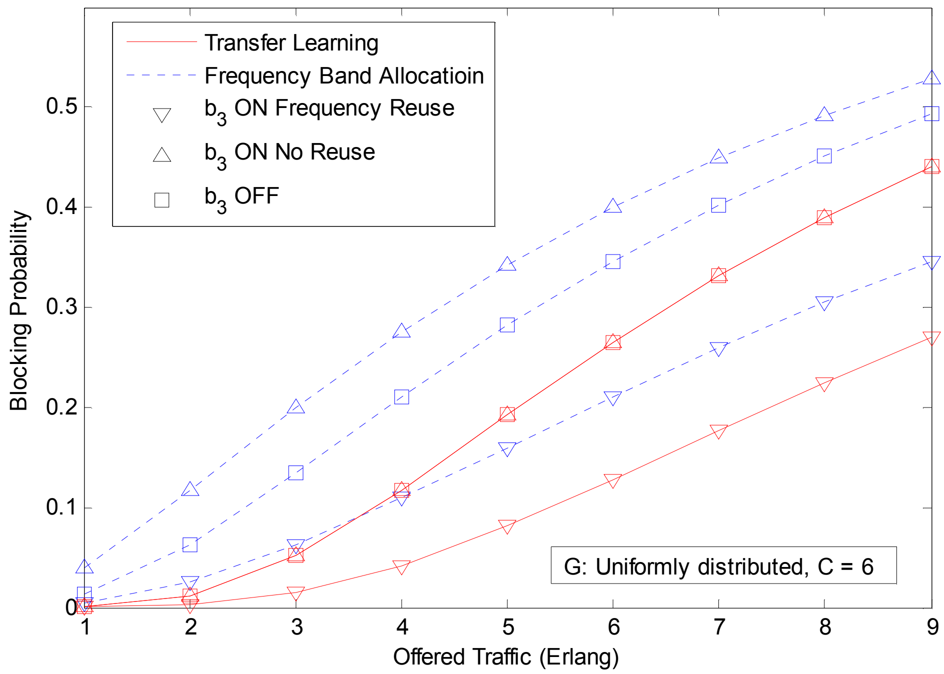

Figure 7.7. Blocking Probability (Topology Transition)

By applying frequency reuse after activating b_3, transfer learning achieves a 30% lower retransmission probability than frequency band allocation on average. This benefits from the dynamic capacity provision between b_1, b_3 or b_2, b_3, when user traffic is generated randomly among these cells. By comparing the topology impact from b_3, it can be seen that both transfer learning and frequency band allocation have a similar improvement in QoS. Their blocking probability difference gradually increases as the offered traffic increases, reaching 20% at 9 Erlangs. It can be concluded that switching on b_3 with appropriate frequency reuse provides effective capacity enhancement to both schemes. Furthermore, under the same blocking probability level, transfer learning supports around 1.2 Erlang higher offered traffic than frequency band allocation in the same topology. This will save energy consumption by keeping the same topology when traffic load increases.

The no reuse strategy indicates that switching on more BSs without frequency reuse provides no benefits to the system QoS. For the frequency band allocation scheme, the blocking probability of three cells is around 30% higher than two cells. This is because the entire spectrum is divided into more frequency bands, which further limits the flexibility of resource utilization in the area. Transfer learning achieves the same level of QoS under dynamic topologies, because the same Pareto efficient resource allocation is achieved without frequency reuse. It can be concluded that for the dynamic spectrum management scenario, a reconstruction of the frequency reuse cluster is necessary to obtain a QoS improvement from topology management.

The second part of the analysis demonstrates the benefit of transfer learning in handling a dynamic variation of offered traffic in the spatial domain. We investigate the scenario where b_3 is switched on and frequency reuse is operated. Figure 7.8 presents a comparison of different spatial traffic proportions between three cells. A cell with three times offered traffic than others is examined throughout the cluster.

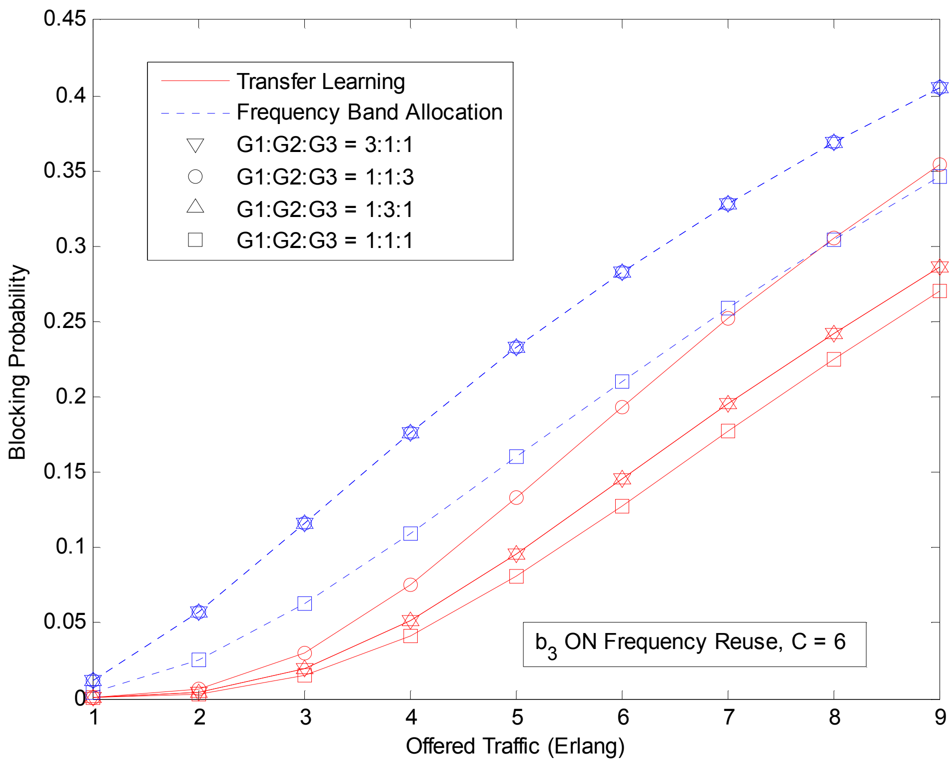

Figure 7.8. Blocking Probability (Spatial Traffic Variation)

In the frequency band allocation scheme, a cell with 3 times offered traffic of the others brings a 30% increase in blocking probability compared to a uniformly distributed traffic scenario. The system has the same blocking probability no matter whether the dominating traffic is generated in b_1, b_2, b_3, because the probability of approaching three boundary planes is the same according to (7.17).

The transfer learning scheme achieves the same blocking probability level no matter where the hotspot traffic occurs in the spatial domain. The scenario with uniformly distributed traffic shows the lowest blocking probability and the scenario with a dominating traffic in b_1 and b_2 is around 10% higher. A gradual increase of the blocking probability occurs when b_3 has a dominate traffic, reaching 30% more than uniformly distributed traffic at 9 Erlangs. This behavior indicates that transfer learning cannot achieve a Pareto efficient resource utilization when the traffic is not equalized in this scenario. Figure 7.5 illustrates that $\{b_1, b_3\}$ and $\{b_2, b_3\}$ are overlapped clusters in the area, with the potential of Pareto efficiency in each under transfer learning. However, the traffic on b_3 significantly constrains the remaining resources in neighbour BSs. Thus only one cluster can achieve Pareto efficiency when traffic is unequal between b_1 and b_2, which causes slightly higher blocking probability. Furthermore, the system capacity reduces when the traffic load on b_3 increases, because in Figure 7.5 the volume of the rectangular cuboids reduces when n_3 increases from 0 to n. This causes a significantly high level of blocking probability when b_3 is dominating the offered traffic. It can be concluded that transfer learning provides QoS improvements to frequency band allocation in general. The Pareto efficiency of overlapped clusters is constrained by resource utilization in the overlapped cells.

The Markov analysis in the three BSs dynamic topology scenario justifies the dynamic capacity provision from the Pareto efficient action space prioritization under transfer learning. The validated QoS levels under different topologies provide the design of traffic or QoS threshold for switching on and off BSs in topology management. Furthermore, a load balancing scheme can be designed for the overlapped cell to maximize Pareto efficiency in the overlapping clusters scenario.

7.4 Dynamic Topology Management

The purpose of topology management is to trade off the QoS and energy consumption. It should define the time and location of activating a base station, and its connection with users. Figure 6.1 and Figure 7.1 illustrate that the coverage areas of base stations have no overlap in this scenario, thus the connections between BSs and MSs are always shortest path. The Pareto efficient resource utilization indicates that resources in a cluster can be dynamically shared based on interim traffic load.

The scenario considered here is based on Figure 6.1, where the 12 Fixed ABSs are permanently activated and the 8 Dynamic ABSs in between can be switched on/off. The activation of an ABS splits the original cluster of two neighbouring ABS and provides frequency reuse to enhance capacity, as demonstrated in Figure 6.1. The model assumes line-of-sight propagation on the same street.

It can be investigated from the cropped model in Figure 7.1 that the activation of an ABS affects the traffic level on neighbouring ABSs. The working mode of dynamic a ABS can be managed by the neighbouring fixed ABSs. Analytical results of the frequency reuse scenario demonstrate that up to $n/2$ capacity can be provided to the cluster by switching on an ABS.

Topology management is carried out between neighbouring Fixed ABSs, and also the dynamic ABS if it is activated. This gives consistent coverage under dynamic topologies, for the measurement of user traffic, QoS, etc. For example in the analytical scenario in Figure 7.1, topology management is carried out by $\{b_1, b_2\}$ when b_3 is off, and by $\{b_1, b_2, b_3\}$ when b_3 is on.

In order to measure the interim traffic or QoS, a sliding time window t_{win} is defined. An ABS remains in its working or sleep model for at least t_{win} period, which stabilizes the topology by avoiding switching on/off too frequently. Furthermore, interim QoS is set to zero after every topology transition, which resets the measurement in the new network topology.

There are multiple rules for switching on/off a BS in the literature related to topology management. However, most of the previous research is based on a frequency band allocation strategy, where capacity usage is defined as a threshold for switching ABSs [118]. This is based on a fixed capacity provision scenario. In the dynamic spectrum management scenario, the capacity of a cell also depends on resource utilization in others. The Markov analysis shows that by applying transfer learning, a cell within overlapped clusters has a dominated impact on the capacity of the neighbouring cells. It is thus difficult to use capacity usage as a parameter for topology management in the dynamic spectrum access scenario.

The objective of switching on/off dynamic ABSs is to keep an adequate QoS for various traffic levels. It is thus possible to directly use QoS (from user requirements) as a threshold to determine the time of triggering an ABS, based on the measurement from the neighbouring ABSs. However, the switching off process cannot be achieved in similar way, because QoS is affected by the variation of topology. Traffic density is a parameter that is affected by the user behaviour rather than the network topology. The traffic load level at which the dynamic ABS is switched on can be used as a threshold for switching off the ABS. In summary, the topology management algorithm is described as follows.

Table 7.1. Topology Management Algorithm

Activation
During the past t_{win}, if average retransmission probability on neighbouring Fixed ABSs $P_b > P_{on}$;
1. Fixed ABSs activate dynamic ABS;
2. Dynamic ABS records traffic level G_{on} from fixed ABS;
3. Fixed ABSs shrink cell sizes, connect to nearest MSs;
4. Restart P_b measurement, keep dynamic ABS on for t_{win}.
Deactivation
During the past t_{win}, if total offered traffic on Dynamic ABS and neighbouring Fixed ABSs < G_{on};
1. Fixed ABSs extend cell size to cover dynamic ABSs;
2. Dynamic ABSs deactivated;
3. Restart P_b measurement, keep dynamic ABS off for t_{win}.

The structure of topology management and transfer, reinforcement learning algorithm is demonstrated in Figure 7.9.

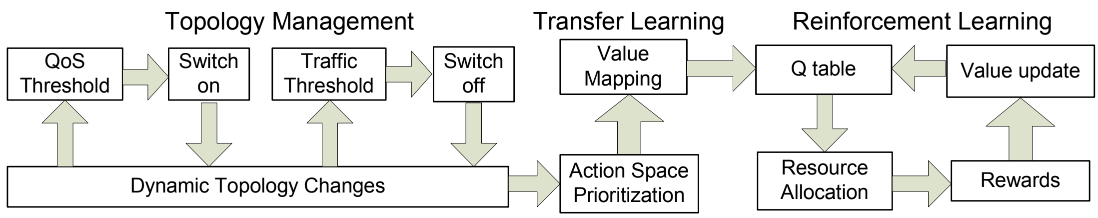

Figure 7.9. Framework of Topology Management with Transfer Learning

It is illustrated that transfer learning provides an interface between topology management and distributed reinforcement learning, which takes network topology information for resource management.

7.5 Simulation

In this section, simulation results are presented for the topology management algorithms together with transfer learning, reinforcement learning and frequency band allocation, respectively. The objective is to validate dynamic capacity provision in transfer learning and its contribution to energy and QoS efficiency. Moreover, the topology management algorithm will be verified in terms of effective energy saving.

The network scenario is based on Figure 6.1 with the parameters listed in Table 6.3. The QoS threshold for triggering a dynamic ABS is set as $P_{on} = 5\%$, following typical QoS requirements in wireless communications [4]. The measurement window is set as $t_{win} = 10s$. In frequency band allocation, the band sizes in each cell are equal. The performance is validated on a long term average basis after the system stabilizes. The 12 fixed ABSs are on initially and the 8 dynamic ABSs are switched on or off according to the topology management algorithm.

The energy efficiency of the network is demonstrated in Figure 7.10, which is evaluated in the format of consumption ratio over the baseline energy level defined by (3.23) in Section 3.5.3. For example, the full deployment of 8 dynamic ABSs over 12 fixed ABSs results in two thirds more energy consumption.

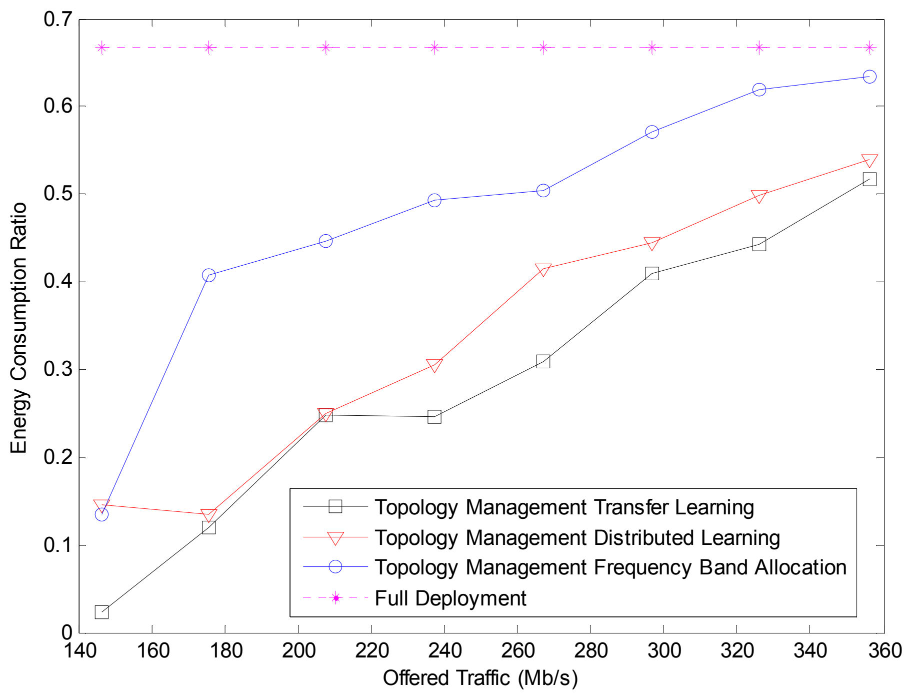

Figure 7.10. Energy Consumption Ratio

It is can be seen that the topology management algorithm effectively reduces energy consumption throughout compared to the full deployment scenario. The energy consumption in frequency band allocation significantly increases when the offered traffic achieves 180 Mb/s, reaching at around 2.3 times higher than the learning strategies. This is because the capacity boundary in each cell largely constrains the flexibility of resource utilization, which triggers significantly more ABSs than other schemes.

The learning schemes achieve similar performance, with continuous lower energy consumption than frequency band allocation. This is because they are designed to select channels from an open spectrum pool. Their long term averaged performance is similar regardless of capacity constraints. However, transfer learning achieves Pareto efficiency much faster than reinforcement learning as illustrated in Chapter 6, which contributes to slightly lower energy consumption. It can be concluded that a significant amount of energy saving is achieved by applying learning technologies to resource and topology management.

A principle of topology management is to maintain QoS at an adequate level. Compared to full deployment, topology management sacrifices a certain amount of QoS for energy saving. However, such reduction is expected to be in control. The probability of retransmission is shown in Figure 7.11.

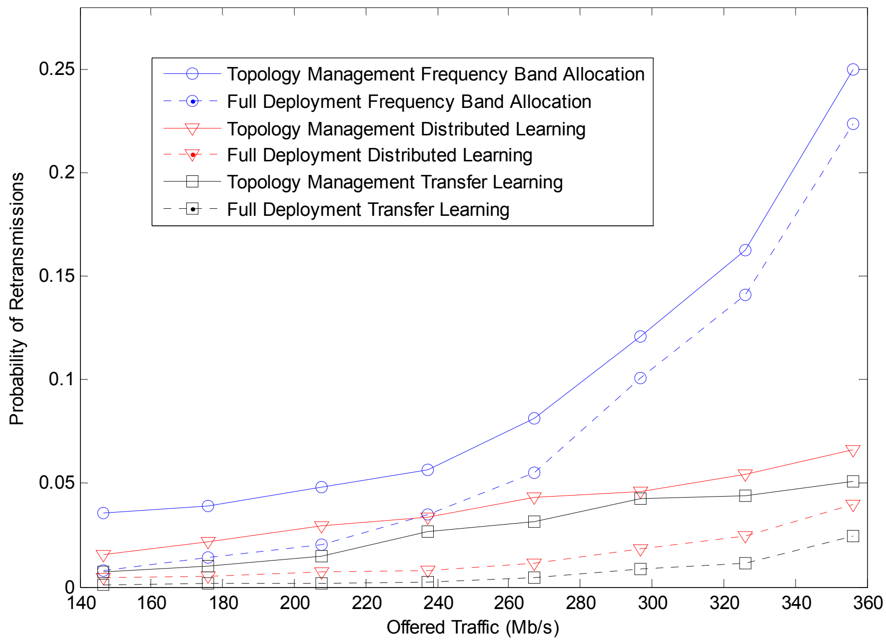

Figure 7.11. Retransmission Probability

Comparing the resource management schemes with and without topology management, it can be seen that a difference of less than 5% retransmission probability reduction is achieved. The QoS threshold $P_{on} = 5\%$ effectively keeps the retransmission probability at an adequate level compared to full deployment. The system with frequency band allocation exhibits a dramatic increase in retransmissions at high traffic levels beyond 240 Mb/s, reaching at 4 times higher than transfer learning. Compared with its energy consumption in Figure 7.10, it can be concluded that fixed capacity provision largely constrains both QoS and energy efficiency in topology management.

The transfer and reinforcement learning strategies achieve similar retransmission performance, because both of them provide flexibility in resource utilization. A slight improvement from transfer learning is shown similar to energy performance, which benefits from fast initial convergence as stated before.

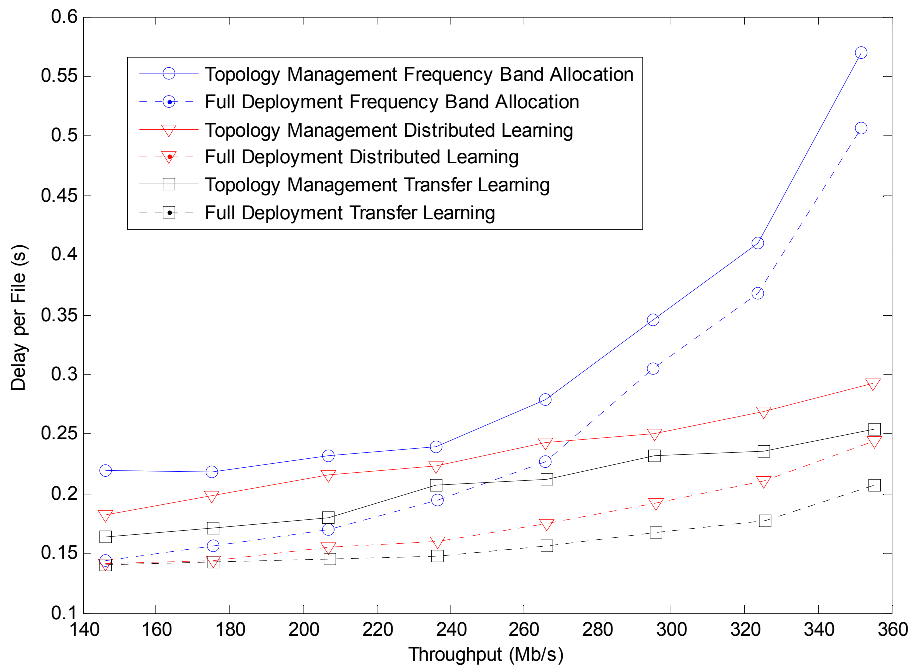

Figure 7.12. Mean Delay per File

The delay performance in Figure 7.12 shows similar behaviour compared with retransmission probability. The delay reduction from applying topology management is controlled within 0.25s compared to full deployment. Frequency band allocation suffers a significant increase of delay when the offered traffic increases beyond 280Mb/s. Learning based resource and topology management strategies achieve almost steady delay levels between 0.2s and 0.25s. It can be concluded that by effectively utilizing spectrum resources and share capacity, the network supports much higher offered traffic with stable QoS.

Figure 7.13 shows the confidence measurement based on the delay performance. The error bars in Figure 7.13 indicates that the file delays during simulation are in reasonable small confidence intervals, compared to the overall performance. The configurations of traffic simulations in this thesis are consistent. As a result, the Monte-Carlo events in the simulation are large enough for performance evaluation.

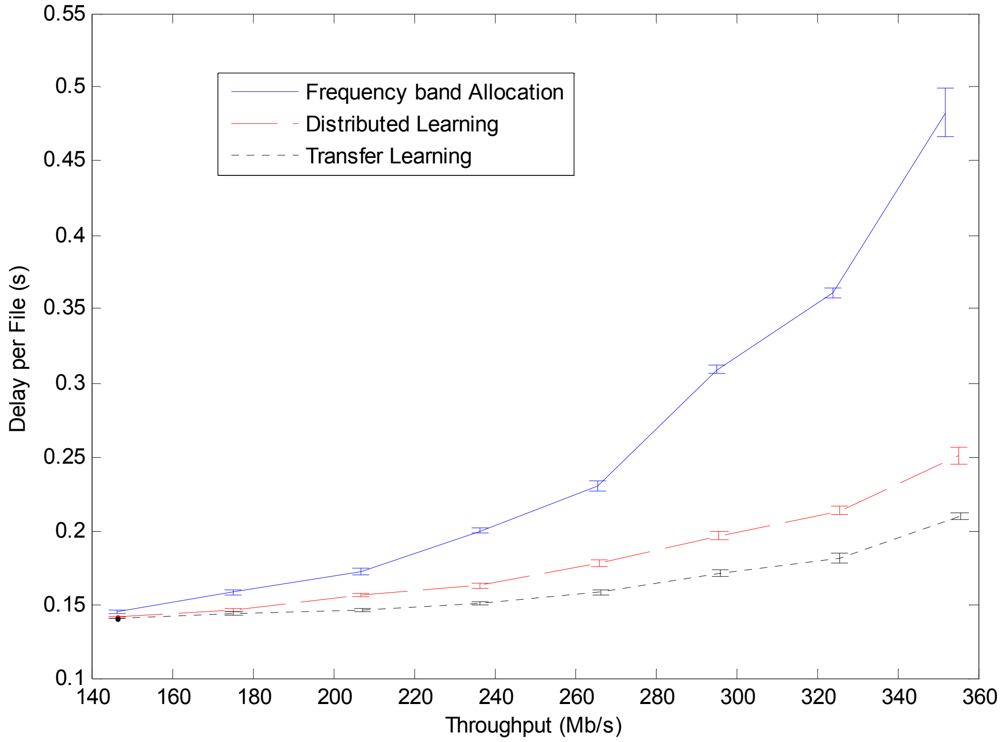

Figure 7.13. Confidence Measurement

7.6 Conclusion

This chapter has investigated dynamic topology management for energy saving on a dense small cell access network. Capacity provision is demonstrated as a principle of operating topology management. Markov analysis is built using a three cell model, to investigate capacity and QoS provided from a conventional Frequency Band Allocation strategy and Pareto efficient Transfer Learning strategy. A QoS and traffic aware topology management scheme is designed in conjunction with several resource management strategies, which is shown to reduce the system energy consumption while keeping an adequate QoS level.

The framework of topology management is based on QoS in a clustered area. Markov analysis shows that by applying transfer learning, a consistent QoS is achieved regardless of traffic variations in neighbouring cells. Moreover, appropriate frequency reuse is vital to provide capacity enhancement from topology management.

The simulation results demonstrate that the QoS parameter used for switching on base stations effectively manages the QoS reduction from topology management to an acceptable level. A significant amount of energy saving is achieved compared to the full deployment. Learning based topology management is shown to improve the QoS and energy efficiency by dynamic capacity provisions between adjacent cells.

Chapter 8. Future Work

This chapter proposes some future work directions based on the work in this thesis.

Dynamic Spectrum Access (DSA) plays an important role for ultra-high capacity network in the future 5G communication systems. The transfer learning algorithm proposed in this thesis is demonstrated as an effective approach to implement docition [77] in spectrum management scenario, which further enhances cognitive radio. In this context, transfer learning can also be applied to other aspects in communication systems that use distributed learning, (i.e. power management, topology management), to improve the system reliability, QoS/capacity, energy consumption, etc. Furthermore, the algorithm can be optimized in terms of flexibility, applicability, convergence efficiency, etc., to enable more general applications.

Dynamic Topology Management can be applied not only to green communications but also for flexible network architectures in many scenarios, such as disaster relief and temporary events. Future communication systems tend to be hyper-dense networks with massive small cell base stations serving different types of traffic. Topology management can be used to effectively manage such complex architecture, in order to provide system capacity and reduce power consumption.

8.1 Implementation of Machine Learning for RRM

This thesis is based on a theoretical research in the Radio Resource Management aspect of dense capacity wireless networks, with performance validation on system level simulations. Machine learning techniques including distributed reinforcement learning and transfer learning has been applied to operate RRM on a self-organized multi-hop backhaul and small cell access network. These techniques provide effective network deployment and management for the operators.

The machine learning techniques introduced in this thesis provide the network with distributed decision making, which is especially important for the lack of planning scenarios. By using transfer learning to improve the convergence, the base stations can be effectively configured in a relatively short time after the deployment. By using reinforcement learning to improve the decisions with past experience, the network can deliver reliable QoS in the rapidly changing radio environment. Cooperation management significantly reduces level of control information overhead in current cellular systems. Furthermore, the integration of reinforcement learning and transfer learning can effectively identify the change of scenario, architecture, topology, etc., and configure the RRM parameters. Thus the reliability of the RRM function can be significantly improved.

Transfer learning is a partially distributed or centralized technique, which introduces the philosophy of conventional frequency planning and interference coordination into the distributed learning algorithm. The cooperation management mechanism as introduced in Chapter 5 allows the operators effectively control the degree of transfer learning. Similar scheme is also proposed in topology management that, transfer learning is carried out when the operators decide to switch on/off the base stations. It should be noted that although the intelligent RRM algorithm proposed in this thesis deliver an effective self-organized solution, the network managers can still monitor the system parameters, such as QoS, spectrum usage, throughput, as demonstrated widely in this thesis. The operators can then control the degree of learning used in the system based on their performance requirement. A typical example is the cooperation management mechanism developed in this thesis which is based on monitoring the stability of learning. Other network control mechanisms can be developed in similar ways.

There are several steps that can be done to implement this academic work to the practical communication systems.

The major functionality of RRM in wireless systems is to schedule the data packets on Resource Blocks with high QoS provision. In this context, the implementation of intelligent RRM algorithms can be based on the design of protocols. 3GPP specifies a Radio Resource Control layer in the control plane protocol stack of the LTE system, which is in control of radio resource usage [13]. It manages UE's signalling and data connections. The distributed learning algorithm can be implemented in this layer's protocol stack to improve the channel assignment process. Furthermore, the X2 Application Protocol (X2AP) is responsible for overall maintenance of the relation between neighbouring eNBs. The signalling messages transfer on X2AP can be modified to include the learning information discussed in this thesis.

The implementation of Dynamic Spectrum Access in the practical communication system has been a hot topic for a long period. The major challenge is the current spectrum allocation policies in most countries. However, DSA can still be operated in some emerging areas that require non-commercialized spectrum allocation, which include but not limited to, high speed WI-FI service through TV White Space, public safety networks, smart grid systems, machine-to-machine communications. In the future 5G systems, DSA is expected to play a key role in solving the spectrum issue for ultra-high data rate services. This has been investigated in a number of EU projects. In this context, the machine learning techniques enable effective interference avoidance in a distributed manner, which assist the operator to reduce the complexity of network architecture and management.

8.2 Intelligent RRM for LTE Systems

Radio Resource Management in LTE is carried out on the eNBs, with signalling information exchange on the control plane over X2 interface. It allows each individual eNB to use the entire frequency band. The Inter-Cell Interference Coordination (ICIC) protocol is designed to reduce interference between neighbouring cells and improve QoS on cell edge users. In ICIC, interference indicators are sent from a eNB that schedules UEs on Resource Blocks (RBs). The neighbouring eNBs that receive such indicators will avoid the occupied RBs when scheduling their UEs [44]. The ICIC is designed in a static or semi-static manner. Static ICIC is based on a frequency planning, which large reduces signalling overhead on X2 interface. Semi-static ICIC carries out information exchange periodically, which is beneficial in dynamic traffic load.

Transfer learning is designed as exchanging learning information between base stations that potentially cause interference to speed up the distributed learning process, which implements the idea of docition into RRM [92]. It has the benefits from both conventional frequency planning and distributed learning, as discussed in Chapter 6. As a result, transfer learning can be employed to effectively improve QoS provided by ICIC. It significantly reduces information exchange overhead on the control interface [92], and provides flexible utilization of the frequency band with effective interference management in a distributed manner [119]. This improves both static- and semi-static ICIC schemes in the LTE standard [44].

Fractional Frequency Reuse (FFR) is an effective spectrum management scheme proposed for the LTE architecture. However, it is designed as a Frequency band Allocation strategy where each fractional zone is assigned a channel set to prevent interference, which is illustrated in Figure 2.1. In this context, FFR is inefficient in managing the dynamics of traffic load, because the capacity is constrained by the frequency size allocated in each fractional zone.

The transfer learning model designed in Chapter 6 is ideally suited for this problem. The learning agent can be implemented in each fractional zone in Figure 2.1 rather than in the whole cell, which separates the MSs into different Q tables based on the local interference environment. In the second step, the action prioritization process is carried out between neighbouring zones, to initially provide an effective policy on each agent. Finally the value mapping scheme is conducted to associate the Q values learnt in the past with the newly prioritized action space. After transfer learning, the BSs start reinforcement learning from Q tables with transferred knowledge. The value training method can also be applied to further reinforce the Q table. Transfer learning enhances the system capacity by providing flexible utilization of radio spectrum in different zones, as well as mitigating inter zone interference by applying FFR information into the knowledge base. Furthermore, the information exchanged between cells on the X2 interface can be minimized. It is expected to be a highly efficient resource management scheme for LTE which can be easily implemented on standardized protocol architectures.

8.3 Intelligent RRM for Ad hoc Networks

The machine learning techniques proposed in this thesis can also be applied for Radio Resource Management in a fixed or mobile ad-hoc network.

In Chapter 4 and Chapter 5, reinforcement learning and transfer learning algorithms have been developed and investigated on the multi-hop backhaul network using tree architecture, as illustrated in Figure 3.2. This can be easily extended to a mesh architecture where the ABSs on different “branches” are allowed to directly connect with each other. In this context, an ABS has multiple transmitters and receivers, as implemented on the HBS. The spatial reuse scheme developed for multiple branches should be extended to the ABS, which considers the antenna directionality and beamwidth when exchanging channel usage information or learning Q tables. The transfer learning and cooperation management algorithms proposed in Chapter 5 are expected to improve QoS and reduce signalling overhead in such mesh network.

The mobile ad hoc network introduced in Section 2.2.4 has more challenges in RRM. Firstly the neighbouring links cannot reuse the same time-frequency Resource Blocks (RBs) because of the omni interference range. Moreover, interference range on a moving UE is highly unpredictable, which may cause excessive interference to a number of other UEs. In this context, the learning engine could be applied on each individual UE. Transfer learning allows a UE to obtain learnt information from others in vicinity, and train its own knowledgebase to avoid interference. The philosophy of Q table transfer and cooperation management can be applied to the mobile ad hoc network, though the “source agent selection” and “target agent training” strategies proposed in Section 5.3 and 5.4 should be adjusted to the interference environment.

8.4 Intelligent Topology Management

The topology management algorithm designed in Chapter 7 uses the expected blocking probability as a threshold to activate the BSs. It is difficult to use their traffic load for such operation in a fully dynamic spectrum access scenario, because the system capacity is mainly constrained by interference rather than spectrum size. However, Pareto improvement action space prioritization in transfer learning effectively eliminates interference between cells involved in knowledge transfer. As a result, Markov analysis has be done in Section 7.3 based on a group of cells. The Markov model can be extended to include the topology management operation, in order to improve the threshold for activating/deactivating BSs.

Furthermore, a function that includes learning from the past TM experience can be designed to improve the threshold for switching on/off and the location of BSs, as well as the connection between BSs and MSs. The TM system can be modelled as a state-action-reward cycle, which is applicable for most of the classical reinforcement learning algorithms such as Q learning and SARSA. In this context, the states can be modelled as a set of traffic levels, such as low, medium and high; the actions can be modelled as the selections of BSs in an overlapped coverage area; the rewards can be modelled as loading, QoS and SINR on each BS. The learning algorithm is aimed at clustering MSs onto a minimum number of cluster heads (BSs) in the low traffic level, in order to switch off other BSs. On the other hand, learning is aimed at balancing traffic load on BSs at the high traffic level.

The disadvantage of applying reinforcement learning to topology management is that a large number of actions are desired to achieve an effective network topology. It has been demonstrated in [120] that switching on/off a BS consumes certain amount of time and energy. Transfer learning is a promising technique to improve the convergence. Moreover, the changes in the deployment map can be carried out through handing over traffic between overlapped cells rather than literately switching on/off BSs. Handover for topology management is expected to effectively reduce the fluctuation of network topology and improves network reliability.

8.5 Dynamic Link Selection

The role of topology management in a wireless cellular network is to carry out dynamic network planning, but also to select the links between BSs and MSs. Dynamic link selection is proposed here as another major part of topology management. It can be used to assist dynamic network planning, such as the learning model for TM in previous section. In addition, QoS and capacity optimization under dynamic topologies is a major target of link selection, which can be achieved from the following operations.

8.5.1 *Load Balancing and Load Unbalancing*

Load balancing and load unbalancing are two load management methods that handle traffic between neighbouring or overlapped cells. Load balancing is a widely investigated technique in the research papers. The conventional definition is to equalize the traffic load in each cell. However, this is not effective in a flexible self-organized network where the amount of radio resources may be unequal between cells. In such scenarios, the ideal traffic load in each cell is to match the number of available resource blocks, where load balancing will be used. This methodology could maximize the capacity provision from activated BSs, and improves QoS on access links. Load balancing will also reduce the number of BSs required in the high traffic level, by improving resource utilization under existing topology.

Load unbalancing on the other hand is designed to speed up the deactivation process. It is carried out in cells with low traffic levels, to transfer local MSs to neighbouring or overlaid cells. This technique is designed to clear the MSs connected to the BS that should be switched off. The objective of load unbalancing is to save energy by quickly switching off BSs, as well as maintaining sufficient QoS.

8.5.2 *Handover and Admission Control*

Link selection can be achieved through handover and admission control in a practical cellular network. These two techniques have been conventionally applied in a static network for user mobility and congestion control, respectively. In the context of topology management, they are used to transfer user traffic. Handover is a reactive technique that transfers ongoing connections from one BS to another, whereas admission control is a proactive technique that selects the BS before connection.

Handover and admission control are implementations of load balancing and load control, with the same target of transferring traffic load between BSs. However, these two operations have a different impact on user experience and system performance. At the user level, handover may delay or even interrupt on-going connections, whereas admission control could delay connection setup or even block the connection. In this case, admission control has less impact on the user experience than handover because the users are more tolerant to a blocked connection rather than an interruption.

At the system level, handover can quickly transfer the traffic load during data transmission. On the other hand, admission control can be carried out only after the transmission has finished. The protracted feature of admission control is thus not as effective as handover in load management. However, it is easier to be implemented because there is no need of extra protocols to protect on-going connections.

Figure 8.1 shows a Markov model for handover and admission control between two BSs. Admission control is operated by a_1 and a_2, which indicates the proportion of traffic allowed in each BSs. Handover is operated by h_{12} and h_{21}, which is the probability of a user being transferred between the two BSs.

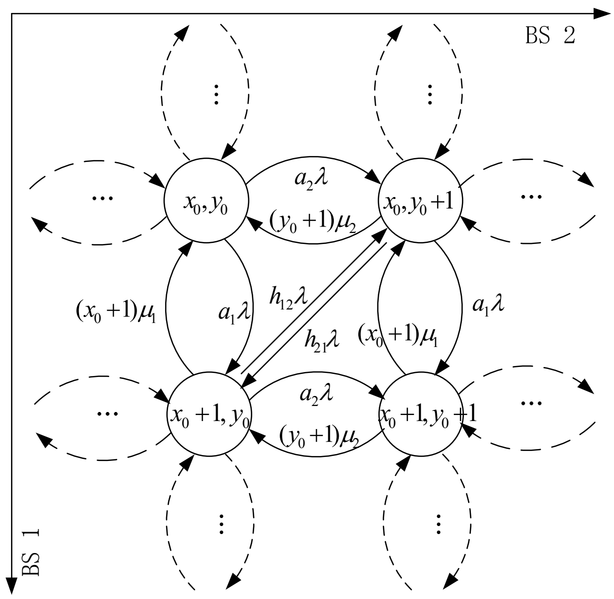

Figure 8.1. Markov model for Link Selection

8.5.3 ***Mobility of Aerial Base Stations***

Aerial platforms have been widely explored for the use of wireless base stations providing mobile broadband access [121]. The FP7 ABSOLUTE project will study the Low Altitude Platform (LAP) for coverage and capacity in disaster relief scenarios [33]. Google carries out Project Loon to provide broadband internet services through a balloon based High Altitude Platform (HAP) [122]. The challenge of aerial BSs compared to conventional BSs is that the location of the aerial platform is not steady, which largely depends on unpredictable air conditions. In this scenario, dynamic link selection is vitally important to guarantee reliable communication and stable QoS.

The mobility and navigation of aerial BSs can be controlled in broadband access scenarios. For example in Project Loon, the balloons travel in the stratosphere where the wind varies in direction and magnitude. A number of balloons form a complete network by moving themselves with the wind in different directions. In this scenario, a stationary user will also be continuously handed over from one balloon to another. The link selection algorithm should guarantee a steady link quality.

In the scenario for opportunistic communications, aerial platforms are supplements for coverage and capacity where conventional BSs are not available. The location of aerial BSs will be managed by dynamic network planning algorithms. Overlapped coverage is expected in this scenario. The link selection algorithm should also consider the traffic load and capacity on each BS to avoid congestion.

8.6 Entropy in Transfer Learning

In this thesis, transfer learning has been studied for resource management in static backhaul and dynamic access networks. Two types of knowledge transfer methods have been developed: value training and value mapping. The value training method continuously updates the learner's Q table until reaching a stable state. The value mapping method associates the Q table with prioritized action space when topology changes are carried out. The major difference between them is that knowledge from source agents dominates the Q table in the value training method, but only initializes the Q table in the value mapping method. This indicates that a learning agent is able to get more reliable knowledge from source agents in a static network topology but less in a dynamic network topology.

Entropy evaluation on knowledge transferred from source agents remains a crucial issue in transfer learning. It can be used to decide which algorithm to use and effectively control the amount of transferred knowledge applied on the target agent. For example, a discount function can be set on $Q_{(SA)}$ in the value training function (5.4), which varies with the entropy of $Q_{(SA)}$ with respect to $Q_{(TA)}$. Furthermore, value training function can also be applied after mapping Q values to the prioritized action space, in order to provide more expert knowledge over late iterations. However, these dynamic knowledge transfer methods should be based on entropy evaluation, which needs substantial research in future.

Source agent selection is the baseline component in transfer learning that determines the entropy of information exchanged and the efficiency of multi-agent cooperation. It should be based on the scenario where transfer learning is applied. The transfer learning methods developed in Chapter 5 for the backhaul network and in Chapter 6 for the access network transfer the neighbor agents' Q table equally to the target agent. However, the entropy of these agents could be different. As a result, a discount factor can be applied on each agent to control the transferred knowledge based on interference. Moreover, Q tables on other agents in vicinity can also be transferred with different strategies applied on the target agent. In this context, an intelligent algorithm is desired to learn the effective discount factors on different agents and various learning stages.

Chapter 9. Summary and Conclusions

9.1 Conclusions of Work

This thesis has studied the use of intelligent learning algorithms for radio resource management in dense capacity wireless networks. Transfer learning has been investigated to improve reinforcement learning by applying knowledge transferred from multi-agent cooperation. In multi-hop backhaul networks with a static topology, transfer learning has been demonstrated to largely improve QoS and reduce cooperation overhead. In small cell access networks with dynamic topologies, transfer learning has been shown to significantly reduce QoS fluctuations during environmental changes. In addition, dynamic topology management with transfer learning has been examined to effectively reduce energy consumption from base stations and enhance network capacity.

The conclusions for the major chapters are listed as follows.

Chapter 1 provided a general introduction to whole work. In Chapter 2, background information related to the area of mobile broadband networks, radio resource management, cognitive radio and machine learning, and energy efficiency of wireless network have been presented. The high capacity density broadband access can be achieved through a small cell access network. Meanwhile, wireless backhaul networks support flexible deployment of small cell base stations. Spectrum management strategies can be categorized into Frequency band Allocation (FA) and Dynamic Spectrum Access (DSA). FA provides effective interference avoidance but has bandwidth constraints regarding traffic dynamics. DSA supplies up to optimal spectrum utilization but has challenges in interference management. Cognitive radio technology is designed to intelligently select radio resources for data transmission in a distributed manner. The learning engine is the core module of cognitive radio that applies historical experience for future decisions. Reinforcement learning makes a distributed agent converge to a part of the spectrum, thereby providing lower interference to others. Multi-agent cooperation can improve decisions made by reinforcement learning and speed up its convergence. Transfer learning has the potential to improve QoS reliability in a dynamic radio environment. Furthermore, dynamic network planning is an effective way to reduce energy consumption in a cellular network, by deactivating base stations that make little capacity contribution.

The modelling, simulation and analysis methodologies used in this thesis have been presented in Chapter 3. The network is modelled with appropriate topology, antenna, propagation and traffic models. Matlab is selected to carry out Monte Carlo simulation in this work. The complete simulator is built upon the architecture, physical layer, traffic, spectrum and topology management modules. Results are evaluated in a long term averaged manner to obtain steady state performance, and on a temporal snapshot basis to trace intermediate performance. Furthermore, Markov modelling has been discussed as an effective tool to analyse system capacity and QoS in different spectrum and topology management algorithms.

Chapter 4 presented a fully cooperative interference coordination and a fully distributed reinforcement learning strategy for resource management in the multi-hop backhaul network. The novel space division channel assignment scheme has been developed, which provides effective channel reuse on multiple transmitter or receiver antennas on the same base station. This scheme has been demonstrated to significantly reduce relay burden and interference in the multi-hop backhaul network, which thus largely improve the throughput and QoS.

Distributed reinforcement learning strategies have been demonstrated on the mutli-hop backhaul network. Convergence analysis showed that linear reinforcement learning performs better than single state Q learning to achieve a stable solution. An improved decision making scheme that selects channels by the weight of interference is developed. The results showed that the interference weighted strategy improves QoS by effectively using information from spectrum sensing. Linear reinforcement learning has been demonstrated to deliver efficient long term performance without interference measurement.

In Chapter 5, a novel intelligent cognitive radio technology – transfer learning has been proposed, which is based on a combination of reinforcement learning and interference coordination schemes studied in the previous chapter. In the multi-hop backhaul network with a static topology, a learning task is modelled as the learning target on a cognitive agent. Transfer learning allows an agent to obtain Q tables learnt by others, with a value training function designed to reinforce the learner's knowledge base. It has been demonstrated that knowledge transfer significantly improves QoS and throughput compared to distributed reinforcement learning. Furthermore, a novel stable state evaluation method has been designed to appropriately define the convergence of learning in dynamic radio environment. Cooperation management strategies have been developed to control transfer learning when stable states have been achieved by either value training function in transfer learning or action-value function in reinforcement learning. It has been demonstrated that cooperation management significantly reduces the amount of information exchanged between multiple agents, meanwhile delivering a high level of QoS as achieved in fully coordinated strategies.

Transfer learning was also examined in a small cell access network with dynamic topologies in Chapter 6. In this context, a learning task has been modelled as the learning target in a network topology. Transfer learning is carried out during topology transitions, to utilize knowledge bases learnt in the past for new scenarios. Knowledge transfer is conducted by prioritizing action spaces among coordinated agents. The training process introduces a value mapping strategy, by associating Q values with a prioritized action space. Pareto efficient resource utilization can be achieved among coordinated cells, which effectively eliminates interference before the action space is fully occupied. The transfer learning algorithm can be carried out with various reinforcement learning algorithms, to improve learning reliability in dynamic network topologies. It has been demonstrated to significantly mitigate the QoS fluctuation incurred by the changes of network topology and user traffic.

Chapter 7 validated the dynamic capacity provision feature in transfer learning through Markov analytical models, and presented the design of the corresponding topology management algorithm to reduce energy consumption and maintain sufficient QoS. Multi-dimensional Markov chains have been used to analyse the achievable system capacity with different QoS levels. Transfer learning that delivers Pareto efficient resource utilization has been proven to significantly enhance the system capacity compared to the frequency band allocation strategy. Moreover, topology management has been designed to dynamically control the number of base stations in the network based on QoS and traffic level. It has been demonstrated that topology management achieves significant energy saving compared to the full deployment scenario. QoS reduction is also well controlled by the predefined thresholds. Furthermore, transfer learning achieves significantly higher energy and QoS efficiency through dynamic capacity provision.

9.2 Summary of Original Contributions

This thesis has provided an in-depth study of knowledge transfer in wireless networks with both static and dynamic topologies, in order to enhance system capacity, QoS and reduce cooperation overheads, energy consumption. There is relatively limited transfer learning research applied on wireless communication systems before this work. The closest work is docition [73] that applies multi-agent cooperation on reinforcement learning. However, docition is designed to allow a distributed agent to obtain expert knowledge from others, which is only part of transfer learning. This section highlights the original contributions provided in this thesis. Some of the work has been published at, submitted to, or in preparation for a number of conferences and journals, which are listed at the end of this thesis.

Transfer Learning on a Muti-agents basis

This thesis applies transfer learning for the first time to communication systems. In a multi-hop backhaul network with static topology, it is proposed to improve learning speed and draw better decisions. The value training method has been designed to transfer and train the knowledge base on multiple agents.

In the context of transfer learning defined in the computer science domain [79, 80], learning tasks have been modelled as the learning targets on multiple agents. Transfer learning allows a target agent to obtain Q tables from adjacent source agents that may incur excessive interference. The value training function is then carried out to reinforce the target agent’s Q table. A Q value on each action is then gradually “trained” over successive iterations. Effective decisions can be made after the Q table is trained to be mature.

The value training method on multi-hop backhaul network has been published in [91]. It has been demonstrated to significantly improve QoS and achieve better convergence. Compared to the docition strategy investigated in [77], transfer learning benefits from information exchange regardless of the maturity of Q tables. The value training function introduces spatial spectrum reuse information to the learners’ knowledge base.

Cooperation Management

Cooperation management is a novel methodology that effectively controls the level of information exchanged between multiple agents in transfer learning. It is designed to minimize cooperation overhead and to maximize QoS and throughput. This strategy terminates transfer learning when stable states are achieved by different learning functions. Cooperation management on an action-value function has been demonstrated to effectively trade off QoS/throughput and cooperation overhead. It delivers a high level of performance as achieved in a fully coordinated network, by using a very small amount of information exchange. Transfer learning with cooperation management is proposed in [92] as an effective approach to improve QoS in distributed networks.

Stable State Evaluation

Stable state and its probability provide an important method to evaluate the convergence of learning algorithms in cognitive radio scenarios. It has been difficult to define convergence in a cognitive radio network, because the effective solution varies with a highly dynamic radio environment. However, the main target of learning in cognitive radio is to find a stable action space that provides effective decisions, which can be used to define a stable state for convergence evaluation.

The stable state is defined in Chapter 5 and [92] that the ranking of actions (channels) in the Q table stabilizes over iterations, which in turn makes the decision policy stable. Moreover, the Q values are initially generated with arbitrary numbers, thus the ranking evaluation is carried out only on the actions that have been taken and updated by the action-value function.

In a dynamic radio environment, a cognitive agent may occasionally drop out from a stable action space. As a result, the probability of stable state is used in practice to illustrate the convergence behaviour. In this context, a cognitive agent converges when a consistently high stable state probability is achieved.

Stable state evaluation is used not only to investigate convergence performance but also to control the level of information exchanged in transfer learning.

Transfer Learning on a Multi-tasks basis

Transfer learning has been proposed for improving network reliability in rapidly changeable network architectures in [119] and [34]. The dynamics of traffic and topology incur highly fluctuating system performance in conventional learning algorithms. Transfer learning models the learning tasks as effective policy in different network topologies. The knowledge base from previous source tasks is transferred to the new target task. The base stations are shown to quickly adapt to the new radio environment and deliver steady QoS.

The Pareto efficient action space prioritization is used to support mufti-agent coordination. A value mapping strategy is proposed to associate the Q values learnt in previous tasks to the newly prioritized action space. In this manner, experience from past decisions can be retained in the Q table and the new environment information can be applied appropriately. Furthermore, it provides an interface between topology and spectrum management, which allows cognitive agent to learn after the establishment of network topology.

Pareto Efficient Action Space Prioritization

The Pareto efficiency is proposed for resource management, which allows a shared spectrum pool effectively utilized between a cluster of cells without interference. An action space prioritization algorithm has been developed, which provides a Pareto improvement resource allocation strategy.

The Pareto efficient action space prioritization algorithm effectively handles traffic dynamics in both time and spatial domains. An enhanced system capacity can be achieved in a cluster of cells by eliminating inter-cell interference and maximizing resource utilization. This algorithm also supports the design of distributed learning models in wireless cellular networks.

Topology Management with Intelligent Resource Management

Topology management is a novel dynamic network planning methodology in wireless cellular system. It has been studied mainly to reduce energy consumption from base stations, though it can also be used to improve network planning. Previous work in this context is mainly based on fixed capacity provision [118]. In this thesis, a novel strategy is proposed with intelligent resource management.

Dynamic capacity provision from transfer learning effectively utilizes radio resources in a group of cells. Topology management thus evaluate the QoS level on a group of base stations. A new base station is switched on to enhance local capacity when approaching an adequate QoS threshold, and is switched off after traffic load reduces. This method is demonstrated to effectively control the QoS reduction in an adequate range by reducing energy consumption from base stations. Topology management with transfer learning based RRM has been proposed in [34].

Markov Analysis for Multi-Cells with Dynamic Spectrum Sharing

A novel multi-dimensional Markov model has been proposed in Section 7.3 to validate dynamic capacity provision between multiple cells. This model is an extension to the two dimensional Markov model presented in [116], which effectively models dynamic spectrum sharing.

Different models for fixed and dynamic resource management strategy have been developed, which validate capacity enhancement from base station deployments. The Pareto efficient state in transfer learning has been proven to achieve dynamic capacity provision, which provides effective QoS on unbalanced traffic load between cells. The analytical model and related results have been presented in [34].

Spatial Reuse on Multi-hop Backhaul Network

A space-division multiple access scheme has been developed for a multi-hop backhaul network. It allows the transmitter or receiver antennas on the same node to reuse radio resources. Moreover, this scheme is carried out via inter link interference coordination that allows fully dynamic access of the radio spectrum. Interference on multi-hop links can be effectively controlled. Furthermore, it significantly reduces relay burden on multi-hop network. The amount of radio resources required is reduced to the same level as the single-hop architecture.

Convergence Analysis of Reinforcement Learning through Reward States

The convergence behaviour of linear reinforcement learning [64] and single state Q learning [70] has been analysed in this thesis. These two algorithms have been applied to spectrum management in previous work but the decision and value updating behaviour remain unclear. This analysis employs reward states, defined as continuous actions with the same reward value, to investigate the Q value changes. By analysing the converged value of the two algorithms in each reward state, linear reinforcement learning is shown to achieve better convergence whilst Q learning is more sensitive to the environment changes. The reward state analysis provides a method to design the learning model in different scenarios, which has been proposed together with transfer learning in [92].

9.3 List of Publications

Conferernce Proceedings

Q. Zhao and D. Grace, "Application of Cognition based Resource Allocation Strategies on a Multi-hop Backhaul Network," *IEEE International Conference on Communication Systems*, Singapore, November 2012.

Q. Zhao and D. Grace, "Agent Transfer Learning for Cognitive Resource Management on Multi-hop Backhaul Networks," *Future Network & Mobile Summit*, Lisbon, July 2013.

Q. Zhao, T. Jiang, N. Morozs, D. Grace and T. Clarke, “Transfer Learning: a Paradigm for Dynamic Spectrum and Topology Management in Flexible

Architectures”, *IEEE 78th Vehicular Technology Conference (VTC2013-Fall)*, Las Vegas, September 2013.

Journal Articles

Q. Zhao, D. Grace, and T. Clarke, "Transfer Learning with Cooperation Management: Balancing the Quality of Service and Information Exchange Overhead in Cognitive Radio Networks," submitted to *IEEE Journal on Selected Area in Communications (JSAC)*.

Q. Zhao, D. Grace, and T. Clarke, "Intelligent Radio Resource and Topology Management with Transfer Learning for Rapidly Changeable Cellular Networks," in preparation for *IEEE Transactions on Wireless Communications* (draft completed).

Project Deliverable

Q. Zhao, D. Grace, S. Rehan, et al., FP7-ICT-ABSOLUTE/D4.1.1, Detailed Network and Protocol Architecture – First Issue, www.absolute-project.eu, May 2013

9.4 Recommendations for Similar Research Scope

This thesis investigates the topic of applying machine learning techniques in Radio Resource Management function of the communication systems, which is a cross discipline research on both wireless communications and artificial intelligence.

The design of communication network is the basis of this research. This includes the use case, scenario and the network architecture. A good understanding of the use case, such as the type of service, can help to define the scenarios that provide information of the propagation environment, terrain, user mobility, etc. The network architecture, including the topology, transceivers, antennas, can then be designed to satisfy the scenario.

The knowledge on protocol architecture is also an important aspect when starting this research work. It is essential to understand the data packet transmission in the wireless network. The conventional MAC protocols in various systems should be well studied, as reviewed in Chapter 2. The physical layer knowledge is also vital, especially the antenna and propagation scenarios, modulation and coding techniques. Other knowledge in network and transport layers could benefit the research, such as routing protocols for topology management, congestion control algorithm for load management. Last but not least, the protocols in existing standards can effectively enhance the research work for practical implementation purposes.

Machine learning algorithm is another major research area. Firstly it is important to analyse the issues and expected targets in the system, in order to select appropriate learning algorithms. The latency requirement should be considered for improving the convergence of learning. Furthermore, the level of centralized/distributed in the learning algorithm should comply with the protocol architecture.

System level simulation is the essential method to validate the proposed mechanisms. This is based on a well understanding of communication systems described before, and an appropriate design of the learning assisted scheme. Comparison with existing research work is vital to prove the ideas. Theoretical analysis is an effective way to further enhance the developed schemes, although it is usually based on simplified models as demonstrated in Chapter 7. A comprehensive analytical model can significantly improve the work in academic aspects.

Furthermore, the hardware implementation, which has not been investigated in this thesis, can largely help to apply the theoretical research work to practical systems. The RRM functions can be implemented on the demonstration platforms through protocol configurations, as discussed in Chapter 8.

Definitions

Cognitive Agent

a wireless entity which observes the radio environment, makes decisions on radio parameters, takes actions on data transmission, learns from current and previous experiences and trains a knowledge base for future decisions. It refers to a base station in this thesis.

Action Space

a set of actions for a cognitive agent to select and take. It refers to a channel set in this thesis.

Q Value

a value in the knowledge base which stands for the learning knowledge of an action.

Action-Value Function

a reinforcement learning function that updates the Q value based on the environment feedback from a particular action.

Probability of Stable States

the probability that a Q table has consistent action ranking in previous iterations.

Target Agent/Task

an agent that carries out distributed learning to solve an individual learning task.

Source Agent/Task

a cognitive agent/learning task that has potential impact on the target agent/task.

Transfer Learning: Value Training Method

a transfer learning strategy that employs multi-agent cooperation to train the agent's knowledgebase.

Value Training Function

a function that trains the target agent's Q values with those transferred from multiple source agents.

Cooperation Management

a strategy that controls the degree of information exchange and knowledge transfer.

Pareto Improvement

a resource allocation strategy that allows any individual in a group to occupy resources without causing interference to others.

Pareto Efficient

an upper bound of Pareto improvement that the entire resource pool can be occupied by a group of individuals without interference, regardless of the resource occupancy status in each.

Transfer Learning: Value Mapping Method

a transfer learning strategy that maps Q values learnt in a source task to a prioritized action space in a target task.

Action Space Prioritization

a sorting algorithm carried out on action spaces to achieve a Pareto improvement in a cluster of agents.

Action-Value Mapping

a function that associate Q values with a prioritized action space.

Glossary

(A)BS	(Access) Base Station
AMC	Adaptive Modulation and Coding
CDF	Cumulative Distribution Function
CM	Cooperation Management
CN	Core Network
CR	Cognitive Radio
DAI	Distributed Artificial Intelligence
DSA	Dynamic Spectrum Access
eNB	Evolved Node B
FA	Frequency band Allocation
FFR	Fractional Frequency Reuse
FP	Frequency Planning
HAP	High Altitude Platform
(H)BS	(Hub) Base Station
HetNet	Heterogeneous Network
(IC)IC	(Inter Cell) Interference Coordination
LTE	Long Term Evolution
MANET	Mobile Ad hoc Network
MDP	Markov Decision Process
MS	Mobile Station
OFDMA	Orthogonal Frequency-Division Multiple Access
QL	Q Learning
QoS	Quality of Service
RL	Reinforcement Learning

REM	Radio Environment Map
RRM	Radio Resource Management
RNC	Radio Network Controller
SCN	Small Cell Network
SDR	Software-defined Radio
SINR	Signal-to-Interference plus Noise Ratio
TL	Transfer Learning
TM	Topology Management
UE	User Equipment

List of References

[1] H. G. Myung, Towards 4G - Technical Overview of LTE and WiMAX, Tutorial in WCNC WEBCAST SESSIONS, 2010.

[2] Cisco, Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2012–2017, www.cisco.com.

[3] S. Dehghan and R. Steele, "Small cell city," *Communications Magazine, IEEE,* vol. 35, pp. 52-59, 1997.

[4] T. S. Rappaport, *Wireless Communications: Principles and Practice*, December 2001.

[5] BuNGee Deliverable: D1.2, Baseline BuNGee Architecture, January 2010.

[6] A. G. Spilling, A. R. Nix, M. A. Beach, and T. J. Harrold, "Self-organisation in future mobile communications," *Electronics & Communication Engineering Journal,* vol. 12, pp. 133-147, 2000.

[7] H. Holma and A. Toskala, *WCDMA for UMTS: Radio Access for Third Generation Mobile Communications*, Third ed.: Wiley.

[8] M. Song, C. Xin, Y. Zhao, and X. Cheng, "Dynamic spectrum access: from cognitive radio to network radio," *Wireless Communications, IEEE,* vol. 19, pp. 23-29.

[9] D. Grace and H. Zhang, *Cognitive Communications: Distributed Artificial Intelligence (DAI), Regulatory Policy and Economics, Implementation*: John Wiley & Sons Inc, 2012.

[10] FP7 EARTH, Energy Aware Radio and neTwork tecHnologies, www.ict-earth.eu.

[11] Z. Niu, "TANGO: traffic-aware network planning and green operation," *Wireless Communications, IEEE,* vol. 18, pp. 25-29.

[12] BuNGee, Broadband Radio Access Networks (BRAN); Very high capacity density BWA networks; System architecture, economic model and derivation of technical requirements, ETSI TR 101 534 V1.1.1, www.ict-bungee.eu.

[13] H. Holma and A. Toskala, *LTE for UMTS: Evolution to LTE-Advanced*, 2 ed.: Wiley, 2011.

[14] S. Zhang, S. Zhang, X. Chen, and X. Huo, "Cloud Computing Research and Development Trend," in *Future Networks, 2010. ICFN '10. Second International Conference on*, pp. 93-97.

[15] K. Y. Kyun and R. Prasad, 4G Roadmap and Emerging Communication Technologies, Artech House, 1-58053-931-9, 2006.

[16] FP7 BuNGee, Beyond Next Generation Mobile Broadband, www.ict-bungee.eu.

[17] M. H. Ahmed, H. Yanikomeroglu, D. Falconer, and S. Mahmoud, "Performance enhancement of joint adaptive modulation, coding and power control using cochannel-interferer assistance and channel reallocation," in *Wireless Communications and Networking, 2003. WCNC 2003. 2003 IEEE*, 2003, pp. 306-310 vol.1.

[18] F. Wang and Z. Liu, "Adaptive Water-Filling Power Control for Wireless Communications Networks," *Communications Letters, IEEE,* vol. 12, pp. 737-739, 2008.

[19] X. Chen, Z. Zhao, H. Zhang, and T. Chen, "Applying multi-agent Q-learning scheme in cognitive wireless mesh networks for green communications," in *Personal, Indoor and Mobile Radio Communications Workshops (PIMRC Workshops), 2010 IEEE 21st International Symposium on*, pp. 336-340.

[20] M. Bennis and D. Niyato, "A Q-learning based approach to interference avoidance in self-organized femtocell networks," in *GLOBECOM Workshops (GC Wkshps), 2010 IEEE*, pp. 706-710.

[21] K. Akkarajitsakul, E. Hossain, D. Niyato, and K. Dong In, "Game Theoretic Approaches for Multiple Access in Wireless Networks: A Survey," *Communications Surveys & Tutorials, IEEE,* vol. 13, pp. 372-395.

[22] J. Ruofan, D. Grace, and P. Mitchell, "Cognitive Radio for UWB spectrum sharing and power allocation," in *Wireless Communication Systems (ISWCS), 2010 7th International Symposium on*, pp. 1001-1005.

[23] A. Damnjanovic, J. Montojo, W. Yongbin, J. Tingfang, L. Tao, M. Vajapeyam*, et al.*, "A survey on 3GPP heterogeneous networks," *Wireless Communications, IEEE,* vol. 18, pp. 10-21.

[24] K. Balachandran, J. Kang, K. Karakayali, and J. Singh, "Capacity Benefits of Relays with In-Band Backhauling in Cellular Networks," in *Communications, 2008. ICC '08. IEEE International Conference on*, 2008, pp. 3736-3742.

[25] S. Peters, A. Panah, K. Truong, and R. Heath, "Relay Architectures for 3GPP LTE-Advanced," *EURASIP Journal on Wireless Communications and Networking,* vol. 2009, p. 618787, 2009.

[26] H. Hu, H. Yanikomeroglu, D. D. Falconer, and S. Periyalwar, "Range extension without capacity penalty in cellular networks with digital fixed relays," in *Global Telecommunications Conference, 2004. GLOBECOM '04. IEEE*, 2004, pp. 3053-3057 Vol.5.

[27] Y.-D. Lin and Y.-C. Hsu, "Multihop cellular: a new architecture for wireless communications," in *INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, 2000, pp. 1273-1282 vol.3.

[28] Z. V. Masato, O. V. Chenxi, and V. V. Dorin, "Multihop relay extension for WiMAX network - overview and benefits of IEEE 802.16 j standard," *Fujitsu Sci. Tech. J,* vol. 44, pp. 292-302, 2008.

[29] M. Cao, X. Wang, S.-J. Kim, and M. M., "Multi-hop wireless backhaul networks: a cross-layer design paradigm," *Selected Areas in Communications, IEEE Journal on,* vol. 25, pp. 738-748, 2007.

[30] R. Pabst, B. H. Walke, D. C. Schultz, P. Herhold, H. Yanikomeroglu, S. Mukherjee*, et al.*, "Relay-based deployment concepts for wireless and mobile broadband radio," *Communications Magazine, IEEE,* vol. 42, pp. 80-89, 2004.

[31] T. Krag and S. Büettrich, O'Reilly Wireless Dev Center, Wireless Mesh Networking, www.oreillynet.com/pub/a/wireless/2004/01/22/wirelessmesh.html

[32] I. Ho, K. Leung, and J. Polak, "Stochastic Model and Connectivity Dynamics for VANETs in Signalized Road Systems," *Networking, IEEE/ACM Transactions on,* vol. 19, pp. 195-208, 2011.

[33] FP7 ABSOLUTE: Aerial Base Stations with Opportunistic Links for Unexpected & Temporary Events.

[34] FP7-ICT-ABSOLUTE/D4.1.1, Detailed Network and Protocol Architecture – First Issue, www.absolute-project.eu, May 2013.

[35] Z. Jens, K. Seong-Lyun, A. Magnus, and Q. Olav, *Radio Resource Management for Wireless Networks*: Artech House, Inc., 2001.

[36] I. Katzela and M. Naghshineh, "Channel assignment schemes for cellular mobile telecommunication systems: A comprehensive survey," *Communications Surveys & Tutorials, IEEE,* vol. 3, pp. 10-31, 2000.

[37] C. Kae Won, J. Wha Sook, and J. Dong Geun, "Resource allocation in OFDMA wireless communications systems supporting multimedia services," *IEEE/ACM Trans. Netw.,* vol. 17, pp. 926-935, 2009.

[38] Y. Chu, P. D. Mitchell, and D. Grace, "ALOHA and Q-Learning based medium access control for Wireless Sensor Networks," in *Wireless Communication Systems (ISWCS), 2012 International Symposium on*, pp. 511-515.

[39] Y. Chu, P. D. Mitchell, and D. Grace, "Reinforcement learning based ALOHA for multi-hop Wireless Sensor Networks with Informed Receiving," in *Wireless Sensor Systems (WSS 2012), IET Conference on*, pp. 1-6.

[40] F. Tobagi and L. Kleinrock, "Packet Switching in Radio Channels: Part II--The Hidden Terminal Problem in Carrier Sense Multiple-Access and the Busy-Tone Solution," *Communications, IEEE Transactions on,* vol. 23, pp. 1417-1433, 1975.

[41] A. S. Tanenbaum, *Computer Networks, fourth edition*, 2003.

[42] M. Zhang and T. S. Yum, "Comparisons of channel assignment strategies in cellular mobile telephone systems," in *Communications, 1989. ICC '89, BOSTONICC/89. Conference record. 'World Prosperity Through Communications', IEEE International Conference on*, 1989, pp. 467-471 vol.1.

[43] N. Saquib, E. Hossain, and K. Dong In, "Fractional frequency reuse for interference management in LTE-advanced hetnets," *Wireless Communications, IEEE,* vol. 20, pp. 113-122.

[44] S. Sesia, I. Toufik, and M. Baker, *LTE - The UMTS Long Term Evolution: From Theory to Practice*: Wiley.

[45] L. Kleinrock, "Queueing Systems - Volume I: Theory," *John Wiley & Sons,* 1975.

[46] FCC, Notice of proposed rule making and order, ET Docket No 03-222, December 2003.

[47] European Commission, Decision on harmonised technical conditions of use in the 790-862 MHz frequency band for terrestrial systems capable of providing electronic communications services in the European Union, May 2010.

[48] Ofcom, The award of 800 MHz and 2.6 GHz spectrum, 2012.

[49] QINETIQ, Cognitive Radio Technology: A Study for Ofcom, 2007.

[50] A. P. Subramanian, M. Al-Ayyoub, H. Gupta, S. R. Das, and M. M. Buddhikot, "Near-Optimal Dynamic Spectrum Allocation in Cellular Networks," in *New Frontiers in Dynamic Spectrum Access Networks, 2008. DySPAN 2008. 3rd IEEE Symposium on*, 2008, pp. 1-11.

[51] Z. Youping, J. H. Reed, M. Shiwen, and K. K. Bae, "Overhead Analysis for Radio Environment Mapenabled Cognitive Radio Networks," in *Networking Technologies for Software Defined Radio Networks, 2006. SDR '06.1st IEEE Workshop on*, 2006, pp. 18-25.

[52] FP7 FARAMIR, Enabling Spectrum-Aware Radio Access for Cognitive Radios, www.ict-faramir.eu.

[53] IEEE 802.22 Working Group on Wireless Regional Area Networks, www.ieee802.org.

[54] ETSI, White Space Devices (WSD); Wireless Access Systems operating in the 470 MHz to 790 MHz frequency band; Harmonized EN covering the essential requirements of article 3.2 of the R&TTE Directive, Draft ETSI EN 301 598 V1.0.0, July 2013.

[55] T. Yucek and H. Arslan, "A survey of spectrum sensing algorithms for cognitive radio applications," *Communications Surveys & Tutorials, IEEE,* vol. 11, pp. 116-130, 2009.

[56] D. Grace, A. G. Burr, and T. C. Tozer, "Distributed channel assignment strategies using coexistence etiquettes for land based radio environment," *Electronics Letters,* vol. 32, pp. 1956-1957, 1996.

[57] L. E. Doyle, Essentials of Cognitive Radio, April 2009.

[58] J. Mitola, III and G. Q. Maguire, Jr., "Cognitive radio: making software radios more personal," *Personal Communications, IEEE,* vol. 6, pp. 13-18, 1999.

[59] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *Selected Areas in Communications, IEEE Journal on,* vol. 23, pp. 201-220, 2005.

[60] R. W. Thomas, L. A. DaSilva, and A. B. MacKenzie, "Cognitive networks," in *New Frontiers in Dynamic Spectrum Access Networks, 2005. DySPAN 2005. 2005 First IEEE International Symposium on*, 2005, pp. 352-360.

[61] B. Han, D. Grace, and P. Mitchell, "Cognitive greedy-backhaul routing metric exploiting cross-layer design for wireless ad hoc and mesh networks," in *Cognitive Radio Oriented Wireless Networks & Communications (CROWNCOM), 2010 Proceedings of the Fifth International Conference on*, 2010, pp. 1-5.

[62] D. Sarkar and H. Narayan, "Transport Layer Protocols for Cognitive Networks," in *INFOCOM IEEE Conference on Computer Communications Workshops , 2010*, 2010, pp. 1-6.

[63] T. Jiang, D. Grace, and P. D. Mitchell, "Efficient exploration in reinforcement learning-based cognitive radio spectrum sharing," *Communications, IET,* vol. 5, pp. 1309-1317, 2011.

[64] T. Jiang, D. Grace, and Y. Liu, "Two-stage reinforcement-learning-based cognitive radio with exploration control," *Communications, IET,* vol. 5, pp. 644-651.

[65] A. Edwards and W. M. Pottenger, "Higher order Q-Learning," in *Adaptive Dynamic Programming And Reinforcement Learning (ADPRL), 2011 IEEE Symposium on*, 2011, pp. 128-134.

[66] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, Mass., London: MIT Press, 1998.

[67] C. Clancy, J. Hecker, E. Stuntebeck, and T. O'Shea, "Applications of Machine Learning to Cognitive Radio Networks," *Wireless Communications, IEEE,* vol. 14, pp. 47-52, 2007.

[68] N. Junhong and S. Haykin, "A Q-learning-based dynamic channel assignment technique for mobile communication systems," *Vehicular Technology, IEEE Transactions on,* vol. 48, pp. 1676-1687, 1999.

[69] A. Galindo-Serrano and L. Giupponi, "Decentralized Q-Learning for Aggregated Interference Control in Completely and Partially Observable Cognitive Radio Networks," in *Consumer Communications and Networking Conference (CCNC), 2010 7th IEEE*, pp. 1-6.

[70] K. Spiros and K. Daniel, "Reinforcement learning of coordination in cooperative multi-agent systems," presented at the Eighteenth national conference on Artificial intelligence, Edmonton, Alberta, Canada, 2002.

[71] A. B. MacKenzie and S. B. Wicker, "Game theory in communications: motivation, explanation, and application to power control," in *Global Telecommunications Conference, 2001. GLOBECOM '01. IEEE*, 2001, pp. 821-826 vol.2.

[72] H. Lima and Y. Kuroe, "Swarm reinforcement learning algorithms -exchange of information among multiple agents," in *SICE, 2007 Annual Conference*, 2007, pp. 2779-2784.

[73] L. Giupponi, A. Galindo-Serrano, P. Blasco, and M. Dohler, "Docitive networks: an emerging paradigm for dynamic spectrum management [Dynamic Spectrum Management]," *Wireless Communications, IEEE,* vol. 17, pp. 47-54.

[74] A. Galindo-Serrano, L. Giupponi, P. Blasco, and M. Dohler, "Learning from experts in cognitive radio networks: The docitive paradigm," in *Cognitive Radio Oriented Wireless Networks & Communications (CROWNCOM), 2010 Proceedings of the Fifth International Conference on*, pp. 1-6.

[75] P. Blasco, L. Giupponi, A. Galindo-Serrano, and M. Dohler, "Energy benefits of cooperative docitive over cognitive networks," in *Wireless Technology Conference (EuWIT), 2010 European*, pp. 109-112.

[76] A. Galindo-Serrano, L. Giupponi, and M. Dohler, "Cognition and Docition in OFDMA-Based Femtocell Networks," in *GLOBECOM 2010, 2010 IEEE Global Telecommunications Conference*, pp. 1-6.

[77] G. Lorenza, M. G.-S. Ana, and D. Mischa, "From cognition to docition: The teaching radio paradigm for distributed \& autonomous deployments," *Comput. Commun.,* vol. 33, pp. 2015-2020.

[78] T. Ming, "Multi-agent reinforcement learning: independent vs. cooperative agents," in *Readings in agents*, N. H. Michael and P. S. Munindar, Eds., ed: Morgan Kaufmann Publishers Inc., 1998, pp. 487-494.

[79] L. Torrey and J. Shavlik, "Transfer Learning," *Handbook of Research on Machine Learning Applications,* 2009.

[80] P. Sinno Jialin and Y. Qiang, "A Survey on Transfer Learning," *Knowledge and Data Engineering, IEEE Transactions on,* vol. 22, pp. 1345-1359.

[81] L. A. Celiberto, J. P. Matsuura, Lo, x, M. pez de, x*, et al.*, "Using Transfer Learning to Speed-Up Reinforcement Learning: A Cased-Based Approach," in *Robotics Symposium and Intelligent Robotic Meeting (LARS), 2010 Latin American*, pp. 55-60.

[82] L. M. Correia, D. Zeller, O. Blume, D. Ferling, Y. Jading, Go*, et al.*, "Challenges and enabling technologies for energy aware mobile radio networks," *Communications Magazine, IEEE,* vol. 48, pp. 66-72.

[83] J. Hoydis, M. Kobayashi, and M. Debbah, "Green Small-Cell Networks," *Vehicular Technology Magazine, IEEE,* vol. 6, pp. 37-43.

[84] G. Auer, V. Giannini, C. Desset, I. Godor, P. Skillermark, M. Olsson*, et al.*, "How much energy is needed to run a wireless network?," *Wireless Communications, IEEE,* vol. 18, pp. 40-49.

[85] F. Richter, A. J. Fehske, and G. P. Fettweis, "Energy Efficiency Aspects of Base Station Deployment Strategies for Cellular Networks," in *Vehicular Technology Conference Fall (VTC 2009-Fall), 2009 IEEE 70th*, 2009, pp. 1-5.

[86] Y. Wu and Z. Niu, "Energy efficient base station deployment in green cellular networks with traffic variations," in *Communications in China (ICCC), 2012 1st IEEE International Conference on*, pp. 399-404.

[87] L. Suarez, L. Nuaymi, and J.-M. Bonnin, "An overview and classification of research approaches in green wireless networks," *EURASIP Journal on Wireless Communications and Networking,* vol. 2012, p. 142.

[88] Ofcom, Sitefinder: Mobile Phone Base Station Database, www.sitefinder.ofcom.org.uk.

[89] O. Eunsung, B. Krishnamachari, L. Xin, and N. Zhisheng, "Toward dynamic energy-efficient operation of cellular network infrastructure," *Communications Magazine, IEEE,* vol. 49, pp. 56-61.

[90] FP7 EARTH, Energy efficiency analysis of the reference systems, areas of improvements and target breakdown, INFSO-ICT-247733 EARTH Deliverable D2.3.

[91] Q. Zhao and D. Grace, "Agent Transfer Learning for Cognitive Resource Management on Multi-hop Backhaul Networks," in *Future Network & Mobile Summit*, 2013.

[92] Q. Zhao, D. Grace, and T. Clarke, "Transfer Learning with Cooperation Management: Balancing the Quality of Service and Information Exchange Overhead in Cognitive Radio Networks," unpublished.

[93] MathWorks, *Getting Start with MATLAB*, 2005.

[94] N. Metropolis, "The beginning of the Monte Carlo method," *Los Alamos Science (1987 Special Issue dedicated to Stanislaw Ulam),* pp. 125–130, 1987.

[95] S. R. Saunders, *Antennas and Propagation for Wireless Communication Systems*, 1999.

[96] J. Thornton, D. Grace, M. H. Capstick, and T. C. Tozer, "Optimizing an array of antennas for cellular coverage from a high altitude platform," *Wireless Communications, IEEE Transactions on,* vol. 2, pp. 484-492, 2003.

[97] COST-231, www.cost.eu/domains_actions/ict/Actions/231.

[98] P. Kyösti, et al, WINNER II Channel Models, v1.2, IST-WINNER D1.1.2, 2007.

[99] A. Papadogiannis and A. G. Burr, "Multi-beam assisted MIMO - A novel approach to fixed beamforming," in *Future Network & Mobile Summit (FutureNetw), 2011*, pp. 1-8.

[100] 3GPP Technical Specification Group Radio Access Networks, LTE - Evolved Universal Terrestrial Radio Access (E-UTRA) Radio Frequency (RF) system scenarios, 3GPP TR 36.942 version 9.0.1 Release 9, April 2010.

[101] A. Rueda and Kinsner, "A survey of traffic characterization techniques in telecommunication networks," in *Electrical and Computer Engineering, 1996. Canadian Conference on*, 1996, pp. 830-833 vol.2.

[102] J. D. Chimeh, M. Hakkak, and S. A. Alavian, "Internet Traffic and Capacity Evaluation in UMTS Downlink," in *Future Generation Communication and Networking*, 2007, pp. 547-552.

[103] A. Papoulis and S. Pillai, *Probability, random variables, and stochastic processes*, 2002

[104] D. Grace, "Distributed Dynamic Channel Assignment for the Wireless Environment," PhD Thesis, Department of Electronics, University of York, 1999.

[105] P. Mitchell, "Effective Medium Access Control for Geostationary Satellite Systems," PhD Thesis, Department of Electronics, University of York, 2003.
[106] M.R.Spiegel, *Stastics*, 1972.
[107] M. Serizawa and D. J. Goodman, "Instability and deadlock of distributed dynamic channel allocation," in *Vehicular Technology Conference, 1993 IEEE 43rd*, 1993, pp. 528-531.
[108] COST 231, Urban transmission loss models for mobile radio in the 900- and 1800- MHz bands (Revision 2), COST 231 TD (90) 119 Rev 2.
[109] M. Vargas and R. Pallas-Areny, "Thermal noise in a finite bandwidth," *Instrumentation & Measurement Magazine, IEEE,* vol. 4, pp. 23-25, 2001.
[110] Q. Zhao and D. Grace, "Application of Cognition based Resource Allocation Strategies on a Multi-hop Backhaul Network," in *IEEE International Conference on Communication Systems*, 2012.
[111] W. Longwei, C. Xianfu, Z. Zhifeng, and Z. Honggang, "Exploration vs exploitation for distributed channel access in cognitive radio networks: A multi-user case study," in *Communications and Information Technologies (ISCIT), 2011 11th International Symposium on*, pp. 360-365.
[112] FP7 EARTH, Final Report on Green Network Technologies, INFSO-ICT-247733 EARTH Deliverable D3.3, www.ict-earth.eu.
[113] FP7 ABSOLUTE, Aerial Base Stations with Opportunistic Links for Unexpected & Temporary Events, www.absolute-project.eu.
[114] A. Sen, "Markets and Freedoms: Achievements and Limitations of the Market Mechanism in Promoting Individual Freedoms," *Oxford Economic Papers,* vol. 45, pp. 519-541, 1993.
[115] M. A. Marsan, S. Buzzi, D. Ciullo, and M. Meo, "Optimal Energy Savings in Cellular Access Networks," in *Communications Workshops, 2009. ICC Workshops 2009. IEEE International Conference on*, 2009, pp. 1-5.
[116] Y. Liu, D. Grace, and P. D. Mitchell, "Exploiting platform diversity for GoS improvement for users with different High Altitude Platform availability," *Wireless Communications, IEEE Transactions on,* vol. 8, pp. 196-203, 2009.
[117] R. B. Cooper, *Introduction to queueing theory*, 2 ed. New York: North-Holland, 1981.
[118] Y. Han, D. Grace, and P. Mitchell, "Energy efficient topology management for beyond next generation mobile broadband systems," in *Wireless Communication Systems (ISWCS), 2012 International Symposium on*, pp. 331-335.
[119] Q. Zhao, T. Jiang, N. Morozs, D. Grace, and T. Clarke, "Transfer Learning: A Paradigm for Dynamic Spectrum and Topology Management in Flexible Architectures," in *IEEE 78th Vehicular Technology Conference* September 2013.
[120] K. Gomez, C. Sengul, N. Bayer, R. Riggio, T. Rasheed, and D. Miorandi, "MORFEO: Saving energy in wireless access infrastructures," in *World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2013 IEEE 14th International Symposium and Workshops on a*, pp. 1-6.
[121] D. Grace and M. Mohorcic, *Broadband Communications via High-Altitude Platforms*: Wiley.
[122] Project Loon: Balloon-Powered Internet for Everyone, www.google.com/loon.