Towards Intelligent Energy-Aware Self-Organised Cellular Networks ($iSONs$)

Salahedin Rehan

PhD

University of York

Electronics

July 2015

Abstract

This thesis investigates the application of intelligent energy-aware resource management techniques for current and future wireless broadband deployments.

Energy-aware topology management is firstly studied aiming at dynamically managing the network topology by fine tuning the status of network entities (dormant / active) to scale the energy consumption with traffic demands. This is studied through an analytical model based on queueing theory and through simulation to help understand its operational capabilities under a range of traffic conditions. Advanced radio resource management is also investigated. This reduces the number of nodes engaged in the service whenever possible reducing the energy consumption at low and medium traffic loads while enhancing system capacity and QoS when the traffic load is high. As an enabling technology for self-awareness and adaptability, Reinforcement Learning (RL) is applied to manage network resources in an intelligent, self-aware, and adaptable manner. This is complemented with a range of novel cognitive learning and reasoning algorithms which are capable of translating past experience into valuable sets of information in order to optimise decisions taken as part of the radio resource and topology management functionalities. Dependencies between the proposed techniques are also addressed formulating an intelligent self-adaptable approach, which is capable of dynamically deactivating redundant nodes and redirecting traffic appropriately while enhancing system capacity and QoS.

Contents

List of Figures

Acknowledgement

In the name of God The most merciful The most gracious. All praise be to God for the strengths and His blessings in completing this thesis, and for guiding me towards the University of York as a whole and specially towards my supervisor Dr David Grace to whom I express my most sincere gratitude for his exceptional guidance and support throughout my PhD studies.

I would also like to thank my entire family, brothers, and sisters. Special thanks to my dear mother for her unconditional love and prayers, my father for his constant support and advice, my wife for her endless encouragement and struggle.

Nevertheless, my friends, colleagues, and members of the communications research group to whom I owe a lot of the knowledge used for the completion of this thesis, I thank you all.

Declaration

This work has not been presented for an award at this, or any other University. All contributions presented in this thesis as original are as such to be the best knowledge of the author. References and acknowledges to other researchers have been given as appropriate.

Some of the research presented in this thesis has resulted in a number of publications and EU research project deliverables. A list of the publications is provided in the next section.

Publications

Conference presentations

1- S. Rehan and D. Grace, "Combined Green Resource and Topology Management for Beyond Next Generation Mobile Broadband Systems," ***Computing, Networking and Communications (ICNC), 2013 International Conference on*** , vol., no., pp.242,246, 28-31 Jan. 2013.

2- S. Rehan and D. Grace, "Macro-Cell Overlaid Green Topology Management for Beyond Next Generation Mobile Broadband Systems," ***Future Network and Mobile Summit (FutureNetworkSummit)***, 2013, vol., no., pp.1,9, 3-5 July 2013.

3- S. Rehan and D. Grace, "Energy-Aware Topology Management for High Capacity Density Temporary Event Networks," ***Advanced Technologies for Communications (ATC), 2013 International Conference on*** , vol., no., pp.307,311, 16-18 Oct. 2013.

4- S. Rehan and D. Grace, "Efficient Joint Operation of Advanced Radio Resource and Topology Management in Energy-Aware 5G Networks," ***IEEE Vehicular Technology Conference*** (VTC – 2015 fall).

Journal articles

1- "Energy-Aware Topology Management for Opportunistic 5G Temporary Event Networks" Salahedin Rehan, David Grace, and Paul D. Mitchell, **submitted to IEEE Transactions on Mobile Computing**.

2- "Hierarchical Learning for Cell Selection and Topology Management in Energy-Aware 5G Cognitive Networks", Salahedin Rehan, David Grace, Luis Suarez, and Loutfi Nuaymi, **submitted to IEEE Transactions on Mobile Computing**.

3- “A Markovian Study into Network Deployment Strategies and Radio Resource Management for Energy-Aware 5G Cellular Systems”, Salahedin Rehan, and David Grace, **submitted to Transactions on Emerging Telecommunications Technologies.**
4- “Managing Resources in Energy Efficient Cognitive 5G Cellular Systems”, Salahedin Rehan, and David Grace, **submitted to Elsevier computer networks.**
5- “A Multi-Criteria BS Switching-Off Algorithm for 5G Heterogeneous Cellular Networks with Hybrid Energy Sources”, Luis Suarez, Loutfi Nuaymi, David Grace, Salahedin Rehan, and Jean-Marie Bonnin**, submitted to IEEE Journal on Selected Areas in Communications Special Issue on Energy-Efficient Techniques for 5G Wireless Communication Systems.**

FP7 ABSOLUTE project deliverables

The work presented in this thesis has also directly contributed to the *EU FP7 Aerial Base Stations with Opportunistic Links for Unexpected and Temporary Events (ABSOLUTE)* and has been published in various project deliverables as follows:

1- FP7 ABSOLUTE D2.6.2, K. Gomez, L. Goratti, T. Rasheed, T. Javornik, A. Svigelj, A. Hrovat, K. Alic, S. Rehan, Y. Han, Q. Zhao, N. Morozs, D. Grace, “System Wide Simulations; Analysis and Results" June 2015, available at: www. http://www.absolute-project.eu/.
2- FP7 ABSOLUTE D4.1.4, Q. Zhao, S. Rehan, D. Grace, A. Vilhar, A. Švigelj, K. Alič, K. Gomez, T. Rasheed, S. Chandrasekharan, K. Sithamparanathan, M. Thakur, A. Munari, L. Reynaud, "Detailed Network and Protocol Architecture: Final Version” June 2015, available at: www. http://www.absolute-project.eu/.

3- FP7 ABSOLUTE D4.2.1, K. Gomez, L. Goratti, K. Sithamparanathan, Q. Zhao, S. Rehan, D. Grace, A. Svigelj, K. Alic, A. Vilhar, T. Javornik, M. Thakur, " System Capacity Assessments" June 2015, available at: www. http://www.absolute-project.eu/.

Chapter 1 **Introduction**

Contents

1.1 **Overview**

The increase in demand for mobile wireless services has remained a strong driving force behind the evolution of mobile wireless network architectures and protocol stacks since its first release in 1947 [1]. Data rates have grown from kb/s to Gb/s and subscribers have evolved to accommodate the communication between things rather than only humans. Starting from the 1980s when 1G, the 1st generation, became available using analogue circuit switching technology where a dedicated path is used to communicate (only voice) between two mobile terminals. By the 1990s, 2G technologies became available denoting the emergence of wireless digital systems by enhancing the air interface. The digital nature of 2G enabled the use of more efficient channel access techniques such as TDMA, FDMA, and CDMA to better separate users within certain frequency bands [1]. 2G also introduced digital encryption and channel coding for error correction and detection in wireless communications offering better voice quality and more reliable data services over fading channels. New data-centric standards were then developed. These can be overlaid upon existing 2G technologies in an effort to increase the data rate which is required to support internet applications such as e-mail traffic, web browsing, and mobile commerce (m-commerce). These standards were considered as 2.5G. 2.5G introduced services like High Speed Circuit switched data. 2.5G was still found to be inefficient in some applications especially data communication due to its spectral inefficiency [2]. However, 2.5G paved the way to 3G which solved the 2.5G limitations by merging wireless communication features with the capabilities of the Internet Protocol based networks to create a network which provides services independent of the technology platform. The first commercial 3G network was launched in Japan in 2001 [1]. 3G offers a wider range of more advanced services while achieving greater network capacity by using WCDMA as the multiple access scheme for the air-interface which improves spectral efficiency.

LTE-3GPP and mobile WiMAX have been on the market since 2006, 2009 respectively and are often referred to as 4G technologies. However, current versions of these technologies do not fulfil the ITU-R prerequisites of 1Gbit/s date rate for 4G systems. LTE can be referred to as 3.5G technology and a transitional step toward 4G technologies. This time the architecture is simplified to a flatter architecture with an Evolved Packet Core (EPC) and an Evolved-UTRAN (e-NodeB). This way achieved by distributing the functionality of the Radio Network Controller (RNC) and Base Station Controller (BSC) to the Base Transceiver Station (BTS) and a set of servers and gateways. This way data transfer is much faster and the network is more cost effective [2]. A 4G system is estimated to deliver a comprehensive and safe all-IP based mobile broadband solutions to all kinds of mobile terminals. The targeted services include broadband internet access, VoIP, gaming services, and streamed multimedia. LTE-Advanced (LTE-A) aims to increase peak data rate, achieve higher spectral efficiency, increase number of simultaneously active subscribers, and improve performance at cell edges. The main functionalities introduced in LTE-A to obtain such performance are Carrier Aggregation (CA), enhanced use of multi-antenna techniques and support for Relay Nodes (RN).

Currently, worldwide mobile data connections have reached 1.1 billion, and are increasing at 60% each year, and are expected to reach 5 billion in 2017 [3]. In addition to this increase in mobile web traffic, mobile terminals are also evolving to accommodate a broader range of applications such as machine-to-machine (M2M) communication, smart objects, and ubiquitous computing creating the Internet of Things. In general, mobile subscriptions are predicted to grow exponentially from 6.2 billion in June 2012 to 9 billion in 2017. This shift to new data-intensive applications is a key factor behind the development of new standards for beyond next generation mobile broadband systems like Long Term Evolution- Advanced (LTE-Advanced) [4], and beyond next generation systems [5]. A lot of effort is being carried out to predict future wireless broadband system requirements as new technologies are being proposed and new scenarios foreseen [6]. There is a broad consensus on the fact that energy efficiency is a key design parameter for 5G systems [7-10]. 5G is considered to be more than a traditional wireless technology evolution given the fact that it is being developed to meet a broader range of requirements compared to previous generations in a cost as well as energy efficient manner. In some cases an ultra-quick response time is required, in other cases an even performance across the service area and cell edge is needed, in some others a high data rate is the predominant criterion. 5G is required to present a wide range of different factors, optimised for different applications from within one common network, in a cost as well as energy efficient manner. Cognitive Radio (CR) is believed to play an important role in enabling effective inter-RAT operations and improve the self-organisational capabilities of 5G networks, enabling the development of intelligent Self-Organised Networks ($iSONs$). Even though it was primarily proposed to deal with the spectrum efficiency problem arising from the increase in demand for mobile wireless services [11-13], CR mechanisms are currently developed to tackle a broader range of wireless network problems such as enhancing its energy efficiency [11, 13, 14]. Cognitive radio is not restrictively applicable to opportunistic spectrum access. As is seen from its definition, a cognitive radio system is "*A radio system employing technology that allows the system to obtain knowledge of its* *operational and geographical environment, established policies and its internal state; to dynamically and autonomously adjust its operational parameters and protocols according to its obtained knowledge in order to achieve predefined objectives; and to learn from the results obtained*" [15]. In fact, a Cognitive Network Management unit is proposed as one of the main functional units in 5G wireless broadband networks and beyond [13]. Also, Third Generation Partnership project ($3GPP$) is emphasising the importance of incorporating self-X features, i.e., self-configuration, self-optimization, self-healing, and plug-and-play, as main functionalities of Long Term Evolution-Advanced ($LTE-A$) network entities such as the evolved Node B (eNB) and these are currently under development [16-18].

1.2 **Motivation**

As previously discussed, energy consumption of future mobile wireless networks is of increasing concern due to the fact that current protocols, and network deployment as well as management strategies have been developed oblivious to their contribution to the network energy efficiency. For instance, among the popular means to address the challenge of meeting the ultra-high capacity demands is the use of high order modulation techniques to achieve higher spectrum efficiency, massive network densification, with particular attention to special event scenarios such as open air festivals and marathons [6, 19, 20], and a range of radio resource management techniques optimised for capacity and QoS [21, 22]. Moreover, the number of Base Stations (BSs) required to serve a given area is predicted to surpass the number of actively transmitting or receiving UEs (User Equipment) [23]. This has resulted in environmental concerns as CO_2 emissions are heading towards critical figures increasing threefold in the next five years which can have a damaging impact on the wellbeing of the ecosystem [8, 24]. The increase in the power required to run a network also concerns industry since the increase in energy demand results in increased CAPEX and OPEX especially with the introduction of environmental related taxes. Hence the challenging task of maintaining the network in an energy efficient manner while delivering the required QoS is not only ecologically, but also economically imperative. According to [25], only 10% of the BSs in a network carry more than 55% of the aggregate network traffic. This reflects the fact that 55% of total data traffic is generated by a limited number of heavy data users that represent only 10% of total subscribers [23]. Intuitively, about 50% of base stations are low loaded, carrying only a load in the range of 1-10MB per day [25]. Current network resource management solutions and protocols do not exploit the temporospatial variation of demand and are only tuned to efficiently work during peak periods. This is proven to be ecologically as well as economically inefficient as previously discussed and new innovative solutions are needed to scale the energy consumption with traffic demand. These solutions in principle strive to trade spectral efficiency for energy efficiency in situations when network resources such as BSs and bandwidth are underutilised. This is done for example in a heterogeneous network by deactivating small cell capacity boosters when the demand is low. Even though this decreases the spectral efficiency, it increases the energy efficiency of the network particularly at night time when most deployments are severely underutilised. Nevertheless, such solutions must first of all not jeopardise the QoS that the network is designed to deliver, and secondly be applicable for a range of different tempo-spatial traffic conditions. For this reason, 5G, unlike its predecessors [1], is facing the challenge of providing the required capacity density in a cost as well as energy efficient manner optimising energy efficiency of the network at low and medium traffic densities, and optimising QoS and capacity when the traffic load is high.

This thesis aims to enhance the energy efficiency of current and future wireless broadband deployments by proposing a range of novel techniques and algorithms to flexibly manage network resources in an intelligent, self-aware, and adaptable manner.

On the network management side, a range of energy-aware topology management schemes are proposed for a next generation wireless broadband system. Energy-aware topology management aims to control the predesigned topology of the system in order to achieve energy savings taking into consideration the level of QoS required. This is achieved by fine-tuning the status of hardware components (active / dormant) depending on the traffic demands. This solution is investigated given the fact that the most energy consuming entity in a wireless broadband system is the base station consuming up to 80 per cent of the total energy consumed by the entire system [26]. Moreover, 70 per cent of the energy consumed by the base stations is used by the power amplifiers, and for cooling purposes which is often independent of the number of users served (just to maintain coverage and connectivity) [26]. The introduction of dormant mode and efficient energy-aware topology management schemes can lead to largely decrease the contribution of base stations to the system energy consumption by exploiting their idle state.

To achieve this, firstly, uniformly distributed traffic is studied to identify the basic requirements for a topology management to operate and identify its potential in enhancing the network energy efficiency. Subsequently, in order to develop a scheme that suits a wide range of scenarios, the impact of high traffic dynamics expected in urban areas during a special event is analysed. The scheme is finally enhanced using Reinforcement Learning (RL) to identify which BS to de-(activate) in an adaptable manner.

This thesis also explores stand-alone radio resource management solutions which can also boost the performance and increase the operational region of the aforementioned topology management paradigm. These solutions can also operate to boost the performance of other energy-aware mechanisms that work on exploiting the idle state of BSs. The operation of the emerging technique of Load Unbalancing is first evaluated and compared with the traditional application of Load Balancing technique. Consequently, a novel multitask hierarchical learning approach based on Reinforcement Learning (RL) is proposed whose output is used to facilitate the development of an adaptive intelligent Radio Resource Control ($iRRC$). The implementation and feasibility of the proposed schemes is also evaluated. Last but not least, the interoperability between the radio resource and topology management techniques is studied and ways proposed to enable their efficient joint operation.

1.3 Thesis outline

The proposed solutions are based on the observed shortcomings, and unsolved challenges as well as existing solutions towards the development of future energy-aware Self-Organised Network (SON) functionalities, such as energy-efficient topology management and radio resource management, aimed at 5G networks. To this end, Chapter 2 firstly summarises the potential technological enablers and latest advancements towards 5G. Subsequently, a detailed state-of-the-art in the active field of energy-aware topology and radio resource management is presented, defined, and classified depending on their employed techniques addressing their advantages and drawbacks.

We subsequently conduct a theoretical evaluation of the two most common network deployment strategies; namely, the homogenous and the heterogeneous deployment strategies and evaluate their optimal occupancy performance under the application of a range of Radio Resource Management (RRM) techniques proposed for energy-aware 5G cellular networks. The evaluation takes into consideration a wide range of aspects. These are the system QoS, the system energy efficiency, and network stability. The RRM techniques evaluated are Load Balancing and the emerging technique of Load Unbalancing. The heterogeneous deployment strategy is of more focus in this work as one of the key aims is to evaluate the importance of the existence of a macro-cell overlay and its effect on the operation of the different RRM techniques. This chapter also indicates the scenarios in which these deployment strategies perform best.

The application of energy-aware topology management for 5G deployments is then studied in Chapter 4. This chapter firstly introduces an energy-aware Neighbour-Based Topology Management (NBTM) scheme for beyond next generation mobile broadband systems which fine-tunes the status of the network nodes (dormant / active) depending on the traffic demands. Secondly, some key elements, such as the existence of a macro-cell overlay, are explored which can contribute to the successful development of feasible energy-aware topology management schemes as concluded from Chapter 3. This leads to the development of a Macro-cell overlaid TM (McTM) which enhances the operation of the NBTM in terms of energy efficiency as well as QoS at low traffic loads. The schemes presented in this chapter as well as the findings from Chapter 3 serve as a baseline for the solutions proposed in the subsequent chapters.

The next step towards the development of flexible and adaptable solutions is the study of the impact of high traffic dynamics expected in urban areas during a special event on the operation of topology management schemes. This is presented in Chapter 5 which also introduces an energy-aware topology management scheme for a 5G ultra-dense network deployment taking into consideration the occurrence of a temporary event. This scheme, unlike its counterparts, aims to neutralise the impact of the high traffic dynamics expected in urban areas during a special event by clustering entities which can cooperate to maximise area throughput and network energy efficiency. This chapter also uses the analytical model presented in Chapter 3 to evaluate the scheme mathematically. The scalability of the theoretical framework is also studied. Lastly, the scheme presented in this chapter is compared against the topology management scheme introduced in Chapter 4.

The application of Reinforcement Learning (RL) to enable the proposed solutions to be adaptable and applicable to a wider range of scenarios and deployments strategies is studied in Chapter 6. This chapter introduces a novel multitask hierarchical learning and reasoning approach based on Reinforcement Learning (RL) whose output is used to facilitate the development of an adaptive intelligent Radio Resource Control ($iRRC$) as well as an intelligent Topology Management (iTM) scheme. The enhanced RL-algorithm enables cross-layer optimisation and facilitates the interpretation of past experience into two sets of useful information which is exploited by the $iRRC$ and iTM to enhance their decision-making and consequently their performances. To enable the efficient joint operation, this chapter also takes into consideration interoperability issues between energy aware radio resource and topology management techniques. Last but not least, the implementation as well as feasibility of the proposed $iRRC$ algorithm is studied in detail and shown to be compatible with 3GPP standards.

In Chapter 7, a range of potential ideas on how to continue the work presented in this thesis are discussed. The main conclusions of this work are finally summarized in Chapter 8.

Chapter 2 **Managing Resources in Energy Efficient Cognitive 5G Cellular Systems – A Literature Review**

Contents

2.1 **Introduction**

This chapter provides an overview of the latest developments and challenges in the active field of energy-efficient topology and radio resource management for future 5G networks. Most of the work presented throughout this thesis is aimed at being compatible with future mobile broadband networks taking into consideration future network elements and their capabilities. The use of artificial intelligence based techniques is one example which is detailed in Chapter 6. Yet another example is the use of dense network deployments and the consideration of uneven traffic dynamics during a temporary event. These are detailed in Chapter 5. For this reason, this chapter firstly presents the potential technological enablers and latest advancements towards 5G and the proposed network capabilities in Section 2.2. This is complemented by a detailed state-of-the-art of existing energy-efficient network Topology Management (TM) as well as Radio Resource Management (RRM) in Section 2.3 and 2.4 respectively. These are important as they form part of the functionalities that this thesis aims to develop to enhance the network energy efficiency by dynamically managing the network resources in an intelligent manner. A classification of these techniques is also presented in their respective sections on the basis of their employed techniques addressing their advantages and drawbacks. The techniques presented in this thesis are also included in Section 2.3 and 2.4 to give the reader a better insight into how they fit within existing literature. Last but not least, Section 2.5 concludes the chapter summarising the most relevant findings and recommendations.

2.2 Envisioning the 5G Cellular System

There is a broad consensus on what 5G needs to provide and in what manner, despite the fact that there is not currently a clear technological enabler that dominates the transition towards it. According to [27-29], 5G is aimed at being 1000 times faster than 2010 SOTA deployments [28], accommodating 100 times more connected devices, be 5 times more responsive, cost-effective, as well as being energy self-sustainable. This implies that 5G needs to deliver, in a general sense, peak speeds of the order of several Gbps for mobile users, reduced latency to the order of milliseconds, while supporting a 1000 times increase in traffic in a cost effective, energy self-sustainable, and self-aware manner. Thus, the minimum requirement for a 5G cellular network is to be scalable, and deliver Gbps data rates for slowly moving mobile users. The energy efficiency as well as the cost aspects are not of paramount importance for early deployments, which are envisioned to begin as soon as 2015, for example as a prototype by ZTE [30], and predicted to be commercially deployed by 2020 [31].

Some of the challenges are already being preliminarily addressed. Early efforts to lay the foundations of 5G come from METIS, which is considered to be the first official international research activity on 5G co-founded by the European Commission. METIS were able to create the first 5G radio channel model [32] which is expected to gain a wide acceptance from industry and academia. The BuNGee FP7 project is also one of the earliest efforts towards 5G, so much so that it was known as a 'beyond next generation broadband system' [5]. The BuNGee project developed one of the first 5G architectures, introducing Hub and Access Base Stations, HBS and ABS respectively. This is a dual-hop design with self-backhaul capabilities enabled by wireless point-to-point backhaul links between HBSs and ABSs. BuNGee successfully delivered 1Gbps data rate in a demo conducted in Barcelona in 2011.

A more current approach is taken by Artemis recently announcing the development of what they call the personal Cell (pCell) [33]. The pCell technology is claimed to achieve the largest increase in capacity in the history of wireless technology sidestepping Shannon's Law according to its inventor who describes it as the wireless equivalent of fibre optic. Despite conducting several demonstrations by wirelessly streaming a 4K UltraHD video to multiple devices at once, experts across the industry seem rather sceptical about the real performance of pCell, as the technology is yet to be demonstrated outside of a carefully controlled environment. The technology purports to be a Distributed-Input-Distributed-Output (DIDO) cloud based system in which multiple transmitters are deployed and centrally controlled by a DIDO Data Centre which fine-tunes the radio signals creating pockets of constructive interference around all devices. This way, independent data channels are created making the whole bandwidth available for every user to utilise [34]. The technology is aimed to be commercialised in the fourth quarter of 2015. Despite initial steps in the right direction towards the realisation of 5G, some researchers are doubtful about its operational status by 2020. In fact, there are currently concerns that there will be several 5G standards [35] hence the risk of interoperability problems. This is due to the fact that various bodies are currently developing their variant of the technology and are rushing towards its commercialisation to the extent of making the choice on the 5G vendors [35-37]. Others, operators mainly, seem rather satisfied with the use of LTE technology beyond 2020, as LTE-A is still in the rolling out phase and it is highly unlikely that vendors are going to jeopardise LTE-A deployment with any signs of network replacements for 5G. However, and since 5G is still undefined technically, while LTE-A is still being deployed, it is more than likely that the 2020 time target will be for elements of the 5G to start being gradually adopted by operators, after the 5G standardisation battle comes to an end. This will mean that a full 5G network will eventually become operational by 2030. Figure 2.1 illustrates the forecast timeline in detail. At this stage, researchers worldwide have a clearer view on the objectives for 5G even though they seem rather discordant regarding the manner in which these goals will be achieved due to the vast pool of potential technological enablers.

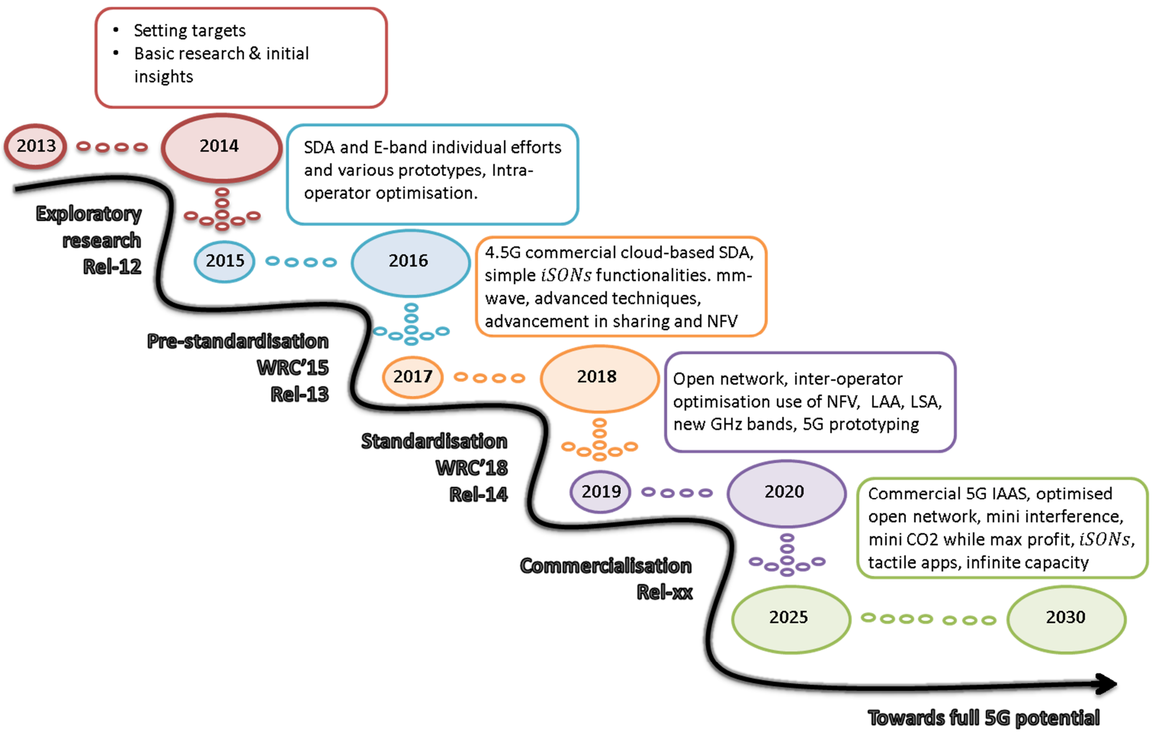


Figure 2.1 Estimated time line towards 5G

2.3 **Energy-Aware Topology Management**

Energy-aware topology management schemes are considered as an important enabler for green 5G systems, enabling the intelligent deactivation of network nodes partially or fully while delivering the required QoS. These are highly important in the case of specific 5G Ultra-Dense Network (UDN) deployments, as fast activation and deactivation of cells are required. Energy-aware topology management aims to dynamically control the predesigned topology of the system in order to achieve energy savings taking into consideration the level of QoS required. This is achieved by fine-tuning the status of hardware components (active / dormant) depending on the traffic demands. It is important to point out the difference between sleep-modes and topology management. A sleep-mode is an intermediate state in which a given network entity minimises its activity by switching some hardware components off in order to save energy while keeping only the essential circuitry operational acting as a state controller. The state controller must be kept operational at all times as it controls the transit from / to the sleep-mode. This term however is vaguely defined in the literature and used interchangeably to sometimes mean switching of eNBs and others to indicate the usage of traffic redistribution with the aim to switch free eNBs to the dormant mode. Topology Management (TM), on the other hand, deals with the control that helps the state controller determine when and what network entity needs to enter or leave sleep-mode, and it can be either distributed or centralised. Advanced topology management consists of three essential management units. These are the Topology Management Information Unit ($TMIU$), the Topology Management Control Unit ($TMCU$), and the Topology Management Sleep-mode Unit ($TMSU$). These are illustrated in Figure 2.2.

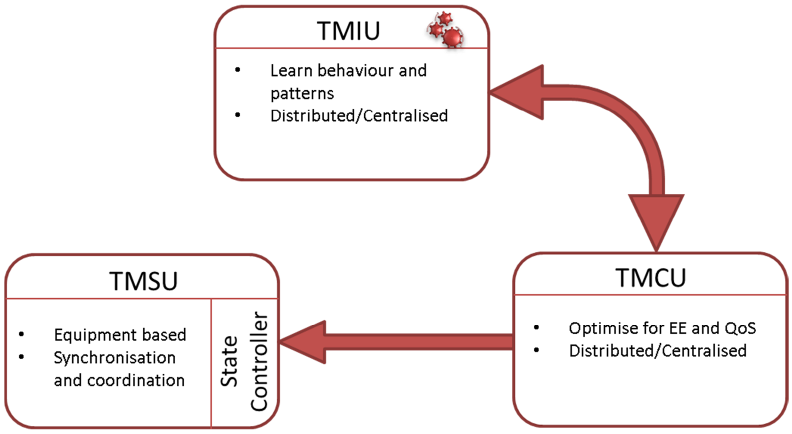


Figure 2.2 Intelligent Topology Management Mechanism

2.3.1 **Topology Management building blocks**

2.3.1.1 *TM Sleep-mode Unit (TMSU):*

The $TMSU$ is responsible for internal equipment coordination in which the functional dependencies between hardware components is of paramount importance. The $TMSU$ decides what component needs to switch off (on) and in what order. This operation is optimised, minimising the transit time required for the equipment to be fully operational, while reducing the power requirements while in the sleep-mode. In current equipment, hardware dependencies are large (i.e., a component cannot be activated until a series of other components are functioning), requiring a long transit time. This is mainly due to the fact that equipment is being developed by manufacturers with little attention currently placed on the importance of flexible equipment functions. It is worth mentioning that there is not an optimal sleep-mode for a given equipment, however there are various depending on the Topology Management Control Units ($TMCU$). As will be shortly described, different $TMCUs$ require different state controller interfaces to communicate with the equipment, hence different circuitry is needed to stay always operational. For this reason current hardware development is undertaken to allow for a more flexible operation [38].

2.3.1.2 *TM Control Unit (TMCU):*

The $TMCU$ is the unit responsible for deciding on when and what enters a certain state. In other words, it is the agent which communicates with the equipment state controller of its choice, given certain rules in order to suggest the equipment to change status (i.e., from / to a sleep-mode). There are a number of agents located at different levels capable of performing such function:

1- ***Self-Aware Control agent (SAC):*** in this case the control agent is embedded into the equipment which is made aware of its surroundings via the TM Information Unit ($TMIU$), providing regular updates on the environment, meaning that it does not require the aid of an external agent in order to change status. The performance of the TM in this case highly depends on the goodness of the information provided by the $TMIU$. The advantage is that it enables the TM Sleep-mode Unit ($TMSU$) to choose the least energy consuming mode boosting its energy efficiency. On the other hand, it becomes highly susceptible to inaccurate information and unpredictable traffic fluctuations which might cause serious degradation of the QoS over the area.

2- ***Backhaul-Based Control agent (BBC):*** this agent is located in a centralised control entity such as a Mobility Management Entity (MME) which controls the operation of the TM coordinating between several network entities. The agent sends TM control signals through the $S1$ link to the equipment state controller. Even though the QoS could be guaranteed given its centralised nature, energy savings might be limited since the equipment needs to maintain the backhaul circuitry operational to listen to control signals.

3- ***Neighbour-Based Control agent (NBC):*** this agent is distributed among all neighbouring nodes and uses the $X2$ interface to coordinate them and send TM-

related control signals. The use of this setting provides a middle ground between a distributed and centralised operation and a trade-off can be found between energy-efficiency and QoS, taking into consideration the amount of information exchange required. The use of this agent can be interpreted as a help signal sent from a neighbouring node which is struggling to keep up with traffic demands and is looking to offload some traffic if possible. However, it might be challenging for the TMInformation Unit ($TMIU$) to keep all nodes up to date, particularly on the information regarding which neighbour sends a control signal to when it is generated (i.e, which neghbour to reactivate).

4- ***User-Based Control agent (UBC):*** this is probably one of the most appealing ways of generating TM control signals as there is no information needed to be exchanged. It also could guarantee the QoS, as the control agent is based at the user equipment (UE) and could request service if this is denied by other nodes. However, the $TMSU$ energy consumption is relatively high compared to previous settings, as it needs to maintain the air interface operational to listen to UE control signals constantly. In this case, the TM Sleep-mode Unit ($TMSU$) should be optimised to minimise the time required to transit to the operational mode and serve the UE, which is another factor that rends this agent the most energy inefficient among the rest.

2.3.1.3 *TM Information Unit (TMIU):*

This is the cognitive unit of the TM which observes its operation and makes use of different learning algorithms to predict future traffic behaviour and potential consequences of certain actions. This unit is considered to be the most important as the optimal operation of the TM highly depends on the goodness of the information extracted from past experience, which is provided to the $TMCU$ for decision making. The $TMIU$ can be made centralised or distributed depending on the network requirements and capabilities.

2.3.2 **Theoretically Derived Approaches**

Starting from the theoretically derived schemes, in [39] the authors modelled the flow-level dynamics of a node within the access network as an M/G/1 processor-sharing model. Two basic strategies are compared, the first assumes that the BS is asleep when there are no users to be served then it systematically switches to the active mode as a user enters service, which could be considered as a User-Based Control scheme. For the second strategy, the authors extended the Markov chain to a 2-Dimensional chain, as the BS is assumed to be controlled by a neighbouring BS. In both cases, the authors strive to optimise the energy efficiency taking into consideration the queue length of files in the system, and derive the optimal triggering threshold. The authors in [40] provided a theoretical framework to calculate the potential energy savings by using sleep-modes and bandwidth (BW) variation (using the minimum BW possible, while ensuring user QoS requirements are satisfied). The framework does not take into consideration traffic variation and is assumed to use a Backhaul-Based Control Agent. The results show that the use of sleep-mode is more favourable in terms of energy efficiency than exploiting BW variation. Another Backhaul-Based Control approach is taken in [41] where the authors use offline optimisation to derive a sleep-mode mechanism taking into account user throughput, the transit delay, and the reactivation frequency of the eNBs associated with load-based schemes. The authors analysed one BS having multiple carriers and these carriers, which they call resources, are adaptively activated depending on traffic conditions. In fact, even though the transit delay and reactivation frequency of eNBs are issues to consider, they are theoretically negligible when de-(activating) the carriers as the transit time is too low, in the order of micro seconds, compared to the user interarrival rate on which load-based schemes depend on for their operation. In [42] however, a HetNet network is taken into account initially deriving an optimal policy taking capacity requirements into consideration to deactivate the entire underlay small-cell network when it is considered to be redundant. The authors then extend the results to have an independent policy for each small-cell to model more spatially distribute traffic. In [43] on the other hand, a two-threshold control scheme is evaluated considering both activation and deactivation of nodes unlike the aforementioned schemes which only consider the transit to the dormant-mode. The authors consider a HetNet which uses a scheme based on two distinct load-based thresholds for activation and deactivation of the small cells. These thresholds are firstly set to be the same for all small cells and optimised using exhaustive searching. Afterwards, a heuristic scheme is compared with a Reinforcement Learning approach to study the effect of having different thresholds for each cell to optimise for cell-edge performance. The scheme assumes the use of Neighbour-Based Control Agent in which the macro BS is responsible for the computation of the various thresholds and then communicates these to the underlying small-cell network via the $X2$ interface. The performance favours the use of heterogeneous thresholds for the small-cells network. All cases so far agree on the fact that the use of sleep-modes is beneficial for the network energy efficiency under different network configurations. For the sake of completion, it is worth mentioning the efforts in [44] as the energy consumption is optimised while ensuring coverage using a random and a strategic approach. In the random approach, the decision is based on Bernoulli trials such that a BS continues to operate with probability q, which is set heuristically, and is turned off with probability $1 - q$, independently. In the strategic approach, q is calculated using the utilisation of the each BS. Given the challenge and inherent dynamics of sleep-modes, it is hard to theoretically capture all the aspects in terms of requirements, advantages and drawbacks within the same theoretical framework. A more complete analysis can be achieved through simulation.

2.3.3 **Neighbour-Based Control schemes (NBC)**

Starting from [45], the authors proposed an NBC scheme based on information exchange and coordination between cells controlled by the same Baseband Unit (BBU) reflecting a centralised BBU deployment which is a potential scenario for 5G [46]. Each cell computes the likelihood of being active and communicates this information to the rest. Consequently, each cell calculates a response depending on the information received using the Sum Product Algorithm (SPA). This is done for a fixed number of iterations after which a final decision is made as which cell is deactivated. The algorithm is tested for $E-UTRAN$ access network composed of 7 BSs each having 3 cells reducing the energy consumption by up to 36% in the situation when there is no traffic. However, no indication of the consequences of applying such a scheme in terms of QoS was presented nor the amount of information that need to be exchanged which seems to be large depending on the number of iterations required for the algorithm to converge. [47] proposes a technique for SONs that uses Delaunay Triangulation (DT) to keep track on the network topology formed by the BSs which are kept operational and the BSs which are within n hops away from each other form a cluster. BSs within a cluster exchange load related information via the $X2$ interface and the lowest utilised BS is switched to the dormant mode if it can redistribute its current load to cluster members. In [48] a fixed time interval optimisation process is also proposed for green SONs, which calculates the number of BSs required to be kept operational depending on traffic demands. The time interval is set to an hour, on which the number of BSs required is recalculated with their respective transmission power to compensate for the cell outage of the dormant cells. An interesting approach is taken by [49], which uses the blocking probability as feedback to indicate the need to modify the topology. When the BP of a given region is above (below) a predefined threshold, a BS is activated (deactivated). A similar approach is found in [50] to control the topology of a disaster relief network. A preventive technique is seen in [51] using an exponential smoothing strategy for traffic prediction, the Holt-Winter (H-W) algorithm in particular, to forecast traffic loads and proactively act on any change in demand. A day is split into 24 time slots at which the scheme takes into account the outcome of the H-W algorithm and decides the output power of the macro BS and the number of small cells needed. In [52], the authors try to make the most of latest advancements in cellular networks by using the $CoMP$ transmission alternative to power control to compensate for the deactivation of certain underutilised nodes. The usage of $CoMP$ transmission is proven to be advantageous in terms of energy as well as spectral efficiency. The authors also highlighted the feasibility of such scheme in terms of sensitivity to channel estimation errors which can lead to inefficient use of resources. In [53], the authors monitored the states and activity of neighbouring eNBs to control the topology of the systems. eNBs are switched to the dormant mode if their load and the load of neighbouring nodes is low over a period of time.

2.3.4 **Backhaul-Based Control schemes (BBC)**

Commonly, centralised approaches tend to be less popular compared to their distributed counterparts. However, providing careful consideration to future directions in terms of potential RAN architectures and assumed capabilities of 5G networks and beyond, it is believed that centralised schemes will be gaining ground in terms of attractiveness [54, 55]. Nonetheless, a number of Backhaul-Based Control schemes have been proposed, which are quite logically driven by the utilisation of the nodes in a given area [56, 57]. [56] is a good approach towards green SONs, as the energy efficiency is centrally optimised by switching the underutilised BSs into sleep mode, while maintaining a minimum set of requirements namely user throughput and coverage. Prior to switching to the sleep mode, users are associated with BSs that could offer the highest throughput rates. This can only lead to a suboptimal solution from the energy efficiency point of view as further discussed in the subsequent section. A similar approach aimed at SONs as well is taken in [57] in which the coverage area is optimised by putting to sleep as many BSs as possible while guaranteeing coverage.

2.3.5 **Self-Aware Control scheme (SAC)**

So far, all the introduced schemes are either based on a Neighbour-Based or a Backhaul-Based controller for feasibility. From an initiative perspective and taking into consideration the wide pool of scenarios for 5G, even though centralised approaches might be more feasible than ever, effective self-aware and fully-distributed algorithms need to be further developed taking into consideration possible network capabilities. In fact, it is of paramount importance for future deployments to have self-aware capabilities to empower distributed decision making to meet delay constrain for certain set of applications such as tactile internet applications [6] [58]. An appealing approach of what can be introduced as a Self-Aware based Control scheme is proposed in [59] where an ultra-small cells deployment is taken into consideration and each cell monitors a circular area, having an adaptive radius depending on the nearest active node, and a predefined maximum monitoring radius. eNBs are made aware of any activity within its monitoring area by means of sensing. The eNBs then transit between the active and dormant mode depending on the number of active links and fulfilling a minimum inactivity period before switching to the dormant-mode. A similar self-aware approach is taken by [60] proposing to minimise the number of sectors in use to maximise energy efficiency in SONs. The algorithm is distributed and each BS senses its environment and sets the number of sectors necessary to meet the estimated number of RBs required and a target BP. Also in [61], a self-aware algorithm is introduced as small cells react to users' activity in the vicinities and alternate between the active and dormant mode accordingly having a macro-cell overlay to ensure coverage. An early effort was conducted in [53] setting a range of possible rules to control the topology of the systems depending on the degree of centralisation and neighbouring activity. Firstly, the authors tested a Neighbour-Based Control scheme varying the load threshold of the node in question and its surrounding nodes, as well as the number of neighbouring nodes to consider before switching to the dormant mode. Secondly, a Self-Aware Control scheme was proposed in which a number of coverage holes resulted as a consequence degrading the system QoS at low loads.

As part of our work, the authors opted to develop a range of low complexity, Neighbouring-Based Control mechanisms aiming at delivering the required QoS and achieving the most energy savings possible. Backhaul-based approaches are in most cases centralised and even though it is feasible to apply techniques based on this approach in some architectures such as cloud-based deployments, the amount of information that needs to be exchanged as well as the single point of failure makes these less attractive. Self-Aware based approaches on the other hand, are of great interest but the least efficient in terms of energy as BSs implementing these techniques are required to have more circuitry active such as a sniffer to listen to potential reactivation signals from UEs. The authors hence studied a potential 5G architecture [62] for a range of feasible yet effective Neighbour-Based TM schemes [63-65]. The investigation of the application of topology management mechanisms for this particular type of architecture is important as it largely differs from systems studied in the literature as detailed in Chapter 4. Apart from the architectural difference of the investigated network, some previously proposed techniques suggest an hourly estimation of traffic or base their decisions on predicted traffic behaviour. Both approaches, particularly the latter, are of questionable effectiveness in areas with medium to high traffic fluctuations as localised traffic behaviour is challenging to capture accurately. Alternative measures will need to be taken if actual traffic behaviour does not match the prediction. Other proposed techniques either require a minimum number of information exchange iterations between neighbouring cells, with further status estimation calculations and information processing, or require keeping track of the network topology formed by BSs that are operational. A further set of schemes is based on QoS feedback from the network such as the blocking probability. First of all, the calculation of an accurate estimation of the network QoS requires a minimum period of time which can affect the response time of the schemes to medium and high traffic fluctuation. This is more challenging when estimating regional QoS instead of the network QoS. Also, the estimated QoS is not reliable as in the case of BP for instance, the network, or the BS, is not aware of all the UEs that are blocked if blockage happens in the MAC layer when UEs are contesting for access. For this reason, the schemes proposed as part of the work presented in this thesis differ from the ones proposed in the literature in the following:

1- All schemes are designed to be compatible with the type of information exchanged as part of existing X2 procedures in the standards, hence not imposing further burden on the X2 link.
2- Instantaneous traffic measurements are used to derive different policies hence the schemes can adapt to low, medium, and high localised traffic fluctuations with no prior information.
3- The schemes are designed to work in a range of different architectures and network deployment strategies, hence no need to track the network topology formed by BSs that are operational.
4- The schemes do not require any QoS feedback such as the blocking probability for efficient and reliable operation.

An early effort was conducted as detailed in [63] setting load-based rules to control the topology of the systems. The rules are based on the instantaneous loading status of eNBs to meet the required response time to regional traffic fluctuation. The authors proposed a Neighbour-Based Control scheme varying the load-based threshold of the node in question to reactivate a dormant neighbouring node. The authors also introduced a minimum required service time between two consecutive dormant periods as a function of the observed user arrival rate, and the load-based threshold to minimise the number of reactivations. A relative deterioration in the QoS at very low loads was noted which is caused by the aggressive mechanism pushing nodes to the dormant mode creating a number of coverage holes. Hence, the authors pursue further in [64] proposing a modification by introducing a macro-cell overlay tier on top of the original architecture to mitigate the QoS problem at low loads and render the system more energy efficient by enabling the complete deactivation of the overlaid small-cells tier at low loads. These schemes are detailed in Chapter 4. A macro-overlay is shown to be essential in making the system more energy efficient and preserving the QoS as proven in Chapter 3 using a theoretical framework based on queueing theory.

Last but not least, illustrating the importance of the TM Information Unit (TMIU) in [65] we took into consideration a temporary event scenario such as an open air festival. To the best of the authors' knowledge, the work presented in [65], which is detailed in Chapter 5, is the first study of energy-aware topology management in the context of unexpected temporary events. Previous work suggests the clustering of eNBs that are no more than n hops away from each other. However, the architectural differences as well as the spatially unbalanced traffic densities require better clustering techniques. Hence, the authors developed a Neighbour-Based Control scheme to neutralise the high traffic dynamics expected during a special event by clustering entities which can cooperate to enhance the energy efficiency of a given area and subsequently fine-tune the dormant / active states of cluster members depending on localised traffic demands. In other words, a cluster is formed by eNBs that can cooperate in covering a given area and enhance the regional energy efficiency. As a result, eNBs might be few meters apart but can belong to different clusters. Chapter 5 details the scheme used and most relevant outcomes. It was primarily concluded that the management of Ultra-Dense Network (UDN) deployments require special attention particularly when the network suffers from a highly unbalanced traffic distribution as traditional schemes fail to operate under these conditions, resulting in low QoS and high energy consumption. Hence, more flexible and intelligent schemes need to be developed to fit a wide range of scenarios. This result emphasises the importance of developing efficient Topology Management Information Units (TMIU) within a network in order to optimise the performance of the Topology Management Control Unit (TMCU) for a wide range of scenarios and traffic conditions using advanced techniques and incorporating Artificial Intelligence (AI). Even though the advancements in this area are quite modest, a range of intelligent algorithms have been proposed. In [66] the authors combined reinforcement learning and case-based reasoning in order to enable learning in dynamic environments with several identifiable phases improving the temporal performance and avoiding performance degradation during the transition period from one phase to another. An interesting approach was also taken in [67] suggesting that more expert network entities could help a newly deployed node to learn its duties more efficiently. This is done by integrating traditional reinforcement learning with Transfer Learning (TL) to neutralise the performance degradation in early stages of a newly deployed node due to the trial-and-error fashion of the learning mechanism by minimising its convergence speed. The scenario is well fitted to reflect the rapidly changing environment and traffic fluctuation of future deployments. The technique is based on clustering source agents (expert nodes) which could help a given target agent (training node) to initialise its knowledge base. Even though the use of intelligent techniques and reasoning strategies seem appealing, the area requires further efforts before achieving an effective real time and fast adaptable knowledge base to enable autonomous SON operation. To mention a few of the challenges, for instance, accurate identification of different periods and mapping the right piece of information to the right period in case-based reasoning and, generally, knowing when a piece of information expires as well as an effective update formula for it. These are particularly challenging in a high dynamic rapidly changing environments. Transfer Learning is also not trivial as tight rules need to be set to decide, first of all, what node can be considered as an "expert node", consequently identify what information needs to be transferred and how this information is processed to render it useful to the target node. The use of AI to achieve the aforementioned goals solving some of the mentioned challenges is detailed in Chapter 6.

Radio Resource Management (RRM) is another important aspect which most literature fails to cover when studying topology management schemes. As previously mentioned, RRM is of paramount importance in developing effective yet green systems, for example by using advanced traffic management. Traffic management has a direct influence on the performance of most TM schemes and it needs to be taken into account to effectively measure their performance. The study of a range of green RRM schemes and their effect on TM performance are studied in the subsequent section. Chapter 6 also details and solves interoperability issues between these two approaches using artificial intelligence to enhance their standalone as well as joint operation.

2.4 Energy-Aware Radio Resource Management

Radio Resource Management techniques and protocols have always been a topic of debate and constant development given the rapidly changing system requirements, as well as users' demand. 5G is no exception having to accommodate a vast range of applications for an increasing variety of connected devices while taking additional design constraints into account. One of these constraints is power consumption and hence schemes currently underway need to incorporate more than just system capacity and required QoS, but also the manner in which these services are offered. To this end, innovative solutions are required with different mind-sets towards more energy efficient use of radio resources. Traditionally, Radio Resource Management (RRM) mechanisms are optimised to balance the load in order to achieve higher capacity and enhance the QoS [68-73]. However, taking into consideration the potential usage of green SON techniques as well as the increasing usage of renewable energy, advanced RRM techniques adaptively strive to 1) optimise energy efficiency by minimising the number of transmission components in use at low traffic loads, 2) optimise energy efficiency by minimising the number of transmission components in use at medium traffic loads given the capacity and QoS constrains, and 3) optimise capacity and QoS enhancements at high traffic loads. The split into three operational regions is evident. As shown in Figure 2.3, advanced RRM techniques need to have a cohesive approach that concurrently optimises both energy consumption and QoS in the three operational regions in a seamless adaptive way.

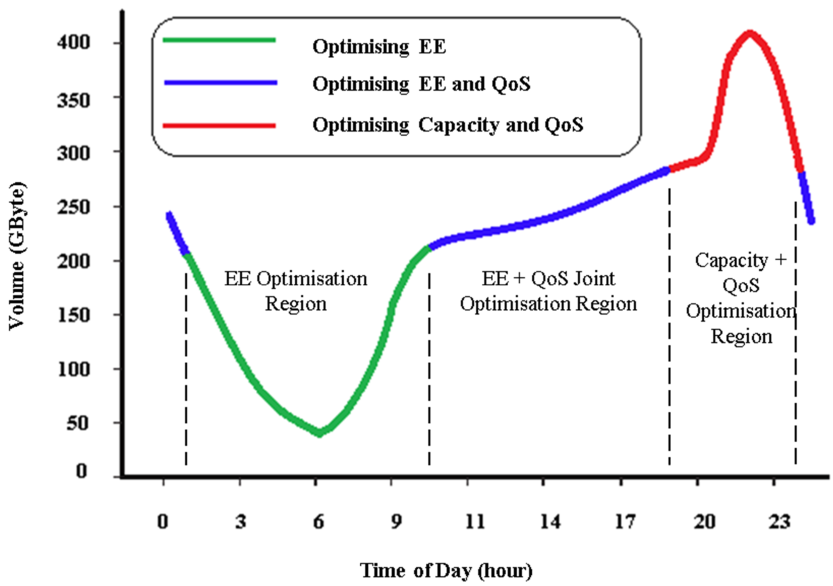


Figure 2.3 Traffic variations throughout a day [56, 74]

The limiting of the number of active components is done to maximise the idle time and hence maximise the network energy efficiency by means of deactivation. However, conventional traffic management tries to spread the demand over as many cells as possible to provide an even traffic distribution over the whole network. This is energy consuming as it requires a large number of nodes to be active and engaged in transmission. Figure 2.4 illustrates the operation of energy-efficient $RRMs$ at different traffic load levels. Normally, the red-coloured users in low and medium traffic loads would camp on one of the idle nodes as the $RSRP$ and $RSRQ$ values are higher due to their proximity. However, advanced $RRMs$ intend to keep as many nodes in the idle mode and for as long as possible if the QoS requirement of all users can be met by the active nodes taking into consideration the presence of any base station running on renewable energy and also the efficiency of all nodes. At high traffic loads on the other hand, the load is balanced optimising for capacity and QoS enhancement.

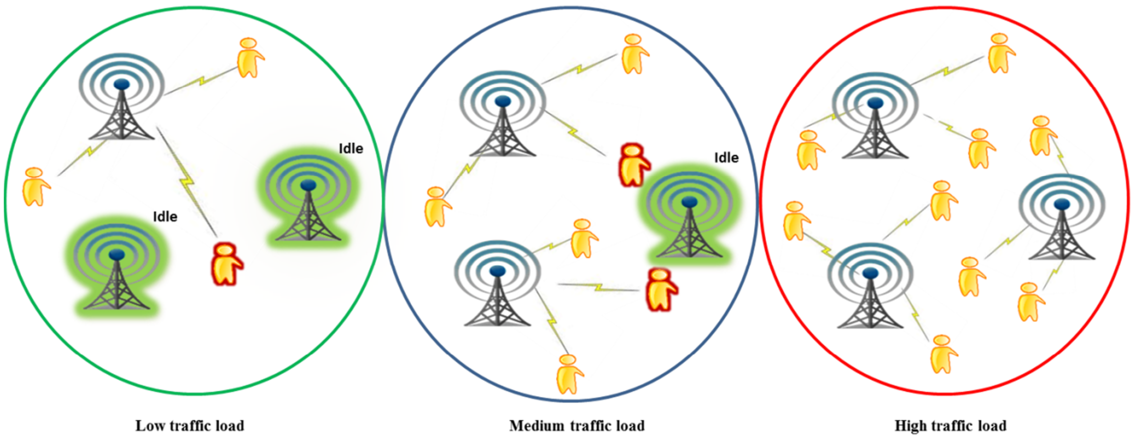


Figure 2.4 Advanced Radio Resource Management operation at different traffic conditions

For these reasons, load balancing is no longer seen as the most efficient manner in which the traffic should be managed in all traffic conditions and a range of advanced energy-aware RRM techniques are being developed. These techniques can be divided into Reactive and Proactive Load Management schemes, RLM and PLM respectively. Figure 2.5 illustrates the proposed classification of load management schemes.

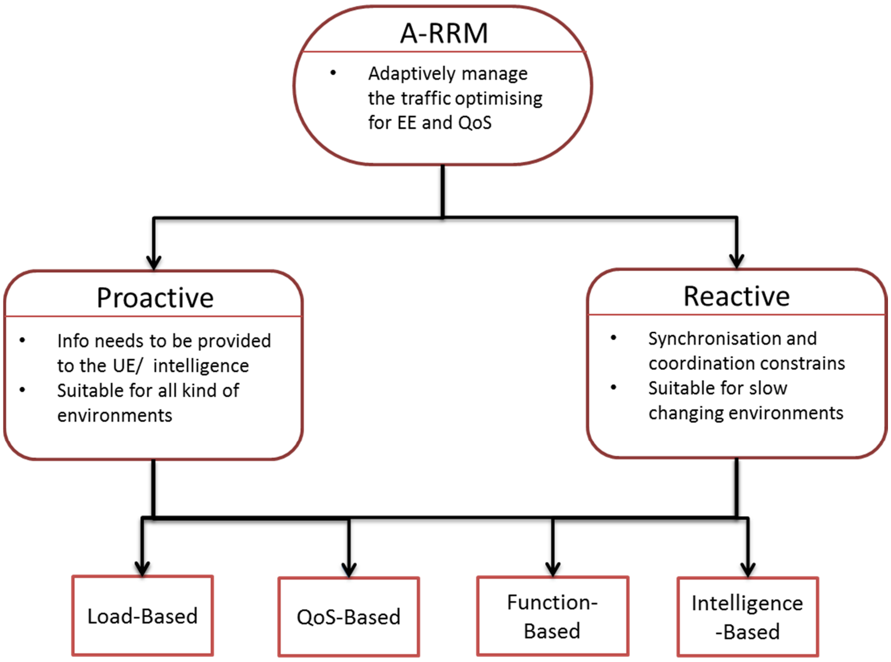


Figure 2.5 Proposed classification of advanced load management schemes

2.4.1 **Reactive Load Management (RLM):**

RLM techniques are initiated by the network in a reactive manner to the incoming traffic in order to redistribute the demand following its acceptance. In most cases, RLM schemes require information to be exchanged between neighbouring nodes which limits the frequency at which they operate rending it inefficient in highly dynamic scenarios. In other words, RLM schemes can be considered as a cell reselection mechanism in which cells are primarily selected according to the Radio Resource Control RRC protocol in place, and then the RLM scheme will be used for the reselection process. RLM mechanisms are usually developed to tailor-fit and boost the performance of a certain energy saving mechanism such as sleep-modes and cell breathing, also known as cell zooming. From a green wireless system perspective, cell breathing and cell zooming are essentially identical in definition despite the difference in terminology. In fact, the term cell zooming has emerged only recently introduced by Zhisheng Niu et al in [49]. Cell breathing on the other hand, despite its restrictive usage to network capacity optimisation, can be traced well back before the cell zooming terminology was first used [75]. In other words, cell zooming can be considered as a green cell breathing technique.

2.4.1.1 *Load-Based schemes:*

A well detailed load-based scheme is proposed in [76], which is a continuation of the work presented in [77] and [78]. The authors relied on the inter-BS cooperation capabilities of emerging deployments to develop a distributed load management scheme. The algorithm is essentially based on the loading status, having three different loading thresholds namely; lower threshold L_f, upper threshold H_f, and acceptable threshold A_f related as $A_f > H_f \geq L_f > 0$. The algorithm is assumed to run consecutively at each node every T_S time units. The periodicity and the sequence of nodes at which the redistribution is carried out are left to the operator to optimise. The operation of the scheme depends on the aggregate traffic of the network denoted as $\tilde{\rho}(t)$ and is divided into low and high traffic periods. In low traffic periods when $\tilde{\rho}(t) < L_f$, the algorithm tries to identify potential acceptors for the traffic of a given node without the necessity to activate any sleeping node. If acceptors are identified, the load is redistributed and the node switches to the dormant mode. During high traffic periods when $\tilde{\rho}(t) > H_f$ and redistribution of traffic among active nodes is not possible, so a partial redistribution is considered in order to achieve better load balancing between active nodes. In the case where the node is incapable of sharing its load, it extends its searching to include dormant nodes. In this case, if a suitable set of acceptors is found, the load is redistributed and the node switches to the dormant state whereas if no acceptors are found, the node strives to share at least a part of its load and activates the target node(s). In all cases, the traffic is redistributed taking into consideration the required QoS, monitoring the blocking probability and recalculating the transmission power of active nodes to ensure coverage. A cluster based approach can be seen in [79] and [80] where a node in a cluster sequentially evaluate the possibility of offloading traffic to one or more neighbouring nodes which might belong to a different cluster. Clusters are formed to synchronise and monitor the sequential execution of the scheme and avoid potential conflicts between nodes with common neighbours. Cluster heads are responsible for informing adjacent clusters of the termination of their redistribution activity. In [79], the execution evaluation of the scheme requires load related information to be exchanged between neighbours. If the node is not the most loaded among its direct neighbours and its entire load can be offloaded, it chooses the highest loaded neighbour first as the target node, followed by the second most loaded if necessary and so on. If a user is not accepted by any of the target BSs, the initial association of all users is maintained. This means that a user who has been previously accepted by a target node will have to be re-associated again to their initial state. The algorithm was further enhanced in [79] by the introduction of a capacity threshold limiting the number of nodes that could perform the redistribution algorithm to those that have a capacity lower than the threshold. The aforementioned two algorithms are jointly used in [80] to suit a HetNet having a combination of open and closed femto cells by firstly targeting neighbouring small cells and then resourcing to neighbouring macro cells. Macro cells are firstly avoided to prevent them from increasing the downlink transmission power which is necessary to accommodate new users.

2.4.1.2 *QoS-Based schemes:*

A QoS-energy efficiency trade-off user association scheme is presented in [81] by taking into consideration the user throughput and the loading status of the base stations. An optimisation problem is formulated by forming a network weight matrix as a function of the user normalised data rate and BS normalised load which is solved iteratively. Users are firstly assigned to their best serving BS depending on the SINR value. Subsequently the network weight matrix is revaluated assuming the redistribution of a given user to a neighbouring BS. The BS which maximises the weight matrix is eventually chosen to serve the user. The process is repeated for all active users. A similar approach is taken in [82] where the user association is optimised taking into account users' requirements in terms of per-bit delay. The authors modelled the user perceived performance as a function of the density of users for three different deployment strategies namely; hexagonal, Manhattan, and Poisson layouts. The result is later used to optimise the user association in order to minimise the required density of BSs in the access network.

2.4.1.3 *Decision Function Based schemes:*

A centralised approach is taken in [83] where the load is redistributed periodically among the access network by a central entity which optimally determines the minimum number of nodes that can be used to serve the offered traffic. The authors formulated the total energy consumption as a function of the number of nodes required, traffic load and cell size. In [84] authors developed a user association mechanism taking into consideration both spectral as well as energy efficiency of the system. The objective function is a division of the system spectral efficiency, which is calculated using the amount of traffic that could be carried by a single subcarrier in the network, and the network energy consumption taking into consideration user throughput and coverage availability. Firstly all users are associated with their best choice of base station depending on the spectrum-energy efficiency values. Consequently, the spectrum-energy efficiency for individual nodes are evaluated and used to sort BSs in descending order. At this stage, starting from the least efficient BS, the algorithm tries to reassign users to a more efficient BS. BSs are then rearranged and the least efficient BS is checked again and users reassigned. This procedure continues until the best solution is reached.

2.4.1.4 *Intelligence-Based schemes:*

The authors in [85] based their scheme on the prediction whether the application of the load management algorithm would increase the number of nodes that can be switched to the dormant mode before performing the redistribution. The scheme works in a centralised manner in which nodes are arranged in an ascending order based on their loading status and these are checked in succession to verify that the current loading state have been below a predefined threshold for a predetermined period of time. The capacity threshold is a trained value taking into consideration previous thresholds and their outcome, however there is no detail on how long this value needs to be trained for before delivering acceptable results. If the aforementioned conditions are met, the load is redistributed among its direct neighbouring nodes. Renewable energy was taken into account in [86] as authors studied a heterogeneous network in which a number of nodes use some degree of renewable energy to operate. The authors propose two approaches to guide more users towards base stations which are powered by green energy sources. Firstly, BSs are identified by what they called the Green degree (Gd) which represents the ratio of green energy that is consumed by the BS to its total consumption. The Gd value can be updated according to the amount of green energy produced. Taking a solar power supply for example, the Gd value is set to 100% in a sunny day, and it is recalculated to be only 10% or lower depending on weather conditions and time of the day. The first suggested approach is the usage of the Gd to tune the cell specific offset of the green cell (Ocn), which is used for handover decision, in order to encourage users to handover to green cells. The second approach suggests the usage of the Gd for power control having the green BS increasing its Downlink (DL) transmission power extending its coverage to accommodate more users. This action is followed by the neighbouring BSs adjusting their transmission power accordingly to avoid interference. A similar approach is seen in [87] as the authors also point out the importance of unbalancing the traffic in a scenario having heterogeneous energy sources. The algorithm proposes periodically adjusting the beacon power level to manage the traffic in an energy efficient manner. Similar to [86], the authors introduced a parameter to indicate the instantaneous percentage of green energy in use which they called the Energy Depletion Rate (EDR). The EDR is the ratio of the instantaneous total energy consumption, which depends on the number of users being served, to the allocated green energy and is recalculated periodically. At first, all nodes use their maximal beacon power level and EDR is calculated. BSs with an EDR above a given threshold are grouped and forced to decrease their beacon power so users are offloaded to a more energy efficient nodes. Users are only allowed to switch to more energy efficient nodes therefore the iteration keeps reducing the maximum EDR till the optimal solution is reached.

2.4.2 **Proactive Load Management (PLM):**

PLM schemes on the other hand direct users to the most appropriate cell from the early process of cell selection and no reselection or redistribution is required. Despite the fact that PLM schemes keep information exchange between nodes to its minimum, extra information, such as loading status, needs to be sent to the UEs for the decision making. Unlike RLM, PLM schemes fit a wider range of energy-aware applications and are used regardless of the energy saving mechanism in place.

2.4.2.1 *Load-Based schemes:*

Authors in [88] studied the dynamics of the system when the network traffic is low and suggested a round robin base station access algorithm. The scheme is based on the loading status of the active cells in which less loaded cells are associated with a larger number of users that are in the region of dormant cells. An interesting scenario is presented in [63] taking into consideration a potential 5G architecture having below rooftop nodes with directional antennas. The authors proposed two load management schemes. The first depends on the loading status of nodes in which the most loaded node is chosen after meeting the minimum QoS requirements. A maximum capacity threshold is used for this scheme at which nodes are considered sufficiently loaded and another node should be considered as the target node. The second approach is based on a predefined priority figure which is assigned to each node. Nodes with higher priority have greater chance to be chosen upon meeting the QoS requirements. These priorities have been assigned heuristically suggesting the use of more advanced techniques to identify the priority list.

2.4.2.2 *Decision Function and QoS -Based schemes:*

Two similar approaches are taken in [89], where authors assume prior knowledge of the topology of the network to derive a more sophisticated priority figure for each node. Firstly, a Clustering Capability Rating (CCR) value is calculated using the potential coverage area, location relative to a centralised location, and loading status. The CCR value is normalized to a value between 0 and 1, thus this value is referred to as the Normalized CCR (NCCR). A user connects to and uses resources of a node with the highest NCCR value as long as the call admission threshold is satisfied. Subsequently, a QoS parameter is taken into account to in a QoS-aware NCCR (CQ-CCR) based scheme to trade-off the high clustering and low SINR of the NCCR for lower clustering and higher SINR whenever this is desirable. This is particularly important under high traffic load conditions when interference in the network is intense. The QoS parameter is called the Quality Factor and is an exponential ratio of the perceived SINR between the user and a potential serving node to the highest SINR perceived by the user. The priority figure CQ-CCR is then computed from the product of the NCCR value and the QF for a given node. A node with higher priority has greater chances of being chosen by an incoming user and thus the traffic is concentrated on certain nodes, thereby increasing the idle state probability of the neighbouring nodes. The authors complemented their work in [90] proposing an Interference Aware NCCR (IA-CCR) scheme to mitigate inter-cell interference in addition to energy efficient resource management resulted from associating users to a distant BS. The IA-CCR limits the choice to a certain number of nodes within the ranked list depending on the SINR value perceived by the user. The final choice of BS is the one with the highest NCCR value among the elements of the permissible BSs. This approach ensures the consideration of both energy efficiency and inter-cell interference mitigation in the association decision. A theoretical approach is taken in [91] proposing an admission control mechanism taking into account the dynamic operation of the BSs. The authors formulated the problem as a Semi-Markov Decision Process (SMDP) using linear programming to obtain an optimal solution targeting a predefined blocking probability. The criterion to maximise is chosen to be the bit-per joule-capacity (bits/J) which is defined as $C_J = C/P_t$, where C is the capacity and P_t is the total energy consumed by the BS. An important aspect of future deployments was studied in [92] taking the backhaul network energy consumption into consideration in a dense small cell deployment. The backhaul network is assumed to be hierarchical (switched) in which every eNB is connected at least to one switch and a switch can serve more than one eNB. The association criterion produces a limited active set of eNBs in which the least number of switches are used up to their maximum capacity. The scheme assigns users to active nodes if the SINR is above a certain threshold. If the SINR does not meet the predefined level, a dormant eNB is chosen provided this eNB is served by a switch that is already active receiving traffic from another eNB. Lastly, if the user is still not served, traditional association is performed regardless of the eNB status or the activity of the backhaul network.

As previously mentioned, RRM has a direct influence on the performance of most TM schemes and it needs to be taken into account to effectively measure their performance. Given the fact that advanced RRMs tend to unbalance the load at low traffic levels, they can have a negative effect on the performance of TMs if not properly coupled. This is due to the fact that the RRM keeps loading certain nodes exceeding the capacity threshold at which the TM, load-based TMs in particular, would unnecessarily trigger the activation of a dormant node. There is also the need to develop efficient distributed RRM approaches despite the high performance of centralised techniques as a prior knowledge of the network is assumed making it unfeasible for a large range of scenarios and not easily scalable failing to meet the scalability and design constrains of 5G deployments. In addition, the focus needs to be shifted towards developing proactive approaches as reactive schemes tend to be less stable and energy consuming having to redistribute most of the previously associated users.

Chapter 6 details a proactive RRM mechanism that sets an adequate clustering degree of user onto access nodes depending on the traffic intensity. The scheme is based on artificial intelligence so as to suit a wider range of scenarios and traffic dynamics. However, more study is needed taking into consideration heterogeneous energy sources such as renewable energy as this is becoming increasingly attractive for future deployments. As load unbalancing translates into highly dynamic traffic observed from the backhaul network point of view, backhaul network flexibility and capabilities to deal with traffic fluctuation needs to be carefully studied. Techniques such as Software Defined Networking (SDN) can be considered for flexible resource management and flexible capacity provisioning in the backhaul network to meet the demand of the high traffic fluctuation at the access network. Last but not least, given the computational capabilities of future deployments, more intelligent schemes need to be considered taking into consideration tempo-spatial traffic predictions, user location and mobility dynamics.

Even though the schemes and techniques proposed in the literature can be applicable to particular systems for certain traffic profiles and for specific QoS constraints, there are a number of concerns regarding their performance in a prolonged undetermined period of time and how these schemes would perform outside their designated environment. In other words, a system under current schemes needs to be reconfigured periodically to update certain parameters and QoS requirements. This is reflected in Figure 2.6 by the highlighted areas. There is a clear gap highlighted of intelligence-based proposals that need to be covered in order to provide schemes which will suit a wider range of scenarios and be as self-organised as possible, while achieving optimum performance with minimum information exchange. Hence, a range of proactive distributed and centralised intelligent as well as effective green RRMs and TMs need to be further studied in order to achieve $iSONs$ meeting 5G requirements and beyond.

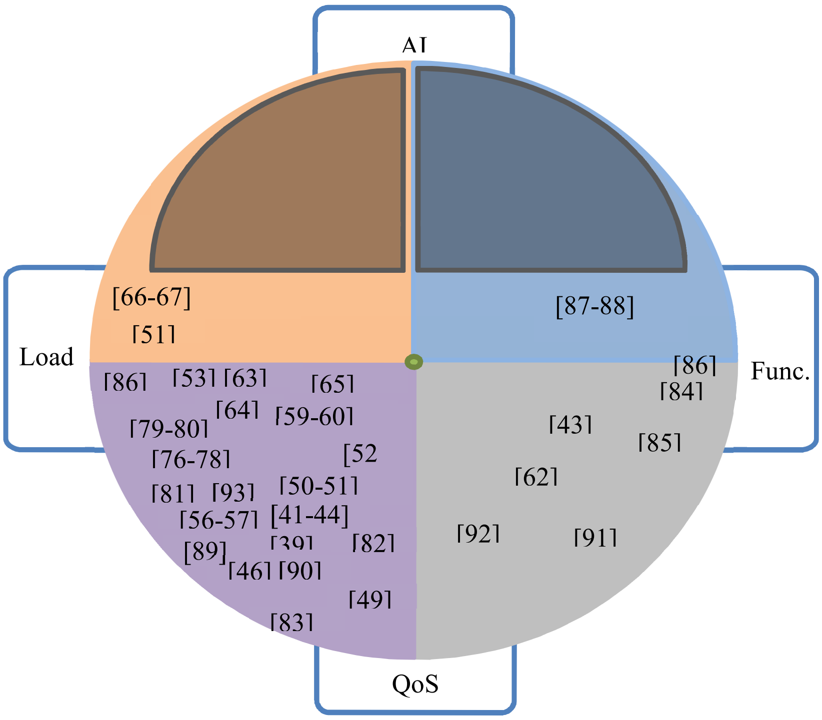


Figure 2.6 Classification Map of the References by Subject Area

2.5 Conclusion

Innovation is proven to dominate a big part of the transition towards 5G where the seamless interoperability between multiple radio access technologies with different standards is inevitable in order to converge towards service oriented heterogeneous networks. Cognition and intelligence are among the most sought after capabilities to achieve these goals, due to their ability to enable real time self-adaptability and self-awareness. In this chapter, enabling technologies towards 5G systems have been presented, highlighting the latest advances. 5G systems are considered to be more than just a wireless network evolution and are aimed at providing perceived unlimited capacity and serve as an interoperability platform for previous cellular network generations and other wireless communication standards. A 5G system is essentially composed of a range of technologies such as next generation cellular broadband networks, WiFi, and WAN cross-optimised for optimal operation appearing as one system to the end user. Essential features need to be adopted to reach the ultimate goal of delivering the services in an energy efficient manner such as energy-efficient self-organisational capabilities. To this end, a detailed state-of-the-art in the field of green topology and radio resource management has been presented, defined, and classified depending on their employed techniques. Despite being early days into the development of green topology and radio resource management mechanisms, an increasing effort is being noticed towards novel approaches to suit 5G requirements. Load management is being studied, aimed at reducing the number of nodes engaged in service. Also, topology management is being introduced as an enabling technology towards intelligent green SONs aiming at dynamically managing the network topology by fine tuning the status of network entities (dormant / active) to scale energy consumption with traffic demands. Unluckily, despite clear dependencies, current proposed topology and radio resource management mechanisms are being addressed separately. A scalability and adaptability issue is also evident across proposed solutions. The main challenge remains the formulation of an intelligent self-adaptable scheme capable of dynamically deactivating redundant nodes and redirecting traffic appropriately across a wide range of network conditions while enhancing or at least meeting the QoS requirements. The rest of the work presented in this thesis is aimed at addressing this challenge by firstly understanding fundamental requirements for the standalone as well as joint operation of radio resource management and topology management techniques. Subsequently, in Chapter 6, artificial intelligence is applied as a mean to develop self-adaptive self-organised techniques suiting a wide range of scenarios.

Chapter 3 A Markovian Study into Energy-Aware Cellular Systems

Contents

3.1 Introduction

This chapter investigates the effect of applying a range of radio resource management mechanism on the use of energy-aware topology management techniques in two of the most common 5G network deployment strategies; namely, homogenous and the heterogeneous deployment strategies. This helps understand the fundamental requirements, and operational limits of these techniques in both deployment strategies. These findings are utilised in subsequent chapters to design feasible techniques towards energy-aware 5G networks. To this end, a thorough evaluation is conducted taking into consideration a wide range of aspects. These are the system QoS, the system energy efficiency, and network stability. The latter is an important design consideration for 5G deployments as it indicates the possible level of complexity and feasibility of the application of topology management mechanisms jointly with Radio Resource Management (RRM) techniques. The RRM techniques evaluated for both deployment strategies are the Load Balancing and the emerging technique of Load Unbalancing. Another important aim of this chapter is to evaluate the importance of the existence of a macro-cell overlay and its effect on the operation of the different RRM techniques. Subsequently, the scenarios at which these deployment strategies perform best are indicated from the energy efficiency point of view. Section 3.2 provides a general as well as a scenario-specific theoretical framework based on queuing theory to validate and investigate the effectiveness of the evaluated techniques. The findings are verified by means of Monte Carlo simulation and presented in Section 3.3. Section 3.4 concludes and summarises the most important results.

3.2 **System Model**

This section introduces the scenarios that are taken into account in this chapter, design parameters, constraints, the mathematical framework as well as the Key Performance Indicators.

3.2.1 **Design parameters and constraints**

As the work presented here assesses a range of RRM solutions under different scenarios and network topologies, design parameters need to be set in such a way that a fair comparison is carried out. The first is the number of accessible eNBs that a given UE can access at any given time. This is denoted as Ψ. This parameter is fixed and equal for all modelled scenarios and network architectures. Ψ can also be thought of as the degree of overlap between neighbouring nodes. The greater the overlap between neighbouring nodes, the more accessible eNBs a given UE has. Another design constraint is the use of cell-wrapping for mitigating the edge effect and converting the finite simulated area into an unbounded surface. Cell-wrapping is essential to limit the number of eNBs required to model the system. Interference is undoubtedly an important design constraint of cellular networks. However, since this study assesses and compares the optimal operation from the system occupation stand point, interference is not taken into account.

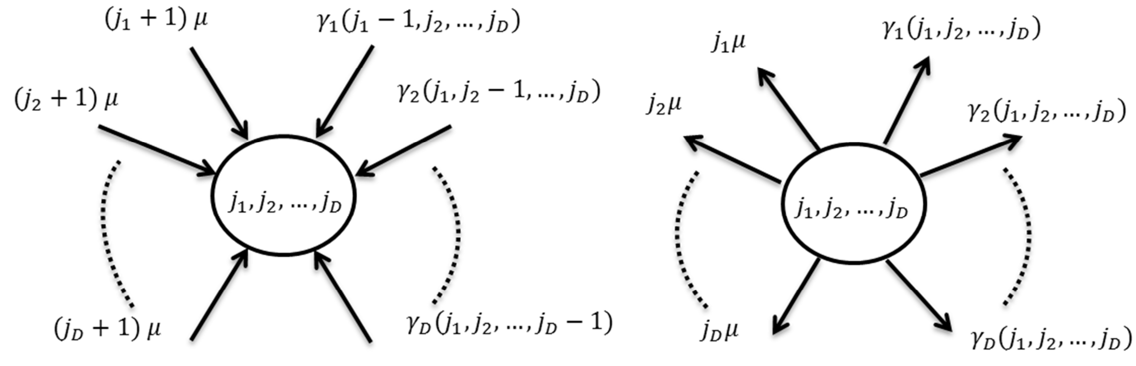
3.2.2 **Traffic density distribution**

In this section we introduce a generic model for different spatial distribution of traffic that is used to model all the RRM schemes assessed in this work. This is also used in Chapter 5 to model a spatially unbalanced traffic distribution representing a temporary event.

For a network with D eNBs, assuming that $\lambda_1, \lambda_2, \dots, \lambda_\mathrm{D}$ $[s^{-1}]$ are the arrival rates for users whose candidate eNBs are $eNB_1, eNB_2, \dots, eNB_\mathrm{D}$ respectively, depending on the design constraints and occupancy of the eNBs. Similarly, λ_{12} is the arrival rate for users capable of accessing eNB_1, and eNB_2, where the order of subscripts in λ_{12} indicates the preferred eNB (i.e., users of λ_{12} only access eNB_2 if eNB_1 is not capable of providing service and vice-versa for λ_{21}). Hence, the total arrival rate for users whose set of candidates is $\{eNB_1,\ eNB_2\}$ is noted as $\lambda_{\{12\}}$ where $\lambda_{\{12\}} = \lambda_{12} + \lambda_{21}$. All other combinations of $\binom{D}{1}$, $\binom{D}{2}$,…,$\binom{D}{D}$ eNBs are calculated similarly, where D is the number of eNBs in the system. Hence, by denoting E as the set of size D representing the eNBs in the network, the overall arrival rate of users into the system (Λ) can be calculated as:

$$\Lambda = \sum_{i=1}^{D} \sum_{k=1}^{\binom{D}{i}} \lambda_{ik} \quad ; \qquad \lambda_{ik} = \{\lambda_{\{E\}} \,|E \subset \{1, \dots, D\}, |E| = i\} \tag{3.1}$$

The system state model of a generic state $(j_1, j_2, \dots, j_D)$ is illustrated in Figure 3.1 where μ is the user departure rate and $\gamma_n(j_1, j_2, \dots, j_D)$ means the fraction of Λ that is directed to eNB_n when the system is in state $(j_1, j_2, \ldots, j_D)$, and j_x is the number of users being served by eNB_x.



(a) Arriving at state $(\boldsymbol{j_1}, \boldsymbol{j_2}, \ldots, \boldsymbol{j_D})$ (b) Departing from state $(\boldsymbol{j_1}, \boldsymbol{j_2}, \ldots, \boldsymbol{j_D})$

Figure 3.1 System state-transition-rate diagram

The values of $\gamma_n(j_1, j_2, \ldots, j_D)$ are dependent on the scenario as well as the deployment strategy, thus a set of non-negative traffic coefficients, denoted by ρ, are defined to compute $\gamma_n(j_1, j_2, \ldots, j_D)$ as follows:

$$\gamma_n(j_1, j_2, \ldots, j_D) = \sum_{i=1}^{D} \sum_{k=1}^{\binom{D}{i}} \rho_{n,\{ik\}}(j_1, j_2, \ldots, j_D)\, \lambda_{ik} \tag{3.2}$$

Where $\rho_{n,\{ik\}}(j_1, j_2, \ldots, j_D)$ is the portion of λ_{ik} that is served by eNB_n when the system is in state $(j_1, j_2, \ldots, j_D)$. λ_{ik} and $\rho_{n,\{ik\}}$ are calculated as follows

$$\lambda_{ik} = \lambda_{\{E\}}\,;\; \rho_{n,\{ik\}} = \bar{\delta}_{j_n m_n} \rho_{n,\{E\}} \mid \; E \subset \{1, \ldots, D\}, |E| = i \tag{3.3}$$

δ_{jm} is the Kronecker's delta which has a value 1 if $j = m$ and 0 otherwise where m is the maximum number of Resource Block Groups (RBGs) available.

In the case when all UEs have an equal number of accessible eNBs, $\gamma_n(J_I)$ can be calculated as follows:

$$\gamma_n(J_I) = \sum_{i=1}^{\binom{D}{\Psi}} \rho_{n,\{E\}}(J_I)\, \lambda_{\{E\}} \mid \quad E \subset \{1, \dots, D\}, |E| = \Psi \tag{3.4}$$

Depending on the RRM scheme in use, the details of the prioritized radio distribution of users differs. The normalised user arrival rates can be defined as:

$$\emptyset_{E_i} = \frac{\lambda_{E_i}}{\lambda_{\{E_i\}}} \quad \mid \quad E \subset \{1, \dots, D\}, |E| = i \tag{3.5}$$

$\emptyset_{E[*]}$ is the normalised user arrival rate which can be also expressed as the probability of user arrival of type $\lambda_{E[*]}$ from zone $\lambda_{\{\, E[*]\,\}}$. Hence, traffic coefficients can be calculated as:

$$\begin{bmatrix} \rho_{1,\{12\}} \\ \rho_{2,\{12\}} \end{bmatrix} = \underbrace{\begin{bmatrix} \rho_{1,12} & \rho_{1,21} \\ \rho_{2,12} & \rho_{2,21} \end{bmatrix}}_{\boldsymbol{A_2}} \begin{bmatrix} \emptyset_{12} \\ \emptyset_{21} \end{bmatrix} \tag{3.6}$$

$$\begin{bmatrix} \rho_{1,\{13\}} \\ \rho_{3,\{13\}} \end{bmatrix} = \underbrace{\begin{bmatrix} \rho_{1,13} & \rho_{1,31} \\ \rho_{3,13} & \rho_{3,31} \end{bmatrix}}_{\boldsymbol{A'_2}} \begin{bmatrix} \emptyset_{13} \\ \emptyset_{31} \end{bmatrix} \tag{3.7}$$

Similarly, $\rho_{n,\{xy\}}$, and $\rho_{n,\{xyz\}}$ are computed where $x, y, z \in \{1, 2, 3\} \mid x \neq y \neq z$ and $\rho_{n,\{xy....D\}}$ is calculated as:

$$\begin{bmatrix} \rho_{1,\{xy..D\}} \\ \rho_{2,\{xy..D\}} \\ \rho_{3,\{xy..D\}} \\ \rho_{4,\{xy..D\}} \\ \cdot \\ \cdot \\ \cdot \\ \rho_{n,\{xy..D\}} \end{bmatrix} = \underbrace{\begin{bmatrix} \rho_{1,xy..D} \\ \rho_{2,xy..D} \\ \rho_{3,xy..D} \\ \rho_{4,xy..D} \\ \cdot \\ \cdot \\ \cdot \\ \rho_{n,xy..D} \end{bmatrix}}_{\boldsymbol{A_D}} \emptyset_{xy....D}; x, y, \dots, D \in \{1, 2, \dots, D\} \mid x \neq y \neq \cdots \neq D \tag{3.8}$$

Where equations (3.5) - (3.8) are a solid way to express the traffic coefficients as a function of matrices $A_2, A'_2, \dots, A_D$. Users that have only one serving eNB are served by that eNB unless it is saturated as expressed in equation (3.3). Matrices $A_2, A'_2, \dots, A_D$ can be computed by taking into account the fact that a certain eNB receives traffic from a given user only when (a) it is not saturated and (b) better candidate $eNBs$ (i.e., candidates with higher priority figure) are saturated. For instance, the probability that a given user with a candidate set of $eNBs$ $\{eNB_1\ eNB_2\ \ eNB_3\}$ is directed to eNB_1 is given by:

$$\rho_{1,\{123\}} = \bar{\delta}_{j_1 m_1}\Big(\emptyset_{123} + \emptyset_{132} + \delta_{j_2 m_2}\emptyset_{213} + \delta_{j_3 m_3}\emptyset_{312} \\ + \delta_{j_2 m_2}\delta_{j_3 m_3}(\emptyset_{231} + \emptyset_{321})\Big) \tag{3.9}$$

Where $\bar{\delta}_{\mathrm{jm}}$ is the inverse of the Kronecker's delta. Kronecker's delta as defined earlier has a value of 1 if $j = m$ and 0 otherwise. Generally, we can write $\rho_{x,xy} = \rho_{x,xyz} = \bar{\delta}_{j_x m_x}$; $\rho_{x,yx} = \rho_{x,yxz} = \bar{\delta}_{j_x m_x}\,\delta_{j_y m_y}$, and $\rho_{x,yzx} = \bar{\delta}_{j_x m_x}\delta_{j_y m_y}\delta_{j_z m_z}$ where $x, y, z \in \{1, 2, 3\}; x \neq y \neq z$. Thus, each column in A_2, $A'_2, \dots, A_D$ sum to unity, and there is a single non-zero element in each column and the other elements are zero. However, when all $eNBs$ are saturated, the matrix is null. The general expression of the steering coefficients for users having M number of accessible eNBs and prioritised as $\{eNB_x, eNB_y \dots eNB_z, eNB_k \dots\dots eNB_M\}$, depending on the scheme in use, for traffic of type $\lambda_{xy,..zk,..M}$ is $\propto_{k,xy\dots zk\dots M} (S_1, \dots \dots . S_D) = \bar{\delta}_{j_n m_n}\delta_{j_x m_x}\delta_{j_y m_y} \dots \dots \delta_{j_z m_z}$. In the case that the distribution of users is uniform and traffic distribution is considered symmetric, then $\emptyset_{xyz\dots M} = \lambda_{xyz\dots\dots M}/\lambda_{\{xyz\dots\dots M\}} = 1/M!$. In this case, the steering coefficients can be calculated as:

$$\rho_{k,\{xy\dots k\dots M\}}(S_1, \dots \dots . S_D) = \frac{\bar{\delta}_{j_x m_x}}{\delta_{j_x m_x} + \delta_{j_y m_y} + \cdots + \delta_{j_k m_k} + \cdots + \delta_{j_M m_M}} \tag{3.10}$$

Expression (3.10) indicates an equal distribution of traffic type $\lambda_{\{xy\dots n\dots M\}}$ among all active and non-saturated candidate $eNBs$.

3.2.3 **Multi-dimensional Markov system**

The system model is based on multidimensional Markov chains. Using queuing theory naming conventions, an eNB can be modelled as a system with no queue having m servers (i.e., it can serve m simultaneous users). The model has been developed, along with the traffic model described in the previous subsection, to assess the performance of different Radio Resource Management mechanisms in this chapter. However, this is also used in Chapter 5 to study the effect of uneven traffic distribution and develop a suitable topology management technique to neutralise it. The model is first introduced for a particular scenario of five eNBs and subsequently generalised for any number of eNBs. The state of the system of five eNBs is described as $(j_1, j_2, j_3, j_4, j_5)$ where j_i denotes the number of occupied channels in $eNB_i | i = 1, 2, 3, 4, 5$. Given that eNBs are of equal capacity m, the number of possible states is $S = (m + 1)^5$. We define $\frac{1}{\mu}[s]$ as the mean service time, which is equal for all $eNBs$, and $\mu_i = u\mu$; $u = 1, 2, \dots, m$ where u is the number of users being served by the i^{th} eNB. γ_i $[s^{-1}]$ is the arrival rate of users into eNB_i.

A system state model of a generic state $(j_1, j_2, \dots, j_D)$ is illustrated in Figure 3.1. The values of $\gamma_n(j_1, j_2, j_3, j_4, j_5)$ $(n = 1, 2, 3, 4, 5)$ are dependent on the scenario and are calculated using the set of non-negative coefficients (ρ) computed in the previous sub-section equation (3.4). These are calculated as shown in equation (3.5). As we cannot have a departure if there are no users in the system, $\mu_0 = \mu_{-1} = \mu_{-2} = \gamma_{-1} = \gamma_{-2} = \gamma_{-3} = \dots = 0$. Also, since it is logical that we cannot have a negative number of users in the system or more users than the system capacity, we can use the condition that $P(-1) = P(-2) = P(m + 1) = P(m + 2) = 0$. Let $P(j_1, j_2, j_3, j_4, j_5)$ be the equilibrium probability that there are j_n users being served by eNB_n. We can derive the corresponding conservation-of-flow equation as follows:

$$
\begin{aligned}
&(\gamma_1 + \gamma_2 + \gamma_3 + \gamma_4 + \gamma_5 + j_1\mu_1 + j_2\mu_2 + j_3\mu_3 + j_4\mu_4 + j_5\mu_5)P(j_1, j_2, j_3, j_4, j_5) \\
&\quad = \gamma_1 P\,(j_1 - 1, j_2, j_3, j_4, j_5) + \gamma_2\, P\,(j_1, j_2 - 1, j_3, j_4, j_5) \\
&\quad + \gamma_3\, P\,(j_1, j_2, j_3 - 1, j_4, j_5) + \gamma_4\, P\,(j_1, j_2, j_3, j_4 - 1, j_5) \\
&\quad + \gamma_5\, P\,(j_1, j_2, j_3, j_4, j_5 - 1) + (j_1 + 1)\mu_1 P\,(j_1 + 1, j_2, j_3, j_4, j_5) \\
&\quad + (j_2 + 1)\mu_2 P\,(j_1, j_2 + 1, j_3, j_4, j_5) \\
&\quad + (j_3 + 1)\mu_3 P\,(j_1, j_2, j_3 + 1, j_4, j_5) + (j_4 + 1)\mu_4 P\,(j_1, j_2, j_3, j_4 \\
&\quad + 1, j_5) + (j_5 + 1)\mu_5 P\,(j_1, j_2, j_3, j_4, j_5 + 1)
\end{aligned} \tag{3.11}
$$

As the system must be in one of the states described by conservation-of-flow equation (3.11), the state probabilities must satisfy the normalisation equation:

$$\sum_{j_1=0}^{m}\sum_{j_2=0}^{m}\sum_{j_3=0}^{m}\sum_{j_4=0}^{m}\sum_{j_5=0}^{m} P\,(j_1, j_2, j_3, j_4, j_5) = 1 \tag{3.12}$$

The system state probability vector can be obtained by solving the $(m + 1)^5$ equations derived from equation (3.11) in conjunction with the normalization equation. Let A be the $(m + 1)^5 \times (m + 1)^5$ coefficient matrix, P the $(m + 1)^5 \times 1$ state probability vector and B the $(m + 1)^5 \times 1$ constant vector:

$$AP = B \tag{3.13}$$

We can obtain the state probability vector P by solving the matrix equation as:

$$P = A^{-1}B \tag{3.14}$$

Equation 3.14 can be solved knowing that the coefficient matrix A is invertible satisfying the following conditions:

a. The rank of A is $n = (m + 1)^5$.

b. The determinant of A is not zero.

c. The columns of A form a linearly independent set. This is true as the number of UEs arriving at a given time period form an independent, identically distributed sequence.

To serve the purpose of validating and assessing the schemes presented here, it is not necessary to present a closed form solution for the problem described above thus we use numerically derived results in subsequent sections. The method used to obtain the probability vector consists of an iterative process in which the probability vector P is initialised to an assumed value. If no prior information is available about the possible values of elements in P, the values are set to $1/(m+1)^5$ (i.e., all system states have equal probabilities). Subsequently, P_i is computed as $P_i = P_{i-1}A$ until convergence. Convergence is decided when the sum of the changes of the probabilities from one iteration to the next is less than a certain value. In the results presented here, a value of 10^{-6} is used. A detailed description of the method used can be found in [93].

The generalisation of the five-dimensional Markov model derived for the five-eNB network is straightforward. Figure 3.1 illustrates the state transition diagram for the corresponding D-dimensional Markov model for D-eNB network.

$$\sum_{i=1}^{D}(\gamma_{\mathrm{i}} + j_{\mathrm{i}}\mu_{\mathrm{i}})\, P(j_1, j_2, \dots, j_{\mathrm{D}}) \\ = \sum_{i=1}^{D}\left[\gamma_{\mathrm{i}} P\left(j_{\mathrm{i}} - 1, j_{\mathrm{j}}, \dots, j_{\mathrm{D}}\right) + (j_{\mathrm{i}} + 1)\mu_{\mathrm{i}} P\left(j_{\mathrm{i}} + 1, j_{\mathrm{j}}, \dots, j_{\mathrm{D}}\right)\right] \tag{3.15}$$

As the system must be in one of the states described by conservation-of-flow equation (3.15), the state probabilities must satisfy the normalisation equation:

$$\sum_{j_1=0}^{m} \sum_{j_2=0}^{m} \dots \sum_{j_{\mathrm{D}}=0}^{m} P\,(j_1, j_2, \dots, j_{\mathrm{D}}) \;=1 \tag{3.16}$$

The system state probability vector can be obtained by solving the $(m+1)^D$ equations derived from equation (3.15) in conjunction with the normalization equation. Each state is described by the $(j_1, j_2, \dots \dots, j_D)$ coordinates, where j_x $(n = 1, 2, \dots . D)$ is the number of users being served by eNB_x , and γ_x is the total user inter arrival rate into eNB_x.

3.2.4 **Key Performance Indicators**

3.2.4.1 *Homogenous deployment strategy: -*

As detailed in the previous sub-sections and in order to keep fair chance for the users of both deployment strategies at accessing the network, each UE has three accessible eNBs out of the five available. The first modelled network is a traditional homogenous network as seen in Figure 3.2. This can be considered as a small-cell network with no macro-cell overlay, or a traditional macro-cell only eNBs with no small-cell capacity boosters.

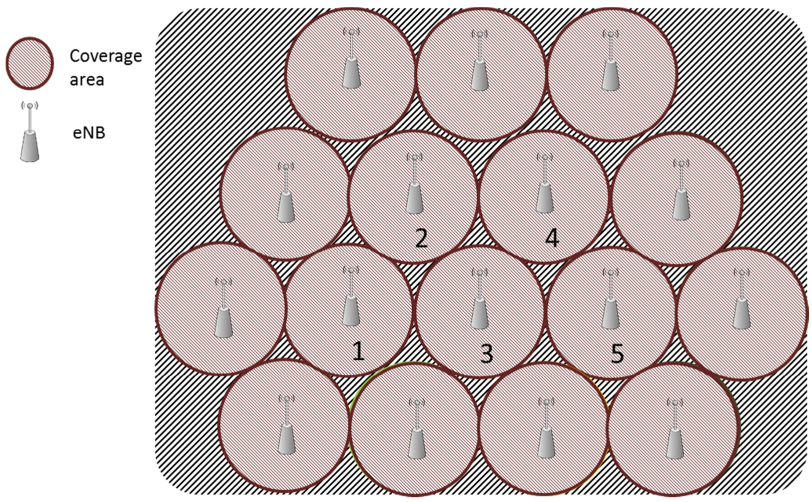


Figure 3.2 Homogenous network

Following the design constraints and parameters, each UE has three accessible eNBs. By applying the cell-wrapping technique, the whole network can be accurately abstracted into a limited number of eNBs forming a D-dimensions Markovian model with $C = \binom{D}{\Psi}$ User Groups (UGs), each group accessing a different set of eNBs. In our model, $D = 5$, and $\Psi = 3$, hence 10 UGs are formed.

- *Blocking Probability Calculation and Coverage limitations*

In this deployment and since each user group can access Ψ eNBs, UEs experience blocking only if Ψ or more eNBs have no free Resource Blocks (RBs). Figure 3.3 and Figure 3.4 illustrate the coverage limitations of one and two eNBs in a 5-eNB network and $\Psi = 3$ (when the other eNBs have no free RBs).

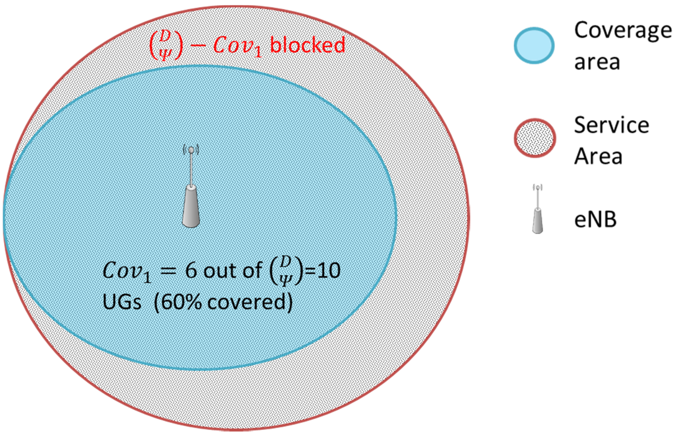


Figure 3.3 Coverage area of a single eNB

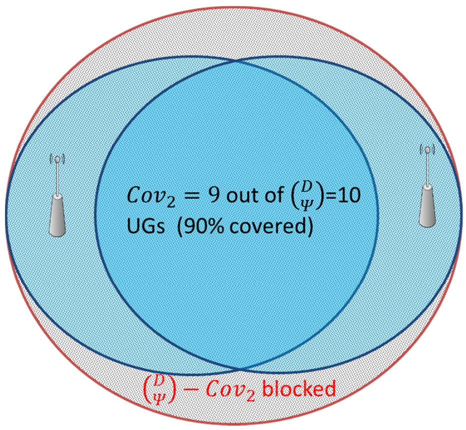


Figure 3.4 Coverage area of any two eNBs

The coverage limitation of any given number of eNBs (n) is calculated as follows:

$$Cov_n = (-1)^{n-1} \bigcap_{i}^{n} Cov_{eNB_i} \tag{3.17}$$

$$\bigcap_{i}^{n} Cov_{eNB_i} = (\mathrm{D} - \mathrm{i})\mathrm{C}(\Psi - i) \tag{3.18}$$

Where xCy is the combination function of subset of length y out of a set of length x, Cov_n is the service area that is covered by n eNBs in terms of number of User Groups (UGs). Given the fact that each UE is served by Ψ eNBs, hence blocking only happens if there are Ψ or more eNBs with no free Resource Blocks (i.e., all RBs are taken hence the eNB is full). The system state probability of having ξ ($\xi \geq \Psi$) full eNBs is calculated as

$$SS_{\xi} = \sum_{\xi=\Psi}^{\xi=D} P(j_I) \;\; | \quad I \subset \{1, \dots, D\}, |I| = \xi, \; j_I = m. \tag{3.19}$$

Where the sum runs over all subsets I of the indices 1,…, D that contain Ψ elements and has the value of m, and m is the number of resource block groups available at each eNB.

$$j_I := \bigcap_{i \in I} j_i \tag{3.20}$$

This denotes the intersection of all j_i with index I. Hence, the system blocking probability for scenario 1 (PB_{sc1}) can be calculated as

$$PB_{sc1} = \sum_{i=\Psi}^{n} \frac{BK_i \lambda_{UG}}{\lambda} SS_i \tag{3.21}$$

$$BK_i = N_{UG} - Cov_{D-i} \tag{3.22}$$

Where BK_i is the number of blocked UGs when any i eNBs are full, N_{UG} is the total number of user groups in the system, Cov_n is the number of user groups served by any n eNBs which is calculated from equation (3.17), λ_{UG} is the User Group (UG) arrival rate into the system and λ is the arrival rate of all UGs into the system (i.e., the system arrival rate).

3.2.4.2 *Heterogeneous deployment strategy: -*

In this scenario, a macro-cell overlay exists which covers the entire service area with a second tier of small-cell eNBs. This type of scenario is becoming increasingly interesting and a subject for intense research as seen as way to boost capacity density and enhance spectrum efficiency [94-98]. Heterogeneous deployment strategies are also considered for providing high speed connectivity and 4G technology globally [97], or even as a disaster relief architecture [96]. The latter scenario is used to evaluate novel radio resource and topology management techniques in Chapter 6 which are developed taking into consideration the findings of this chapter. Again, for fairness, all UEs have Ψ accessible eNBs. Given that the Macro-eNB (MeNB) covers all the service area, all UEs can access $\Psi - 1$ small-cell eNBs and the MeNB. Hence, the system behaviour in terms of blockage is fundamentally different.

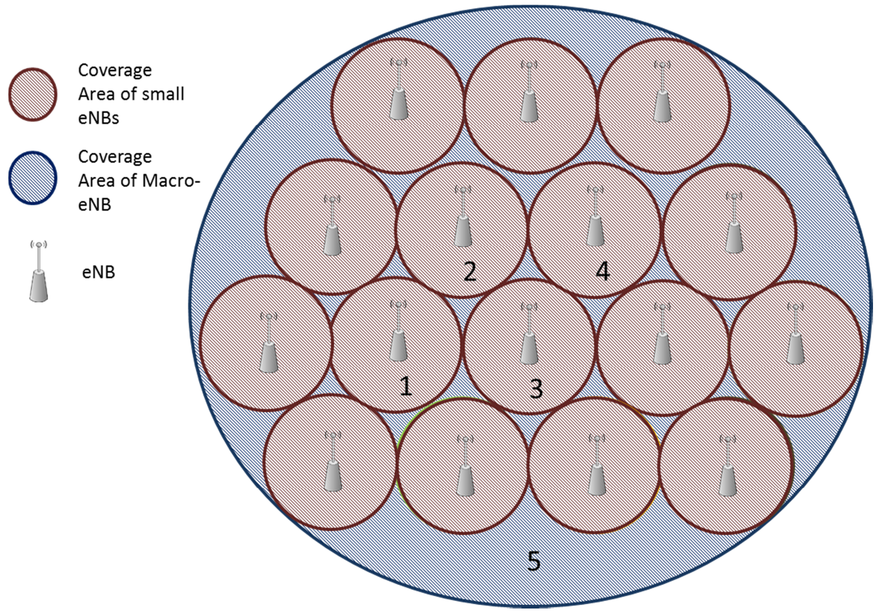


Figure 3.5 Heterogeneous deployment

Also for this deployment strategy the whole network is abstracted into a limited number of eNBs forming a D-dimensions Markovian model with $\mathrm{C} = \binom{D}{\Psi}$ User Groups (UGs), each group accessing a different set of eNBs. This was done following the design parameters and by applying the cell-wrapping technique. The number of UGs is changed given the presence of a macro-cell overlay to only $\mathrm{C} = \binom{D-1}{\Psi-1}$. The number of MeNBs needs to be subtracted also when using equation (3.17) to calculate the coverage limitations of a given eNB. By keeping the parameters D, and Ψ constant, six user groups are formed.

- *Blocking probability calculation and coverage limitations*

In this case, a UE or one UG is blocked if any $\Psi - 1$ of the small-cell eNBs have no free RBs and the MeNB is full. Figure 3.6 and Figure 3.7 illustrate the coverage limitations of one and two eNBs (when the other eNBs have no free RBs).

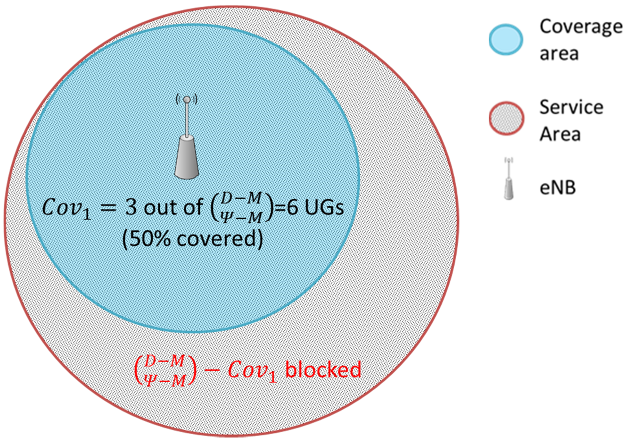


Figure 3.6 Coverage area of a single eNB

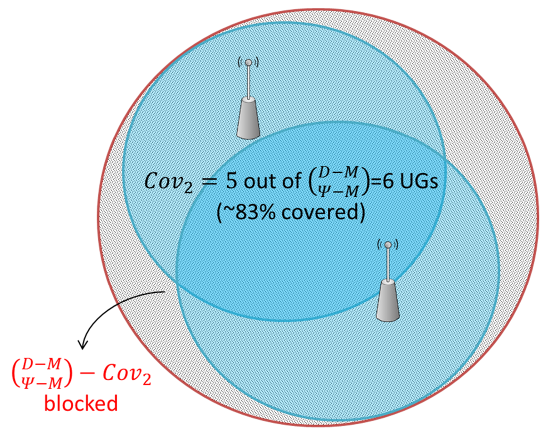


Figure 3.7 Coverage area of any two eNBs

Hence, blocking only happens if the macro-cell eNB is full and there are ψ or more full small-cell eNBs ($d \geq \psi \geq \Psi - 1$), where d is the number of small-cell eNBs. The system state probability of having ψ full small-cell eNBs is calculated as

$$SS_{\xi} = \sum_{\psi=\Psi-1}^{\psi=d} P(j_I) \mid I \subset \{1 \ldots d\}, |I| = \xi,\ j_I = m. \ (23) \tag{3.23}$$

Where the sum runs over all subsets I of the indices $1, \ldots, D-1$ that contain $\Psi - 1$ elements and has the value of m. m is the number of resource block groups available at each eNB.

$$j_I := \bigcap_{i \in I} j_i \tag{3.24}$$

This denotes the intersection of all j_i with index I. Hence, the system blocking probability for scenario 2 (PB_{sc2}) can hence be calculated as

$$PB_{sc2} = \sum_{i=\Psi-1}^{D-1} \frac{BK_i \lambda_{UG}}{\lambda} SS_i \tag{3.25}$$

$$BK_i = N_{UG} - Cov_{D-1-i} \tag{3.26}$$

Where BK_i is the number of blocked UGs when the MeNB and any i small-cell eNBs are full, N_{UG} is the total number of user groups in the system, λ_{UG} is the UG arrival rate into the system and λ is the arrival rate of all UGs into the system (i.e., the system arrival rate), and Cov_n is the number of user groups served by any n eNBs which is calculated from equation (3.17) excluding the MeNB. The exclusion of the MeNB needs to be taken into account when using equation (3.17) by subtracting $D-1$ and $\Psi-1$ as seen in the previous equations. D and Ψ were defined earlier.

3.2.4.3 *Other Key Performance Indicators*

Other Key Performance Indicators (KPIs) considered here are the average number of eNBs in the idle-state, system power requirements, and Idle State Stability (ISS) of nodes. ISS represents the probability of transiting from the idle state to the active state. This particular KPI is not yet common. However, it is becoming an important design parameter of future flexible access networks that implement sleep-mode techniques and topology management algorithms to enhance the system energy efficiency.

Starting from the average number of eNBs in the idle state, the idle state probability, for eNB1 for instance, is calculated as

$$P_{idle\_1} = \sum_{j_2=0}^{m} \sum_{j_3=0}^{m} \dots \sum_{j_d}^{m} P\,(j_1, j_2, \dots, j_{\mathrm{D}}) \mid \mathrm{j}_1 = 0 \tag{3.27}$$

Similarly, we can calculate the idle state probability for eNB_2, eNB_3, eNB_4 and eNB_5. The average number of idle nodes in the system at any given time is the calculated as follows:

$$Idle\_N_s = \sum_{i=1}^{D} P_{idle\_i}\,\mathrm{D} \tag{3.28}$$

Where D is the number of eNBs in the system. In terms of system ISS calculations, the probability that any given eNB in the system transit from the idle to the active state is calculated as

$$ISS_s = \sum_{i=1}^{D} P_{idle\_i}\gamma_i \tag{3.29}$$

System power requirements (PW_s) can be also calculated from this model using the steady-state probabilities as the power model adopted in this study relies on the time spent at each state as follows:

$$PW_s = \sum_{i=1}^{e}\sum_{u=1}^{m} PW_u P_i(u) \tag{3.30}$$

Where PW_u is the power required to serve u number of users which is calculated using the energy model described in the subsequent section, $P_n(u)$ is the steady-state probability of having u number of users being served by the i^{th} eNB, and e is the number of eNBs. The system throughput for both scenarios (TH_{sc1} and TH_{sc2}) can be directly derived from their respective system blocking probability as, by definition, it is the portion of the offered traffic (OT) that is not lost:

$$TH_{sc1} = OT[1 - PB_{sc1}] \tag{3.31}$$

$$TH_{sc2} = OT[1 - PB_{sc2}] \tag{3.32}$$

3.2.5 **Radio Resource Management techniques**

The RRM techniques assessed here are the Load Balancing, Load Unbalancing, and the Macro-cell eNB prioritisation. These are defined as follows:

1. **Load Balancing ($\boldsymbol{LB}$)**: LB tries to balance the traffic demand across the access network. It dynamically sets the priority figure of the eNBs depending on their loading status. The least loaded eNB has a higher priority over other eNBs in the access network. As seen in the previous sections, the probability of a given user with a candidate set of $eNBs$ $\{eNB_1\ eNB_2\ \ eNB_3\}$ is directed to eNB_1 is given by $\rho_{1,\{123\}} = \bar{\delta}_{j_1 m_1}\left(\emptyset_{123} + \emptyset_{132} + \delta_{j_2 m_2}\emptyset_{213} + \delta_{j_3 m_3}\emptyset_{312} + \delta_{j_2 m_2}\delta_{j_3 m_3}(\emptyset_{231} + \emptyset_{321})\right)$. Generally, these probabilities can be written as $\rho_{x,xy} = \rho_{x,xyz} = \bar{\delta}_{j_x m_x}$; $\rho_{x,yx} = \rho_{x,yxz} = \bar{\delta}_{j_x m_x}\delta_{j_y m_y}$, and $\rho_{x,yzx} = \bar{\delta}_{j_x m_x}\delta_{j_y m_y}\delta_{j_z m_z}$ where $x, y, z \in \{1, 2, 3\}; x \neq y \neq z$. LB changes the order of this UE's candidate set according to their loading status assigning highest priority to the least loaded eNB and the lowest to the most loaded eNB.

2. **Load Unbalancing ($\boldsymbol{LUB}$)**: This technique tends to cluster the traffic demand onto as few access eNBs as possible. The LUB technique adopted here uses a fixed priority list to aid the decision of camping on a cell. The priority list, denoted by Ω, is as follows $\Omega_1 > \Omega_2 > \cdots > \Omega_D$ for $eNB_1, eNB_2, \ldots eNB_D$ respectively.

3. **Macro-cell overlay priority (MP)**: This technique sets a priority figure to indicate whether the macro-cell eNB has greater or lower priority than the small-cell eNBs. This is used to identify the role of the Macro-cell eNB (MeNB) in serving the traffic and how it can be used to optimise system performance. Two configurations are used. Assuming that eNB_D is the MeNB, the first configuration sets the priority of eNB_D the highest (i.e., $\Omega_D > (\Omega_1, \Omega_2, \dots, \Omega_{D-1})$). The second configuration is to set the priority of the MeNB the lowest (i.e., $\Omega_D < (\Omega_1, \Omega_2, \dots, \Omega_{D-1})$). Please note that this priority overrides any priority set by the LB or the LUB algorithms.

3.3 Results and discussion

The different RRM mechanisms and deployment strategies are assessed using the mathematical model described in the previous sections for a network of $D = 5$ eNBs. The analytical model assumes that users can only connect to Ψ =3 eNBs out of the available 5 depending on their occupancy as interference is not taken into account. The power model used here is adopted from the EARTH project [99-101]. The model and parameters are based on anticipated improvements in 2014-eNB's components which can be expressed as:

$$\mathrm{P_{supply}} = \begin{cases} \mathrm{P_0} + \beta \mathrm{P_{Tx}} & ; \; 0 \le \mathrm{P_{Tx}} \le \mathrm{P_{max}} \\ \mathrm{P_d} & ; \; \mathrm{P_{Tx}} = 0 \end{cases} \tag{3.33}$$

Where $\mathrm{P_{max}}$ is the maximum transmit power, P_0 and P_d are the idle- and dormant-mode power consumption of the eNB respectively, P_{Tx} is the instantaneous transmit power, and β is the load dependency constant. Table 3.1 contains the detailed simulation parameters.

Parameter	*Value*
Number of channels ($\boldsymbol{m}$)	4
Number of eNBs ($\boldsymbol{D}$)	5
Load-dependency constant ($\boldsymbol{\beta}$)	1.25
Power in the idle-mode ($\boldsymbol{P_0}$)	60 [W]
Power in the dormant-mode ($\boldsymbol{P_d}$)	20 [W]
Maximum transmit power ($\boldsymbol{P_{Tx}}$)	5 [W]

Table 3.1 Simulation Parameters [99], [102]

The results of the mathematical model have been verified using a Monte Carlo simulation. For clarity, Monte Carlo simulation results are marked using a dot only with no line for all graphs presented in this work. The role of the Macro-cell overlay in a heterogeneous network is identified first, and then the two deployment strategies are compared subsequently.

3.3.1 **Role of the Macro-cell overlay**

In this subsection, the importance of the existence of a macro-cell overlay and its effect on the RRM technique is studied using a wide range of terms; namely, QoS, energy efficiency, and system stability. Four different settings have been assessed. These are the load balancing with the MeNB having the lowest priority (LB - MeNB low P), the load balancing with the MeNB having the highest priority (LB - MeNB high P), load unbalancing with the MeNB having the highest priority (LUB - MeNB high P), and lastly load unbalancing with the MeNB having the lowest priority (LUB - MeNB low P). These four settings have been chosen to have a clear insight into the potential of a heterogeneous deployment strategy to deliver the required QoS in a stable and energy efficient manner. Figure 3.8 illustrates the system QoS in terms of blocking probability for the aforementioned settings as marked on the legend.

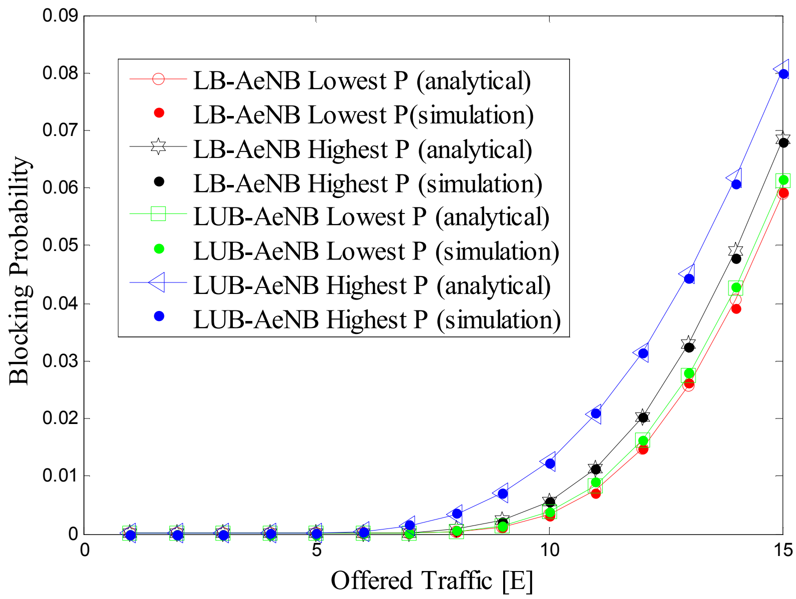


Figure 3.8 System Blocking Probability

The figure indicates that system performance depends heavily on the priority of the macro-cell overlay and that this needs to be configured properly. There are roughly 25% more files blocked by changing the priority of the macro-cell eNB from lowest to highest when using the load unbalancing RRM. This percentage can be further enhanced to surpass the 30% mark when balancing the load at medium and high offered traffic and keeping the MeNB with lowest priority.

On the energy efficiency and network stability side, Figure 3.9 and Figure 3.10 illustrate the system performance in this regard in terms of average idle state probability and average frequency to the active state of network nodes respectively. Again, the importance of the prioritisation of the macro-cell eNB is evident in both figures. In Figure 3.9, when load balancing is in place, the alteration of the priority of the MeNB from highest to lowest decreases the changes for network nodes of being in the idle state by an average of 20% at low traffic loads. Generally, load unbalancing seems to be the most beneficial approach in terms of energy efficiency at all traffic loads. However, taking Figure 3.8 into account, it is obvious that load unbalancing is not the optimal technique to avoid congestion at high traffic loads.

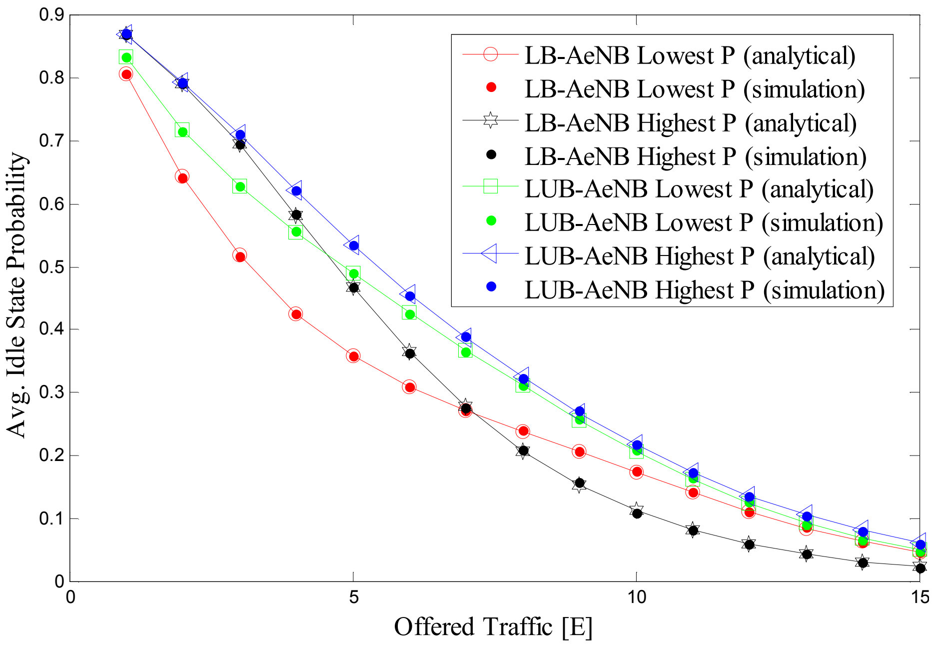


Figure 3.9 Avg. idle state probability of eNBs

Figure 3.10 ought to be also taken into account when assessing any energy efficient mechanism. Precisely, this figure is particularly important when designing sleep-mode mechanisms with special attention to load-based approaches. The lower the frequency to the active state the more stable the system is as nodes are less probable to leave the idle state (i.e., stay dormant). We can see that when the MeNB is set to have the lowest priority, for both RRM techniques, network nodes tend to frequently alternate between the idle and the active state. Hence, when a sleep-mode mechanism is in place, nodes will tend to frequently switch between the sleep mode and the active mode which destabilises the network especially when taking into consideration the transit period between the dormant and the active mode. This shows that the priority of the MeNB plays a more important role than the RRM technique in stabilising the network and avoiding redundant transitions from the idle to the active state at all traffic loads. In other words, different RRM schemes affect largely the energy efficiency of the network but the settings of the MeNB have a larger effect on network stability.

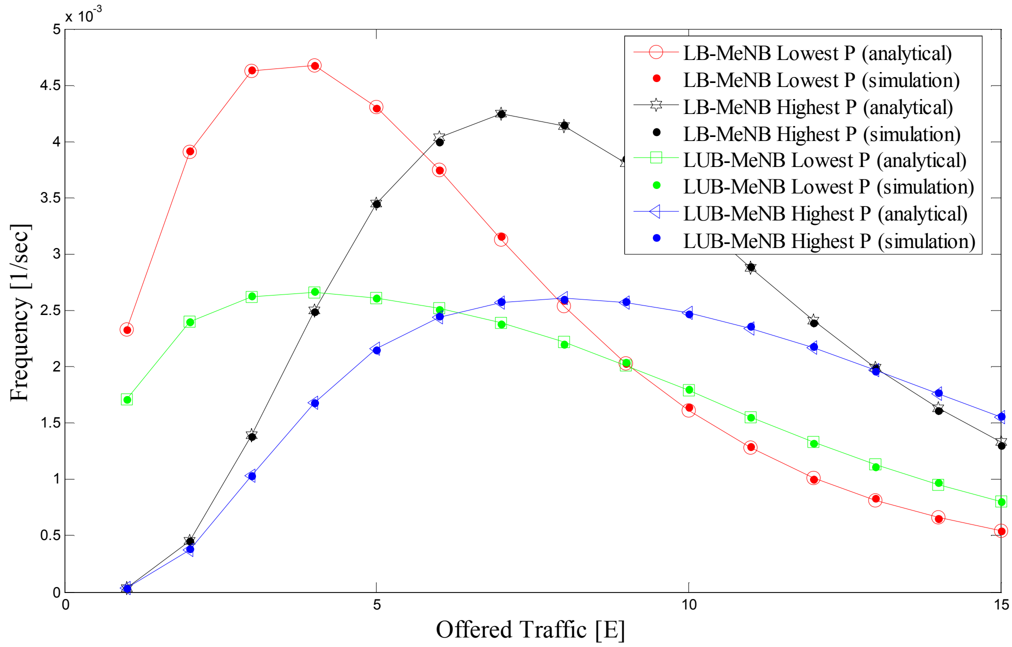


Figure 3.10 Avg. idle to active state frequency

Figure 3.11 is the Idle State Probability of the small-cell layer (ISP-SC) (i.e., the probability of having all small-cell eNBs in the idle state). The difference in performance between the RRMs and MeNB prioritisation is clear from Figure 3.11 that the prioritisation of the MeNB is of great importance to the stability as well as energy efficiency of the network allowing the underlying small-cell eNBs to be kept in the idle state while ensuring the whole service area is covered. Figure 3.12, on the other hand, illustrates the potential network power requirements of each setting, having a sleep-mode mechanism in place that exploits the idle state of the eNBs.

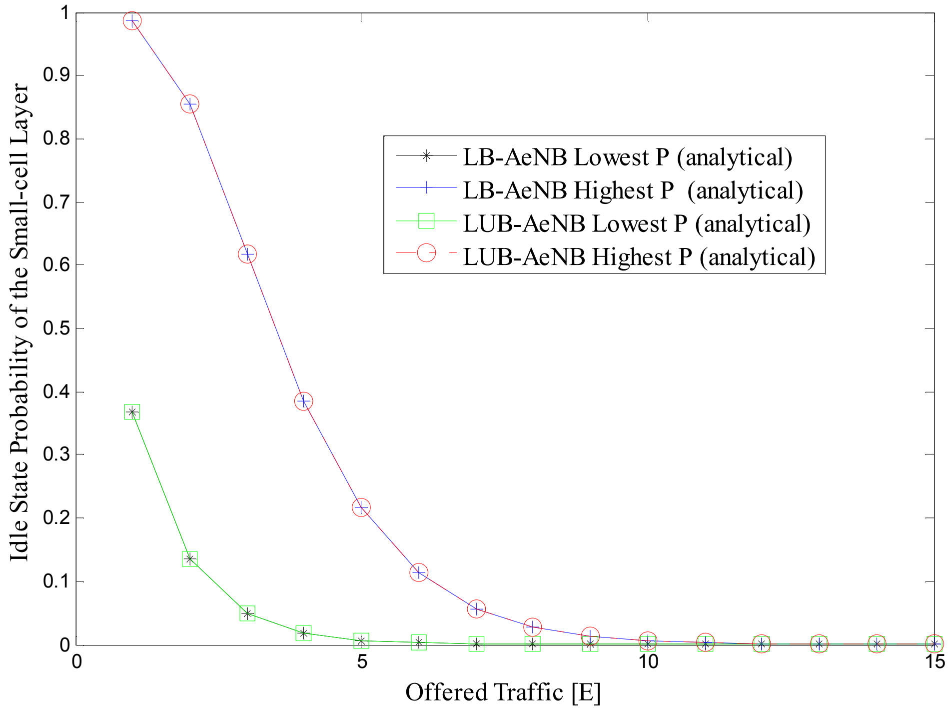


Figure 3.11 ISP of the small-cell layer

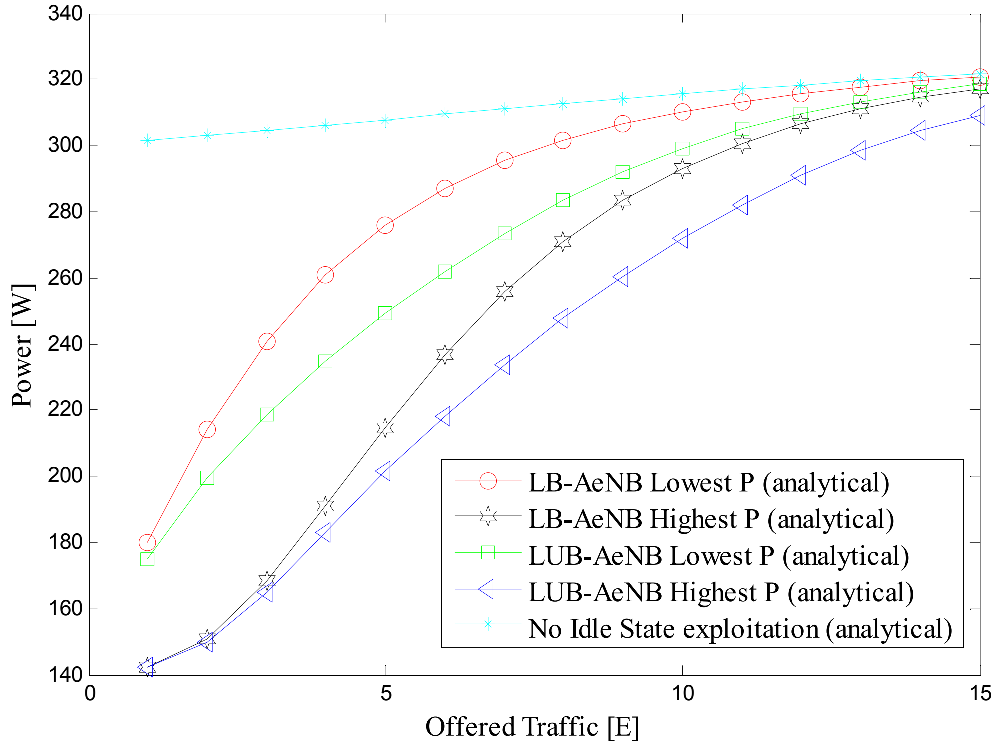


Figure 3.12 System power requirements

Since the study of the functionality of sleep-mode mechanisms is out of the scope of this chapter, the sleep mode mechanism used here ignores the transitional period between the dormant and the active mode and the power associated with it. Also, the MeNB is assumed to be always in the active state and never enters the dormant state. You can notice from Figure 3.11 that despite the fact that both RMMs with lowest priority MeNB have the same idle state probability, Figure 3.12 indicates that LUB outperforms the LB RRM technique in terms of network power requirements. This is also applicable to the case when the MeNB has the highest priority. This is mainly due to the fact that eNBs have longer idle periods when the priority of the MeNB is set high even if these are less frequent as seen in Figure 3.12. The same applies to both load balancing settings. Clearly, the load unbalancing largely outperforms all other RRMs settings in terms of both network stability and energy efficiency when setting the priority of the MeNB to the highest.

3.3.2 **Homogenous VS heterogeneous deployment strategies**

Based on the performance of different settings of the heterogeneous deployment strategy, the load unbalancing with high priority MeNB and the load balancing with low priority MeNB are chosen for the comparison against homogenous deployment strategy.

Figure 3.13 and Figure 3.14 illustrate the system QoS as well stability in terms of blocking probability and transition frequency from the idle to the active state. Even though the homogenous deployment strategy performs well, it is still outperformed by the heterogeneous strategy when the MeNB is set to have a low priority. The heterogeneous deployment is still performing much better than its homogenous counterpart also in terms of network stability keeping the frequency of transiting from the idle to the active state as low as zero at very low traffic loads.

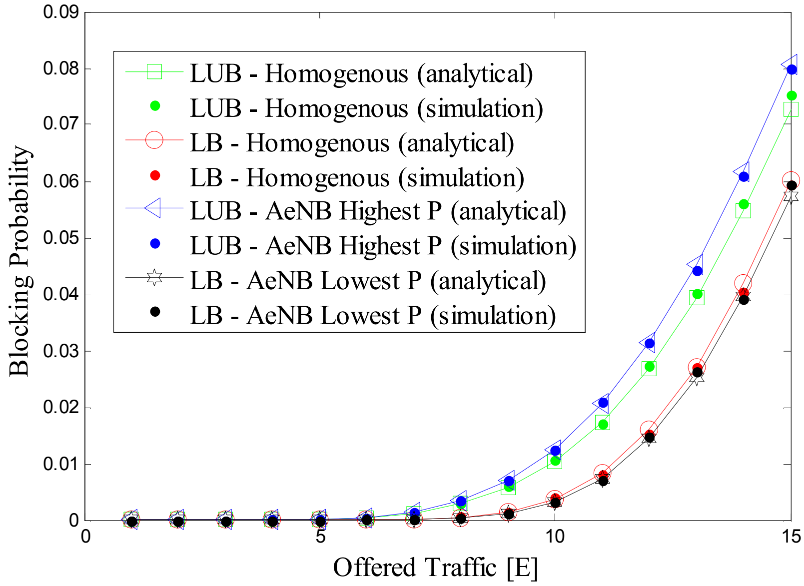


Figure 3.13 System Blocking Probability

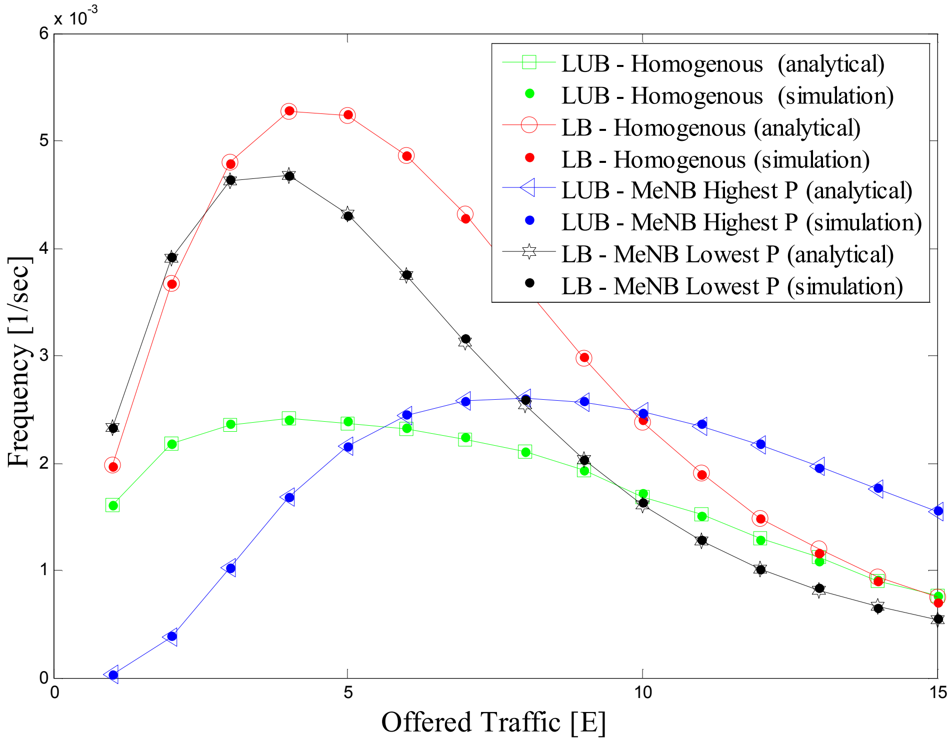


Figure 3.14 Idle to active state frequency

Similar behaviour is noticed in terms of energy efficiency as shown in Figure 3.15 and Figure 3.16 having the load balancing technique in homogenous deployments performing by far the worst. In the case of load unbalancing, both deployment strategies perform highly similar. In fact, even though the average idle state probability is lower, we can notice a slight advantage of the homogenous deployment strategy at very low loads in terms of network power requirements, see Figure 3.16. This is due to the inability of the macro-cell overlay to switch to the dormant state hence its idle state cannot be exploited to save energy.

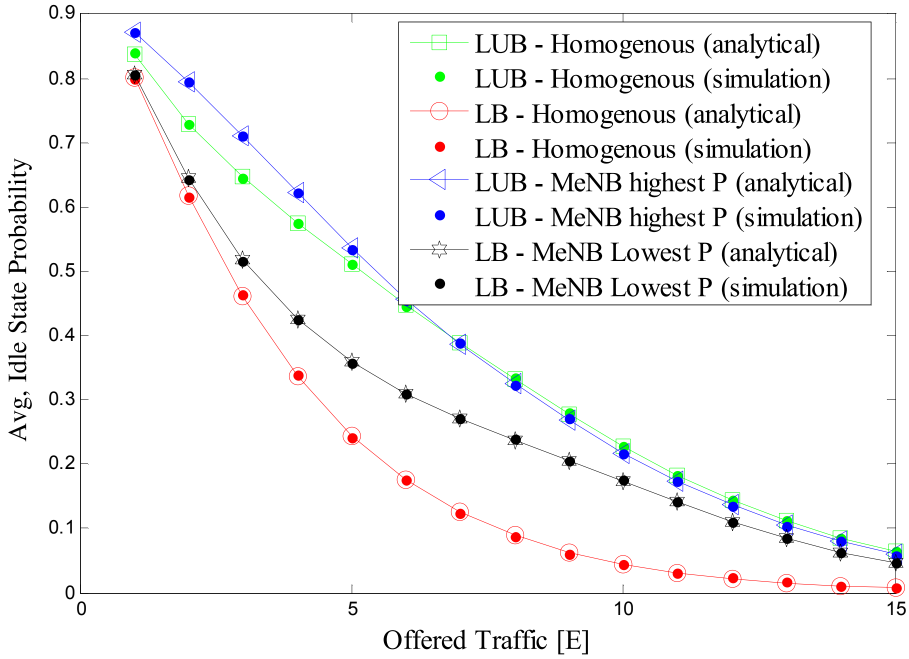


Figure 3.15 Avg. Idle state probability of eNBs

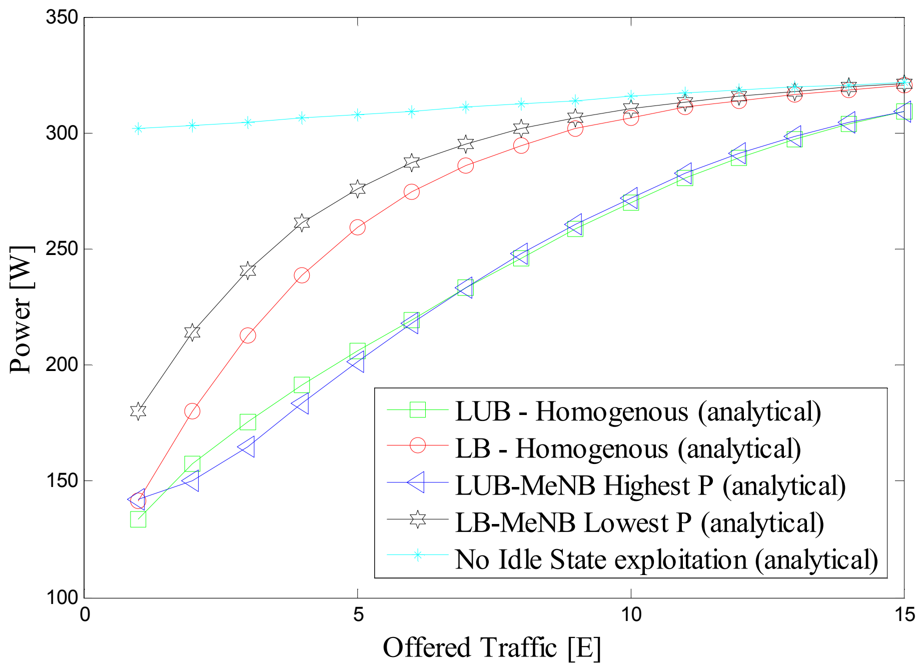


Figure 3.16 System power requirements

A fundamental advantage of the heterogeneous deployment strategy over the homogenous one is the ease at which it is capable of delivering its best performance, best energy efficiency without destabilising the network. As mentioned earlier, It can be seen from Figure 3.16 that it is a lot harder to develop a feasible sleep-mode mechanism, or any green mechanism that exploits the idle state of network nodes, that works efficiently and ensures QoS as well as coverage in homogenous deployments due to the high probability of transiting from the dormant to the active state.

3.4 Conclusion

The performances of homogenous as well as heterogeneous deployment strategies were evaluated using queueing theory. Two main Radio Resource Management techniques; namely, Load Balancing and Load Unbalancing were also evaluated for both deployment strategies. The evaluation was conducted taking into consideration a wide range of system KPIs. These are the system QoS in terms of blocking probability, network energy efficiency as well as network stability in terms of frequency from the idle to the active state. The importance of having a macro-cell overlay was also assessed. It is concluded that even though the RMM technique is important, system performance in all terms relies heavily on the Macro-eNB (MeNB) prioritisation settings. The only disadvantage seen is the fact that MeNBs consume more energy and it is difficult to exploit their idle state. However, having a macro-cell overlay not only enhances system performance but it also adds flexibility to the underlying capacity boosters and adds stability to the network. These findings are directly used in the subsequent chapters to firstly develop an efficient topology management mechanism and then enhance its operation by introducing a macro-cell overlay. Homogenous networks also perform close to their heterogeneous counterparts. However, more effort is required to reach the performance level of heterogeneous deployments in terms of idle to active state frequency which consequently tends to be more destabilised. This translates into less energy harvested and more complexity added to energy-aware mechanisms. For this reason, homogenous deployment strategies are more suitable in areas that have a semi-constant traffic demand or traffic of low fluctuation over time. This could be high, medium or low traffic demand or a mix of these but with a slow changing nature such as busy city centres where traffic is constantly intense even overnight. Heterogeneous deployment strategies on the other hand are proven to have extended capabilities especially in terms of energy efficiency and network stability hence it is recommended that these types of strategies are used in areas with unpredictable traffic demand and of medium to high fluctuation nature. The effect of having an uneven traffic distribution is further investigated in Chapter 5 which implies that, in order to be effective, the topology management needs to have a certain level of complexity compared to the situation when the traffic is distributed evenly.

Also, in the absence of the MeNB, Load Balancing and Load Unbalancing techniques are proven to be of great benefit even though these only relate to a specific region of the offered traffic (i.e., either at low or at high traffic loads). Load Balancing enhances the system performance in terms of QoS. However it tends to consume more energy and make the network less stable. The application of the emerging RRM Load Unbalancing technique on the other hand is seen to add a considerable amount of stability as well as energy efficiency to the network especially when operating in a homogenous network. There is still the need to develop a radio resource management mechanism that is capable of efficiently and dynamically set the optimal clustering degree in order to cluster demand onto as less number of eNBs when the traffic is medium and low while balancing the load appropriately when the demand is high optimising system capacity and QoS. This is investigated in Chapter 6 which introduces a load management mechanism based upon the recommendations and outcomes of this chapter.

Chapter 4 **Energy-Aware Topology Management for Next Generation Mobile Broadband Systems**

Contents

4.1 **Introduction**

The past chapter studied the fundamental requirements and limitations of resource management techniques in different deployment strategies. This chapter firstly investigates the benefits and highlights some key elements that can contribute to the successful development of feasible yet effective ways to manage the topology of a homogenous next generation mobile broadband network. This technique is referred to as the energy-aware or green topology management technique and aims to reduce energy consumption of a given network by switching into a dormant mode as many underutilized base stations as possible while maintaining the QoS at a satisfactory level at all traffic loads. As mentioned in Chapter 1 and Chapter 2, the use of the dormant mode controlled by the topology management algorithm is of increasing importance in exploiting the temporospatial variation of demand, thereby enhancing the energy efficiency of the network. Subsequently, as suggested in Chapter 3, a macro-cell overlay is added forming a two-tier network of small-cell capacity boosters overlaid by a macro-cell eNB. The schemes presented in this chapter serve as a baseline for the work presented in Chapter 5 which considers the occurrence of a temporary event and its consequences on the operation of the topology management scheme.

4.2 Beyond Next Generation Mobile Broadband System (*BuNGee*)

The BuNGee architecture aims at achieving, in a cost efficient way, throughput densities of at least 1 Gb/sec/km^2. This is achieved by introducing Hub and Access Base Stations (HBS and ABS respectively) forming a two-hop network with a wireless point-to-point backhaul between ABSs and HBSs. The HBSs are connected to the operator's backhaul network using multi-beam directional antennas. These antennas provide high capacity self-backhaul links to the access network which is formed by the ABSs. Each HBS serves 20 surrounding below-rooftop ABSs forming a cell and are only used for backhauling without any direct connections to the users from this tier. The communication between ABSs and the associated HBSs is enabled through the Hub Subscriber Station (HSS) which are collocated with the ABSs. HSSs are wirelessly connected to the HBSs using single beam directional antennas and to the ABSs by wired links. Mobile Stations (MSs) communicate with the HBSs through the access network, whereas ABS-to-ABS communication is enabled through the serving HBS [103]. The BuNGee architecture has been recently considered in the European FP7 Beyond Next Generation Mobile Broadband

Project, see Figure 4.1. Different techniques have been developed in order to achieve the targeted throughput density such as the use of below-rooftop access base stations, the use of very high capacity feeding hubs with high order spatial reuse, and the use of directional antennas and advanced MIMO techniques. The entire access network is overlaid with the backhaul link through the HBSs achieved using a combination of in-band backhauling, which is point-to-multipoint in its nature, and 60 GHz point-to-point wireless links.

4.2.1 **Heterogeneous BuNGee (Het-BuNGee)**

As an early contribution of this thesis is a proposed modification to the original BuNGee architecture by placing an additional ABS on top of each HBS tower. These additional ABSs, which are called macro-ABSs, are equipped with one omnidirectional antenna so the entire architecture is overlaid by macro-ABSs providing macro level coverage. The use of macro-ABSs was firstly suggested in Chapter 3 primarily to ensure the service area is fully covered and to provide connectivity at low traffic loads enabling the complete switching to the dormant mode of the entire street-level access network. Although the proposed architecture encompasses some traditional features of Heterogeneous Networks (HetNets), there are nevertheless a range of distinct contrasts. Het-BuNGee is a holistically managed hierarchical structure with the access network being entirely overlaid by the macro-ABS layer with backhaul links through the hub base stations achieved using a combination of in-band backhauling and 60 GHz point-to-point wireless links. More to the point, although Het-BuNGee is also a multi-tiered architecture consisting of a high-powered base station tier, and a low-powered base station tier of small cells, the HBSs, the high-powered base stations, are only used for backhauling and there are no direct connections to the users from this tier. Last but not least, street-ABSs form part of the main network whereas small cells in 3G or LTE are used to provide enhanced coverage in areas that are poorly served by an existing cellular infrastructure and improve capacity incentives [104, 105].

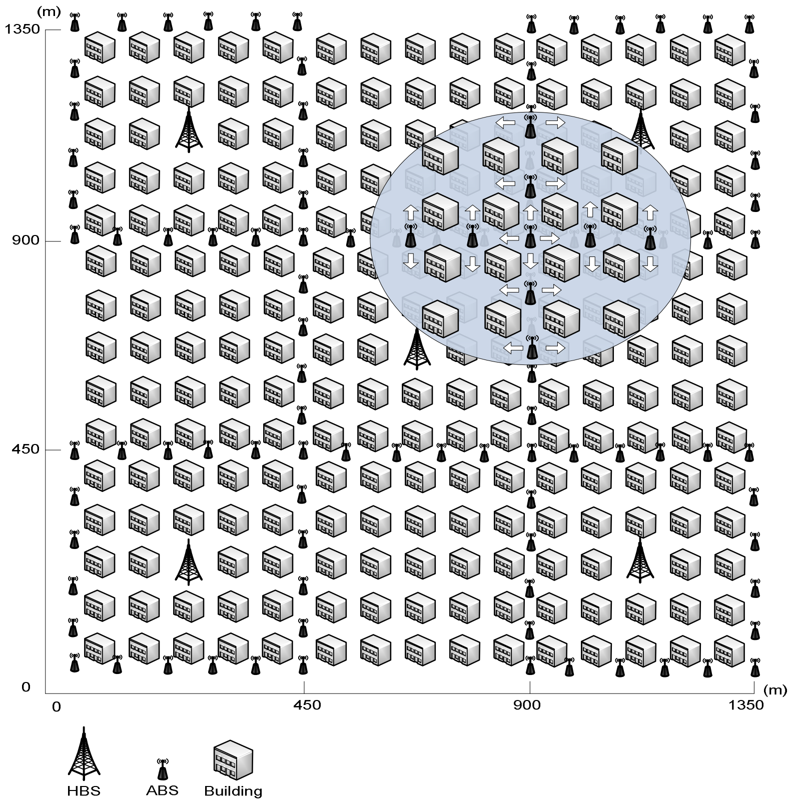


Figure 4.1 BuNGee architecture

4.3 **Green Topology Management for Beyond Next Generation Mobile Networks**

Green topology management aims to dynamically control the predesigned topology of the system in order to achieve energy savings without prejudicing the QoS. This is achieved by fine-tuning the status of the ABSs (dormant / available) depending on the traffic demands. This technique is sometimes referred to as exploiting the sleep modes of a base station [26].

4.3.1 **Neighbour-Based Green Topology Management**

The Neighbour-based Green Topology Management (*NBTM*) algorithm is shown in Figure 4.2.

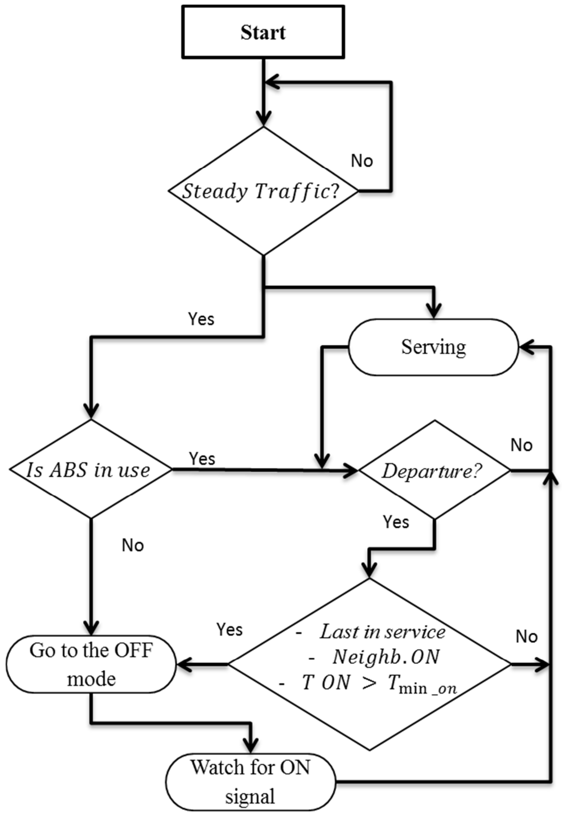


Figure 4.2 Neighbour-Based Topology Management Algorithm

Initially, we assume all ABSs are active within a network. As a first step, ABSs which are not in use switch to the dormant mode. Afterwards, the remaining ABSs cannot switch to the dormant mode unless they satisfy several conditions; namely, an ABS must have at least one neighbouring ABS in the active mode, and it must have less than C_{off} customers in service, where C_{off} is the number of customers in service at which the ABS is switched to the dormant mode, and finally it must have been in the active mode for a period of at least $T_{\min\_on}$. This period is used to stabilize the system by avoiding frequent switching between the active and dormant modes. $T_{\min\_on}$ is calculated from the offered traffic as:

$$T_{min\_on} = C_{off} N\, T \tag{4.1}$$

Where, N is the number of ABSs used in the architecture, T is the perceived average user inter-arrival time, and $T_{\min\_on}$ is the average time needed for C_{off} customers to request for service.

The switch to the dormant mode is not immediate as an ABS enters a transitional period from the active to the dormant mode in which it does not admit any new customers. It will finally switch to the dormant mode when the last user in service departs. Additionally, when the capacity of a particular ABS exceeds C_{on}% of the maximum capacity, the ABS selects a dormant ABS to wake up. The ABS further along the street, which covers the same street, is given priority in this selection. If this is found to be active, one of its other neighbours is selected randomly. The selection of the most suitable ABS at this stage is important. Chapter 6 investigates an adaptable mechanism to optimise the identification of nodes to be switched to the active mode and those to be left dormant.

4.3.2 Macro-cell overlaid Topology Management

The Macro-cell overlaid Topology Management (*McTM*) scheme makes use of macro-level coverage provided by the proposed incorporation of macro-ABSs to the original BuNGee architecture as discussed in Section 4.3. Initially it is assumed that all ABSs are active within a network. Then, the street-ABSs which are not in use switch to the dormant mode. From this point on, each street-ABS can be in two different states:

1- ***Macro-ABS is in the dormant mode:*** in this case, a street-ABS only switches to the dormant mode when the last customer in service finishes transmission. At this point the street-ABS must also satisfy two conditions; namely, it must have at least one neighbour street-ABS in the active mode, and finally it must have been at least T_{min_on} seconds in the active mode as specified in equation (4.1).

2- ***Macro-ABS is in the active mode:*** in this case a street-ABS needs to wait until the end of service of the last user to make sure it has been at least T_{min_on} seconds in the active mode. The street-ABS switches to the dormant mode regardless of whether there is a neighbouring street-ABS in the active mode or not.

We can notice from the two points above the difference in complexity and rules imposed onto the street-ABSs depending on the state of the macro-ABS. As suggested in Chapter 3, the presence of a macro-cell overlay adds flexibility and lowers complexity of the topology management scheme.

A street-ABS can be switched to the active mode by either another street-ABS or the serving HBS. Street-ABSs select another street-ABS to switch to the active mode when they exceed a predefined capacity threshold (C_{on}% of the maximum capacity). The street-ABS which covers the same street is given priority in this selection. When this is found to be active, one of its neighbours is selected randomly. An HBS wakes up a street-ABS if the macro-ABS of that cell exceeds a predefined capacity threshold and the cell-delay is above a predefined threshold too. This can be expressed as follows;

$$\mathrm{HBS} \in \{\mathrm{HBS_i} : \forall \mathrm{i} \in [1,5]\} \tag{4.2}$$

$$\mathrm{M_{ABS}} \in \{\mathrm{M_{ABS_i}} : \forall \mathrm{i} \in [1,5]\} \tag{4.3}$$

$$\mathrm{ABS} \in \{\mathrm{ABS_{i,j}} : \forall \mathrm{i} \in [1,5], \forall \mathrm{j} \in [1,20]\} \tag{4.4}$$

$$\text{if} \quad \mathrm{C_{M_{ABS_i}}} > \mathrm{C_{thrshld}} \text{ and } \mathrm{D_{HBS_i}} > \mathrm{D_{thrshld}} \tag{4.5}$$

$$\mathrm{S_{ABS}} = \mathrm{ABS_{i,j}} : \mathrm{ABS_{i,j}} \in \{\mathrm{ABS} : \mathrm{i}, \forall \mathrm{j} \in [1,20]\} \tag{4.6}$$

Where $\mathrm{HBS_i}$ is the serving HBS in the ith cell, $\mathrm{M_{ABS_i}}$ is the macro-ABS in the ith cell, $\mathrm{ABS_{i,j}}$ is the jth ABS in the ith cell, $\mathrm{C_{M_{ABS_i}}}$ is the capacity of the macro-ABS in the ith cell, $\mathrm{D_{HBS_i}}$ is the delay experienced by the serving HBS in the ith cell, $\mathrm{C_{thrshld}}$ and $\mathrm{D_{thrshld}}$ are the capacity and delay thresholds respectively, and finally $\mathrm{S_{ABS}}$ is the street-ABS to wake up. Macro-ABSs can also switch to the dormant mode depending on traffic loads. If dormant, macro-ABSs are switched back to the active mode in two cases as follows:

1- When a macro-ABS is experiencing delay above the predefined delay threshold and the loading is less than $\mathrm{C}_{\mathrm{thrshld}}$% of the maximum capacity. At this point it activates another macro-ABS.
2- When the loading of a street-ABS is less than C_{on}% of the maximum capacity while experiencing delays above the threshold level.

In the latter case, macro-ABSs double check with their respective serving HBS if the cell-delay exceeds the threshold or it is just a particular ABS case. If the cell-delay is above the threshold, it is activated. As far as the switching to the dormant mode is concerned, macro-ABSs only enter the dormant mode when more than 40% of the in-cell street-ABSs (8 street-ABSs out of 20) are active. Again, the switching to the dormant mode is not immediate. If the aforementioned condition is met, the macro-ABS enters a transitional period in which it does not admit any new customers and will finally enter the dormant state when the last customer in service departs.

4.4 Frequency planning and channel assignment

The BuNGee fixed frequency planning approach has been adopted for this scenario. This approach is proposed by the BuNGee project as in [106]. The access network uses four 10 MHz dedicated frequency bands. Street-ABSs serving North-South (N-S) streets use two different frequency bands from those serving East-West (E-W) streets. Beams pointing in opposite directions use two different bands. Macro-ABSs use one dedicated frequency band each. The capacity of both street-ABSs and macro-ABSs is the same. Channels are assigned based on a best SINR (Signal-to-Interference plus Noise-Ratio) basis.

4.5 **Energy model**

ABSs are the major source of energy consumption in the access network as previously discussed in Chapter 1. The energy consumed by an ABS depends mainly on the time spent in each mode. Equation (4.7) shows the energy consumption of the access network [53] and Table 4.1 provides the values for the energy model parameters.:

$$E_{ABSs} = \sum_{i=1}^{N} (t_{off,i} P_{off} + t_{RX,i} \frac{P_{RX}}{\mu_{RF}} + t_{TX,i} \frac{P_{TX}}{\mu_{RF}} + t_{idle,i} P_{idle} + \eta_i E_{wakeup}) \left(\frac{1}{1-\mu_{sl}} \right) \tag{4.7}$$

Parameter	***Value***
Power in receiving mode	5 [W]
Power in idle mode	5 [W]
Power in dormant mode	0.25 [W]
Efficiency of RF	20%
Efficiency of supply loss	10%
ABS max transmit power	5 [W]

Table 4.1 Energy Model Parameters [53, 107, 108]

Where, in general, t refers to time and P to power. N is the total number of ABSs, $t_{off,i}$ P_{off} is the energy consumed by the i^{th} ABS in the dormant mode, η_i is the number of times the i^{th} ABS switches to the active mode, $E_{mwakeup}$ is the energy consumed in the waking up process, μ_{RF} is the efficiency of the RF amplifier, and finally μ_{sl} is the battery back-up and power supply loss. As shown, the energy is consumed even when an ABS is in the dormant state as it still needs to constantly listen to the serving HBS and the neighbouring ABSs to know when they need to be activated. The energy consumed in the active mode is divided into three parts, the energy consumed in transmitting $t_{TX}P_{TX}$, the energy consumed in receiving $T_{RX}P_{RX}$, and the energy consumed doing neither, but remaining idle providing coverage and connectivity $t_{idle}P_{idle}$.

4.6 **Results**

4.6.1 **Theoretical bounds for energy savings**

As mentioned earlier, the amount of energy savings is highly dependent on the time each ABS spends being in the dormant and active modes. Here we present a performance bound for the maximum number of ABSs that can be deactivated. From [104], the number of transmitters needed at a given traffic level for an infinite wireless network is as follows:

$$OT_{total} = \xi . U . \sqrt{n} \tag{4.8}$$

Where U is the capacity of an ABS in terms of number of users and OT_{total} is the offered traffic in Erlangs and n is the number of transmitters and ξ is the constant factor of the capacity behaviour of the wireless network system. Hence, the number of transmitters needed to serve a certain amount of traffic determined by OT_{total} is,

$$n = \left(\frac{OT_{total}}{U_{ABS}}\right)^2 \cdot \frac{1}{\xi^2} \tag{4.9}$$

Then,

$$n_{off} = N - n \tag{4.10}$$

Where n_{off} is the number of transmitters which can be deactivated out of the total number of transmitters deployed N. n is upper-bounded by the idealised ABS capacity, hence,

$$\xi \leq \sqrt{\frac{OT_{total}}{U_{ABS}}} \tag{4.11}$$

An approximate value of ξ for the system in question is calculated using the Erlang-B blocking probability formula to estimate the maximum traffic load which an individual ABS can serve while maintaining the blocking probability less than a certain level as follows [2],

$$BP = \frac{(OT_{total}^{U}/U!)}{\sum_{i=0}^{U} \frac{OT_{total}^{i}}{i!}} \tag{4.12}$$

Intuitively, we can calculate the maximum offered traffic in Erlangs that can be supported by an ABS while keeping the level of blocking less than *bp* as follows,

$$OT_{BP=bp} = \max_{BP=bp}(OT) \tag{4.13}$$

Hence, the value ($OT_{BP=bp}$) can be used to calculate the constant ξ as follows:

$$\xi = \sqrt{\frac{OT_{total}}{U_{ABS} \cdot \frac{U_{ABS}}{OT_{BP=bp}}}} \tag{4.14}$$

The value of 0.002 is chosen for bp. This value is chosen to have a comparable theoretical indicator to our results as the schemes presented here aim to keep the blocking to the minimum level (approximately zero) when not congested.

4.6.2 **Simulation results**

Simulation has been used to assess the performance of the schemes described above. The simulated architecture is composed of 105 ABSs deployed in a 13.5×13.5 km area having 5000 users randomly distributed along the streets. Users arrive into the system with exponentially distributed inter-arrival times. File sizes are also exponentially distributed with an average size as described in the Table 4.2. The WINNER II propagation model has been used as detailed in [109].

Parameter		***Value***
Street width		15 [m]
Building block size		75 × 75 [m]
Street-ABS max antenna gain		17 [dBi]
Carrier Frequency	Street-ABS	3.5 [GHz]
	Macro-ABS	2.0 [GHz]
SINR threshold		1.8 [dB]
Noise floor		-114 [dBm/MHz]
Average transmit rate		4 [Mbps]
Average size of a file		1 [MB]
Antenna height	Macro-ABS	25 [m]
	Street-ABS	5 [m]
	MS	1.5 [m]
MS transmission power		23 [dBm]
Street ABS-MS propagation model		WINNER II B1
Macro ABS-MS propagation model		WINNER II B5
$OT_{BP=0.002}$		10 [Erlangs]
Delay threshold ($\boldsymbol{D_{thrshld}}$)		2.1 [sec]
Deactivation threshold ($\boldsymbol{C_{off}}$)		1
Street-ABS activation threshold ($\boldsymbol{C_{on}}$)		90%
Macro-ABS activation threshold ($\boldsymbol{C_{thrshld}}$)		90%

Table 4.2 Simulation Parameters [108]

Different scenarios have been developed in order to clearly evaluate the proposed scheme. The first scenario is where macro-ABSs stay always in the active mode. This increases the maximum throughput of the system and enhances the QoS at high offered traffic and adds flexibility at lower traffic intensities for the TM to operate. The second scenario is where the macro-ABSs are used only at low traffic loads. These scenarios are compared with the Neighbouring-Based Topology Management (NBTM) scheme introduced in Subsection 4.3.1. Figure 4.3 presents the average number of dormant ABSs for the various schemes. The Infinite Area Upper-Bound (IAUB) gives a theoretical indication about the average number of ABSs which can be deactivated while keeping the BP to its minimum. This represents a good upper bound model at low traffic levels, but is less accurate at high offered traffic levels (>1000 Erlangs in this case). The main reason for the discrepancy is that this is an infinite area upper bound whereas the simulation architecture is finite having edge ABSs which are dormant most of the time.

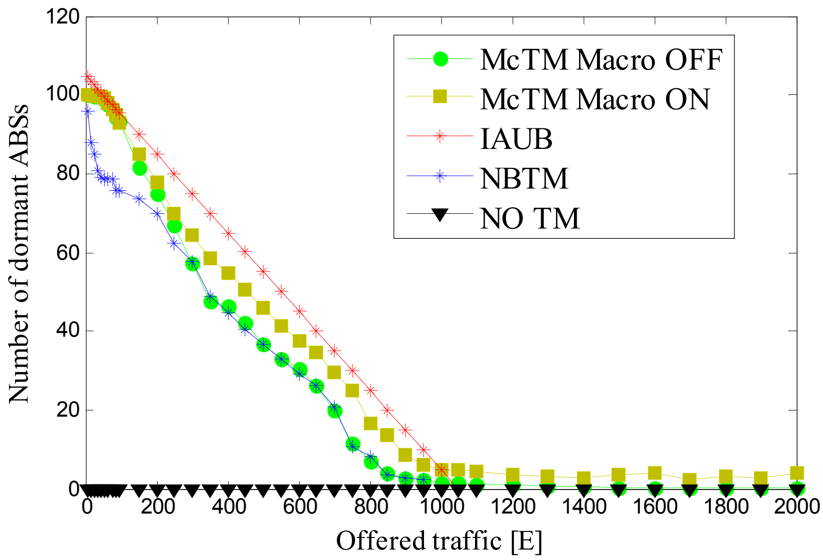


Figure 4.3 Average number of ABSs in the dormant mode vs Offered Traffic

Figure 4.4a illustrates the percentage of energy savings, while the energy consumed per file delivered is shown Figure 4.4b.

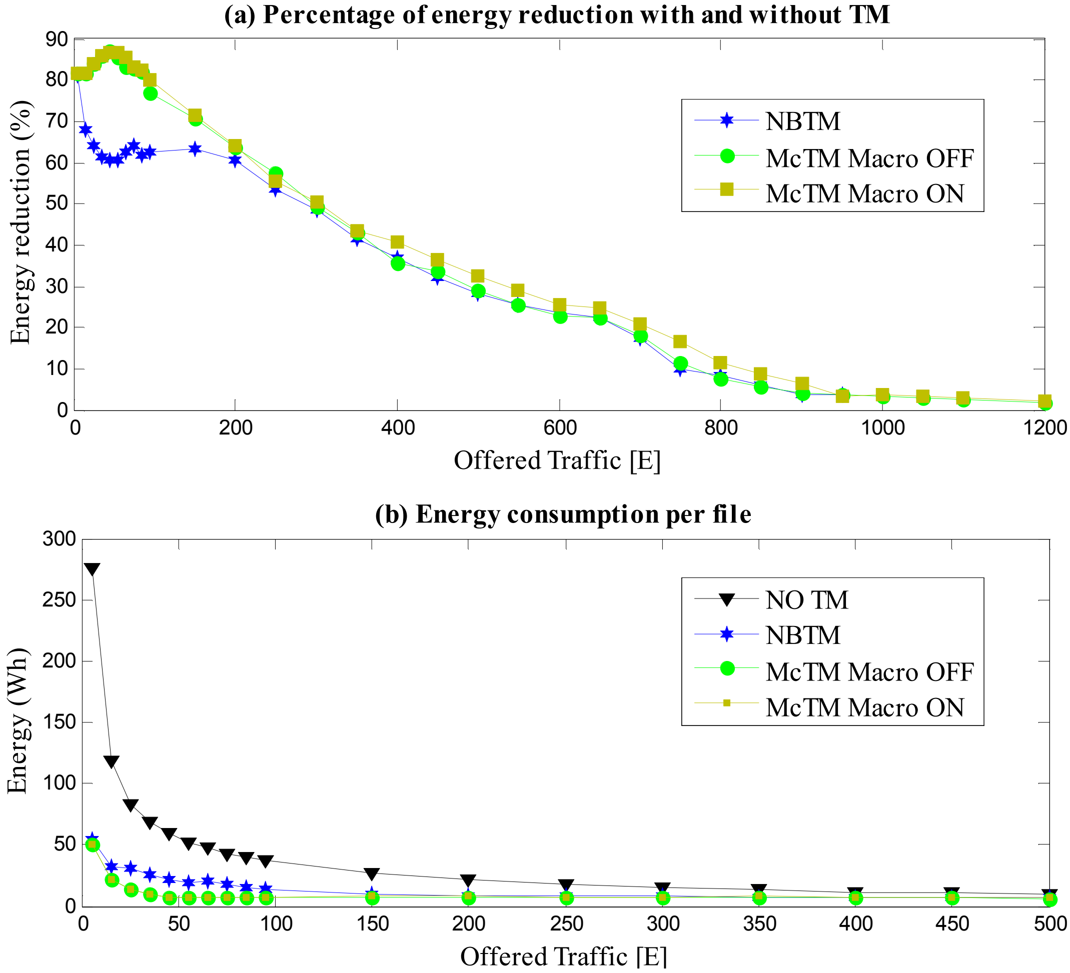


Figure 4.4 Energy Savings and Consumption vs Offered Traffic

The percentage of energy savings exceeds the 85% when using the macro aided topology management scheme at low traffic loads. The difference in energy savings between the macro aided scheme and the neighbour-based scheme is quite noticeable saving the former around 20% more on average at low traffic loads and an average of 5% afterwards. This difference is attributed directly to the number of dormant ABSs at each traffic level. When using omnidirectional macro level antennas, only the macro-ABSs are needed to totally cover the area, whereas when using below roof-top directional antennas it is hard to cover all the area using a small number of street-ABSs as these would switch to the dormant mode at low traffic load resulting in coverage holes. The percentage of energy saving reduces to almost zero at high traffic loads as the system needs all ABSs to be in the active mode. In addition, the macro aided scheme still outperforms the neighbour-based scheme by having the lowest energy consumption per file delivered ensuring that the energy has not been saved by prejudicing the QoS (the number of files delivered in this case).

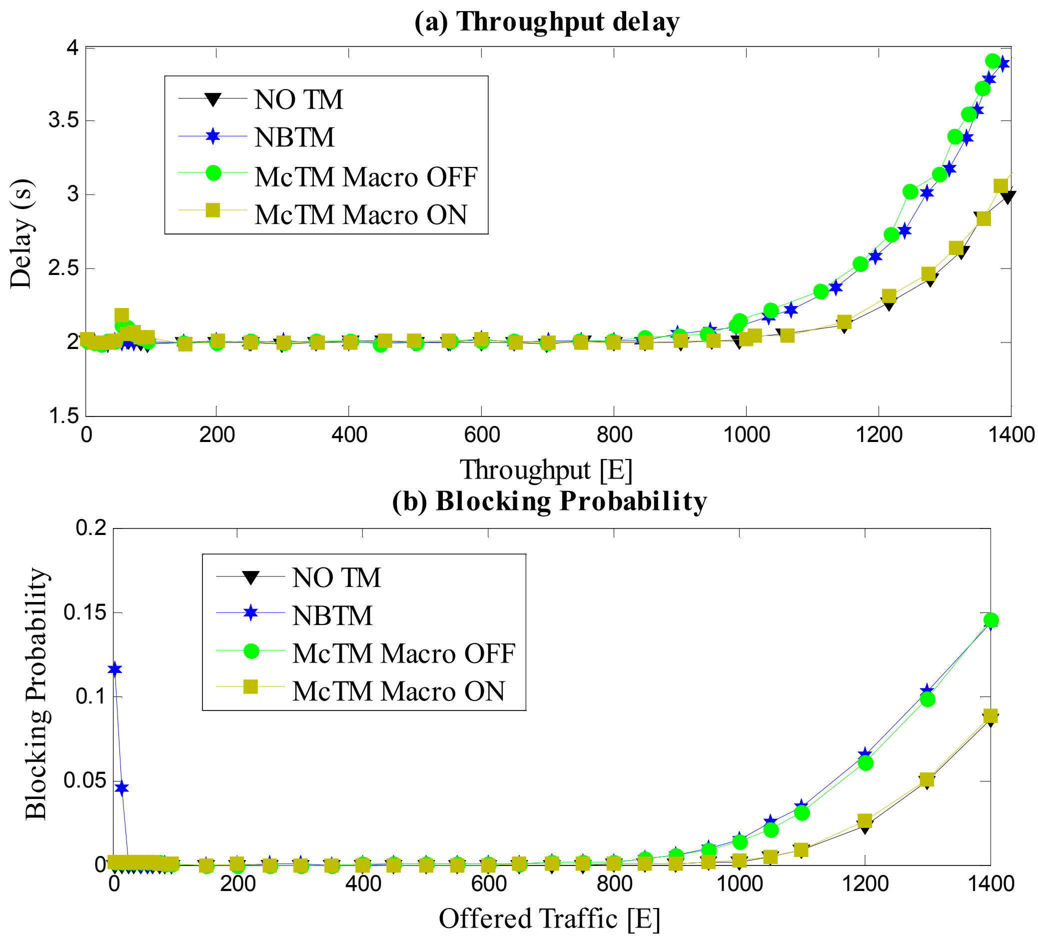


Figure 4.5 QoS vs Throughput and Offered Traffic

Figure 4.5 shows the system QoS when using these different topology management schemes in terms of blocking probability and end-to-end delay. In the case of the NBTM, deterioration in the system QoS is observed in terms of blocking probability at low traffic loads. This is attributed to the coverage holes resulted from deactivating an excessive number of ABSs indicating that a more complex scheme needed to be used. However, this is resolved by the introduction of the macro-cell overlay and the use of the McTM.

The McTM results in some delays at low traffic loads having a peak of just above 2.1 seconds. This is due to the delay threshold (2.1 seconds) used by the macro-ABSs as a condition to activate street-ABSs. This shows that the scheme can respond quickly enough to deal with medium and high traffic fluctuations. We can also notice the difference between the throughput with and without the macro-ABSs active at high traffic loads as the macro-ABSs increase the system capacity. The usage of macro-ABSs at high traffic loads depends on the preferences of the service provider and / or area requirements. In addition, the McTM works well at low traffic loads, having no impact on the QoS as only the macro-ABSs are needed to fully serve customers and cover the service area.

4.7 **Conclusion**

This chapter investigated a range of green topology management schemes for beyond next generation mobile broadband systems. The study is important as; firstly, the system taken into consideration is different from traditional deployments thereby imposing different challenges to the application of energy-aware topology management. Previously proposed techniques require accurate traffic density estimation, a number of information exchange iterations between neighbouring cells, with further status estimation calculations and information processing, or require keeping track of the network topology formed by BSs that are operational. The schemes proposed here do not require any prior information about the nature of the traffic in the area and are designed to cope with rapid changes in traffic conditions by using instantaneous traffic measurements to derive different policies and switching conditions. The schemes are also able to perform in an efficient manner without any QoS feedback from the network and are designed to work in different network deployment strategies, hence no need to track the network topology formed by BSs that are operational. These schemes have been tested under spatially even traffic distribution and as Chapter 3 suggested, the macro cell-overlaid scheme outperforms the neighbour-based scheme, which outlines the benefits of having a macro level coverage, especially at low traffic loads from both the energy saving as well as QoS perspectives. The macro aided scheme is able to maintain the system QoS (end-to-end delay and blocking probability) while managing to save an average of 80% of energy consumed without topology management at low traffic loads. At medium offered traffic loads the energy savings are around 30%. In the next chapter, spatially uneven traffic distribution is investigated to develop a topology management scheme that can be applied to a wider range of different temporospatial traffic conditions.

Chapter 5 **Energy-Aware Topology Management for Opportunistic Temporary Event Networks**

Contents

5.1 **Introduction**

Thus far, the application of energy-aware topology management has been assessed under a spatially even traffic distribution. However, as discussed in Chapter 1, the fact that only 10% of the BSs in a network carry more than 55% of the aggregate network traffic indicates the common existence of spatially uneven traffic distributions across the access network. Also, special events such as open air festivals and marathons pose yet additional complications in the operation of energy-aware topology management schemes. Opportunistic networks are being proposed to augment and provide additional capacity density at the occurrence of such events containing eNBs powered by batteries of limited capacity. Such scenarios are taken into consideration in this chapter in order to develop an efficient topology management scheme that controls the roll out and roll back process of opportunistic networks taking into consideration a wider range of traffic conditions, including the occurrence of a temporary event. A hotspot of variable densities is designed to simulate the occurrence of a temporary event. The impact of having such uneven traffic distributions is explained in Section 5.2. Section 5.3 describes the scheme which, unlike its counterparts, aims to neutralise the impact of the high traffic dynamics expected in urban areas or during a special event by clustering entities which can cooperate to maximise the area throughput and network energy efficiency. The network model and KPI calculation are described in Section 5.4. Lastly, Section 5.5 assesses the scheme introduced in this chapter and compares its performance to the performance of the topology management scheme introduced in Chapter 4.

5.2 5G Network Model for Temporary Event Scenario

The city model temporary event scenario represents a scenario where an event such as a festival or city Marathon is held in a dense populated city. A conventional cellular network is well deployed in most cases but a capacity density constraint occurs during the event period [110]. The opportunistic network will be densely deployed in conjunction with the operators' conventional networks to enhance the regional capacity density by providing more access points and resource blocks to the UEs (User Equipment). Dynamic roll out / back and energy consumption is a crucial issue for portable light-weight eNBs densely deployed in a wide city area. The topology management algorithm controls the number and service time of these eNBs based on the local traffic variations and rolls back the network when the event is ended. Figure 5.1 presents the BuNGee architecture which is described in detail Chapter 4. The BuNGee architecture [111] has been adopted for the temporary event city scenario in the European FP7 ABSOLUTE project to alleviate traffic in a commercial or residential area during a temporary event. ABSOLUTE aims to design and validate an innovative rapidly deployable future network architecture which is resilient and capable of providing broadband multiservice, secure, and dependable connectivity for large coverage areas affected by large scale unexpected events, which leads to the demand for very high throughput and augmented network capacity [112]. The study of the disaster relief network is presented in Chapter 6. The temporary event is modelled as a hotspot as illustrated in Figure 5.1. The user density in the hotspot area is defined by the HotSpot Densification Factor ($HSDF$) which determines the number of users in the hotspot as follows:

$$U_{hs} = \frac{\mathrm{U_{SA}.HSDF}}{(\text{number of blocks} + \mathrm{HSDF} - 1\,)} \tag{5.1}$$

$$U_{hsd} = \frac{U_{hs}}{Area_{hs}} \tag{5.2}$$

$$U_{nhsd} = \frac{U_{SA} - U_{hs}}{Area_{nhs}} \tag{5.3}$$

Where U_{hs} is the number of users in the hotspot, U_{hsd} and U_{nhsd} are the hotspot and out-of-the-hotspot user density respectively, U_{SA} is the number of users in the service area (including the hotspot), $Area_{hs}$ and $Area_{nhs}$ are the hotspot area and normal traffic area respectively, and the number of blocks is 9 as illustrated in Figure 5.1. The value of HSDF indicates the density of users inside the hotspot with respect to the density of users outside the hotspot. A value if HSDF=1 indicates equal densities in both areas. A values of HSDF=2 indicates that user density in the hotspot area is twice as much the density of users outside the hotspot area.

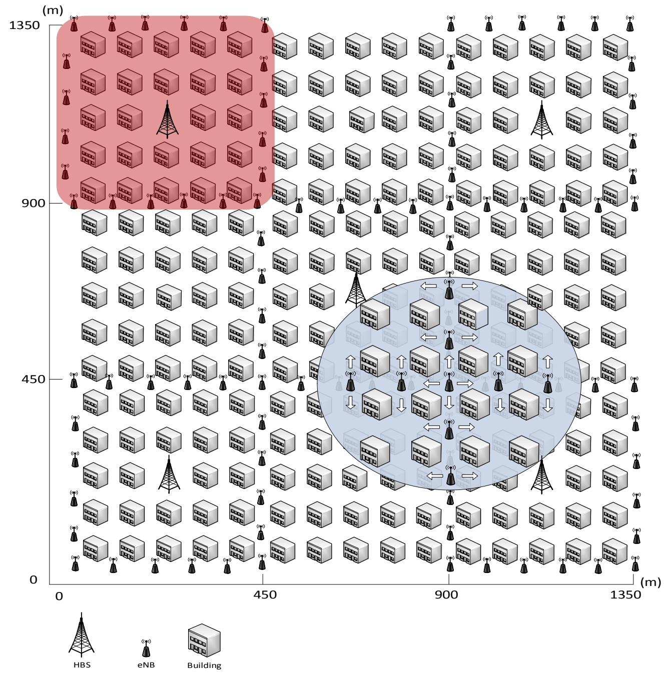


Figure 5.1 BuNGee architecture for temporary event Scenarios

5.3 Energy-Aware Topology Management for High Capacity Density Temporary Event Networks

As described in Chapter 4, energy-aware topology management aims to dynamically control the predesigned topology of the system in order to achieve energy savings while meeting the required QoS. In the context of opportunistic deployments, the topology management is also responsible for the roll out and roll back of the network. This is achieved by fine-tuning the status of the eNBs (active / dormant) depending on the traffic demands. The scheme presented here, the Cluster-Based Topology Management ($CBTM$) scheme [65], is based on the identification of entities which help provide service to a shared area and which can cooperate to maximise the area throughput as well as network energy savings. In other words, a randomly chosen area from the service area is covered by multiple eNBs. These eNBs, which are called a cluster, cooperate in order to deliver the required QoS at minimum energy cost. For the architecture presented here, the service area has 14 North-South and 14 East-West streets each street representing a cluster. Figure 5.2

shows the clusters that could alleviate the impact of having a hotspot by making use of the directional antennas. The $CBTM$ scheme introduced here also uses the observed instantaneous traffic density for the derivation of its policies as proven effective in chapter 4. However, this is applied at the cluster level as detailed in the next subsection.

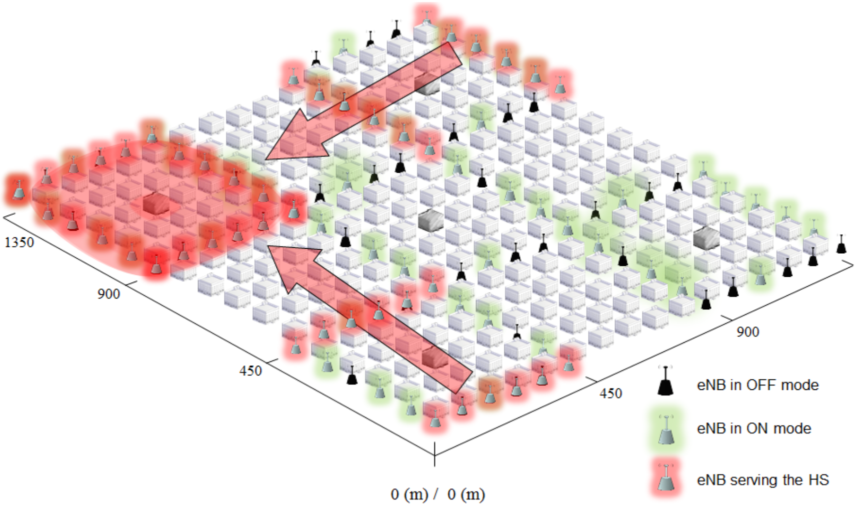


Figure 5.2 Cluster formed in the presence of a hotspot

5.3.1 **Topology management operation**

At least one eNB in each cluster is permanently in the active-mode in order to ensure street coverage. When the cluster capacity exceeds a predefined capacity threshold, a cluster member is signalled to wake up by an active eNB in the cluster. The maximum capacity of a cluster is a function of the number of members in the active-mode. Since the proposed scheme is based on load-derived policies, an updated neighbour list as well as the loading status of each node needs to be available at the time of the decision making for the algorithms to be feasible. The neighbour list is provided and kept up to date by the *Automatic Neighbour Relation (ANR)* management function which deals with automatic *Neighbour Relation (NR)* removals or additions. The ANR function is essential to Hand Over (HO) operations and to the $X2$ interface setup which enables information exchange between nodes. Loading statuses are made available by using the Load Reporting function which supports the mobility load balancing function. The *Load Reporting* function enables the exchange of cell specific load information between neighbouring $eNBs$. This function is run over the $X2$ and $S1$ interfaces depending on the RAT (Radio Access Technology) of the neighbouring node. $X2$ interface is used for intra-LTE and $S1$ is necessary for inter-RAT communications [113].

5.3.1.1 *Switching to the active mode:*

As the opportunistic network is deployed, the roll out is triggered when the traffic load increases. Initially, there is one eNB active at each cluster in order to ensure all streets are covered. This means that a minimum of 28 eNBs are active in the service area at any giving time. Afterwards, when the cluster capacity exceeds C_{on}% of the maximum cluster capacity, the cluster head uses the $X2$ interface to request the activation of other cluster-members by means of the *Cell Activation* procedure [113]. The capacity of the cluster is a function of the number of members being active. If an inter-RAT $eNBs$ is configured and capable of requesting the activation of an eNB, the *MME Direct Information Transfer* procedure is used over the $S1$ link. This procedure can be described as follows:

$$CeNB \in \{CeNB_{i,j} \; : \; \forall i \in [1,28] \;\; \forall j \in [1,4]\} \tag{5.4}$$

$$if \quad C_i > C_{th\_on} \tag{5.5}$$

$$D_{eNB} = CeNB_{i,j} \;\; \in \{CeNB : \; \mathrm{i}\, , \forall \mathrm{j} \in [1,4]\} \tag{5.6}$$

Where $CeNB_{i,j}$ is the jth eNB in the ith cluster, C_i is the actual capacity of the ith cluster, C_{th_on} is the capacity threshold to activate a dormant eNB, and finally D_{enb} is the eNB in the dormant-mode to reactivate. C_{th_on} is calculated from equation (5.10).

5.3.1.2 *Switching to the dormant mode*

An eNB is switched to the dormant mode only when the cluster has excess resources. The concerned eNB needs to inform all peer $eNBs$ about the action taken over the $X2$ interface by using the *eNB Configuration Update* procedure. The inter-RAT eNB can also be informed by using the *eNB Direct Information Transfer* procedure over $S1$. For quicker reconfiguration and synchronisation, all $eNBs$ need to maintain the configuration data of the dormant eNB such as neighbour relationship configuration [113]. This can be expressed as:

$$CeNB \in \{CeNB_{i,j} \ : \ \forall i \in [1, 28] \ \ \forall j \in [1, 4]\} \tag{5.7}$$

$$if \quad C_i < C_{th\_off} \tag{5.8}$$

$$C_{eNB} = CeNB_{i,j} \ \in \{CeNB : \ \mathrm{i}, \forall \mathrm{j} \in [1, 4]\} \tag{5.9}$$

However, in this case, C_{eNB} is the eNB in the active mode to deactivate. C_{th_off} is the capacity thresholds to switch to the dormant mode and can be calculated as follows:

$$C_{th\_on} = C_{on} C_{cmax} \tag{5.10}$$

$$C_{th\_off} = C_{off} C_{cmax} \tag{5.11}$$

Where C_{on} and C_{off} are percentage values to set the activation and deactivation thresholds respectively, and C_{cmax} is the maximum capacity of the cluster which is dependent on the number of active eNBs as:

$$C_{cmax} = \ Nm \tag{5.12}$$

Where N is the number of eNBs in the active-mode which can still admit new users, and m is the number of channels of an eNB. Switching to the dormant-mode is not immediate. If an eNB is chosen to transit to the dormant-state, the eNB enters a transitional state in which it does not admit any new customers and will finally switch to the dormant-mode when the last customer in service departs. At the end of the temporary event, the entire opportunistic network will be rolled back as all eNBs enter the dormant mode.

5.4 **Analytical model**

Due to the spatial coverage isolation provided by high buildings and the use of below-roof top directional antennas, the system is split into 28 isolated service areas forming 28 clusters, each providing service to a different street. The analytical model is used to analyse two clusters separately, one hotspot cluster and one non-hotspot cluster. Each cluster contains 4 cluster members representing a street. Hotspot clusters are involved in serving the traffic at the hotspot. Multidimensional Markov analysis is used to analyse each cluster separately and expand the results to predict system behaviour of a larger network of similar characteristics having a combination of hotspot and non-hotspot clusters. Since the model is four-dimensional as each BuNGee street is covered by 4 eNBs, an incoming user has up to four choices of eNBs in the access network depending on the number of eNBs in the active-mode. The mathematical representation of the traffic distribution and markovian network model is described in Chapter 3.

5.4.1 **Key performance indicators**

Key Performance Indicators (KPIs) considered here are similar to those derived in the previous chapters. These are the average number of eNBs in the dormant-state, system power requirements, blocking probability, and throughput. In our calculations, we assume that eNBs are ranked as eNB_x, eNB_y, and eNB_z. eNB_x being the most suitable to switch-on and eNB_z the least. The cluster head is denoted as eNB_{ch}. For eNB_x, the time spent in the dormant-mode (T_{dx}) is equivalent to the sum of all state probabilities where the number of channels occupied (J) is less than C_{th_off} and eNB_x is serving no users:

$$T_{dx} = \sum_{j_{ch}}^{m}\sum_{j_y}^{m}\sum_{j_z}^{m} P\left(j_{ch}, j_x, j_y, j_z\right)|_{[\,J<C_{th\_off}\ \&\ j_x=0]} \tag{5.13}$$

Similarly, we can calculate the time spent in the dormant-mode for eNB_y (T_{dy}), and eNB_z (T_{dz}) as follows:

$$T_{dy} = \sum_{j_{ch}}^{m}\sum_{j_x}^{m}\sum_{j_z}^{m} P\left(j_{ch}, j_x, j_y, j_z\right)|_{[\,J<C_{th\_off}\ \&\ j_y=0]} \tag{5.14}$$

$$T_{dz} = \sum_{j_{ch}}^{m}\sum_{j_x}^{m}\sum_{j_y}^{m} P\left(j_{ch}, j_x, j_y, j_z\right)|_{[\,J<C_{th\_off}\ \&\ j_z=0]} \tag{5.15}$$

The system blocking probability (PBs) is the probability of the system being in the (m, m, m, m) state

$$PB_s = P\ (m, m, m, m) \tag{5.16}$$

The system throughput (TH_s) on the other hand is directly derived from the blocking probability as, by definition, it is the portion of the offered traffic (OT) that is not lost:

$$TH_s \ = \ OT[1 - PB_s] \tag{5.17}$$

System power requirements (PW_s) can be also calculated from this model using the steady-state probabilities as the power model adopted in this study relies on the time spent at each state as follows:

$$PW_s = \sum_{n=1}^{e}\sum_{u=1}^{m} PW_u P_n(u) \tag{5.18}$$

Where PW_u is the power required to serve u number of users which is calculated using the energy model described in Section 5.5, $P_n(u)$ is the steady-state probability of having u number of users being served by the n^{th} eNB, and e is the number of eNBs.

5.4.2 **Scalability of the KPIs**

The four-dimensional Markov chain describes the behaviour of a cluster which is a street in the BuNGee architecture. Thus, the results from the analytical model can be used to predict an accurate estimation of the behaviour of a larger network containing a larger number of streets by making use of the isolation provided by the high buildings and the deployment of below roof-top directional antennas. This is used as an alternative way to generalise the results obtained from the four-dimensional model given the fact that the computation time increases exponentially as the number of dimensions increases. In this section, the scaling-up procedure of different KPIs is described.

- *Blocking probability*

The system blocking probability of a network composed of k streets, or clusters, can be calculated by combining the individual clusters blocking probability by taking into consideration the different average inter-arrival times of users into different clusters. This can be expressed as:

$$BP_s = \sum_{i=1}^{k} \frac{\lambda_{ci}}{\lambda} BP_{ci} \tag{5.19}$$

Where BP_{ci} is the blocking probability obtained by using the analytical mode of cluster i, and k is the number of clusters. λ is the average user arrival rate into the system. λ_{ci} is the average user arrival rate into cluster i where $\sum_{i=1}^{k} \lambda_{ci} = \lambda$. λ_{ci} is calculated from the user density in individual clusters:

$$\lambda_{ci} = (U_{hsd}Area_{hsi} + U_{nhsd}Area_{nhsi})\,\lambda_u \tag{5.20}$$

Where U_{hsd} and U_{nhsd} are calculated from equation (5.2) and (5.3) respectively. $Area_{hsi}$ and $Area_{nhsi}$ are the hotspot and none-hotspot areas that cluster i covers, and λ_u is the average user arrival rate into the system which is equal for all users and depends on the offered traffic.

- *Power and average number of eNBs in the dormant-mode*

Unlike the blocking probability, the system power requirements as well as the average number of eNBs in the dormant-mode ($DeNB_s$) is a direct summation of the results obtained from the analytical model as follows:

$$PW_s = \sum_{i=1}^{k} PW_{ci} \tag{5.21}$$

$$DeNB_s = \sum_{i=1}^{k} DeNB_{ci} \tag{5.22}$$

Where PW_{ci} and $DeNB_{ci}$ are the power requirements and average number of eNBs in the dormant-mode for cluster i, and k is the number of clusters. The scaled-up results obtained from the analytical model are compared with simulation results in the subsequent section.

5.5 Results and discussion

The topology management introduced in this chapter is assessed using the mathematical model described in Chapter 3 which assumes that all UEs within a cluster can access the four cluster members, hence $\Psi = 4$. In addition, and for a better understanding of the limitation of the proposed topology management scheme, a Monte

Carlo simulation is also conducted modelling the propagation channel and taking interference into account. The results of the mathematical model are presented first.

5.5.1 **Numerical results**

The topology management algorithm is tested and validated under different values of HotSpot Densification Factor (*HSDF*). A network of 20 eNBs is taken into consideration as illustrated in Figure 5.3. This network is extracted from the original BuNGee architecture in Figure 5.1. Each street forms a cluster with 4 cluster members creating a 5-cluster network. A hotspot is created at the top-left corner of the network as described in Section 5.2. Values of 1, 2, 3, and 4 for $HSDF$ are used to represent even, moderately uneven, highly uneven traffic distribution, and a temporary event respectively. Users are uniformly distributed along the streets with exponentially distributed inter-arrival times. The analytical model assumes that users can only connect to the cluster which is covering their street depending on the occupancy of the eNBs.

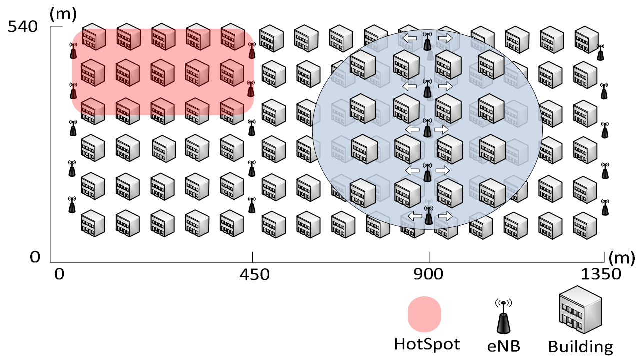


Figure 5.3 Network Model

The topology management scheme introduced in Chapter 3 is used as a benchmark algorithm for comparison. The Neighbour-Based TM (*NBTM*) as previously introduced is a capacity based algorithm in which an eNB is deactivated if it is idle and activates a random neighbour upon reaching 90% of its maximum capacity, while having no or little knowledge about the capabilities of the latter in alleviating the traffic demand.

Parameter	***Value***
Service rate (μ)	0.01
Number of channels (m)	4
Number of eNBs (D)	20
Load-dependency constant (β)	1.25
Power in the idle-mode (P_0)	60 [W]
Power in the dormant-mode (P_d)	20 [W]
Maximum transmit power (P_{Tx})	5[W]
C_{off} and C_{on}	10%, 90%

Table 5.1 Simulation Parameters [99], [102]

The power model used here is adopted from the EARTH project [99]. The model can be expressed as:

$$\mathrm{P_{supply}} = \begin{cases} \mathrm{P_0} + \beta \mathrm{P_{Tx}} & ; \quad 0 \leq \mathrm{P_{Tx}} \leq \mathrm{P_{max}} \\ \mathrm{P_d} & ; \quad \mathrm{P_{Tx}} = 0 \end{cases} \tag{5.23}$$

Where $\mathrm{P_{max}}$ is the maximum transmit power, $\mathrm{P_0}$ and $\mathrm{P_d}$ are the idle- and dormant-mode power consumption of the eNB respectively, $\mathrm{P_{Tx}}$ is the instantaneous transmit power, and β is the load dependency constant. Table 5.1 contains the detailed simulation parameters. Note that in all graphs to be presented the scaled-up results obtained from the analytical model are labelled by a circle marker and no line.

Figure 5.4 illustrates the system-average number of eNBs in the dormant-mode for various configurations. It can be seen from Figure 5.4 that the scaled-up results obtained from the analytical model are consistent with the simulation results at all traffic loads and different values of $HSDF$. Also, it is clear that the CBTM outperforms the state-of-the-art topology management scheme as it successfully identifies eNBs which could help serve the increase in traffic demand keeping the rest in the dormant-mode. To have a more realistic comparison with fewer constraints on boundary conditions, users in a street are allowed to connect to one cluster member of a neighbouring cluster. This is denoted as “5 acc” in the legend meaning 5 accessible eNBs instead of only 4. As shown, there is very little difference as far as the number of eNBs in the dormant-mode is concerned, making our assumptions regarding boundary conditions valid. A surprising fact however, is that the CBTM makes use of fewer eNBs at high offered traffic when $HSDF$ is set to 4 in comparison with $HSDF = 2$. Again, this shows the effectiveness of the algorithm in responding to localised traffic fluctuations.

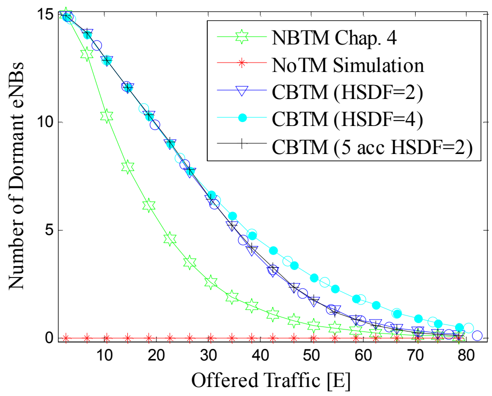


Figure 5.4 Average no. eNBs in the dormant-mode

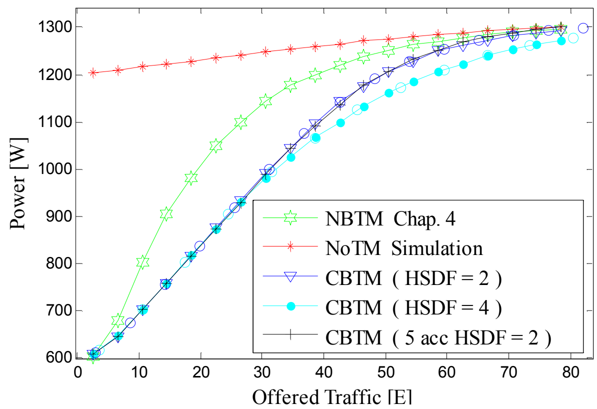


Figure 5.5 System power requirements

This behaviour is attributed to the shift in traffic towards the hotspot area at high values of *HSDF*, which can be served by a limited number of eNBs. The CBTM makes use of this traffic shift and keeps more eNBs in the dormant-mode in clusters which are not involved in serving the hotspot area, showing its capability of identifying the eNBs which can alleviate the demand from highly loaded eNBs unlike the NBTM scheme which, due to the lack of information and its capacity-based nature, reactivates eNBs irrespective of their real contribution in covering the high density traffic area.

Figure 5.5 presents the system power requirements at different traffic loads for different system settings. This figure gives an insight into the energy reduction capabilities of both schemes when using the energy model described above. The CBTM can achieve up to 50% reduction in energy consumption at low load conditions. Also, the power consumption scales more uniformly with traffic when using the CBTM over the NBTM scheme. Figure 5.6 on the other hand, compares the system QoS in terms of system blocking probability when using the CBTM with the QoS when no topology management is used and with the state-of-the-art scheme introduced in the previous chapters for different system configurations and $HSDF$ is set to 2.

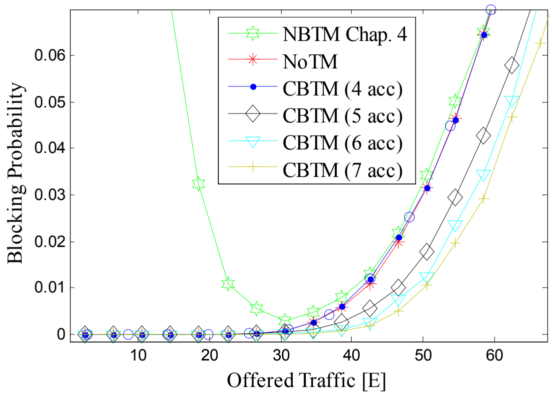


Figure 5.6 Blocking probability (*HSDF*=2)

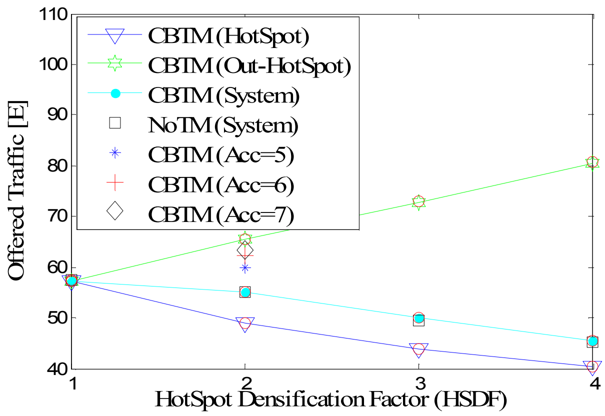


Figure 5.7 Offered Traffic Supported at 0.05 Blocking Probability

It is shown in Figure 5.6 that the QoS delivered by the CBTM is no different from the QoS delivered when the whole system is kept in the active-mode under the constraint of having one cluster accessible per user (4 accessible eNBs). However, if more flexibility is allowed and users connect to neighbouring clusters (5, 6, and 7 accessible eNBs), a slight difference in the performance is noticed. Also, the CBTM clearly outperforms the NBTM scheme at medium and low traffic conditions. This is mainly due to the fact that the

CBTM keeps at least one eNB in the active mode per cluster ensuring the whole service area is covered whereas most eNBs in the benchmark scheme are deactivated after meeting the requirements to switch-off leaving large coverage holes resulting in high blocking probability. Figure 5.7 gives more localised QoS measurements for different values of HSDF in terms of the system offered traffic supported at 0.05 blocking probability of individual clusters. When HSDF is set to 1, there is no difference between the system and localised blocking probability as traffic is equally distributed between different areas. However, as traffic starts to concentrate around the hotspot area with densification factor of HSDF, the localised behaviour becomes different from the system behaviour. Hotspot areas reach the 0.05 blocking probability constraint at an earlier stage than the none-hotspot areas. This difference increases as HSDF reaches the value of 4. System behaviour is more affected by the hotspot than the rest of the service area especially for high values of HSDF as more traffic is generated at the hotspot. Again, the close behaviour of our results compared with a more flexible system having 5, 6, and 7 accessible eNBs can be seen.

5.5.2 **Simulation results**

In order to accurately assess the effectiveness of the topology management introduced in this chapter, a Monte Carlo simulation is also used under different values of $HSDF$ for the network shown in Figure 5.1. These scenarios are compared with the scheme introduced in the previous chapter, the Neighbour-Based TM (*NBTM*). Table 5.2 summarises the simulation parameters. Figure 5.8 presents the system and cluster-average number of dormant eNBs for the various schemes and values of HSDF. The cluster-average dormant eNBs are the number of eNBs in the dormant mode which are members of the clusters that cover the hotspot area. It is clear from this figure that the CBTM performs better than the scheme presented in Chapter 3 under uneven traffic distribution.

Even though both schemes need more eNBs in the active mode when HSDF is 4 compared to when HSDF is 2 at low and medium traffic loads, they have different behaviours at high traffic loads.

Parameter	***Value***
Street width	15 [m]
Building block size	75 × 75 [m]
eNB max antenna gain	17 [dBi]
Frequency	3.5 [GHz]
SINR threshold	1.8 [dB]
Noise floor	-114 [dBm/MHz]
Average transmit rate	4 [Mbps]
Average size of a file	1 [MB]
Standard deviation of the normal distribution	170
Coordinates of mean point of the Normal distribution	(225 m,1800 m)

Table 5.2 Simulation Parameters [108]

Surprisingly, the CBTM uses fewer eNBs when HSDF is 4 compared to the eNBs usage when HSDF is 2. This is due to the fact that most traffic is shifted to the hotspot area with a high value of HSDF which can be served by a limited number of eNBs. The scheme only activates eNBs in clusters that help serve the hotspot area whereas the neighbour-based (*NBTM*) scheme activates eNBs oblivious to their contribution in covering the high density traffic area. In addition, the CBTM uses more eNBs at low traffic loads (<200 E) than the NBTM scheme which can be contributed to the fact that the CBTM uses a minimum of one eNB in each cluster to ensure the whole service area is covered (28 eNBs). Figure 5.8(b), on the other hand, illustrates the behaviour of the eNBs in the clusters which serve the hotspot area.

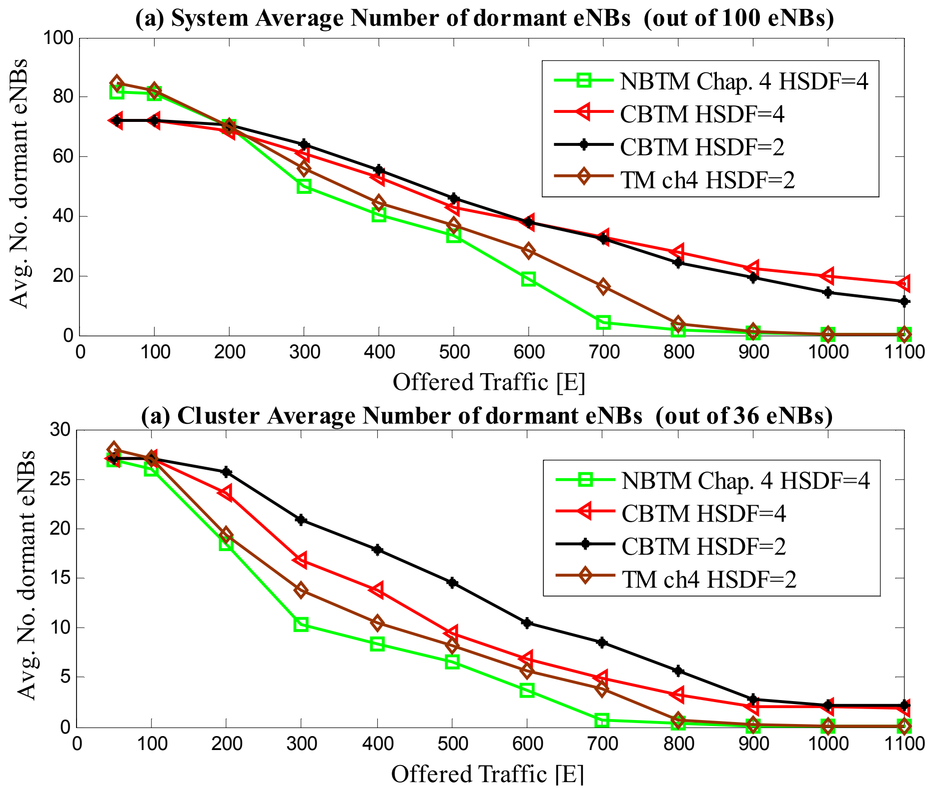


Figure 5.8 Average number of eNBs in the OFF mode vs Offered Traffic

It is clear that when increasing the HSDF to 4 both schemes use additional eNBs compared to the amount of eNBs used when HSDF is 2. However, the CBTM uses more additional eNBs than the NBTM scheme. This illustrates the ability of the CBTM in identifying the eNBs which can alleviate the demand from highly loaded eNBs. Moreover, CBTM when HSDF is 4 still uses less eNBs than the NBTM when HSDF is 2. Figure 5.9 illustrates the percentage of energy savings. This figure provides an insight onto the energy reduction capabilities of both schemes. Energy reductions highly depend on the number of entities kept in the dormant mode. Still, the CBTM outperforms the NBTM scheme at traffic loads greater than 200 Erlangs saving around 20% more of energy on average. However, as Figure 5.9 suggests, the CBTM saves less energy than the NBTM scheme at low offered traffic due to the minimum number of eNBs needed to ensure coverage to the service area.

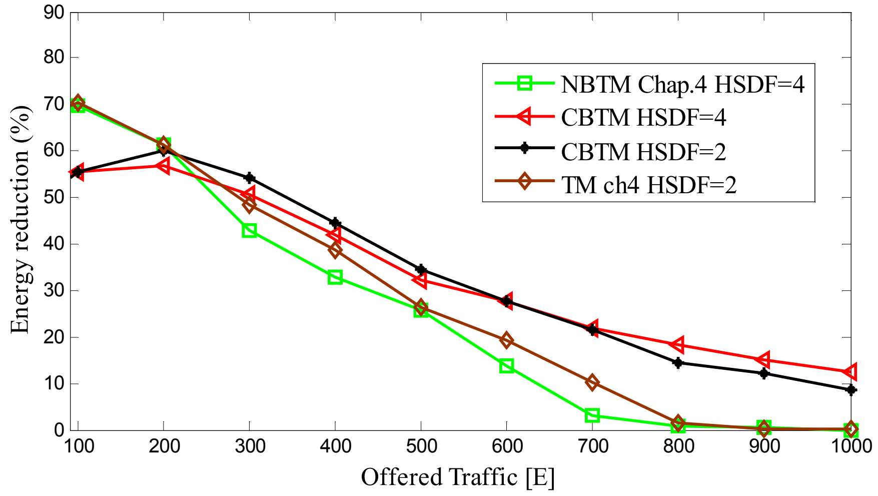


Figure 5.9 Energy Savings vs Offered Traffic

Figure 5.10 and Figure 5.11 compare the system QoS when using different topology management schemes with the QoS when no topology management is used for different values of HSDF in terms of end-to-end delay and blocking probability respectively.

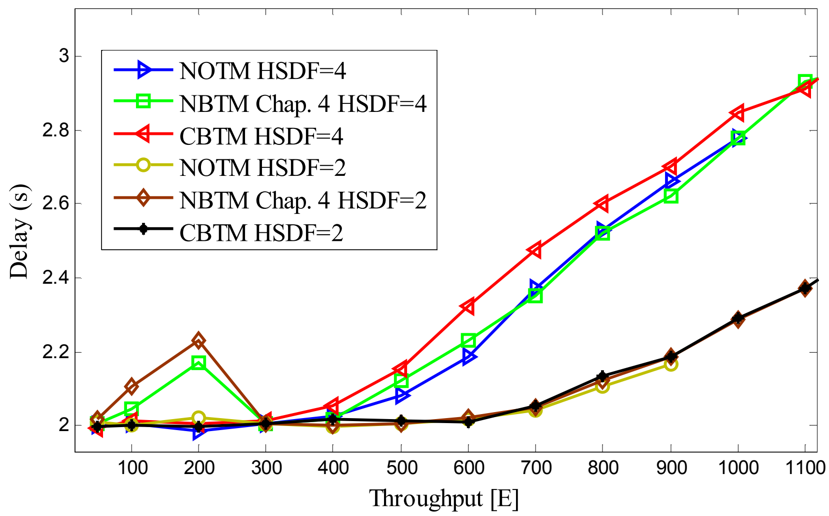


Figure 5.10 Delay vs Throughput

At low HSDFs (when HSDF is 2), the CBTM performs better than the NBTM scheme and as good as when having all eNBs in the system in the active mode. However, when using a high HSDF the CBTM performs slightly worse.

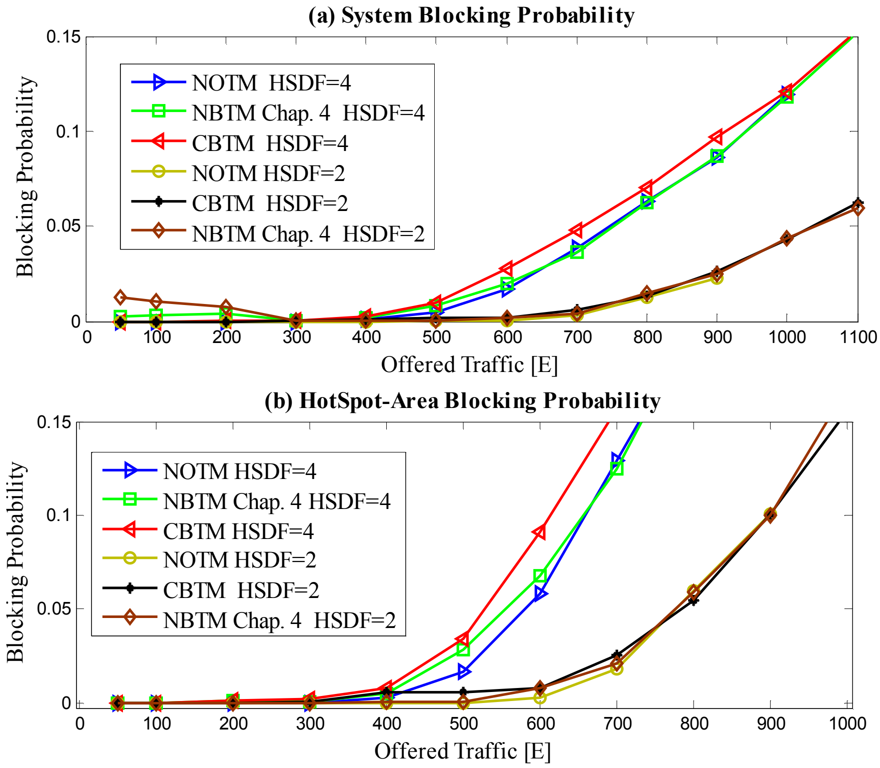


Figure 5.11 Blocking Probability vs Offered Traffic

The slightly poor performance of the CBTM at high HSDF can be attributed to the inflexibility of cluster members to reactivate eNBs that belong to a different cluster that can alleviate some of the traffic load.

5.6 Conclusion

The effects of having a spatially uneven traffic distribution in an ultra-dense urban area have been studied developing a Neighbour-Based Control scheme to neutralise the high traffic dynamics expected during a special event. Unlike its counterparts, which use distance as a criterion for grouping of eNBs, the scheme introduced in this chapter clusters entities which can cooperate to serve the demand of a given area. Upon the identification of different clusters, the scheme fine-tunes the dormant / active states of cluster members depending on localised traffic demands. The scheme controls the roll out / back of the ABSOLUTE system following localised traffic increase caused by temporary events and only activates the infrastructure which could provide and augment the network capacity density while maintaining the desirable QoS. An analytical model based on queuing theory has been used to further investigate the effectiveness of the scheme proposed, while Monte Carlo simulation is used to further assess the scheme under different configurations. The analytical tool is demonstrated to be scalable and accurate under a range of configurations despite the rather pessimistic estimate of the achievable QoS in terms of blocking probability in more flexible eNB selection scenarios. This can also serve as a performance indicator and can be used as an accurate lower bound on performance. It is shown that traditional schemes fail to operate under spatially unbalanced traffic conditions resulting in low QoS and high energy consumption whereas the proposed scheme can maintain up to 75% of eNBs in the dormant-mode which translates into a 50% reduction in the power required to operate the network at low loads. In the subsequent chapter the usage of artificial intelligence is investigated in an attempt to develop intelligent topology and resource management schemes that can adapt to a range of scenarios and traffic intensities.

Chapter 6 **Hierarchical Learning for RRC and Topology Management in Green Cellular Networks**

Contents

6.1 Introduction

Previous chapters investigated a range of energy aware topology management solutions under various traffic conditions and commented on the suitability and efficiency of the use of different network deployment strategies. The aim of this chapter is to complement the work presented in previous chapters by addressing some unsolved challenges, such as the choice of base station to de-(activate), and study interoperability issues between advanced radio resource and topology management. In Chapter 3, the technique of load unbalancing was studied in depth and concluded that its use is highly beneficial from the network energy efficiency perspective. However, this is only true for certain traffic intensities. This is mainly because that at high traffic loads, the use of energy efficient radio resource management technique can jeopardise the QoS provided to the UEs. Hence the well investigate technique of load balancing is then seen as a better way to manage the load. This chapter investigates a novel hierarchical learning approach which provides two sets of useful information, thereby incorporating past experience in the decision making process. These sets of information can be used by multiple entities of a wireless network and associated UEs to control network radio resources and topology. In fact, the outcome of the learning algorithm is used to develop an adaptive intelligent Radio Resource Control ($iRRC$) algorithm to control the cell selection and reselection procedure of UEs while in the idle mode. The $iRRC$, unlike other algorithms, aims to (1) maximise the energy efficiency of the network whenever possible and (2) maximise system capacity and enhance the QoS. In addition, an intelligent Topology Management (iTM) scheme is proposed which restricts the access of UEs to certain eNBs in order to maximise the time spent in the idle mode while ensuring the required level of QoS. The iTM is also responsible for managing the roll out and roll back of the ABSOLUTE disaster relief opportunistic network. Even though the operation of iTM is based on the outcome of the techniques presented in previous chapters, it takes a wider range of requirements into account such as the time required for the base station to be fully operational after being in the dormant mode. The reactivation frequency of base stations and interoperability issues between the advanced radio resource and topology management is investigated. The chapter is organised as follows; Section 6.2 introduces the hierarchical learning algorithm and explains how past experience can be translated into useful information, Section 6.3 presents the intelligent Radio Resource Control of UEs in the idle mode and explains how the learnt information is used and when, whereas Section 6.4 presents the functionality of the proposed intelligent topology management. Section 6.5 discusses the feasibility of the proposed $iRRC$ algorithm with current technology. The system model and dynamics are presented in Section 6.6. Results and discussion are provided in Section 6.7, and, finally, Section 6.8 concludes.

6.2 Multitask Hierarchical Learning

Two important aspects of a Cognitive Network (CN) are learning and reasoning. The first sets the manner of interaction of the agent with the environment and interprets that in a concrete knowledge base, whereas the second decides how the knowledge base should be used and when. In other words, the learning and reasoning agents within a CN agent decide how to learn, what to learn, and what to use the learnt knowledge for and when. This is one of the most difficult set of tasks in the CN cycle and is of great importance. So far, different learning and reasoning schemes have been put forward to perform different tasks. That implies the presence of multiple learning and reasoning agents within the same CN agent, each gathering a set of information, which can probably be correlated, to optimise or perform different tasks at the same layer. However, the flexibility of a CN agent to perform cross-layer optimisation by using the same knowledge base to enhance the performance of multiple tasks at different layers suggests that the interdependencies between different layers can be positively used [114]. To this end, novel multitask hierarchical cognitive learning and reasoning techniques are introduced, see Figure 6.1. The agent is based at the UE which collects information from its interaction with the system using these techniques to enhance the efficiency of Reinforcement Learning (RL) in exploiting wireless mobile environments, and hence improving decisions taken at different layers. Two important sets of information are produced. The first set is localised (i.e., it differs from one UE to another) and is a direct outcome of the enhanced reinforcement learning algorithm. The second set is globalised (i.e., it is the same for a large number of UEs) and is a result of processing the localised information. The novel action selection strategy, which is called Staged-Action Selection (SAS), builds an impartial knowledge base of a ranked list of possible actions at a given state. The SAS strategy ensures better post-convergence performance of the chosen best action and ranks the remaining suboptimal actions accurately. The other novel auxiliary scheme to RL is the Multitask Hierarchical Reasoning (M-HR). The M-HR allows for interpreting users' interactions with the system into a multitask knowledge base allowing for its utilisation to enhance multiple tasks. This way, it not only assures better converged values, but it also enables its reusability by other agents to enhance their operation.

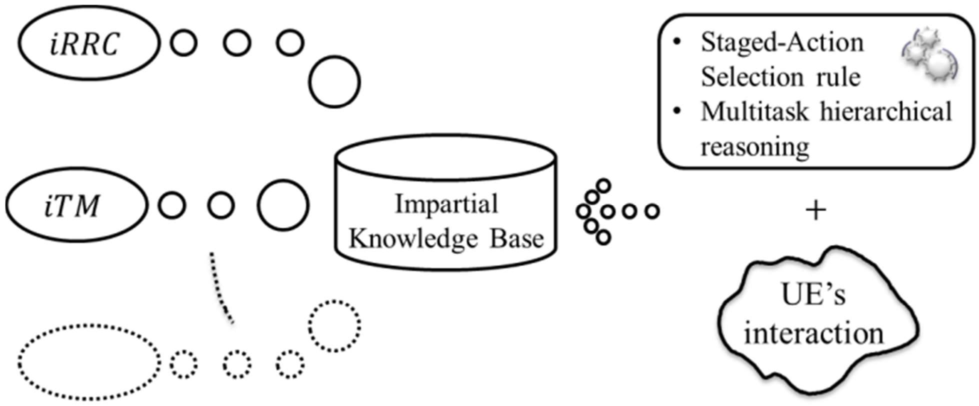


Figure 6.1 Multitask hierarchical learning mechanism

6.2.1 **Reinforcement Learning**

Reinforcement Learning (RL) is a machine learning technique widely used in cellular networks to enable the implementation of intelligent self-organisational functionalities for current or next generation systems only through trial-and-error [115]. The policy is represented by the Q-value which is the expected cumulative reward for an action $a \in A$ taken by the agent while in state $s|s \in S$, where A is a set of actions that can be taken while in the state $s|s \in S$. The task is to learn a policy $\pi : S \rightarrow A$ without having prior knowledge of the transition probability from one state to another and the goodness of these actions. For this reason a model-free RL type of machine learning is used in this thesis [116].

Mobile wireless environments are best formulated as stateless decision problems and therefore it is more beneficial to use the stateless Q-learning algorithm. The reason is mainly twofold; first, the number of state-action pair that needs to be estimated is reduced by a factor of $|S|$, hence speeding up the learning process. Second, it increases adaptability and re-configurability capabilities of the network dealing with unpredicted traffic patterns or unexpected system behaviour in general. In this case, the RL algorithm only needs to estimate an expected value of one reward for each action converting the Q-table into a 1-dimensional table Q(a) [117]. The algorithm hence chooses among actions based on the Q-value $\pi : Q \rightarrow A$, so the information regarding the system states is not needed which makes Q-learning suitable for the purpose of the work presented here. $Q(a)$ can be defined recursively as follows:

$$Q_t(a) = (1 - \alpha)Q_{t-1}(a) + \alpha R_t(a) \tag{6.1}$$

Where α is the learning rate, and $R_t(.)$ is the reward function which is detailed in the subsequent subsection. It is proven that $Q_t(a)$ converges to $Q^*(a)$ as $t \rightarrow \infty$ with probability 1 if each action a is visited infinitely often and the learning rate is small enough $\alpha \rightarrow 0$ [118].

6.2.2 **Independent Stateless Q-Learning**

The scheme proposed here is based on multiple independent single agents where each agent is trained to maximise the total reward of their actions in the long run regardless of the actions taken by other agents. In other words, a learning agent is oblivious of the existence of the other agents. Thus, the optimal policy π^* corresponds to the action $\pi(.)$ that maximises $Q\big(\pi(.)\big)$:

$$\pi^* = \underset{\pi(.)\in A}{arg\,max}\ Q^*\big(\pi(.)\big) \tag{6.2}$$

The algorithm is used to collect useful information from user interactions with the network and their experience in the manner set by the SAS strategy. The information provided is processed by the *M-HR* to derive a Cell Association Policy (CAP) between UEs and $eNBs$. This policy is used by the UEs to select which cell to camp on while in, or returning to the idle mode. The UE interacts with the environment and learns the cell association policy which maximises the Q-value.

Each UE (agent) maintains a 1-dimentional Q-table which is used to rank accessible cells $Q_i = q(i,c)|c \in C_i$ where C_i is the set of accessible cells for the i^{th} UE and Q_i is the Q-table of the i^{th} UE. The Q-values of all accessible cells are initialised to zero $Q_{t=0}(i,j) = 0\ \forall i\ \in UE, j \in C_i$, where UE is the set of UEs.

6.2.3 **Reward function**

A tailored-fit reward function has been developed for this algorithm specifically. This is done for the purpose of differentiating between serving cells by the QoS provided. The reward function used here returns a value depending on the QoS provided by the serving cell as follows:

$$\mathrm{r} = \begin{cases} 1 & ;\ if\ best\ QoS\ guaranteed \\ QoS_{best} - QoS & ;\ acceptable\ QoS\ provided \\ QoS_{best} - QoS_{worst} & ;\ if\ service\ is\ denied \end{cases} \tag{6.3}$$

Where the QoS parameter used for this work is the end-to-end delay. Users calculate the $\mathrm{QoS_{best}}$ and $\mathrm{QoS_{worst}}$ from the user throughput which in turn can be calculated from the SINR at which the user is transmitting their files. Hence, if a user is capable of transmitting at the highest rate possible, a reward of $r = 1$ is given, whereas if the user is capable of transmitting but at a lower rate a reward of $\mathrm{r = QoS_{best} - QoS}$ is given. In the case where the user is denied service (i.e., blocked) a reward of $\mathrm{r = QoS_{best} - QoS_{worst}}$ is given.

6.2.4 **Staged-Action Selection strategy (*SAS*)**

The SAS is developed to enable the optimisation of current and future cellular technology by making the most out of users' interaction with the system. There are a number of action selection strategies the most prominent ones being the greedy, Ɛ-greedy, and the $softmax$ action selection strategies (ASS) [119].

- **The greedy ASS** is known for its simplicity of selecting the action with highest estimated action value. This strategy capitalises on exploiting present knowledge base to maximise immediate reward paying no attention to less valued actions. For this reason, optimality cannot be guaranteed.
- **The Ɛ-greedy ASS** is a greedy rule in its essence with a small probability of selecting an action independently of action-value estimates. The advantage of this strategy is that if actions are sampled an infinite number of times, convergence can be guaranteed.
- **The $softmax$ ASS** is a derivative of the greedy scheme that uses a probabilistic approach during the exploitation phase. The probability of choosing an action during the exploitation phase is derived from the action value.

These approaches are not appropriate for exploring wireless mobile environments to gather a suitable knowledge base that can help improve a range of applications and procedures within cellular systems for the following reasons:

1- All the aforementioned approaches strive to reach that one and only ideal action even if there are other equally good actions.
2- The difference in goodness between one action and another is not captured. This is because some actions are sampled relatively more than others.
3- Given the greedy or semi-greedy behaviour of these action selection rules, it requires a large number of samples to have an accurate estimation of the goodness of all sub-optimal actions as these are visited very infrequently. This affects the convergence and reconvergence speed which is of paramount importance for cellular applications.
4- Due to the cumulative nature of the action value update function for these strategies, a long time is needed to re-converge when the best valued action is no longer optimal.
5- Ɛ-greedy and $softmax$ strategies can only perform the optimal action with $(1 - \mathcal{E}) + \mathcal{E}/|A|$ probability, where $|A|$ is the number of actions available in the action space. This is because they constantly choose an action randomly with probability of Ɛ (assuming $softmax$ explores with probability Ɛ as well).

The strategy proposed here, which is called the Staged-Action Selection strategy (SAS), first of all, converges to a better action, or actions if more than one exists. Secondly, it accurately ranks all sub-optimal actions resulting not only in one single optimal action but in a ranked list of possible actions depending on their goodness. This type of outcome is increasingly important as will be seen in the subsequent sections.

6.2.4.1 *Operation and convergence*

Traditionally, convergence is not decided but observed. In other words, when using an action selection strategy such as the greedy strategy, convergence is observed when an agent chooses an action $a \in A$ as their optimal action entering a steady state. However, the Staged-Action Selection (SAS) strategy offers two approaches, the aggressive and the relaxed approaches, each with their distinctive features.

- **Aggressive SAS:** This approach assumes that there are no completely bad actions and that most actions can deliver acceptable results or it is affordable to perform bad actions for a limited period of time. Hence, the algorithm explores all possible actions aggressively which intensely accelerates convergence. This is most suitable at low to medium traffic load conditions. This approach requires making a decision about when action values are mature enough to be used (i.e., moving from the exploration to the exploitation phase). One way is to observe the ranking of the action values and declare convergence if the ranking is seen to be stable which reflects the fact that $Q(s,a)_t = Q^*(s,a)\ \forall a \in A$. However, some agents are likely to struggle to reach this stage as multiple actions might have similar outputs in terms of reward. This disturbs the stability of the rank of the possible outcome. To solve this problem, a maximum number of episodes permitted is defined. Hence, agents are said to have converged when the $Q(s,a)_t = Q^*(s,a)\ \forall a \in A$ or the number of episodes reaches Ep_{max}. All actions are sampled successively for an equal number of times. This way, equal opportunity is given to all actions to become the optimal choice and an accurate degree of goodness of all sub-optimal actions is guaranteed. This is also valid when convergence is decided by the maximum number of episodes permitted.

- **Relaxed SAS:** the agent can start from the aggressive approach to build a knowledge base as quick as possible, or can choose to take a more preventative although time consuming tactic which is the relaxed approach. In this approach the knowledge base is updated with probability ξ, and exploited with probability $1 - \xi$. The knowledge base must remain untouched while in the exploitation phase to keep impartial action values. While in the update phase, the algorithm makes it compulsory to update all action values by sampling each action at least once. The parameter Ep_{max} can be also used here to allow for more than one update episode per update phase to take place. Ep_{max} allows to tune the aggressiveness of the exploration / update phase to the desired level.

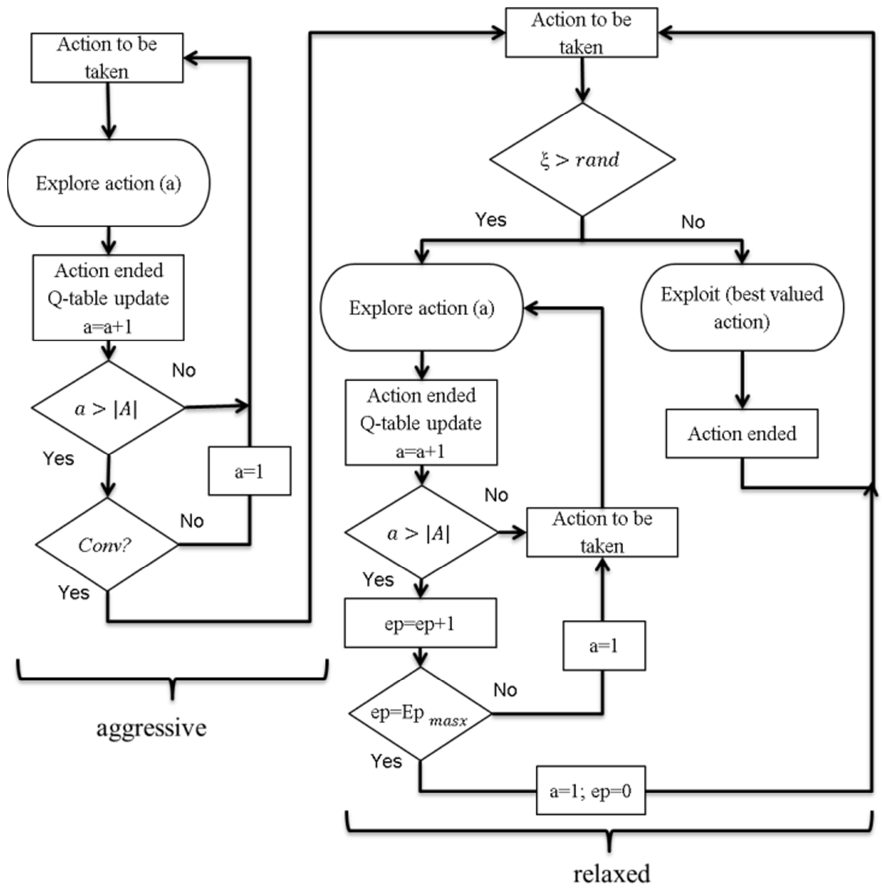


Figure 6.2 Staged-Action Selection strategy

The Ep_{max} parameter along with the fact that the knowledge base remains untouched while in the exploitation phase makes it possible for this strategy to cope with radical changes in the dynamics of the environment enabling quick re-convergence. The SAS strategy is illustrated in Figure 6.2. The application of the SAS is further discussed in details in the subsequent subsections with attention to its usage to enhance RRC procedures and green topology management.

6.2.5 **Multitask Hierarchical Reasoning**

The main piece of information is directly generated from the interaction of the UEs with the system. This is subsequently processed to complete the knowledge base which can be used not only by the UEs but also by the system itself to enhance a range of operations. The knowledge base generation process is described in Figure 6.3 and can be summarised as follows:

1- During the exploration phase, all UEs perform the cell selection / reselection procedure based on a randomised list of available cells $R_c = randperm(C_i)$. These cells are tested periodically based on the list order until either the steady state is reached (i.e.,$Q(s,a)_t = Q^*(s,a)\ \forall a \in A$) or a maximum number of episodes are sampled as specified earlier. This will result with each UE having a ranked list of accessible cells and their degree of goodness described by the weight value of the Q-table $Q_i \in Q$. This table is called the Personal Cell Association Policy table ($PCAP$).

2- At this stage, UEs report their resulting $PCAP$ tables to their serving cells which in turn reports it to a local Mobility Management Entity (MME). The MME combines all Q-tables by performing a basic arithmetic aggregation to form one table as

$$GCUP = \sum_{i=1}^{n_{ue}} PCAP_i \tag{6.4}$$

Where n_{ue} is the number of UEs in the region of the local MME. The result of this addition is a weighted list of all serving cells in the region which is called the Global Cell Utility Policy table ($GCUP$).

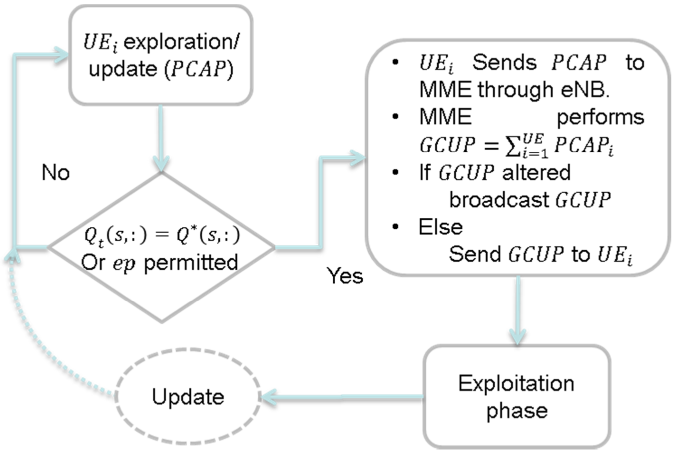


Figure 6.3 Flowchart of the learning mechanism

1- The MME processes the information provided by the UEs each time it receives a $PCAP$ table. However, if seen that the $GCUP$ table remains unaltered in terms of ranking of eNBs, the MME only sends the $GCUP$ to the UE from whom a $PCAP$ was received. In the case when the ranking of eNBs is changed, the $GCUP$ is broadcast to all UEs in the area.

This process results in two crucial pieces of information that can be used by a range of cognitive mechanisms and applications to enhance their performance. The first piece is the $PCAP$ table which is different from one UE to another. Hence this table is considered as personal localised information regarding the network used by the UE in decisions such as selecting a cell. The second is the $GCUP$ table which is globalised information that gives an insight into the global degree of goodness of various cells in the access network in terms of QoS provided and their contribution in serving the offered traffic. The $GCUP$ can be used by the network to take smart decisions and to support cognitive mechanisms such as green topology management as well as by the UEs to reinforce their decision when it concerns the access network.

The local MME is an entity in next generation networks that is envisioned to be responsible for a smaller area than current $MMEs$ as future networks are being designed to be highly distributed [13]. In fact, the FP7 ABSOLUTE project is developing a distributed MME which they call a Flexible Management Entity (FME) [94]. This entity is designed to be incorporated along with other eNB entities and functionalities. The FME is a distributed element which is designed to enable virtual EPC ($vEPC$) support for 4G-LTE cellular networks and beyond. The FME enables $eNBs$ to perform most tasks independently from the core network (EPC) by embedding fundamental EPC functionalities so as to reduce signalling overhead and make the network less centralised and vulnerable to a single point of failure. Even though this is developed in the context of emergency and disaster relief rapidly deployable networks, it is nonetheless in line with highly distributed features intended for 5G networks and beyond. For current deployments, given the fact that the complexity of the proposed learning process is low, a cluster-head eNB ($CeNB$) can easily process the information and distribute it to neighbouring cells.

6.3 Intelligent Cell Selection and RRC in Idle Mode ($iRRC$)

In most work on optimising the cell selection mechanism, balancing the load was the most important criterion in finding a suitable cell in order to achieve higher frequency efficiency, better capacity, or enhanced QoS. In practice as reported in Chapter 2 and according to [120], the cell selection and reselection procedure is done using the Reference Signal Received Power ($RSRP$). In [121] however, the authors propose a cell selection scheme based on the Expected bit rate ($E[B]$). The $E[B]$ scheme outperforms a range of cell selection mechanisms as reported in [121]. We have taken these two cell re-(selection) schemes as state-of-the-art benchmark algorithms in order to compare the results. As suggested in Chapter 3, balancing the load is not always the optimum solution. Also, next generation access networks and beyond need to deal with greater peak-to-average loads and traffic fluctuations and uneven traffic distribution as concluded in Chapter 5. In such circumstances, the offered traffic needs to be managed in a way that uses the least resources possible when the traffic is relatively low (load unbalancing), and to balance the load for capacity optimisation when the system is pushed to its limits. This is done for the purpose of enhancing the energy efficiency by trading in the spectral efficiency whenever possible. Such techniques need to be taken into consideration when developing next generation and beyond next generation backhaul networks, as they need to be intelligent and flexible enough to deal with high traffic with high traffic fluctuations arising from the load management techniques in the access networks [62, 122].

Due to the high density deployment of next generation access networks, it is certain that at low traffic loads, when the interference level is at its minimum, a UE can have a number of accessible cells that can guarantee the required QoS. Traditionally, these UEs are spread throughout the access network for load balancing which in turn will maintain an unnecessarily large number of active cells. The algorithm proposed here, the $iRRC$, differs from the aforementioned techniques in the following ways:

1- The $iRRC$ tries to appropriatly cluster users at low traffic load onto a minimum number of cells, while still meeting the demanded QoS.
2- As the traffic load increases, the degree of clustering decreases adaptively and the $iRRC$ starts to balance the load appropriately to match the load increment.
3- At high traffic load, the $iRRC$ completely de-clusters the users and tries to balance the load in order to enhance the spectral efficiency and system performance.

This is achieved with the help of the learning algorithm introduced in the previous section which provides both the global ($GCUP$) and local ($PCAP$) pieces of information.

Figure 6.4 illustrates an example of a cell selection / reselection procedure using assumed values for $PCAP$ and $GCUP$ tables at different traffic loads.

UE_1	$PCAP_1$	$GCUP$
$Cell_1$	2	5
$Cell_2$	2	0
$Cell_3$	2	1
$Cell_4$	2	-6
$Cell_5$	2	2
$Cell_6$	-3	-14
....	....	.....
$Cell_c$	$PCAP_{1,c}$	$GCUP_c$

$\Rightarrow Cell_1$

(a) Low traffic load

UE_1	$PCAP_1$	$GCUP$
$Cell_1$	0	5
$Cell_2$	2	0
$Cell_3$	2	1
$Cell_4$	-1	-6
$Cell_5$	2	2
$Cell_6$	-4	-14
....	....	.....
$Cell_c$	$PCAP_{1,c}$	$GCUP_c$

$\Rightarrow Cell_5$

(b) Medium traffic load

UE_1	$PCAP_1$	$GCUP$
$Cell_1$	-3	5
$Cell_2$	0	0
$Cell_3$	4	1
$Cell_4$	-1	-6
$Cell_5$	2	2
$Cell_6$	-3	-14
....	....	.....
$Cell_c$	$PCAP_{1,c}$	$GCUP_c$

$\Rightarrow Cell_3$

(c) High traffic load

Figure 6.4 Cell reselection procedure using the knowledge base

Cells that can perfectly serve a certain UE_i will have equal Q-values in the $PCAP_i$ table. Hence, the $PCAP_i$ table alone cannot be used to make a holistically optimum decision, so the $GCUP$ table is used concurrently. When a cell selection / reselection procedure is triggered, UE_i checks the $PCAP_i$ table for cell candidates (cells having the highest Q-value). These cells are considered to be the ones that can perfectly serve UE_i and deliver the best QoS possible. Afterwards, the $GCUP$ table is used to select the one that is considered to be the most beneficial from the network energy and spectrum efficiency point of view. Due to the clustering behaviour of the algorithm at low and medium traffic loads, $iRRC$ also monitors the loading statuses of candidate eNBs in order to not overload the eNBs unnecessarily. This is done by re-prioritising the $GCUP$ table depending on the loading status of each eNB, see figure 6.5. The $iRRC$ divides the $GCUP$ table into two groups, the High Loaded and the Light Loaded groups. The High Loaded (HL) group consists of eNBs that exceeded the maximum loading threshold (M_l%). The Light Loaded (LL) group consists of eNBs whose load is less than the maximum loading threshold. UEs try to camp on eNBs from the LL group if possible. If unsuccessful, the UE moves back to the HL group for selection.

$$\left(eNB_i \in HL \,\middle|\, L_{eNB_i} \geq L_{thr} \; else \;\; eNB_i \in LL\right) \forall\, i \in eNB \tag{6.5}$$

L_{thr} is a percentage M_l of the maximum capacity C in terms of number of resource blocks as given in equation (6.6).

$$L_{thr} = floor[max(C).M_l\%] \tag{6.6}$$

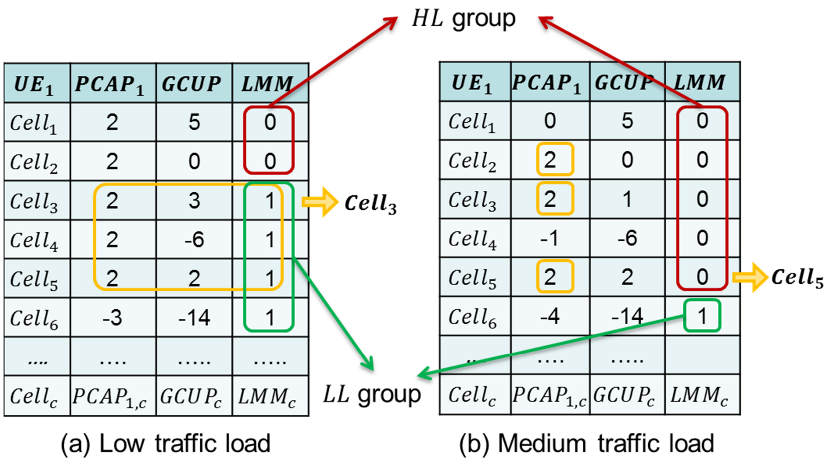


Figure 6.5 $iRRC$ cell reselection procedure

This mechanism can be used as a standalone middleware to solve interoperability issues between advanced radio resource management and energy-aware topology management techniques.

6.4 Intelligent Energy-Aware Topology Management (iTM)

Chapter 1 defines energy-aware topology management mechanism as a way to dynamically control the predesigned topology of the system in order to achieve energy savings by trading off spectrum efficiency whenever possible while meeting the required QoS. This is achieved by fine-tuning the status of the eNBs (dormant / active) depending on the traffic demands. This chapter decouples the operation of the TM scheme into three independent units as detailed in Subsection 2.3 of Chapter 2. These units are the Topology Management Information Unit ($TMIU$), the Topology Management Control Unit ($TMCU$), and the Topology Management Sleep-mode Unit ($TMSU$). The TMIU is represented by the learning mechanism introduced in the previous sections. The TMSU is modelled as described by the EARTH project in [99]. This model is slightly modified incorporating a transit period from the dormant to the operational mode. The Macro-cell overlaid TM scheme proposed in Chapter 4 is used here as the TMCU to control the roll out and roll back of the network. This is decoupled from the TMSU by introducing the accessible and restricted states on top of the already defined active and dormant modes. The accessible state means that an eNB is active and accessible as normal. The restricted state on the other hand represents a state in which an eNB is active but not accepting customers. The use of the accessible and restricted states makes it possible for the TMCU to be used to boost the performance of any green algorithm that relies on the idle state of the eNBs to obtain energy savings such as Discontinuous Transmissions (DTX) [123] and sleep modes [124]. The TMCU proposed in Chapter 4, the Macro-cell overlaid TM (McTM), which randomly decides on which cells need to be accessible and which need to be restricted, is compared with an enhanced version in which the information provided by the learning algorithm is used by the TMIU for the identification of such cells.

6.4.1 **Policy of the Topology Management Control Unit**

The TM algorithm uses loading status as the main policy to control the topology. For example, the absence of UEs is a good indicator to when an eNB needs to switch to the dormant mode. The topology management best fits heterogeneous deployment as it relies on macro-level coverage of macro-eNBs ($MeNB$) to provide basic coverage and small-cell eNBs ($SeNB$) as capacity-density boosters which can be switched to the accessible state on a need basis. This is detailed in Chapter 4.

6.4.1.1 *Switching to the restricted state:*

$SeNBs$ switch to the restricted state only when these are underutilised and have spent a minimum time in the accessible state as specified in Chapter 4. An eNB is said to be underutilised when $L_{eNB} \leq L_{rest}$ where L_{eNB} is the instantaneous load and L_{rest} is the threshold to switch to the restriction state which is a percentage C_{rest} of the maximum capacity C of the eNB as

$$L_{rest} = floor[max(C) . C_{rest}\%] \tag{6.7}$$

At this point, the restricted state can be exploited by the TMSU. The concerned eNB needs to inform all peer eNBs about the action taken over the X2 interface by using the eNB *Configuration Update procedure*. If the eNB is switched to the dormant mode, all neighbouring eNBs need to maintain the configuration data of the dormant eNB such as neighbour relationship configuration for quicker reconfiguration and synchronisation. The switch to the restricted state of an eNB is not immediate. If the above conditions are met, the eNB enters a transitional period in the restricted state in which it does not admit any new customers. It will finally switch to the restricted state when the last user in service departs. If a UE is updating their $PCAP$ table, eNBs in the restricted state are rewarded according to equation (6.3) as "$service\ denied$". As legacy eNBs are restored, more eNBs from the opportunistic network will be in the restricted state. The network is said to be rolled back when all eNBs are in the restricted state.

6.4.1.2 *Switching to the accessible state:*

The on-demand services of $SeNBs$ is triggered by accessible eNBs ($SeNB$s or $MeNBs$) when the traffic load increases resulting in $L_{eNB} \geq L_{acss}$ where L_{acss} is a percentage C_{acss} of the maximum capacity C of the eNB as

$$L_{acss} = floor[max(C) . C_{acss}\%] \tag{6.8}$$

$\mathrm{C_{rest}}$ and $\mathrm{C_{act}}$ are, as mentioned, a percentage of the maximum capacity C in terms of number of resource blocks. Both the $MeNB$ as well as $SeNBs$ are responsible for triggering the reactivation of restricted $SeNBs$ by using the $X2$ interface by means of the *Cell Activation* procedure [125]. The $SeNB$ to be reactivated is chosen by the TMIU using the global information gathered by the learning algorithm (i.e., the $GCUP$) in the case of the iTM, whereas in the case of the TMIU in Chapter 4 this $SeNB$ is chosen randomly from the neighbouring set of restricted eNBs.

6.4.2 **Topology management operation**

An eNB checks its loading status whenever a UE demands service (the arrival block in Figure 6.6) or finishes service (departure block in Figure 6.6). In other words, eNBs having any UE in the $RRC\_CONECTED$ mode do not switch to the restricted mode until all UEs terminate their activity or, as expressed in Figure 6.6, depart from service, hence no $RRC\_CONNECTED$ handover is performed. However, UEs that are idle and camped on an eNB that is restricting its access are forced to perform a reselection procedure and camp on an accessible eNB. The TMSU makes use of the time spent by the eNBs in the restricted state to switch them to the specified sleep mode to save energy. This is described in Figure 6.6. Since the TMCU uses load derived policies, an updated neighbouring list needs to be available at the time of the decision making. The neighbouring list is provided and kept up to date by the Automatic Neighbour Relation (ANR) management function which deals with automatic Neighbour Relation (NR) removals or additions. ANR function is essential to Hand Over (HO) operations and to the X2 interface setup which enables information exchange between nodes. Loading statuses are made available by using the Load Reporting function which supports the mobility load balancing function. The Load Reporting function enables the exchange of cell specific load information between neighbouring eNBs. This function is run over the X2 and S1 interface depending on the RAT of the neighbouring node. X2 interface is used for intra-LTE and S1 is necessary for inter-RAT communications.

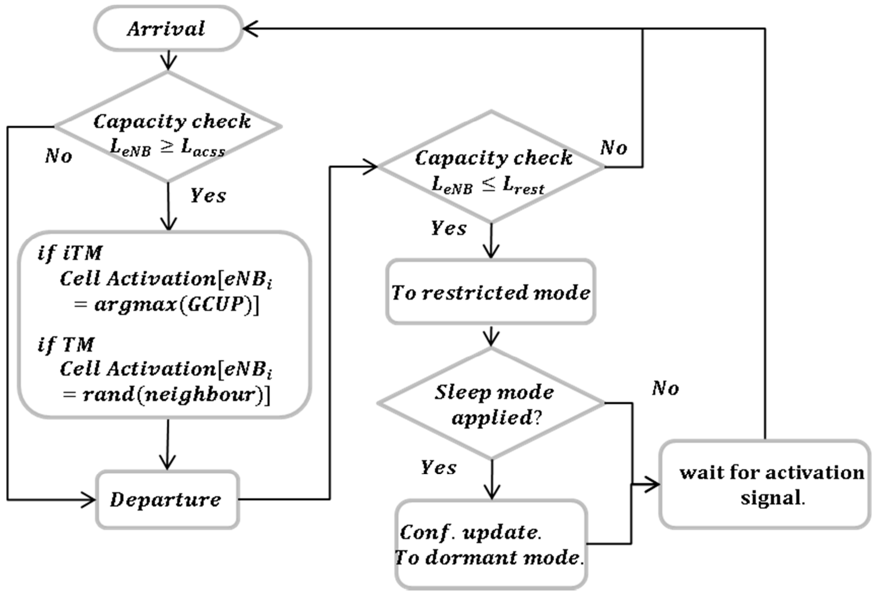


Figure 6.6 Topology management algorithm

6.5 Implementation & Compatibility

6.5.1 Implementation

6.5.1.1 *Association of Q-values*

All UEs will receive two essential signals as soon as they register with a suitable Public Land Mobile Network (PLMN). These signals are called the Primary and Secondary Synchronisation Signals, PSS and SSS respectively. These carry the physical layer identity and the physical layer cell identity which are essential for the learning algorithm proposed here to work properly. This information is available at this stage in line with current LTE standards to perform cell search and cell selection procedures. The Evolved-Universal Terrestrial Radio Access Network (E − UTRAN) Cell Identifier (ECI) which is a 28 bit ID, formed by adding the Physical Cell ID (PCI) 20 bits and the 8-bit Cell ID (CID), is used by the UE to associate the Q-values with their respective cells as shown in Figure 6.7. We consider one PLMN hence the use of the ECI as the cell identifier. Also, the use of ECI is appropriate for this purpose as ECI information is carried in the System Information Block Type 1 (SIB1) which is broadcast every 80ms and accessible to all UEs after initial cell synchronisation and reading the Master Information Block (MIB). If multiple PLMNs are considered, the Evolved Cell Global Identifier (ECGI) can be used to identify different cells.

6.5.1.2 *Cross-layer optimisation and protocol operation*

Figure 6.7 illustrates where the association between Q-values and different $ECIs$ take place as well as how upper layer information is fed-back to lower layers to assist in the cell selection and reselection procedure. As can be seen, a user initially needs to be camped on a suitable cell before obtaining service. Subsequently, taking the uplink into account, the user contests for accessing the network (L2/MAC layer). If successful, resources are reserved in L1 to transmit / receive data. The proposed scheme takes the outcome of this process (which is the QoS received) to the RRC sublayer in L3 and enhances its operation.

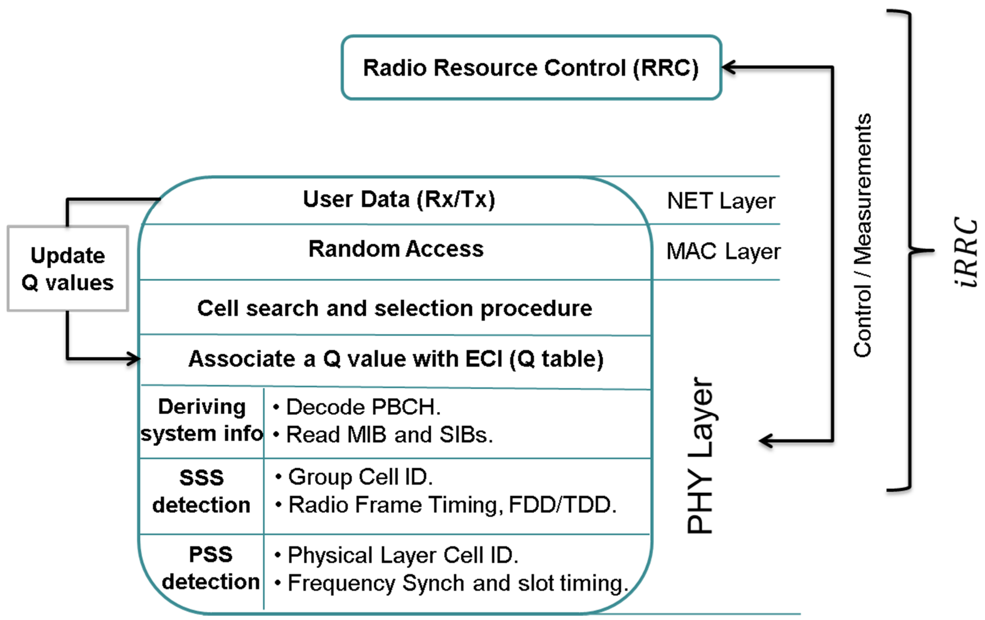


Figure 6.7 $iRRC$ cross-layer optimization and protocol stack

The signalling messages *PCAP*, *GCUP,* *GCUP_REQUEST*, and *GCUP_REQUEST_ACK*, are illustrated in Table 6.1. The protocol end-to-end message flow is demonstrated in Figure 6.8 and described in Table 6.2.

GCUP_REQUEST

Message purpose	Indicates that a PCAP policy is derived and waiting for a GCUP
msg. generation trigger	When a PCAP policy has been derived
Source	UE
Destination	Serving eNB and CeNB
List of Information Elements	
IE Name	Description
Transmitter_ID	Identifier of the entity sending the message
Receiver_ID	Identifier of the entity receiving the message
IMEI	UE Identifier that generated the message

GCUP_REQUEST_ACK

Message purpose	Acknowledges the GCUP_REQUEST and completes the GCUP request process
msg. generation trigger	After a GCUP_REQUEST has been received
Source	CeNB
Destination	UE / serving eNB
List of Information Elements	
IE Name	Description
Transmitter_ID	Identifier of the entity sending the message
Receiver_ID	Identifier of the entity receiving the message
IMEI	UE Identifier that generated the message

PCAP

Message purpose	Notify the eNB with the Personal Cell Association Policy (PCAP)
msg. generation trigger	When a PCAP policy has been derived
Source	UE
Destination	Serving eNB / CeNB
List of Information Elements	
IE Name	Description
PCAP	Ranked list of accessible eNBs

GCUP

Message purpose	Notify the UE of the Global Cell Utility Policy (GCUP)
msg. generation trigger	When CeNB receives the GCUP_REQUEST from the UE
Source	CeNB
Destination	UE / serving eNB
List of Information Elements	
IE Name	Description
GCUP	Global ranked list of all eNBs in the region of the CeNB

Table 6.1 ***iRRC*** downlink signalling messages

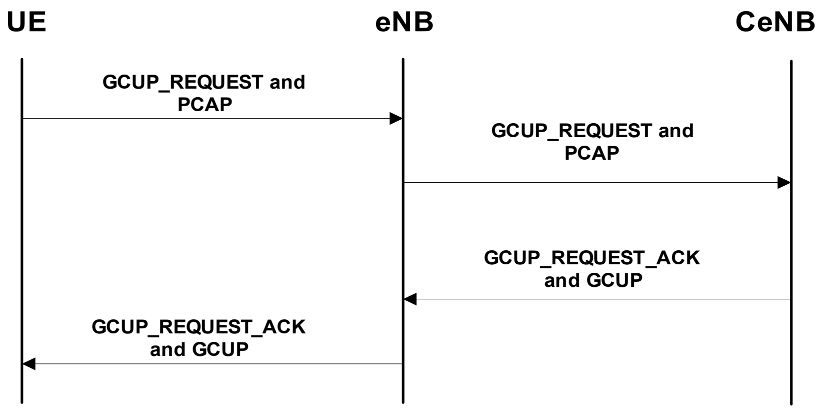


Figure 6.8 End-to-End message flow

UE-eNB link: 1. The UE sends a $GCUP$ request to the serving eNB containing a ranked list of accessible eNBs and their Q-values ($PCAP$). 2. The eNB sends an acknowledgement back to the UE which at the same time contains the Global Cell Utility Policy table ($GCUP$). eNB-CeNB link: 1. The eNB forwards the $PCAP$ to the CeNB for processing. 2. The CeNB sends an acknowledgement back to the eNB containing the Global Cell Utility Policy table ($GCUP$).

Table 6.2 End-to-End procedure

As seen from Figure 6.8 and Table 6.2, the scheme proposed is of low complexity as it requires limited information exchange between a UE and the system.

6.5.2 **Compatibility**

According to [120] and [125], the UE periodically assesses the serving cell at least every Discontinuous Reception (DRX) cycle which is 2.56 seconds maximum. If the serving cell is not considered to be a suitable cell to be camped on, the UE performs a cell search to look for a suitable cell. As a first step, standards clearly specify the possibility of having stored information that could help assist with the cell reselection procedure. The $iRRC$ provides information on cell performance from previously selected or attempted cells which is used to perform the cell reselection procedure without altering the standards. Regarding the required message exchange, such as the $GCUP$ and load related information, it is also specified in [120] as part of the reselection procedure that the serving cell may provide this information to the UE when transiting from the RRC_CONNECTED mode to the RRC_IDLE mode. Hence, the GCUP table and load-related information can be attached to existing Information Elements (IEs) such as the $redirectedCarrierInfo$ which is part of the $RRCConnectionRelease$ message. Alternatively, these tables can also be sent as part of the System Information Block 4 (SIB4). SIB4 is a dedicated block to cell-reselection parameters regarding neighbouring intra-frequency cells. In fact, the $GCUP$ can replace the $intraFreqNeighCellList$ IE which is a list of neighbouring cells with related cell reselection parameters [126].

6.6 **System Model**

The network considered here, referring to Figure 6.9, is a heterogeneous network of low altitude platforms which are called Aerial eNBs (AeNBs) and multiple Terrestrial eNBs (TeNBs) on the ground. A broad consensus [94-98] is that these type of heterogeneous architectures are likely to become dominant in the near future. Some examples are its use as an opportunistic network for short-term events such as the Olympics, for rapidly deployable post-disaster networks, to provide services to remote areas, and to provide high speed connectivity and 4G technology globally [97]. The network considered here is the disaster relief FP7 ABSOLUTE project architecture. ABSOLUTE aims to design and validate an innovative, rapidly deployable, future network architecture which is resilient and capable of providing broadband multiservice, secure and dependable connectivity for wide coverage areas affected by large scale disasters such as tsunami or hurricanes. Energy consumption is a crucial aspect in all these scenarios as a large number of portable light-weight $eNBs$ are densely deployed in a wide city area, as in the case of special event scenario, and there is also need to prolong the battery life of $AeNBs$. The learning algorithm presented here, as well as its application in the Radio Resource Control and Topology Management, aims to minimise energy consumption by setting the appropriate degree of clustering of UEs onto different $eNBs$ as well as determining the number and service time of these $eNBs$ based on the local traffic variation.

6.6.1 **Power model**

The power model for the TM Sleep-mode Unit (TMSU) used here is adopted from the EARTH project [99-101] and can be expressed as:

$$P_{supply} = \begin{cases} P_0 + \beta P_{Tx} & ; \quad 0 \le P_{Tx} \le P_{max} \\ P_d & ; \quad P_{Tx} = 0 \end{cases} \tag{6.9}$$

Where P_{max} is the maximum transmit power, P_0 and P_d are the idle- and dormant-mode power consumption of the eNB respectively, P_{Tx} is the instantaneous transmit power, and β is the load dependency constant. Since micro $eNBs$ are considered in this work, the results from [102] are also taken into consideration to set the values of the power model parameters.

6.6.2 **System dynamics and parameters**

In this work, we consider the access network having a single $AeNB$ and 19 $TeNBs$ where the $AeNB$ uses a dedicated spectrum band. The transmission rate (R) is calculated using the Truncated Shannon Bound (TSB). TSB, equation (6.10), provides an estimation of the link throughput that can be achieved in practice given an adaptive modulation and coding (AMC) code-set [108].

$$R = \begin{cases} 0 & ; \quad SINR < SINR_{min} \\ \gamma . S(SINR) & ; \quad SINR_{min} < SINR < SINR_{max} \\ R_{max} & ; \quad SINR > SINR_{max} \end{cases} \tag{6.10}$$

Where $SINR_{min}$ and $SINR_{max}$ are the minimum and maximum supported $SINR$ of the code-set in dB. γ is the attenuation factor representing implementation losses, (R_{max}) is the maximum achievable transmission rate, and $S(SINR)$ is the Shannon bound as:

$$S(SINR) = log_2\,(1 + SINR) \quad [bps/Hz\,] \tag{6.11}$$

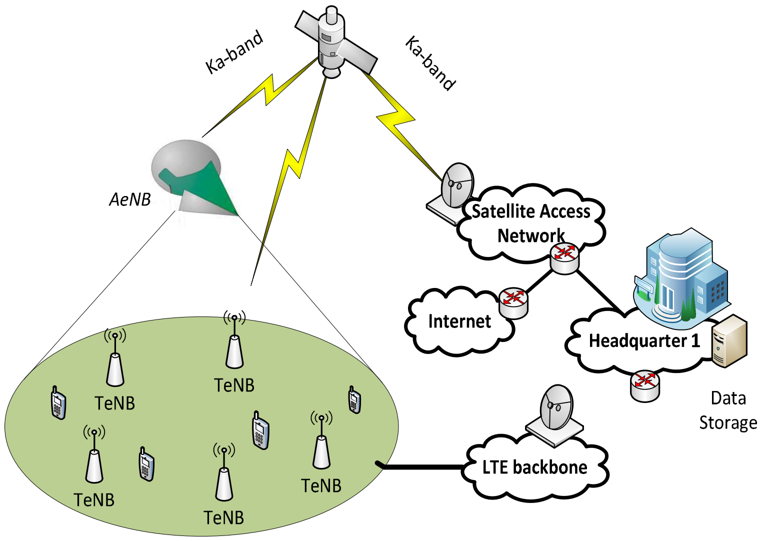


Figure 6.9 ABSOLUTE disaster relief architecture

An event-driven simulation has been used to assess the performance of the aforementioned schemes. Users are uniformly distributed within the service area and arrive into the system with exponentially distributed inter-arrival times. File sizes are also exponentially distributed with an average size as described in Table 6.3.

The results were averaged across 5 simulation runs, each run simulating 2 million trials. Users are stationary and no mobility is taken into account. This is a valid assumption for the scope of this work as in heterogeneous networks it is assumed that capacity-booster nodes are intended for static and slow moving users and that the macro-overlay, the $AeNB$ in this case, is responsible for providing service and connectivity for nomadic and rapid moving users [20, 127, 128].

Parameter		Value/Description
Access Network		
eNB type	Aerial eNB	Omnidirectional
	Terrestrial eNB	3-sector antenna
Inter-site distance		500 [m]
Carrier frequency		2600 [Mhz]
Spectrum bandwidth		5MHz
Noise floor		-114 [dBm/MHz]
Propagation model	Aerial eNB	Free space path loss; lognormal shadowing std dev 8 dB
	Terrestrial eNB	L=127 + 30log(d[Km]); lognormal shadowing std dev 4 dB
Antenna height	Aerial eNB	300 [m]
	Terrestrial eNB	15 [m]
Transmission power		25 dBm
UEs distribution		Uniformly
$SINR_{min}$ / $SINR_{max}$		1.8 / 21 [dB]
Attenuation factor (γ)		0.65
Max. transmission rate (R_{max})		4.5 [bps/Hz]
Average file size		1 [MB]
Learning algorithm		
Learning rate (α)		0.01
Maximum episodes (Ep_{max})		50
Power model		
Load-dependency constant (β)		2.4
Power in idle mode (P_0)		107 [W]
Power in dormant mode (P_d)		55 [W]
Reactivation power (P_r)		214 [W]
Topology Management		
Activation threshold (C_{acss})		0.9
Restriction threshold (C_{rest})		0.1
Reactivation period		4 [sec]
Minimum active period		60 [sec]

Table 6.3 Simulation parameters [99, 102, 108, 129, 130]

6.7 Results and Discussion

6.7.1 Convergence analysis

Convergence analysis has been carried out to illustrate the convergence speed as well as system stability before, during and after convergence is reached. This is done for low and high traffic intensities. At low traffic loads, energy efficiency has been chosen as the parameter to show convergence in terms of average number of idle eNBs, see Figure 6.10. At low loads, users using the $iRRC$ with the Staged Action Selection (SAS) rule require just above 10 file trials per eNB to identify each eNB's position in the table. The scheme is file-based rather than session-based to increase its efficiency and convergence speed. In any case, since the scheme has no negative effects on the QoS at low and medium traffic loads, convergence speed is not of paramount importance. The figure also compares the performance of the $iRRC$ using the proposed Staged-Action Selection rule (SAS) to the performance when the Ɛ–greedy action selection rule is used instead. These two are also compared with a pure Ɛ–greedy approach and an RSRP-based approach as a base line. We can notice that the pure Ɛ–greedy approach convergences to near the RSRP performance, both performing the poorest in terms of number of idle eNBs. A user adopting the Ɛ–greedy action selection strategy initially chooses an action randomly as all values in the Q-table are initialised to zero (i.e., there is no prior information stored). If the randomly chosen action is rewarded positively, then this action will be used repeatedly with probability $1 -$ Ɛ. Hence, the Ɛ–greedy based scheme tends to spread users semi-randomly across the access network. The pure Ɛ–greedy approach is enhanced when using the $iRRC$ hierarchical reasoning. This approach uses the Ɛ–greedy as an action-selection strategy but reasoning in a hierarchical manner using the GCUP and PCAP tables to make the final decision as detailed earlier. The $iRRC$ supervises the Ɛ–greedy guiding it towards the selection of the eNB that has the highest GCUP value. In the case of the Staged-Action Selection (SAS) strategy, it first forces users to test all accessible eNBs equally before deciding to use the Q-table. This way, a user has a better insight into which eNBs they want to choose or if there are multiple equally valued actions that they can choose from. In other words, the SAS does not limit the action-selection space to a single optimum action.

Instead, it gives the user all possible optimum actions if more than one exists. The GCUP table is subsequently used to choose one action from the action-space.

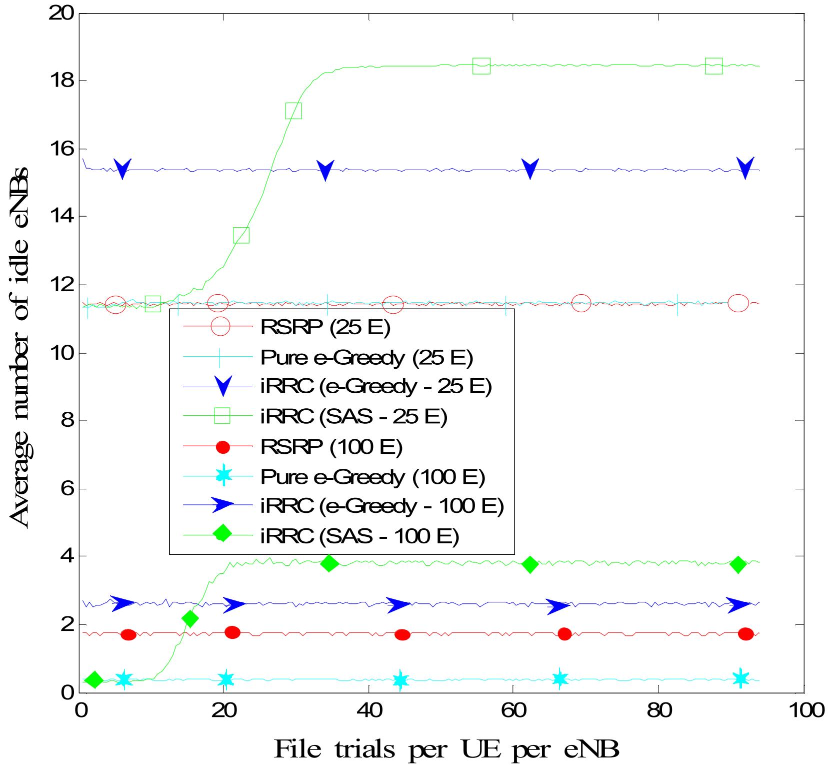


Figure 6.10 System convergence at low and medium traffic loads

At high traffic loads, blocking probability has been chosen instead, see Figure 6.11. Convergence is reached quickly with under 10 trials per eNB. Although the initial performance of the $iRRC$ is the poorest, it convergences to a better solution than the other schemes. The difference in post-convergence performance is around 20% compared to the E[B]-based approach and 40% compared to RSRP and Ɛ–greedy based schemes. Also, at these traffic intensity levels, 8 trials can be performed in a matter of seconds. For this reason, the poor initial performance of the $iRRC$ is traded-off by its long term benefits. The poor initial performance of the $iRRC$ compared to the Ɛ–greedy approach can be attributed to the fact that the aggressive Staged-Action Selection rule is used which explores blindly until convergence is reached whereas the Ɛ–greedy approach only explores with probability $1 - \mathcal{E}$. The initial performance can be enhanced by using the relaxed Staged-Action Selection strategy and tuning the ξ parameter as the operator desires.

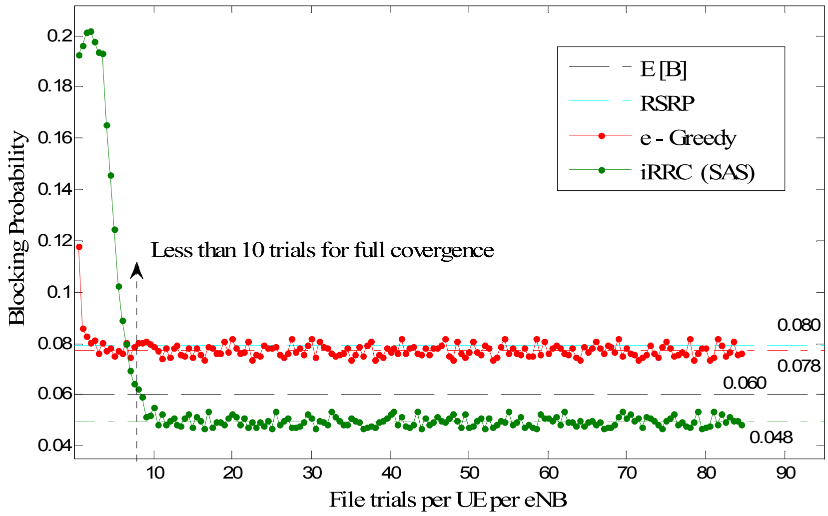


Figure 6.11 System convergence at high traffic loads

The hierarchical processing of the output of the SAS action selection strategy and its usage by the $iRRC$ can be appreciated in Figure 6.10 and Figure 6.11. These figures emphasise the importance of converging to an impartially derived policy. This policy has multiple actions in the action-space. If more than one optimal action exists, it appropriately ranks all sub-optimal actions. This results not only in one single action, as in the case of the Ɛ–greedy rule, but in a ranked list of possible actions depending on their goodness. This way, the $iRRC$ is capable of using the knowledge base more efficiently in terms of QoS as well as energy efficiency. The Ɛ–greedy based solution is omitted from the results hereafter as it is outperformed by the SAS scheme.

6.7.2 **Post-convergence system performance**

Figure 6.12 shows the network idle state probability which reflects the potential energy savings that can be gained by exploiting the idle state of the eNBs. It clearly proves that the proposed intelligent Radio Resource Control ($iRRC$) outperforms the scheme based on Expected bit rate ($E[B]$) as well as the Reference Signal Received Power ($RSRP$) based scheme. Even though the difference is around 10% at very low loads, the $iRRC$ is capable of doubling the number of eNBs in the idle mode for most traffic load levels compared to the $E[B]$ and $RSRP$ based schemes, hence doubling the amount of potential energy savings. Figure 6.12 also shows the benefit of using the restricted mode imposed by the topology management on top of the radio resource management. As shown, the topology management scheme is compatible with all radio resource control schemes and it has great potential in enhancing the system energy efficiency. The figure compares the performance of the topology management (TM) introduced in Chapter 4, the intelligent topology management (iTM) as well as the combination of the these with the $iRRC$ and the $E[B]$. The performance of the TM is enhanced by around 10% when the TMIU uses the information gathered by the learning algorithm to enhance the decision of which eNB to make available and which to restrict its service. Even though the usage of the topology management limits the possible increment in the number of idle eNBs, the performance is further enhanced by combining it with the intelligent radio resource control mechanism.

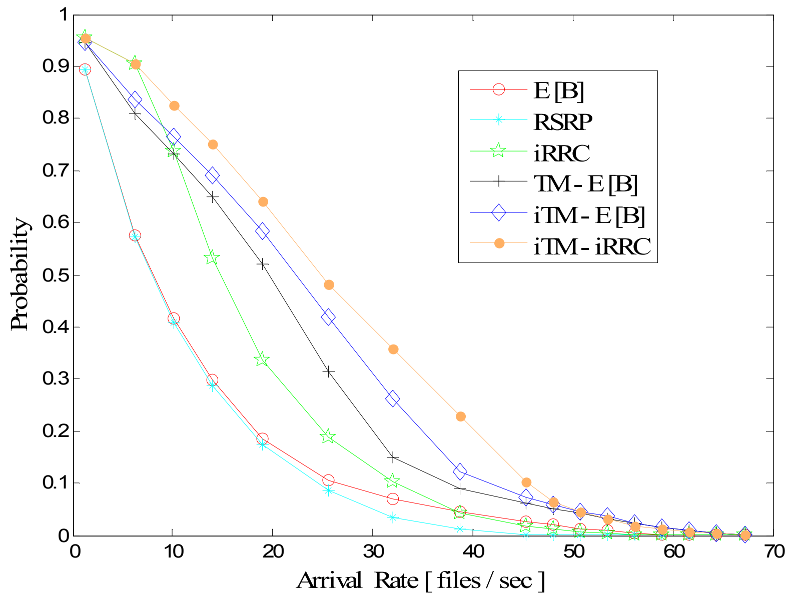


Figure 6.12 Network idle state probability

Figure 6.13 shows the network power requirements which provide an indication of the potential of the proposed schemes to enhance the network energy efficiency by applying a sleep-mode algorithm by the TMSU. The sleep mode scheme is used in order to exploit the idle state of the eNBs. Clearly, the topology management is capable of reducing the network energy consumption by up to 40% at low traffic loads.

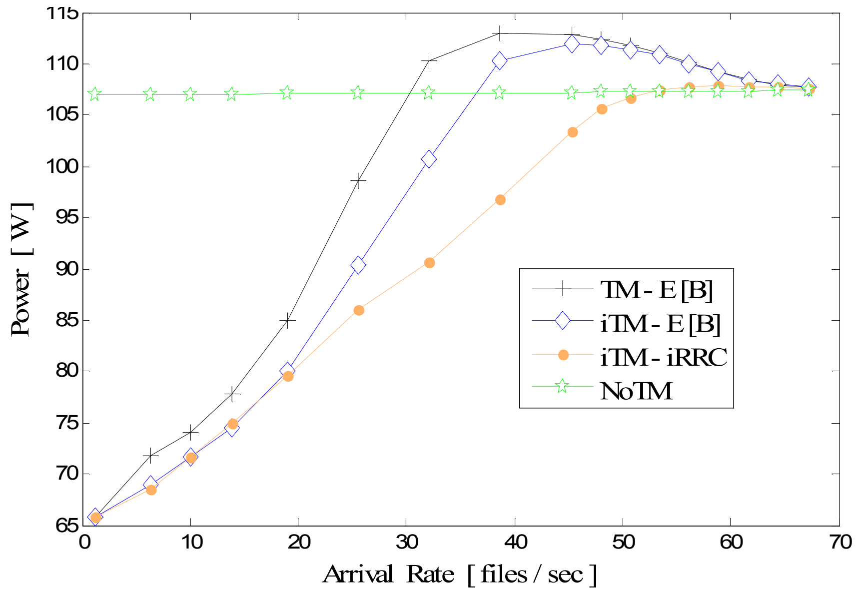


Figure 6.13 Network Power Requirements

As Figure 6.12 suggests, better identification of which eNB to switch to the dormant mode allows for further increase in energy gains. Also, by adequately clustering the users using the $iRRC$, the sleep-mode algorithm is capable of keeping more eNBs in the sleep mode and for longer periods which further enhances the network energy efficiency.The figure also shows that at high traffic loads the network requires more power to operate when using the topology management jointly with the $E[B]$. The TMCU switches additional eNBs to the accessible state even though it does not necessarily result in frequent switching from the dormant to the active mode which requires additional energy for the reactivation. This is resolved by the $iRRC$ by dividing the $GCUP$ table into two groups, the High Loaded and the Light Loaded groups. This prevents the condition at which the TMCU activates an eNB to be met unless necessary. Figure 6.14 and Figure 6.15 show the system QoS and give an insight into the spectral efficiency in terms of probability of retransmission and end-to-end delay respectively. First of all, both figures reflect the fact that the topology management has no or little impact on the system QoS.

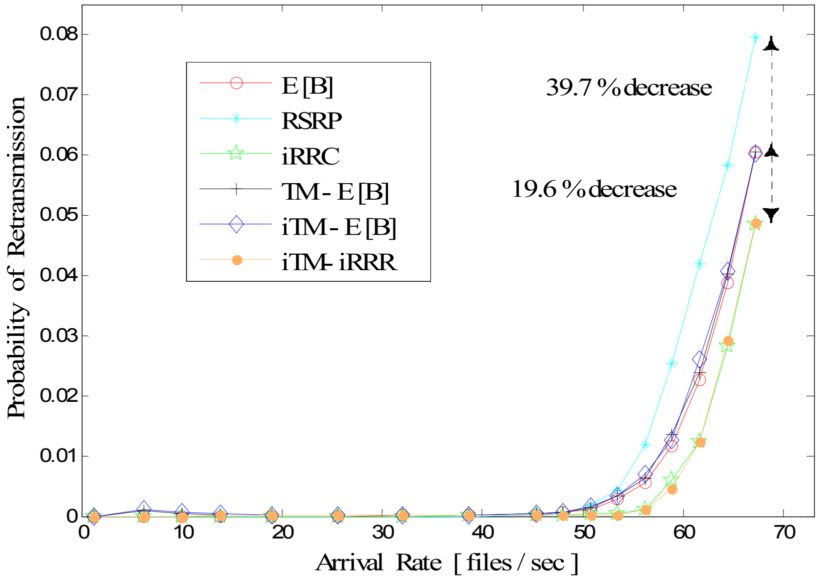


Figure 6.14 Probability of Retransmission

Figure 6.14 and Figure 6.15 also show that the $iRRC$ not only has the ability to obtain energy gains at low and medium traffic loads but it also has the potential of increasing the system spectral efficiency at high traffic loads by reducing the retransmission probability by around 40% and 20% in comparison with the $RSRP$ and $E[B]$ based schemes respectively.

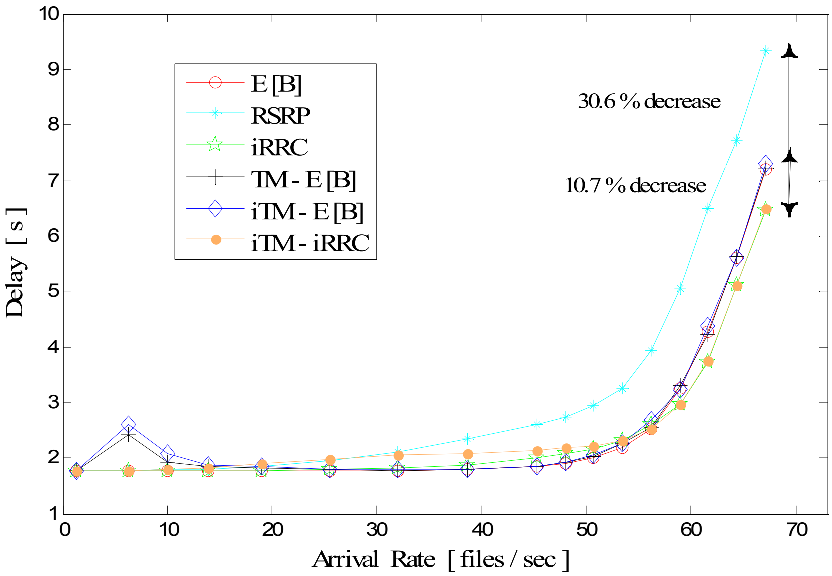


Figure 6.15 System End-to-End Delay

The peaks in delay at low traffic loads for TM-$E[B]$ and iTM-$E[B]$ is because users use the $E[B]$ to decide which eNB to camp on. Users selfishly choose the eNB which is predicted to provide the best data rates which in most cases increases the level of interference hence reducing spectrum reusability.

Figure 6.16 provides a more detailed view of how each scheme is able to enhance the system energy efficiency as well as spectral efficiency. At high traffic loads, it is interesting to note that the $iRRC$ delivers better performance in terms of both energy efficiency as well as QoS by requiring less effort in terms of the required number of UE-cell reselections when compared with the $E[B]$ and $RSRP$ based schemes. In fact the $iRRC$ has the least UE-cell reselection frequency even when coupled with the topology management compared with the $E[B]$ based approach. This is mainly because at high traffic loads a user has only one acceptable action in the action-space, hence UEs tend to be associated with their optimal choice unless this is no longer available. This minimises signalling overhead which is of great concern at high traffic intensities.

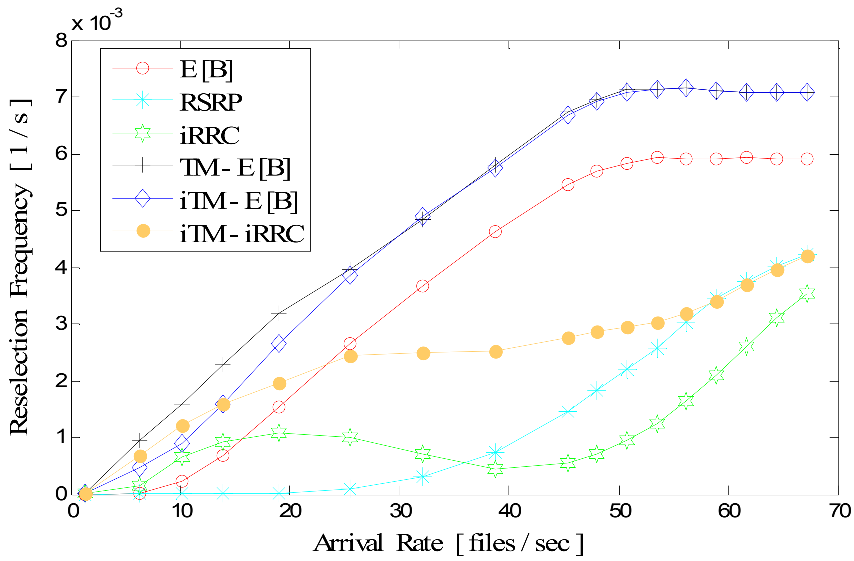


Figure 6.16 Cell Reselection Frequency

At low and medium traffic loads, however, the $iRRC$ effectively associates and dissociates users with and from different cells in the access network as appropriate in order to cope with traffic fluctuations, reduce the interference and enhance the network energy efficiency. The use of the topology management introduces an additional dimension to the dynamics of the system as it switches the state of the eNBs from accessible to restricted and vice-versa to deal with temporal and spatial traffic fluctuation; hence users change their serving cell more frequently. Given that the $RSRP$ based scheme is making no efforts to try to mitigate the energy efficiency problem, it has the lowest cell reselection frequency among the presented schemes at low loads.

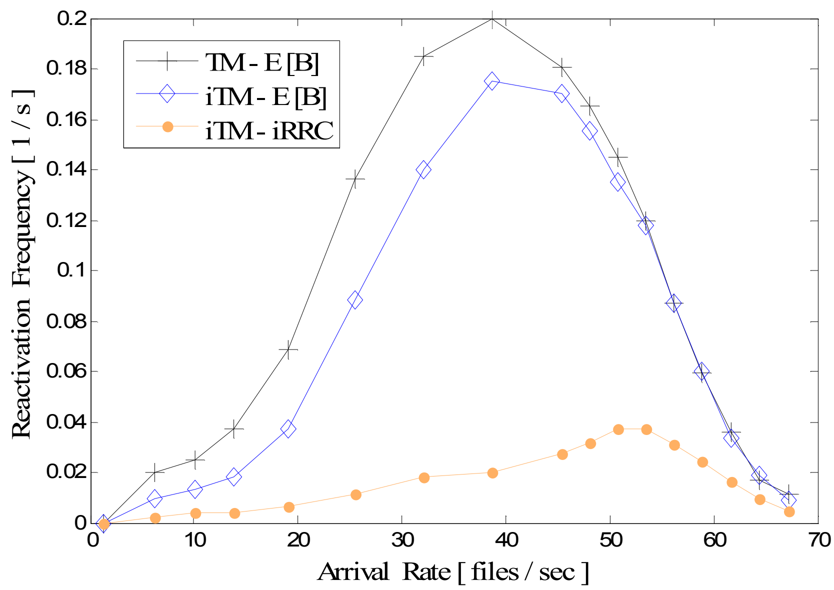


Figure 6.17 Cell Reactivation Frequency

Figure 6.17 reveals an important aspect of the application of sleep modes in current and future networks; that being the cell reactivation frequency. Cell reactivation occurs when a cell in the dormant mode is reactivated due to traffic increase. It is important that the network is stable and cells do not change their status too frequently as this can worsen the network energy efficiency as seen in Figure 6.13. It is clear from Figure 6.17 that the network is much more stable when using intelligence to optimise different decisions. First, the decision of which cell needs to be kept accessible and which need to be switched to the restricted state is improved by the iTM which shows a decrease in the number of times cells are switched back to the accessible state. The reactivation frequency is further decreased by a significant 10-fold figure at medium traffic loads when the iTM is combined with the $iRRC$. This is mainly due to the load monitoring capabilities of the

$iRRC$ algorithm when clustering users onto eNBs to avoid overloading the eNBs which in turn avoids triggering the reactivation of restricted cells unnecessarily.

6.8 **Conclusion**

This chapter investigated novel cognitive learning and reasoning algorithms which are capable of translating past experience into valuable sets of information in order to optimise decisions taken at different layers of the protocol stack (i.e., core network decisions and RRC layer decisions). These are the Staged Action Selection strategy (SAS) and the Multitask Hierarchical Reasoning mechanism (*M-HR*). The SAS strategy builds an impartial knowledge base assuring 1) better post-convergence performance of the chosen best action, or best set of actions if more than one exists, 2) accurate ranking of all sub-optimal actions. The multitask hierarchical reasoning scheme, on the other hand, allows for interpreting users' past experience and the output of the novel action selection strategy into a multitask knowledge base. This knowledge base is subsequently used to develop two important techniques to enhance the system energy efficiency as well as QoS. The first is an intelligent Radio Resource Management ($iRRC$) which sets the adequate degree of clustering of users onto eNBs. This carefully balances and unbalances the offered traffic across the access network depending on the traffic level. This is a crucial requirement for 5G deployments as set in Chapter 2 and Chapter 3. The $iRRC$ is capable of doubling the probability of eNBs to be idle, which in turn increases the energy gains that can be achieved when exploiting the idle mode of the eNBs. Also, by appropriately balancing the traffic at high traffic loads, the system QoS is enhanced by an average of 30% in comparison with standard techniques. In addition, an intelligent Topology Management (iTM) scheme is proposed that aims to control the operational (accessible / restricted) of eNBs of an opportunistic network in order to gain energy savings while delivering the required QoS and enable dynamic roll out / back. The performance of the TM introduced in Chapter 4 is enhanced by around 10% when the Topology Management Information Unit (TMIU) uses the information gathered by the learning algorithm to enhance the decision of which eNB to make available and which to restrict its service. The performance is further enhanced by combining both intelligent schemes the iTM and the $iRRC$ to keep the reactivation of cells at its lowest. The $iRRC$ enhances the performance of the iTM at all traffic loads by an average of 15%. This figure translates to up to 50% in energy savings when exploiting the idle state of the eNBs using the sleep mode mechanism. Last, but not least, the implementation as well as feasibility of the proposed $iRRC$ algorithm is studied in detail and shown to be compatible with 3GPP standards.

Chapter 7 **Future work**

Contents

This chapter discusses potential research avenues related to or as a continuation of the work presented in this thesis. These avenues are identified based on the outcomes of different techniques introduced in this work to explore additional capabilities of the protocols, and to help improve their performance further.

7.1 **Traffic of the future**

Even though the schemes presented here are aimed at being flexible and adaptable to different traffic profiles, Machine Type Communication (MTC) traffic needs to be studied, and their requirements taken into consideration. This is to investigate its effects on the developed techniques and how these can be improved to accommodate the traffic of the future. MTC is predicted to penetrate quite heavily into cellular deployments which might impose a range of different requirements for the topology and radio resource management to work efficiently. The study of MTC traffic in this context can be generally split into two main scenarios. The first represent a suburban or residential area in which most machines are static and with predictable traffic profiles. The second scenario is more challenging as it also contains nomadic machines with unpredictable demand such as in urban areas.

7.2 Efficient Energy-Aware Exploration

Even though this thesis proposed ways to explore, process and interpret past experience into useful information that can help the functioning of a range of energy-aware mechanisms, the proposed cognitive techniques themselves are not energy-aware. The work presented in this thesis can be complemented by developing an efficient energy-aware exploration process which would require less energy while exploring. A potential approach to achieve this is to heuristically restrict the action space by ruling out actions of low likelihood to be effective. Another approach is the use of Transfer Learning (TL) to minimise the number of agents required to explore the environment. Mature agents would then share their knowledge with the rest. Both techniques are of great potential and can be applied at the same time achieving greater performance. A number of challenges exist in the application of these approaches such as the accurate identification of actions to dismiss, or the kind, amount and frequency at which knowledge needs to be transferred and to which agent.

7.3 Location-Aware Resource Management

The work presented in this thesis assumes that each UE builds their own knowledge base learning independently from other agents. However, as a way to tackle mobility issues and enhance convergence speed as well as initial performance, location-aware Transfer Learning (TL) techniques can be utilised by associating the knowledge base with a particular location rather than an agent. This has a range of potential benefits. Firstly, it can limit the action space of individual agents as more than one agent can contribute to the same location-based knowledge base. Also, this application supports the mobility of agents as the information is associated with specific locations. Nomadic agents will be given a totally different knowledge base upon leaving their current location and entering the location of another knowledge base. Moreover, immature or newly deployed agents will not require exploring the environment as they will be given a mature knowledge base that can be utilised instantly. The application of location aware techniques have also the potential of solving or aiding the Hand Over (HO) decision of choosing a target cell taking into consideration the speed and direction of the agent. Challenges in this regard include the update frequency of such information and the size and shape of the area of each knowledge base which can be derived from localised traffic profiles and average speed of agents.

7.4 **Joint Intelligent Resource and Spectrum Management**

The work presented in this thesis tightly couples radio resource with topology management. However, spectrum management is another dimension to the problem which has not been taken into account. Optimally, radio resource, topology and spectrum management needs to be controlled and optimised boosting the performance of the system not only in terms of energy efficiency and QoS but also in terms of spectrum efficiency. The choice of the sub-channel can be further studied to enhance spectrum reusability on top of the choice of eNB which is performed by the iRRC. The operation of the iRRC can be enhanced firstly using standard interference mitigation techniques such as ICIC or eICIC. Also, Reinforcement Learning (RL) can be used to aid the iRRC with the decision of the sub-channel to maximise spectrum reusability. A trade-off between the iRRC clustering capability and spectrum reusability needs to be investigated.

7.5 Energy-Efficient Backhaul Networks

The work presented here splits the access network from the backhaul network focusing on energy related aspects only in the access network. For example, in BuNGee, each ABS has only one HBS for backhauling making it impossible the switch to the dormant mode of any HBS without largely degrading the system QoS. Also, when clustering UEs onto a limited number of ABSs in the access network, a risk of backhaul bottleneck exists. Hence the backhaul network needs to be further investigated and designed to be responsive to large traffic tempo-spatial fluctuations and minimise latency, while guaranteeing cost and energy efficient ubiquitous access. Techniques such as Software Defined Networking (SDN) can be considered for flexible resource management and flexible capacity provisioning in the backhaul network to meet the demand of the high traffic fluctuation at the access network.

Chapter 8 **Summary and Conclusions**

Contents

8.1 **Summary and conclusions**

This thesis has investigated radio resource and topology management schemes and protocols taking into consideration their interoperability. The techniques developed are showed to largely enhance the energy efficiency as well as system capacity and QoS of current and future wireless mobile deployments operating under unpredictable and spatially uneven traffic demand.

This thesis demonstrated the requirements to implement advanced radio resource and topology management in various deployment strategies highlighting the scenarios in which these techniques can be feasibly implemented and substantially improve performance. The application of advanced radio resource and topology management, such as the intelligent Radio Resource Control ($iRRC$), the intelligent Topology Management (iTM), and the Cluster-Based Topology Management ($CBTM$), are also found to perform significantly better than standard techniques doubling the energy efficiency of the network while providing a better QoS and increasing the system capacity. These techniques are also designed to be more flexible than existing proposals and are shown to outperform state of the art techniques particularly when dealing with spatially uneven traffic distribution. These techniques are capable of efficiently operate within the same network which further enhances their individual performance. Despite the dynamic nature of mobile wireless environments, Artificial Intelligence (AI) was also shown to be applicable in a much more flexible and beneficial way for such environments with the help of innovative auxiliary tools that have been developed as part of the work presented here. These tools, namely the Staged Action Selection strategy and the Multitask Hierarchical Reasoning mechanism, define efficient ways for exploring mobile wireless networks and how the information gathered from the exploration phase can be processed to benefit and enhance the operation of multiple tasks within the network. The application of such techniques opens new ways for the development of flexible and self-organising protocols at different layers that can efficiently adapt to mobile wireless environments.

Chapter 1 presented a brief overview of the evolution of wireless mobile networks and stated the reasons for the importance of the research presented in this thesis.

Background information and a related literature review of energy-aware resource management techniques were provided in Chapter 2. Cognition and intelligence are among the most sought after capabilities to reach the ultimate goal of delivering the services in an energy-efficient and adaptable manner, due to their ability to enable real time self-adaptability and self-awareness. A detailed state-of-the-art in the field of energy-aware topology and radio resource management was presented, defined, and classified according on their employed techniques, addressing their advantages and drawbacks. It is noticed from the literature that there is neither a clear definition nor function related to the term “sleep modes”; it has a range of different meanings. Therefore, this chapter also introduced the functional building blocks as part of the operation of topology management schemes. Despite clear dependencies, current proposed topology and radio resource management mechanisms are being addressed separately. A scalability and adaptability issue is also noted across the proposed solutions.

Chapter 3 evaluated the application of Load Balancing and the emerging technique of Load Unbalancing for the two most common network deployment strategies; namely, the homogenous and the heterogeneous deployments. Queueing theory was used to conduct an evaluation taking into consideration a wide range of system KPIs. These are the system QoS in terms of blocking probability, network energy efficiency as well as network stability in terms of frequency from the idle to the active state. The importance of having a macro-cell overlay was also assessed.

It was concluded that having a macro-cell overlay not only enhances system performance but it also adds flexibility to the underlying capacity boosters and adds stability to the network. The priority of the Macro-eNB (MeNB) plays a more important role than the RRM technique in stabilising the network and avoiding redundant transitions from the idle to the active state at all traffic loads. In other words, different RRM schemes largely affect the energy efficiency of the network but the settings of the MeNB have a larger effect on network stability. It was also concluded that homogenous deployment strategies are more suitable in areas that have a semi-constant traffic demand or traffic of low fluctuation over time. Heterogeneous deployment strategies on the other hand were recommended to be deployed in areas with unpredictable traffic demand and of medium to high fluctuation nature.

The application of energy-aware topology management for 5G deployments was studied in Chapter 4. This chapter firstly introduced an energy-aware Neighbour-Based Topology Management (NBTM) scheme for beyond next generation mobile broadband systems which fine-tunes the status of the network nodes (dormant / active) depending on the traffic demands. Secondly, Macro-cell overlaid TM (McTM) which enhances the operation of NBTM in terms of energy efficiency as well as QoS at low traffic loads was introduced. The macro cell-overlaid scheme outperforms the neighbour-based scheme, which outlines the benefits of having macro level coverage as suggested in Chapter 3, especially at low traffic loads from both the energy saving as well as QoS perspectives.

The study conducted in Chapter 4 was extended in Chapter 5 to investigate the effects of having a spatially uneven traffic distribution in an ultra-dense urban area. It was showed that the management of Ultra Dense Network (UDN) deployments requires special attention particularly when the network is suffering from spatially unbalanced traffic distribution as traditional schemes fail to operate under these conditions, resulting in low QoS and high energy consumption. To this end, a traffic-aware topology management scheme was introduced. The scheme successfully follows localised traffic increase caused by temporary events and only activates the infrastructure which can provide and augment the network capacity density while maintaining the desirable QoS.

Finally, the application of cognitive intelligent mechanisms was studied in Chapter 6. As suggested in Chapter 2, these techniques enable the delivery of the required QoS in an energy-efficient, adaptable manner. Auxiliary tools such as the Staged Action Selection (SAS) strategy and the innovative Multitask Hierarchical Reasoning (M-HR) mechanism were proposed for the efficient application of Reinforcement Learning (RL) to mobile wireless network environments. Particularly, RL was used to facilitate the development of adaptive intelligent Radio Resource Control ($iRRC$) as well as enhance the operation of the TM proposed in the previous chapters. The $iRRC$ adaptively manages the offered traffic to improve the energy efficiency at medium and low traffic loads while maximising system capacity and enhancing the QoS at high traffic loads. To enable the efficient joint operation, this chapter also took into consideration interoperability issues between energy aware radio resource and topology management techniques. The implementation as well as feasibility of the proposed $iRRC$ algorithm was studied in detail and shown to be compatible with 3GPP standards.

8.2 **Novel contributions**

8.2.1 **Development of cognitive tools for the effective application of Reinforcement Learning in mobile wireless communication environments**

On the Artificial Intelligence (AI) side, this thesis has introduced a range of tools to adapt Reinforcement Learning (RL) to the requirements of mobile wireless communication networks. Specifically, the proposed tools define efficient ways for exploring mobile wireless networks and how the information gathered from the exploration phase can be processed to benefit and enhance the operation of multiple tasks within the network. These are the Staged Action Selection strategy (SAS) and the Multitask Hierarchical Reasoning mechanism (*M-HR*). As detailed in Chapter 6, learning and reasoning mechanisms in the literature strive to converge on one action from the action space without paying much attention to the goodness of the remaining actions. The novel action selection strategy proposed here builds an impartial knowledge base assuring 1) better post-convergence performance of the chosen best action, or best set of actions if more than one exists, 2) accurate ranking of all sub-optimal actions. This results not only in one single optimal action but in an optimal ranked list of possible actions depending on their goodness. The multitask hierarchical reasoning scheme, on the other hand, allows for interpreting users' past experience and the output of the novel action selection strategy into a multitask knowledge base. The scheme translates the users' interaction with the access network into regional and global information about the network enabling its reusability by a range of agents to enhance their operation at the RRC layer and system level. To the best of our knowledge, this thesis is the first exploration of such auxiliary cognitive tools for the application of RL to enhance system level performance and RRC protocols for wireless mobile networks. These contributions have been **submitted to IEEE Transactions on Mobile Computing**.

8.2.2 **Intelligent radio resource control based on RL**

This thesis successfully used an enhanced RL technique to develop an intelligent Radio Resource Control ($iRRC$) mechanism. The $iRRC$ adaptively balances and unbalances the offered traffic depending on the traffic level to improve the energy efficiency at medium and low traffic loads while maximising system capacity and enhancing the QoS at high traffic load. Radio Resource Control mechanisms available in the literature mostly focus on balancing the load to enhance system capacity and user experience. These also ignore the performance of the MAC layer, the RACH channel more specifically, which can be important as in the case of Machine Type Communications (MTC) for instance. The $iRRC$ performs cross-layer optimisation as the cell (re)-selection procedure takes into account the outcome of past interactions with the MAC layer and the system user plane as a whole.

8.2.3 **Real time energy-aware topology management for opportunistic deployments**

A range of topology management schemes were proposed in various chapters of this thesis for the ABSOLUTE temporary event and disaster relief systems to control the dormant / active states of the eNBs. Apart from the architectural difference of the investigated systems, previously proposed techniques require a number of information exchange iterations between neighbouring eNBs to make different decisions. The schemes proposed here are based on existing information exchange procedures using locally available information (i.e., at the eNB) to derive different policies without requiring any further information or a QoS feedback. These techniques firstly enhance the system energy efficiency and prolong the battery life of the eNBs of ABSOLUTE systems while delivering the required QoS. Secondly, the schemes are also used to control the roll out and roll back of ABSOLUTE systems minimising the need for periodic human intervention. To the best of our knowledge, this is the first study of energy-aware topology management in the context of unexpected temporary events. Other techniques were developed to manage the ABSOLUTE disaster relief network using a centralised QoS-based approach. Our approach uses a distributed clustering technique based on the identification of nodes that can cooperate in serving a given area to neutralise the effect of having spatially unbalanced traffic distribution. These various contributions have been presented at the ***International Conference on Advanced Technology for Communications (ATC2013),*** the ***International Conference on Computing, Networking and Communication (ICNC2013)***, and the ***Future Network & Mobile Summit (FNMS2013).*** Part of these contributions has been also **submitted to IEEE Transactions on Mobile Computing**.

8.2.4 **Application of RL to energy-aware topology management**

A novel RL-based topology management mechanism was proposed that intelligently identifies which eNBs to activate and deactivate. Previous solutions to identify eNBs to activate and deactivate are mostly based on complex centralised calculations or heuristic approaches. The most common is the choice of a neighbouring eNB as this is seen as contributing to the regional traffic demand. However, even in a homogenous deployment of eNBs with omnidirectional antennas an eNB has a number of neighbouring nodes. Therefore further information is needed to make an accurate decision. Moreover, the use of sectored antennas and the increasing popularity of directional antennas adds complexity to the choice of the most suitable eNB to reactivate. The scheme proposed here bases its decision on the degree of goodness of the eNBs which is collected from the interaction of UEs with the access network. These contributions have been also **submitted to IEEE Transactions on Mobile Computing**.

8.2.5 **Efficient joint operation of advanced radio resource and topology management**

An energy-aware network management middleware was proposed to enable the efficient joint operation of energy-aware radio resource and topology management schemes. To the best of our knowledge, there is no approach or study in the literature that takes into consideration interoperability issues when coupling advanced radio resource and topology management mechanisms. The middleware monitors the loading state of the eNBs and uses a capacity-based threshold as a function of the reactivation policy of the topology management to harmonise its joint operation with advanced radio resource management techniques. These contributions have been accepted for presentation in the ***IEEE Vehicular Technology Conference (VTC 2015-Fall).***

8.2.6 **Performance estimation of radio resource, topology management, and the importance of a macro-cell overlay**

A novel contribution of this thesis is the development of a theoretical framework based on queueing theory capable of considering a range of spatial traffic distributions. The framework was used to accommodate the evaluation of the application of a range of load management schemes and assess their requirements. This was done for homogenous as well as heterogeneous deployments to have a clear insight on how the macro-cell overlay can contribute to solving the energy efficiency problem and also to specify the scenarios in which these perform best. These techniques are assessed in the literature from the QoS and capacity point of view but fail to give a clear insight into the contribution of various entities and radio resource protocols to the energy efficiency problem neither do they take the interoperability of energy aware solutions into account. These contributions have been **submitted to Transactions on Emerging Telecommunications Technologies**. The theoretical framework was also used to validate and investigate the effectiveness of various proposed topology management schemes under spatially uneven traffic conditions such as the cluster-based topology management. These contributions have been **submitted to IEEE Transactions on Mobile Computing**.

8.2.7 **Definition and introduction of the building blocks of energy-aware topology management**

A fundamental, novel aspect of this thesis is the introduction of various functional building blocks as part of the operation of topology management schemes. These are the Topology Management Information Unit ($TMIU$), the Topology Management Control Unit ($TMCU$), and the Topology Management Sleep-mode Unit ($TMSU$). This is necessary as the terms are vaguely defined in the literature and restrict their usage to energy related schemes. However, even though topology management is used to primarily enhance energy efficiency of wireless mobile networks in this thesis, it can be used, for instance, to dynamically control the predesigned topology of a network for QoS purposes. In addition, the term used to describe the switching of eNBs between the dormant and active modes is "sleep modes". This term is in fact used interchangeably to sometimes mean switching of eNBs and others to indicate the usage of traffic redistribution with the aim to switch free eNBs to the dormant mode. The identification and clear definition proposed in this thesis helps facilitate the development and practical implementation of topology management schemes as its function as well as the function of its different units is clearly identified within a mobile wireless system. These contributions have been **submitted to Elsevier computer networks**.

Glossary

3GPP	3rd Generation Partnership Project
ABS	Access Base Station
AeNB	Aerial eNB
AI	Artificial Intelligence
ANR	Automatic Neighbour relation
BBC	Backhaul-Based Control
BBU	Base Band Unit
BP	Blocking Probability
BTS	Base Transceiver Station
BSC	Base Station Controller
BS	Base Station
BuNGee	Beyond Next Generation Mobile Broadband System
BW	Bandwidth
CAP	Cell Association Policy
CBTM	Cluster Based Topology Management
CCR	Clustering Capability Rating
CDMA	Code Division Multiple Access

CeNB	Cluster-head eNB
CID	Cell ID
CN	Cognitive Network
CoMP	Coordinated MultiPoint
CQ-CCR	QoS-aware Clustering Capability Rating
CR	Cognitive Radio
DRX	Discontinuous Reception
DT	Delaunay Triangulation
DTX	Discontinuous Transmission
E[B]	Expected Bitrate
E-UTRAN	Evolved Universal Terrestrial Access Network
E-W	East-West
ECGI	Evolved Global Cell Identifier
ECI	E-UTRAN Cell Identifier
EDR	Energy Depletion Rate
EE	Energy Efficiency
eNB	Evolved Node B
EPC	Evolved Packet Core

FDD	Frequency Division Duplex
FDMA	Frequency Division Multiple Access
FME	Flexible Management Entity
GCUP	Global Cell Utility Policy
Gd	Green degree
HBS	Hub Base Station
Het-BuNGee	Heterogeneous Beyond Next Generation Mobile Broadband System
HetNets	Heterogeneous Networks
HL	High Loaded
HO	Hand Over
HSDF	Hot-Spot Densification Factor
HSS	Hub Subscriber Station
HW	Holt-Winter
IA-CCR	Interference-Aware Clustering Capability Rating
IAAS	Infrastructure as a Service
IAUB	Infinite Area Upper-Bound
IEs	Information Element
IMT	Information Management and Technology

iRRC	Intelligent Radio Resource Control
iSONs	Intelligent Self-Organised Networks
ISP-SC	Idle State Probability of the Small Cell layer
ISS	Idle State Stability
iTM	Intelligent Topology Management
ITU-R	International Telecommunication Union-Radiocommunication Sector
KPI	Key Performance Indicator
LAA	License Assisted Access
LB	Load Balancing
LL	Light Loaded
LSA	Licensed Shared Access
LTE	Long Term Evolution
LTE-A	Long Term Evolution-Advanced
LUB	Load Unbalancing
M-HR	Multitask Hierarchical Reasoning
M2M	Machine-to-Machine
MAC	Medium Access Control
McTM	Macro-cell overlaid Topology Management

MeNB	Macro Evolved Node B
MIB	Master Information Block
MIMO	Multiple-Input Multiple-Output
MME	Mobility Management Entity
MS	Mobile Station
MTC	Machine-Type Communication
N-S	North-South
NBC	Neighbour Based Control
NBTM	Neighbour-Based Topology Management
NCCR	Normalised Clustering Capability Rating
NFV	Network Function Virtualisation
OT	Offered Traffic
PBCH	Physical Broadcast Channel
PCAP	Personal Cell Association Policy
PCI	Physical Cell ID
PLM	Proactive Load Management
PLMN	Public Land Mobile Network
PSS	Primary Synchronisation Signal

QoS	Quality of Service
RACH	Physical Random Access Channel
RAN	Radio Access Network
RAT	Radio Access Technology
RBG	Resource Block Group
RL	Reinforcement Learning
RLM	Reactive Load Management
RNC	Radio Network Controller
RRC	Radio Resource Control
RRM	Radio Resource Management
RSRP	Reference Signal Receive Power
RSRQ	Reference Signal Receive Quality
SAC	Self-Aware Control
SAS	Staged Action Selection
SDA	Software Defined Architecture
SDN	Software Defined Networks
SeNB	Small-cell eNB
SIB	System Information Block

SINR	Signal-to-noise-plus-interference ratio
SMPD	Semi-Markov Decision Process
SONs	Self-Organised Networks
SOTA	State-of-The-Art
SPA	Sum Product Algorithm
SSS	Secondary Synchronisation Signal
TDD	Time Division Duplex
TDMA	Time Division Multiple Access
TeNB	Terrestrial eNB
TL	Transfer Learning
TM	Topology Management
TMCU	Topology Management Control Unit
TMIU	Topology Management Information Unit
TMSU	Topology Management Sleep-mode Unit
TSB	Truncated Shannon Bound
UBC	User-Based Control Agent
UDN	Ultra-Dense Networks
UE	User Equipment

UG	User Group
UTRAN	Universal Terrestrial Access Network
vEPC	Virtual EPC
WAN	Wide Area Network
WCDMA	Wideband Code Division Multiple Access
WiMAX	Worldwide Interoperability for Microwave Access

References

[1] L. R. Marcus, A. T. Michael, F. M. Robert, and A. R. Richard, "Evolution of the air interface of cellular communications systems toward 4G realization," *Communications Surveys & Tutorials, IEEE,* vol. 8, pp. 2-23, 2006.

[2] T. S. Rappaport, *Wireless Communication Principles and Practice* Second edition ed.: Prentice Hall 2001.

[3] D. Gilstrap, "Traffic and Market Report, On the Pulse of the Networked Society," Ericsson, SE-126 25 Stockholm, Sweden,, Rep. June 2012.

[4] A. Ghosh, R. Ratasuk, B. Mondal, N. Mangalvedhe, and T. Thomas, "LTE-advanced: next-generation wireless broadband technology [Invited Paper]," *Wireless Communications, IEEE,* vol. 17, pp. 10-22, 2010.

[5] O. Marinchenco, N. Chayat, A. Burr, A. Papadogiannis, M. Dohler, M. Payaró, M. Ware, "BuNGee Project Overview," in *in Proceedings of the 3rd International IEEE Conference on Microwaves, Communications, Antennas and Electronic Systems (IEEE COMCAS Nov 7 - Nov 9)*, 2011.

[6] P. Popovski, V. Braun, G. Mange, P. Fertl, D. G. Serrano, N. Bayer, H. Droste, A. Roos, G. Zimmermann, M. Fallgren, A. Höglund, H. Tullberg, S. Jeux, Ö. Bulakci, J. E. Li, P. Marsch, K. Pawlak, M. Boldi, J. Monserrat, "D6.2: Initial report on horizontal topics, first results and 5G system concept," METIS FP7 Project, [Online] available at: https://www.metis2020.com/documents/deliverables/, Deliverable 2014.

[7] Ruckus Wireless Inc, "Dealing with Density: The Move to Small-Cell Architectures [Online] available at: http://www.ruckuswireless.com/whitepapers)," White Paper2013 [Accessed: Jan 2015].

[8] Alcatel-Lucent, "Making Smart Use of Excess Antennas: Massive MIMO, Small Cells, and TDD," *Bell Labs Technical Journal,* pp. 5–21, 2013.

[9] 4GAmericas, "Meeting the 1000x Challenge: The need for Spectrum, Technology and Policy Innovation [Online] available at: http://www.4gamericas.org/en/resources/white-papers/," *White Paper,* October 2013 [Accessed: December 2014].

[10] Ericsson, "5G radio access: Research and Vision," Ericsson, Stockholm, Sweden,, White paperJune 2013.

[11] G. Gür and F. Alagöz, "Green wireless communications via cognitive dimension: an overview," *Network, IEEE (Volume:25 , Issue: 2)* vol. 25, pp. 50-56, 2011.

[12] Z. Hasan, H. Boostanimehr, and V. K. Bhargava, "Green Cellular Networks: A Survey, Some Research Issues and Challenges," *Communications Surveys & Tutorials, IEEE (Volume: 13 , Issue: 4)* pp. 524-540, 2011.

[13] 5GPPP, "EC H2020 5G Infrastructure PPP Pre-structuring Model RTD & INNO Strands," *Recommendation by 5G Infrastructure Association Version 2.0,* 2014.

[14] S. Bu and F. R. Yu, "Green Cognitive Mobile Networks With Small Cells for Multimedia Communications in the Smart Grid Environment," *Vehicular Technology, IEEE Transactions on,* vol. 63, pp. 2115-2126, 2014.

[15] International Telecommunication Union, "Definition of Software Defined Radio (SDR) and Cognitive Radio System (CRS)," SM.2152, 2009.

[16] K. Watanabe and M. Machida, "Outdoor LTE Infrastructure Equipment (eNodeB)," *Fujitsu Sci. Tech, Vol. 48, No. 1, pp. 27–32,* January 2012.

[17] 3GPP, "Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Self-configuring and self-optimizing network (SON) use cases and solutions TR 36.902 v.9.3.1 R9," 2011.

[18] 3GPP, "Universal Mobile Telecommu nications System (UMTS); LTE; Telecommunication management; Self-Organizing Networks (SON); Concepts and requirements, TS 32.500 R11," vol. 11, ed: 3GPP, 2012.

[19] T. L. Marzetta, "Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas," *Wireless Communications, IEEE Transactions on,* vol. 9, pp. 3590-3600, 2010.

[20] J. G. Andrews, H. Claussen, M. Dohler, S. Rangan, and M. C. Reed, "Femtocells: Past, Present, and Future," *Selected Areas in Communications, IEEE Journal on,* vol. 30, pp. 497-508, 2012.

[21] M. Lopez-Benitez and J. Gozalvez, "Common Radio Resource Management Algorithms for Multimedia Heterogeneous Wireless Networks," *Mobile Computing, IEEE Transactions on,* vol. 10, pp. 1201-1213, 2011.

[22] L. Ying Loong, C. Teong Chee, J. Loo, and A. Vinel, "Recent Advances in Radio Resource Management for Heterogeneous LTE/LTE-A Networks," *Communications Surveys & Tutorials, IEEE,* vol. 16, pp. 2142-2180, 2014.

[23] Ericsson, "Ericsson Mobility Report," Report June 2015.

[24] A. Fehske, G. Fettweis, J. Malmodin, and G. Biczok, "The global footprint of mobile communications: The ecological and economic perspective," *Communications Magazine, IEEE,* vol. 49, pp. 55-62, 2011.

[25] U. Paul, A. P. Subramanian, M. M. Buddhikot, and S. R. Das, "Understanding traffic dynamics in cellular data networks," in *INFOCOM, 2011 Proceedings IEEE*, 2011, pp. 882-890.

[26] L. M. Z. Correia, D. Blume, O. Ferling, D. Jading, Y. Godor, I. Auer, G. Van der Perre, L., "Challenges and enabling technologies for energy aware mobile radio networks," *Communications Magazine, IEEE,* vol. 48, pp. 66-72, 2010.

[27] SEMAFOUR, "Perspectives of Network Management Evolution Towards 5G," in *Presentation FIA 2014 RAS Workshop*, Athens, Greece, 2014.

[28] Nokia Siemens Networks, "2020: Beyond 4G Radio Evolution for the Gigabit Experience," White Paper2011.

[29] METIS. (2014 [Accessed: November 2015]). *Project Overview Presentation: Mobile and wireless communications Enablers for the 2020 Information Society* [Online]. Available: www.metis2020.com

[30] Press Trust of India. (2014 [Accessed: November 2014]). *ZTE expects to introduce 5G network gear prototype in 2015, Press release* [Online]. Available: http://www.business-standard.com

[31] Samsung. (2014 [Accessed: November 2014]). *Samsung Announces World's First 5G mmWave Mobile Technology, Press Release* [Online]. Available: http://global.samsungtomorrow.com

[32] METIS, "D1.2 Initial channel models based on measurement [Online] Available at: https://www.metis2020.com/documents/deliverables/," Public Deliverable2014 [Accessed: Feb 2015].

[33] Artemis. (2014 [Accessed: November 2014]). *What is pCell? Press release* [Online]. Available: http://www.artemis.com/pcell

[34] S. Perlman and A. Forenza, "Distributed-Input-Distributed-Output (DIDO) Wireless Technology A New Approach to Multiuser Wireless," White paper2011.

[35] SKTelecom. (2014 [Accessed: December 2014]). *SK Telecom Starts Joint Research with Ericsson for 5G. Press release* [Online]. Available: http://www.sktelecom.com

[36] DOCOMO, "5G Radio Access: Requirements, Concept and Technologies ", White Paper2014.

[37] M. Campolargo, "5G, the way forward!," presented at the ETSI Summit on Future Mobile and Standards for 5G, 2014.

[38] K. Watanabe and M. Machida, "Outdoor LTE Infrastructure Equipment (eNodeB)," *Fujitsu Sci. Tach,* vol. 48, pp. 27-32, 2012.

[39] W. Jian, W. Yiqun, Z. Sheng, and N. Zhisheng, "Traffic-aware power adaptation and base station sleep control for energy-delay tradeoffs in green cellular networks," in *Global Communications Conference (GLOBECOM), IEEE, Anaheim, CA, 3-7 Dec. 2012*, 2012, pp. 3171-3176.

[40] V. Suryaprakash, A. Fehske, A. Fonseca dos Santos, and G. P. Fettweis, "On the Impact of Sleep Modes and BW Variation on the Energy Consumption of Radio Access Networks," in *Vehicular Technology Conference (VTC Spring), 2012 IEEE 75th*, Yokohama, Japan, 6-9 May 2012, pp. 1-5.

[41] S. E. Elayoubi, L. Saker, and T. Chahed, "Optimal control for base station sleep mode in energy efficient radio access networks," in *INFOCOM, 2011 Proceedings IEEE*, 2011, pp. 106-110.

[42] P. Dini, M. Miozzo, N. Bui, and N. Baldo, "A Model to Analyze the Energy Savings of Base Station Sleep Mode in LTE HetNets," in *Green Computing and Communications (GreenCom), 2013 IEEE and Internet of Things (iThings/CPSCom), IEEE International Conference on and IEEE Cyber, Physical and Social Computing*, Beijing, China, 2013, pp. 1375-1380.

[43] Y. Guanding, C. Qimei, and Y. Rui, "Dual-threshold sleep mode control scheme for small cells," *Communications, IET,* vol. 8, pp. 2008-2016, 2014.

[44] Y. S. Soh, T. Q. Quek, and M. Kountouris, "Dynamic sleep mode strategies in energy efficient cellular networks," in *Communications (ICC), 2013 IEEE International Conference on*, Budapest, Hungary, 2013, pp. 3131-3136.

[45] C. Turyagyenda, K. Al-Begain, and N. Albeiruti, "A Novel Sleep Mode Operation for Energy Efficient LTE Cellular Networks: A Sum Product Algorithm Implementation," in *Next Generation Mobile Apps, Services and Technologies (NGMAST), 2013 Seventh International Conference on*, Prague, Czech Republic, 25-27 September 2013, pp. 159-164.

[46] T. Chih-Hsuan, W. Chi-En, L. Chin-Wei, and L. Chen-Yin, "Network energy efficiency for deployment architectures with base station site model," in *Communications in China Workshops (ICCC), 2012 1st IEEE International Conference on*, Beijing, China, 15-17 August 2012, pp. 85-90.

[47] L. Gilsoo, K. Hongseok, K. Young-Tae, and K. Byoung-Hoon, "Delaunay Triangulation Based Green Base Station Operation for Self Organizing Network," in *Green Computing and Communications (GreenCom), 2013 IEEE and Internet of Things (iThings/CPSCom), IEEE International Conference on and IEEE Cyber, Physical and Social Computing*, 2013, pp. 1-6.

[48] L. Zifan, Y. Peng, L. Wenjing, and Q. Xuesong, "Self-organizing Energy-Saving mechanism with base stations cooperation for heterogeneous cellular networks," in *Network Operations and Management Symposium (APNOMS), 2013 15th Asia-Pacific*, 2013, pp. 1-3.

[49] G. Jie, Z. Sheng, N. Zhisheng, and Y. Peng, "Traffic-aware base station sleeping in dense cellular networks," in *Quality of Service (IWQoS), 2010 18th International Workshop on*, 2010, pp. 1-2.

[50] Z. Qiyang and D. Grace, "Transfer learning for QoS aware topology management in energy efficient 5G cognitive radio networks," in *5G for Ubiquitous Connectivity (5GU), 2014 1st International Conference on*, Levi, Finland, 26 – 28 November 2014, pp. 152-157.

[51] S. Morosi, P. Piunti, and E. Del Re, "Improving cellular network energy efficiency by joint management of sleep mode and transmission power," in *Digital Communications - Green ICT (TIWDC), 2013 24th Tyrrhenian International Workshop on*, 2013, pp. 1-6.

[52] G. Cili, H. Yanikomeroglu, and F. R. Yu, "Cell switch off technique combined with coordinated multi-point (CoMP) transmission for energy efficiency in beyond-LTE cellular networks," in *Communications (ICC), 2012 IEEE International Conference on*, 2012, pp. 5931-5935.

[53] H. Yunbo, 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*, 2012, pp. 331-335.

[54] L. Studer Ferreira and D. H. Pichon, A. Gomes, A. Dimitrova, D. Braun, T. Karagiannis, G. Karimzadeh, M. Branco, M. Correia, L. M., "An architecture to offer cloud-based radio access network as a service," in *Networks and Communications (EuCNC), 2014 European Conference on*, 2014, pp. 1-5.

[55] China Mobile Research Institute, "C-RAN: The Road Towards Green RAN," White paper2013.

[56] H. Klessig, A. Fehske, G. Fettweis, and J. Voigt, "Cell Load-Aware Energy Saving Management in Self-Organizing Networks," in *Vehicular Technology Conference (VTC Fall), 2013 IEEE 78th*, 2013, pp. 1-6.

[57] G. Secinti and B. Canberk, "A spatial optimization based adaptive coverage model for green self-organizing networks," in *Consumer Communications and Networking Conference (CCNC), 2014 IEEE 11th*, 2014, pp. 495-500.

[58] G. P. Fettweis, "The Tactile Internet – Driving 5G," in *ETSI Future Mobile Summit*, 21 November 2013.

[59] Z. Li, D. Grace, and P. Mitchell, "Traffic perception based topology management for 5G green ultra-small cell networks," in *Cognitive Cellular Systems (CCS), 2014 1st International Workshop on*, 2014, pp. 1-5.

[60] M. F. Hossain, K. S. Munasinghe, and A. Jamalipour, "Toward self-organizing sectorization of LTE eNBs for energy efficient network operation under QoS constraints," in *Wireless Communications and Networking Conference (WCNC), 2013 IEEE*, 2013, pp. 1279-1284.

[61] K. Hiltunen, "Utilizing eNodeB sleep mode to improve the energy-efficiency of dense LTE networks," in *Personal Indoor and Mobile Radio Communications (PIMRC), 2013 IEEE 24th International Symposium on*, 2013, pp. 3249-3253.

[62] Z. Roth and M. C. Goldhamer, N. Burr, Alister Dohler, M. Bartzoudis, N. Walker, C. Leibe, Y. Oestges, C. Brzozowy, M. Bucaille, I., "Vision and architecture supporting wireless GBit/sec/km² capacity density deployments," in *Future Network and Mobile Summit, 2010*, Florence, Italy, 2010, pp. 1-7.

[63] S. Rehan and D. Grace, "Combined green resource and topology management for beyond next generation mobile broadband systems," in *Computing, Networking and Communications (ICNC), 2013 International Conference on*, 2013, pp. 242-246.

[64] S. Rehan and D. Grace, "Macro-cell overlaid green topology management for beyond next generation mobile broadband systems," in *Future Network and Mobile Summit (FutureNetworkSummit), 2013*, 2013, pp. 1-9.

[65] S. Rehan and D. Grace, "Energy-aware topology management for high capacity density temporary event networks," in *Advanced Technologies for Communications (ATC), 2013 International Conference on*, 2013, pp. 307-311.

[66] N. Morozs, D. Grace, and T. Clarke, "Case-based reinforcement learning for cognitive spectrum assignment in cellular networks with dynamic topologies," in

Military Communications and Information Systems Conference (MCC), 2013, 2013, pp. 1-6.

[67] Z. Qiyang, J. Tao, N. Morozs, D. Grace, and T. Clarke, "Transfer Learning: A Paradigm for Dynamic Spectrum and Topology Management in Flexible Architectures," in *Vehicular Technology Conference (VTC Fall), 2013 IEEE 78th*, 2013, pp. 1-5.

[68] T. Warabino, S. Kaneko, S. Nanba, and Y. Kishi, "Advanced load balancing in LTE/LTE-A cellular network," in *Personal Indoor and Mobile Radio Communications (PIMRC), 2012 IEEE 23rd International Symposium on*, 2012, pp. 530-535.

[69] L. Fei, P. Mahonen, and M. Petrova, "A handover scheme towards downlink traffic load balance in heterogeneous cellular networks," in *Communications (ICC), 2014 IEEE International Conference on*, 2014, pp. 4875-4880.

[70] A. De Domenico, V. Savin, and D. Ktenas, "A backhaul-aware cell selection algorithm for heterogeneous cellular networks," in *Personal Indoor and Mobile Radio Communications (PIMRC), 2013 IEEE 24th International Symposium on*, 2013, pp. 1688-1693.

[71] A. J. Fehske, H. Klessig, J. Voigt, and G. P. Fettweis, "Concurrent Load-Aware Adjustment of User Association and Antenna Tilts in Self-Organizing Radio Networks," *Vehicular Technology, IEEE Transactions on,* vol. 62, pp. 1974-1988, 2013.

[72] A. Lobinger, S. Stefanski, T. Jansen, and I. Balan, "Load Balancing in Downlink LTE Self-Optimizing Networks," in *Vehicular Technology Conference (VTC 2010-Spring), 2010 IEEE 71st*, Taipei, 16-19 May 2010, pp. 1-5.

[73] T. Kürner, M. Amirijoo, I. Balan, H. Vandenberg, A. Eisenblätter, T. Jansen, L. Jorguseski. R. Litjens, O. Linnell, A. L. Binger, M. Neuland, F. Phillipson, L. C. Schmelz, B. Sas, N. Scully, K. Spaey, S. Ski, J. Turk, U. Türke, K. Zetterberg, "Deliverable D5.9: Final Report on Self-Organisation and its Implications in Wireless Access Networks ", SOCRATES FP7 Project, Deliverable2010.

[74] O. Blume, H. Eckhardt, S. Klein, E. Kuehn, and W. M. Wajda, "Energy Savings in Mobile Networks Based on Adaptation to Traffic Statistics," Published online in Wiley Online Library (wileyonlinelibrary.com)2010.

[75] A. Jalali, "On cell breathing in CDMA networks," in *Communications, 1998. ICC 98. Conference Record. 1998 IEEE International Conference on*, 1998, pp. 985-988 vol.2.

[76] M. F. Hossain, K. S. Munasinghe, and A. Jamalipour, "Distributed Inter-BS Cooperation Aided Energy Efficient Load Balancing for Cellular Networks," *Wireless Communications, IEEE Transactions on,* vol. 12, pp. 5929-5939, 2013.

[77] M. F. Hossain, K. S. Munasinghe, and A. Jamalipour, "A self-organizing cooperative heterogeneous cellular access network for energy conservation," in *Communications (ICC), 2012 IEEE International Conference on*, 2012, pp. 5316-5320.

[78] M. F. Hossain, K. S. Munasinghe, and A. Jamalipour, "On the energy efficiency of self-organizing LTE cellular access networks," in *Global Communications Conference (GLOBECOM), 2012 IEEE*, 2012, pp. 5314-5319.

[79] L. Suarez, L. Nuaymi, and J. M. Bonnin, "Analysis of the overall energy savings achieved by green cell-breathing mechanisms," in *Sustainable Internet and ICT for Sustainability (SustainIT), 2012*, 2012, pp. 1-6.

[80] L. Suarez, L. Nuaymi, and J. M. Bonnin, "Analysis of a green-cell breathing technique in a hybrid access network environment," in *Wireless Days (WD), 2013 IFIP*, 2013, pp. 1-6.

[81] Z. Yutao, K. Tian, Z. Tiankui, and Z. Zhimin, "QoS-aware user association based on cell zooming for energy efficiency in cellular networks," in *Personal, Indoor and Mobile Radio Communications (PIMRC Workshops), 2013 IEEE 24th International Symposium on*, 2013, pp. 6-10.

[82] B. Rengarajan, G. Rizzo, and M. A. Marsan, "Bounds on QoS-constrained energy savings in cellular access networks with sleep modes," in *Teletraffic Congress (ITC), 2011 23rd International*, 2011, pp. 47-54.

[83] X. Lin, F. Pantisano, R. Verdone, G. Xiaohu, and C. Min, "Adaptive traffic load-balancing for green cellular networks," in *Personal Indoor and Mobile Radio Communications (PIMRC), 2011 IEEE 22nd International Symposium on*, 2011, pp. 41-45.

[84] H. Chan-Ching, J. M. Chang, C. Zi-Tsan, and Z. Abichar, "Optimizing Spectrum-Energy Efficiency in Downlink Cellular Networks," *Mobile Computing, IEEE Transactions on,* vol. 13, pp. 2100-2112, 2014.

[85] Z. Yiran, C. Yinghui, L. Wenjing, and Y. Peng, "A Novel Energy-Saving Cell Selection Mechanism for Cellular Access Networks," in *Wireless Communications, Networking and Mobile Computing (WiCOM), 2011 7th International Conference on*, 2011, pp. 1-4.

[86] Z. Juejia, L. Mingju, L. Liu, S. Xiaoming, and C. Lan, "Energy Source Aware Target Cell Selection and Coverage Optimization for Power Saving in Cellular Networks," in *Green Computing and Communications (GreenCom), 2010 IEEE/ACM Int'l Conference on & Int'l Conference on Cyber, Physical and Social Computing (CPSCom)*, 2010, pp. 1-8.

[87] H. Tao and N. Ansari, "ICE: Intelligent Cell BrEathing to Optimize the Utilization of Green Energy," *Communications Letters, IEEE,* vol. 16, pp. 866-869, 2012.

[88] M. Chao, L. Xin, L. Xiaowen, L. Tian, J. Yanxiang, and H. Wei, "A low complex energy saving access algorithm based on base station sleep mode," in *Communications in China (ICCC), 2013 IEEE/CIC International Conference on*, 2013, pp. 491-495.

[89] A. Fisusi, D. Grace, and P. Mitchell, "Energy efficient cluster-based resource allocation and topology management for beyond next generation mobile broadband networks," in *Communications Workshops (ICC), 2013 IEEE International Conference on*, 2013, pp. 576-580.

[90] A. Fisusi, D. Grace, and P. Mitchell, "Interference aware, energy efficient resource allocation for beyond next generation mobile networks," in *Personal Indoor and Mobile Radio Communications (PIMRC), 2013 IEEE 24th International Symposium on*, 2013, pp. 2197-2202.

[91] W. Qingmin, F. R. Yu, and S. Yi, "Optimal joint base station and user equipment (BS-UE) admission control for energy-efficient green wireless cellular networks," in *Global Communications Conference (GLOBECOM), 2012 IEEE*, 2012, pp. 2119-2124.

[92] C. Bottai, C. Cicconetti, A. Morelli, M. Rosellini, and C. Vitale, "Energy-efficient user association in extremely dense small cell networks," in *Networks and Communications (EuCNC), 2014 European Conference on*, 2014, pp. 1-5.

[93] R. B. Cooper, *Itroduction to Queueing Theory*: Elsevier North Holland, Inc, 1981.

[94] ABSOLUTE. (2014 [Accessed: 26 Jun 2014]). *ABSOLUTE FP7* [Online]. Available: http://www.absolute-project.eu/

[95] Google. (2014 [Accessed: Jun 2014]). *Google Loon* [Online]. Available: http://www.google.com/loon/

[96] ABSOLUTE, "Aerial Platforms Study [Online] Available at http://www.absolute-project.eu," ABSOLUTE, Deliverable 2013 [Accessed: Feb 2015].

[97] A. Mohammed, A. Mehmood, F. N. Pavlidou, and M. Mohorcic, "The Role of High-Altitude Platforms (HAPs) in the Global Wireless Connectivity," *Proceedings of the IEEE,* vol. 99, pp. 1939-1953, 2011.

[98] Transparency Market Research, "High Altitude Platforms (HAPs) Technologies Market - Global Industry Analysis, Size, Share, Growth, Trends and Forecast (2012 - 2018)," ed: Transparency Market Research.

[99] H. Holtkamp, G. Auer, and H. Haas, "On Minimizing Base Station Power Consumption," in *Vehicular Technology Conference (VTC Fall), 2011 IEEE*, 2011, pp. 1-5.

[100] G. Auer, V. Giannini, C. Desset, I. Godor, P. Skillermark, M. Olsson, M. A. Imran, D. Sabella, M. J. Gonzalez, O. Blume, A. Fehske,, "How much energy is needed to run a wireless network?," *Wireless Communications, IEEE,* vol. 18, pp. 40-49, 2011.

[101] G. Auer, O. B . Giannini, I. Godor, M. A. Imran, Y. Jading, E. Katranaras, M. Olsson, D. Sabella, P. Skillermark, W. Wajda, "Deliverable D2.3: Energy efficiency analysis of the reference systems, areas of improvements and target breakdown," EARTH FP7 Project, Deliverable2010.

[102] C. Desset, B. Debaillie, V. Giannini, A. Fehske, G. Auer, H. Holtkamp, H. Wajda, W. Sabella, D. Richter, F. Gonzalez, M. J. Klessig, H. Godor, I. Olsson, M. Imran, M. A. Ambrosy, O. Blume, "Flexible power modeling of LTE base stations," in *Wireless Communications and Networking Conference (WCNC), 2012 IEEE*, Paris, France, 2012, pp. 2858-2862.

[103] M. Goldhamer, Z. Gross, N. Haklai, N. Lavi, O. Marinchenco, P. Michalski, A. Pascual-Iserte, M. Payaró, M. Dohler, B. Phillips, M. Ware, P. Hemphill, C. Walker, A. Burr, D. Grace, T. Jiang, I. Bucaille, J. Bazin, Y. Leibe, L. Shousterman, "D1.2 Baseline BuNGee Architecture [Online] Available at: http://www.ict-bungee.eu/page.asp?pID=63," FP7 BuNGee project, Deliverable 2010 [Accessed: March 20113].

[104] L. Huiguang, J. I. Payne, and H. S. Kim, "Femto-Cells: Problem or Solution? A Network Cost Analysis," in *Global Telecommunications Conference (GLOBECOM 2011), 2011 IEEE*, 2011, pp. 1-6.

[105] A. M. Damnjanovic, J. Yongbin, Wei Tingfang, Ji Tao, Luo Vajapeyam, M. Taesang, Yoo Osok, Song Malladi, D. ,, "A survey on 3GPP heterogeneous networks," *Wireless Communications, IEEE,* vol. 18, pp. 10-21, 2011.

[106] N. Khan, C. Oestges, F. Mani, and O. Renaudin, "BuNGee Deliverable: D2.1 Final BuNGee Channel Models," 2011.

[107] Y. L. A. Kumar, T. Singh, and S. S. Khurmi, "Sustainable Energy Optimization Techniques in Wireless Mobile Communication Networks," presented at the The First International Conference on Interdisciplinary Research and Development, Thailand, June 2011.

[108] T. Jiang, A. Papadogiannis, D. Grace, and A. Burr, "BuNGee Deliverable: D4.1.1 Interim Simulation," Deliverable2011.

[109] P. Kyösti, J. Meinilä, L. Hentilä, X. Zhao, T. Jämsä, C. Schneider, M. Narandzić, M. Milojević, A. Hong, J. Ylitalo, V. Holappa, M. Alatossava, R. Bultitude, Y. de Jong, T. Rautiainen, "D1.1.2: WINNER II Channel Models Part I Channel Models," IST-4-027756 WINNER II2007.

[110] Q. Zhao, D. Grace, S. Rehan, A. Munari, R. Hermenier, K. Sithamparanatha, S. Chandrasekharan, L. Reynaud, A. Vilhar, "D4.1.1: Detailed Network and Protocol Architecture [Online] Available at: http://www.absolute-project.eu/," ABSOLUTE FP7 Project, Deliverable2013 [Accessed: Jan 2015].

[111] O. Marinchenco, N. Chayat, A. Burr, A. Papadogiannis, M. Dohler, M. Payaró, M. Ware, "BuNGee Project Overview," *in Proceedings of the 3rd International IEEE Conference on Microwaves, Communications, Antennas and Electronic Systems (IEEE COMCAS 2011),* November 2011.
[112] I. Bucaille, "ABSOLUTE: Aerial Base Stations with Opportunistic Links fir Unexpected and Temporary Events [Online] Available at: http://www.absolute-project.eu/." *General Presentation,* 2013, [Accessed: Jun 2013].
[113] 3GPP, "Lte; Evolved Universal Terrestial Radio Access Network; X2 application Protocol (TS 36.423)," 11.8.0 ed, 2014.
[114] A. J. Goldsmith and S. B. Wicker, "Design challenges for energy-constrained ad hoc wireless networks," *Wireless Communications, IEEE,* vol. 9, pp. 8-27, 2002.
[115] W. Wang, A. Kwasinski, D. Niyato, and Z. Han, "A survey on applications of model-free strategy learning in cognitive wireless networks," *CoRR, abs/1504.03976,* 2015.
[116] L.S. Kaelbling, M.L. Littman, and A.W. Moore, "Reinforcement learning: a survey," *Journal of Artificial Intelligence Research,* vol. 4, pp. 237–285, 1996.
[117] C. Claus and C. Boutilier, "The dynamics of reinforcement learning in cooperative multiagent systems," in *Proceedings of the fifteenth national/tenth conference on Artificial intelligence/Innovative applications of artificial intelligence*, 1998.
[118] C. J. Watkin and P. Dayan, "Machine Learning, technical note; Q-learning," ed. Boston: Kluwer Academic Publishers, 1992.
[119] R. S. Sutton, Barto, A.G, "Reinforcement learning: an introduction," *The MIT Press,* 1998.
[120] 3GPP, "LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) procedures in idle mode (3GPP TS 36.304 version 11.6.0 Release 11)," 2014.
[121] O. Jinyoung and H. Youngnam, "Cell selection for range expansion with almost blank subframe in heterogeneous networks," in *Personal Indoor and Mobile Radio Communications (PIMRC), 2012 IEEE 23rd International Symposium on*, 2012, pp. 653-657.
[122] C. J. Bernardos, A. De Domenico, J. Ortin, P. Rost, and D. Wubben, "Challenges of designing jointly the backhaul and radio access network in a cloud-based mobile network," in *Future Network and Mobile Summit (FutureNetworkSummit), 2013*, 2013, pp. 1-10.
[123] H. Holtkamp, G. Auer, S. Bazzi, and H. Haas, "Minimizing Base Station Power Consumption," *Selected Areas in Communications, IEEE Journal on,* vol. 32, pp. 297-306, 2014.
[124] D. J. Ferling, P. Xin, Yu Bohn, T. Ambrosy, A. Fernandez, Y. Erdem, A. Pascht, A., "Power saving by sleep modes in base station transceivers for LTE," in *Microwave Conference Proceedings (APMC), 2012 Asia-Pacific*, Kaohsiung, Taiwan, 04 Dec - 07 Dec 2012, pp. 947-949.
[125] 3GPP, "LTE; Evolved Universal Terrestrial Radio Access Network (E-UTRAN); X2 Application Protocol (X2AP), TS 36.423 R 11," vol. 11, ed: 3GPP, 2013-2014.
[126] 3GPP, "Universal Mobile Telecommunications System (UMTS); Radio Resource Control (RRC); Protocol specification (3GPP TS 25.331 version 12.4.0 Release 12)," 2015.
[127] Nokia, "Setting up HSPA+ heterogeneous networks for the best customer experience," White paper2014.
[128] GSM Association, "Advancing 3GPP Networks: Optimisation and Overload Management Techniques to Support Smart Phones," White Paper2012.

[129] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA); Further advancements for E-UTRA physical layer aspects," 3GPP, technical Report210-2013.

[130] K. Gomez, T. Rasheed, R. Hermenier, T. Jiang , Q. Zhao , S. Rehan , D. Grace, L. Reynaud, T. Javornik, I. Ozimek, L. L. Garrec, "D2.6.1: System-wide Simulations Planning Document," ABSOLUTE FP7 Project, Deliverable 2013.