2014

Resource optimization, spectrum allocation and fault tolerance planning in broadband wireless networks

Tamer R. Omar
Iowa State University

Resource optimization, spectrum allocation and fault tolerance planning in broadband wireless networks

by

Tamer R. Omar

A thesis submitted to the graduate faculty

in fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Electrical and Computer Engineering

Program of Study Committee:

J. Morris Chang, Co-major Professor

Ahmed E. Kamal, Co-major Professor

Sang W. Kim

Zhao Zhang

Lu Ruan

Iowa State University

Ames, Iowa

2014

DEDICATION

I would like to dedicate this thesis to the soul of my father, to my lovely mother for her infinite support, to my beloved wife whom without her support I would not have been able to complete this work, my brother for always watching my back and to my kids for the joy they bring to my life. I would also like to thank my family and friends for their continuous guidance and assistance during my Ph.D. work.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

In current (4G) and future (5G) broadband cellular networks, new cell coverage planning ideas, network architectures proposals, novel physical resources allocation optimization techniques, and dynamic spectrum allocation optimization frame works provide good opportunities for mobile service providers (MSPs) to improve their return on investments (ROI), and for mobile equipments manufacturers to increase their profit and market share. Despite the attractive opportunities that network architecture, cell planning and resources allocation optimization offers, there are many challenges and difficulties that are facing MSPs when planning and operating networks to cope with the tremendous increase in mobile applications and to satisfy different users requirements. Physical resources allocation, spectrum allocation optimization, network architecture enhancement, and fault tolerance cell planning are major issues in broadband cellular networks.

The work accomplished in this thesis aims at enhancing the network performance by optimizing the planning and operations of the network. Different optimization techniques are used throughout this thesis to help increase the spectral and energy efficiency in 4G and 5G networks. The objectives of this study are four objectives, first to propose a physical resources allocation utility based frame work using a novel utility function that can jointly optimize the maximum normalized spectral efficiency (NSE) and power consumed locally in each cell in order to increase the mobile service providers ROI. The ROI is enhanced by increasing the profits through maximizing the network spectral efficiency and decreasing the operational costs by minimizing the power consumption in the network.

The second objective is to determine the optimal down-link frequency partition configuration that can efficiently allocate the spectrum resources to different network frequency partitions in order to globally achieve the same joint optimization objective

by addressing the DFPCs dynamic behavior according to the network topology, load conditions, and users distribution.

The third objective is to propose a new network architecture that consists of a data collection system that aid as a traffic data repository and a decision support system (DSS) introduced as a new self optimization module within the self organized networks (SON) framework to automate the optimization of the dynamic spectrum allocation.

The last objective is to perform a network planning that aims at placing the optimal number of relay stations that aid in achieving network full coverage and minimum rate requirements with a fault tolerance functionality to avoid network failures and using the self organized frame work to perform the self healing by managing the backup solutions needed in response to the network failures.

In order to achieve the previously mentioned objectives a detailed study to the state of the art in network planning using relay stations, physical resource allocation, dynamic spectrum allocation, network architecture and SON frame work is conducted. Different methodologies such as integer linear programming, stochastic programming and non-parametric estimation analysis are presented to propose a novel physical resources and dynamic spectrum allocation schemes. A plan-do-control-act model is also proposed within the DSS in the new suggested network architecture for continuous improvement of spectrum allocation. A non-linear to linear formulation conversion using an expanded space state is utilized to perform an in-band fault tolerance network planning that consider network interference between relay stations and base stations and avoid network failures. Simulations and results are conducted to validate the proposed methodologies and to compare it against state of the art work.

CHAPTER 1. GENERAL INTRODUCTION

1.1 Introduction

The joint optimal spectral efficiency and power allocation problem in 4G networks is gaining more importance due to the scarcity of the available spectrum and the continuous increase in power costs. A novel utility based frame work is proposed, this includes the formulation of a utility function that is utilized in the optimization problem. The utility can be used by the mobile service providers as one of the system key performance indicator (KPI) to enhance the spectrum and power utilization efficiency in the cellular networks.

The problem is developed in an integer linear program (ILP) with all applicable power and integrality constraints to produce the physical resources (PRs) allocation matrix. The allocation matrix states the assignment of the PRs to the mobile stations or user equipments that achieves the problem objective. The problem objective is the joint minimization of the power consumption to decrease the operational cost and maximization of the network spectral efficiency to increase revenue. The QoS and delay requirements are considered in order to maintain the users satisfaction.

The joint power consumption minimization and spectral efficiency maximization through optimal PRs allocation in the multi-cell broadband cellular systems is ad-

dressed in order to support the tremendous need for more resources that is necessary for today's growing applications. PRs allocation reduces the complexity of the optimization problem by decreasing the number of variables used in the problem in comparison to previous allocation of sub-carriers, this results in faster allocation schemes with minimal waste in bandwidth.

Central (by an RRM higher entity) and distributed (by base station) allocation schemes that can be used to allocate radio resources with the objective to increase the spectral efficiency and minimize the power consumption in the network are proposed for WiMAX networks. A primary central allocation control mechanism is offered to adjust the spectral efficiency by indicating different adaptive fractional frequency reuse (AFFR) mechanism called the down-link frequency partition configurations (DFPC) that can periodically change to adjust the AFFR configuration in the network. This configurations change on semi-static basis according to the base station load conditions and the mobile stations rate requirements. A distributed radio resource management scheme is proposed for PRs allocation. The base stations or macro cells in the network periodically allocate PRs to the mobile stations or user equipments according to an allocation matrix. The allocation matrix is calculated by solving the integer linear programming problem (ILP) to find the optimal utility for PRs allocation in the network. The ILP optimal solution is of high computational complexity that make it impractical for implementation. For practical purposes a suboptimal algorithm using sub-gradient optimization techniques is presented.

A new network architecture for the next generation (5G) of wireless communication systems is proposed to automate the PRs and spectrum allocation. The architecture aims at building a self organizing networks module in the core network. The module will work as decision support system aided by a data collection infrastructure. The infrastructure include a Geo-location resources allocation database in the network control and management system of the access network. The new network

architecture using the DSS is expected to manage the spectrum utilization in order to forecast the best down-link frequency partition configuration to be utilized in the network.

The problem of placement relay stations in heterogeneous network such that any relay station failure will not cause an interruption in the service is addressed to design a fault-tolerant network. This problem is an important design and planning problem that has not been addressed in the literature. The problem of RS locations planning with fault tolerance is investigated using both out-band and in-band modes and an interference model is presented to consider the in-band mode and to address the effect of interference on RSs placement planning.

1.2 Thesis Organization

The thesis consist of six chapters that are arranged in a paper based format and organized as follows:

1.2.1 Chapter 2: Joint spectrum and power optimization in 4G networks

This paper is to be submitted to the Springer© journal of wireless networks. The authors of this paper are Tamer R. Omar and Dr. J. Morris Chang. Tamer is the primary author of the paper that conducts the research, performs the analysis, simulations and the paper writing under the supervision and advice of the Co-Author Dr. Chang.

1.2.2 Chapter 3: Joint spectral efficiency and power allocation optimization in IEEE 802.16m

This paper is published in the IEEE Transactions of Mobile Computing© journal. The authors of this paper are Tamer R. Omar and Dr. J. Morris Chang. Tamer is the primary author of the paper that conducts the research, performs the analysis, simulations and the paper writing under the supervision and advice of the Co-Author Dr. Chang.

1.2.3 Chapter 4: Down-link spectrum allocation in 5G HETNETs

This paper is published in the proceedings of the 2014 International Wireless Communications and Mobile Computing Conference. The authors of this paper are Tamer R. Omar, Dr. Ahmed E. Kamal and Dr. J. Morris Chang. Tamer is the primary author of the paper that conducts the research, performs the analysis, simulations and the paper writing under the supervision and advice of the Co-Author Dr. Kamal and the assistance of the the Co-Author Dr. Chang.

1.2.4 Chapter 5: Fault-tolerant planning of relay station locations in WiMAX networks

This paper is ready to be submitted to the IEEE Transactions of Mobile Computing© journal. The authors of this paper are Tamer R. Omar the primary author of the paper, Dr. Zakhia Abichar the researcher that conduct the out-band model research and analysis, Dr. Ahmed E. Kamal and Dr. J. Morris Chang. Tamer's role in this paper is investigating the in-band relaying model, addressing the interference effect on the network, formulating the new in-band constraints, performing the simulations and writing a new version of the paper to extend Dr. Abichar previous work under

the supervision and advice of the Co-Author Dr. kamal and the assistance of the Co-Author Dr. Chang.

1.2.5 Chapter 6: General conclusions

This chapter concludes the thesis contributions and discusses of the results achieved from the different research topics conducted in this thesis. In this chapter a section is devoted to the suggested recommendations for future research.

1.3 Literature Review

1.3.1 Physical resources allocation

Scarce bandwidth and limited power resources require intelligent resource allocation strategies for broadband wireless networks. Orthogonal frequency division multiple access (OFDMA) is a promising multiple access technique for overcoming these challenges. This technique benefits from the efficiency of orthogonal frequency division multiplexing (OFDM) that mitigates frequency-dependent distortion across the channel band and simplifies the equalization in a multi-path fading environment [1].

Optimization of sub-carriers assignment and power allocation problem for downlink single or multi-cell broadband cellular systems has been addressed before using different approaches. The first approach aims to maximize the network throughput with power constraint. This approach is referred to as rate adaptive (RA) resources allocation [2, 3, 4, 5]. The second approach aims to minimize the total power consumption in the network with rate constraint. This approach is referred to as margin adaptive (MA) resources allocation [6, 7, 8, 9]. RA and MA resources allocation schemes address a simple objective that focuses on enhancing the efficiency by either increasing the throughput or decreasing the amount of consumed power in the system.

The third approach (the focus of this paper) that aims to optimize an objective modeled as a utility function. The utility function is formulated according to the mobile service provider goals. This utility function can be exploited by service providers as one of the Key Performance Indicators (KPIs) monitoring the overall system performance. This approach is referred to as utility adaptive (UA) resources allocation [10, 11, 12].

1.3.2 Dynamic spectrum allocation

RRM include many functions such as frequency partitioning, multi-connection assignment and scheduling resource units. Frequency partitions (FPs) in WiMAX divide the cell area into two regions (cell center and cell edge), each of them utilize a part of the frequency spectrum. There are two models for partitioning the cell; the grid model in which constant signal to interference plus noise ratio (SINR) contours can be defined as concentric circles around the central base station with mobile stations in grid interior radius considered as center area mobile stations and mobile stations in grid exterior radius as edge area mobile stations, or the SINR threshold model in which the base station classifies users with average SINR less than the threshold as edge mobile station, while mobile station with average SINR greater than the threshold as interior users. The advantage of FPs is the mitigation of the interference between the edge mobile station in neighboring cells. This result in enhancing the SINR which achieves better channels quality and increase the the network throughput, however this comes on the cost of degrading the network spectral efficiency.

There are several frequency partitioning techniques such as universal reuse [13], Strict Fractional Frequency Reuse (SFFR) and Adaptive Fractional Frequency Reuse (AFFR) with selected power boosted partitions [14]. FFR concept is used to overcome the deficiency in the network spectral efficiency due to frequency partitioning. AFFR is a promising frequency reuse technique adopted by WiMAX to effectively

balance between the improvement in the spectral efficiency and mitigating the inter-cell interference (ICI) in the network. WiMAX allow flexible RRM through AFFR mechanisms.

1.3.3 Network architecture & self organized networks

A cluster of new wireless enabling technologies are lately offered to support mobile service providers (MSPs) in their mission to satisfy the requirements needed for 4G technology and are expected to act as the baseline for 5G technology. Heterogeneous network (HetNet) and small cells are discussed in [15, 16, 17, 18]. Multiple issues like dynamic spectrum allocation (DSAn), interference management techniques, design principles, integrated network architectures, cell range expansion, beam-forming to allow more small cells to share the spectrum and adaptive resource partitioning techniques to balance the load among nodes are presented and examined.

The idea of a Geo-location databases is previously used for different purposes. TV bands database is used by the federal communications commission (FCC) to protect licensed TV broadcasters from interference from unlicensed white space devices (WSDs) [19]. A protocol layer between the application and the IP layers in the radio units is proposed in [20] to aid in frequency negotiation and dynamic spectrum access (DSA). The protocol includes a client to database part maintained by the network operator to manage a central database for the registration of radio parameters.

Self organizing networks (SON) for HetNets is discussed in [21] with a main driver to reduce capital and operational expenses that increase due to the increased number of network parameters to be monitored. SON architecture alternatives is presented and distributed SON architecture is recommended to avoid scaling issues. Also other different components functions like auto configuration, self optimization, diagnosis, and healing are evaluated.

All previous enabling technologies aims at maximizing the MSPs return on investment (ROI) by maximizing the revenues and minimizing the expenditures. The objectives to maximize the ROI include decreasing the energy consumption and increasing the spectral efficiency. However, non of the previous work present a SON frame work for a decision support system (DSS) that help automate the optimization process of the PRAn and DSAn. Thus, its essential to develop a new architecture and a supporting SON module to optimally distribute and utilize the available resources and frequency spectrum in order to achieve the targeted benefit from this enabling technologies.

The DSAn is a network based central mechanism that determine the optimal amount of resources in the network, and PRAn is cell based distributed mechanism that determine the optimal resources to be scheduled for each user. This central nature grant DSAn and PRAn a special importance over all other enabling technologies that aid in enhancing the spectral efficiency.

DSAn and PRAn are discussed in [22, 23, 24] using eICIC schemes that aim to mitigate interference. An adaptive distributed resource allocation algorithm based on potential game is proposed in [22]. Nash equilibrium existence and improvement in the transmitted power, throughput and system fairness are shown. Resources scheduling using MIMO transmission strategy according to the ICI level experienced by a user is investigated in [23]. The performance of the proposed scheme shows a better performance than schemes employing a single transmission strategy. Multiple-input single-output (MISO) system that jointly design eICIC and inter-cell user scheduling is studied in [24]. Authors proposed a low complexity algorithm to select the best users and their beam-forming strategies in order to maximize the weighted sum rate. The proposed scheme combined with proportional fair scheduling provides high throughput.

1.3.4 Relay stations placement and fault tolerance

Existing 4G Broadband technologies Like 3GPP LTE-A and WiMAX are employed by mobile service providers to operate their wireless networks [25, 26]. Multi-tier cellular networks known as Heterogeneous networks (HetNets) that include both macro cells and small cells are currently considered as the most preferred deployment approach to enhance the network efficiency and decrease the deployment costs. Smalls cells are used in HetNets for either offloading data traffic from macro cells to improve hot-spots throughput in urban areas or to extend the macro cell coverage into rural areas that are out of coverage.

Small cells are characterized by their high density deployment nature due to their coverage range (100m-2km). This hight density nature increase the probability of having more faulty cells in the network. Faulty cells cause degradation in the service level afforded to users once these cells gets out of service. Fault tolerance capabilities is widely used in wireless sensor networks to guarantee service continuity. Surveys in [27, 28] discuss different types of failures to show the different detection techniques for each type of failure and how to overcome these failures.

Automation of fault tolerance detection and correction process can be handled through Self Organized Networks (SON) frame work. A widely spread self healing taxonomy comprises three main phases (detection, diagnosis and compensation) [29]. An LTE networks self healing framework is developed and evaluated in [30] to provide detection and compensation of cell outages. Early fault detection and the impact of failures timely correction are discussed to show their impact on radio links and number of connected users. A reinforcement Learning based algorithm for the self optimization and self healing of coverage holes through antenna tilting is discussed in [31]. The algorithm shows better adaptation to network environments without prior knowledge than other supervised learning and Q-learning based algorithms. However, the proposed solutions in [30, 31] are designed for networks with macro cells (eNodeB)

deployment only. The scalability of such self-healing proposals need to be reexamined for dense HetNet deployments.

1.4 References

[1] R. v. Nee and R. Prasad, *OFDM for Wireless Multimedia Communications*, 1st ed. Norwood, MA, USA: Artech House, Inc., 2000.

[2] R. Iyengar, K. Kar, B. Sikdar, and X. Luo, "QoS provisioning and radio resource allocation in OFDMA based WiMAX systems," ECSE Department, Rensselaer Polytechnic Institute, 2009.

[3] H. Wang and W. Jia, "An optimized scheduling scheme in OFDMA WiMAX networks," *International Journal of Communication Systems*, vol. 23, pp. 23–39, 2010.

[4] A. Biagioni, R. Fantacci, D. Marabissi, and D. Tarchi, "Adaptive subcarrier allocation schemes for wireless OFDMA systems in WiMAX networks," *IEEE Journal on Selected Areas in Communications*, vol. 27, pp. 217–225, 2009.

[5] C. Lengoumbi, P. Godlewski, and P. Martins, "Dynamic subcarrier reuse with rate guaranty in a downlink multicell OFDMA system," in *IEEE 17th International Symposium on Personal, Indoor and Mobile Radio Communications*, 2006.

[6] N. U. Hassan and M. Assaad, "Optimal fractional frequency reuse (FFR) and resource allocation in multiuser ofdma system," in *International Conference on Information and Communication Technologies, ICICT '09*, 2009.

[7] N. Ksairi, P. Bianchi, P. Ciblat, and W. Hachem, "Resource allocation for downlink cellular OFDMA systems — Part I : Optimal allocation," *IEEE Transactions on Signal Processing*, vol. 58, pp. 720–734, 2010.

[8] ——, "Resource allocation for downlink cellular OFDMA systems — Part II : Practical algorithms and optimal reuse factor," *IEEE Transactions on Signal Processing*, vol. 58, pp. 735–749, 2010.

[9] S. Hamouda, S. Tabbane, and P. Godlewski, "Improved reuse partitioning and power control for downlink multi-cell OFDMA systems," in *BWANï¿œ06, International Workshop on Broadband Wireless Access for ubiquitous Networking*, 2006.

[10] G. Song and Y. Li, "Cross-layer optimization for ofdm wireless networks-part i: theoretical framework," *Wireless Communications, IEEE Transactions on*, vol. 4, no. 2, pp. 614–624, 2005.

[11] ——, "Cross-layer optimization for ofdm wireless networks-part ii: algorithm development," *Wireless Communications, IEEE Transactions on*, vol. 4, no. 2, pp. 625–634, 2005.

[12] M. Katoozian, K. Navaie, and H. Yanikomeroglu, "Utility-based adaptive radio resource allocation in OFDM wireless networks with traffic prioritization," *IEEE Transactions on wireless communications*, vol. 8, pp. 66–71, 2009.

[13] K. T. Kim and S. K. Oh, "A universal frequency reuse system in a mobile cellular environment," in *Vehicular Technology Conference, 2007. VTC2007-Spring. IEEE 65th*, april 2007, pp. 2855 –2859.

[14] N. Himayat, S. Talwar, A. Rao, and R. Soni, "Interference management for 4g cellular standards [wimax/lte update]," *Communications Magazine, IEEE*, vol. 48, no. 8, pp. 86 –92, august 2010.

[15] A. Damnjanovic, J. Montojo, Y. Wei, T. Ji, T. Luo, M. Vajapeyam, T. Yoo, O. Song, and D. Malladi, "A survey on 3GPP heterogeneous networks," *IEEE Wireless Communications*, vol. 18, no. 3, pp. 10–21, 2011.

[16] J. Chen, P. Wang, and J. Zhang, "Adaptive soft frequency reuse scheme for in-building dense femtocell networks," in *IEEE International Conference on Communications in China (ICCC)*, 2012, pp. 530–534.

[17] S. Ryoo, C. Joo, and S. Bahk, "A decentralized spectrum allocation and partitioning scheme for a two-tier macro-femtocell network with downlink beamforming," *EURASIP Journal on Wireless Communications and Networking*, 2012.

[18] T. Nakamura, S. Nagata, A. Benjebbour, Y. Kishiyama, T. Hai, S. Xiaodong, Y. Ning, and L. Nan, "Trends in small cell enhancements in LTE advanced," *IEEE Communications Magazine*, vol. 51, no. 2, pp. 98–105, 2013.

[19] T. Baykas, M. Kasslin, M. Cummings, H. Kang, J. Kwak, R. Paine, A. Reznik, R. Saeed, and S. Shellhammer, "Developing a standard for TV white space coexistence: technical challenges and solution approaches," *IEEE Wireless Communications*, vol. 19, no. 1, pp. 10–22, 2012.

[20] T. Hoang, M. Skjegstad, T. Maseng, and T. UlversØy, "FRP: The frequency resource protocol," in *IEEE International Conference on Communication Systems (ICCS)*, 2010, pp. 746–750.

[21] O. Østerbø and O. Grøndalen, "Benefits of Self-Organizing Networks (SON) for Mobile Operators," *Journal of Computer Networks and Communications*, vol. 2012, no. 862527, p. 16, 2012.

[22] D. Ling, Z. Lu, Y. Ju, X. Wen, and W. Zheng, "A multi-cell adaptive resource allocation scheme based on potential game for ICIC in LTE-A," *International Journal Communication Systems*, 2013.

[23] N. Binglai, V. W. Wong, and S. Robert, "Downlink Scheduling with Transmission Strategy Selection for Multi-Cell MIMO Systems," *IEEE Transactions on Wireless Communications*, vol. 12, no. 2, pp. 736–747, 2013.

[24] S.-H. Moon, C. Lee, S.-R. Lee, and I. Lee, "Joint User Scheduling and Adaptive Intercell Interference Cancelation for MISO Downlink Cellular Systems," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 1, pp. 172–181, 2013.

[25] 3GPP, *3GPP TS 36.300 version 10.8.0 Release 10, Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2*, 2012.

[26] IEEE, *IEEE Standard for Local and metropolitan area networks Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems*, May 2009.

[27] A. Samson Arun Raj, K. Ramalakshmi, and C. Priyadharsini, "A Survey on Classification of Fault Tolerance Techniques Available in Wireless Sensor Network," vol. 3, no. 1, 2014.

[28] Q. Lu and C. Zhang, "A Survey on Classification of Fault Tolerance Techniques Available in Wireless Sensor Network," vol. 846-847, 2013, p. 463.

[29] O. Aliu, A. Imran, M. Imran, and B. Evans, "A Survey of Self Organisation in Future Cellular Networks," *Communications Surveys Tutorials, IEEE*, vol. 15, no. 1, pp. 336–361, First 2013.

[30] M. Asghar, S. Hamalainen, and T. Ristaniemi, "Self-healing framework for LTE networks," in *Computer Aided Modeling and Design of Communication Links and Networks (CAMAD), 2012 IEEE 17th International Workshop on*, Sept 2012, pp. 159–161.

[31] A. Thampi, D. Kaleshi, P. Randall, W. Featherstone, and S. Armour, "A sparse sampling algorithm for self-optimisation of coverage in LTE networks," in *Wireless Communication Systems (ISWCS), 2012 International Symposium on*, Aug 2012, pp. 909–913.

CHAPTER 2. JOINT SPECTRUM AND POWER OPTIMIZATION IN 4G NETWORKS

A paper to be submitted to *The Springer Journal of Wireless Networks*

Tamer R. Omar[1] [2]and J. Morris Chang[3]

2.1 Abstract

Satisfying the demand for higher data rate and ensuring the quality of service requirements are gaining more research interest. Investigating the problem of resource management for mobile systems is crucial in solving this issue. This paper aims at optimizing the down-link physical resources allocation in LTE-A multi cell systems. This constitutes the allocation of the physical resources to optimize the spectral and power efficiency under different down-link frequency partition configurations. The allocation objective is to increase the mobile service providers return on investment by increasing revenues and decreasing operational costs, while maintaining the minimum quality of service (QoS) requirements. A new scheme employing a utility-based

[1]Graduate student and Associate Professor, respectively, Department of Electrical and Computer Engineering, Iowa State University.

[2]Primary researcher and author.

[3]Author for correspondence.

framework is proposed to achieve the allocation objectives. The allocation problem is formulated as an integer linear program that jointly maximizes the network spectral efficiency (revenues) and minimizes the power consumption (costs), given minimum required rate, QoS, and delay constraints. The problem is mapped as a multiple choice knapsack problem and the optimal solution is calculated using Cplex©. Since the optimal solution has high computational complexity, an effective suboptimal solution using a greedy heuristic is further presented for practical implementations. Results show that the proposed utility-based methodology achieves more efficient resources allocation compared to the commonly used rate and margin adaptive resource allocation methodologies.

Keywords: 4G, LTE-A, WiMAX 2, resource allocation, utility adaptive optimization.

2.2 Introduction

Broadband wireless access based on LTE-Advanced (LTE-A) is one of the approved 4G technologies by the international telecommunication union (ITU). 3GPP LTE-Advanced [32] is showing an increasing interest as the demand for implementing new networks increases. IEEE 802.16m (WiMAX 2) is also an important amendment [33] to IEEE 802.16e [34] standard which is approved by ITU as a 4G technology. Both technologies meet the international mobile telecommunication system (IMT-Advanced) specifications and are considered by most mobile service provider (MSP) seeking to upgrade their networks to 4G.

Optimization of physical resources assignment and power allocation problem in 4G down-link single or multi-cell systems has been addressed before using different approaches. The first approach aims to maximize the network throughput with power constraint. This approach is referred to as rate adaptive (RA) resources allocation

[35, 36, 37, 38]. The second approach aims to minimize the total power consumption in the network with rate constraint. This approach is referred to as margin adaptive (MA) resources allocation [39, 40, 41, 42]. RA and MA resources allocation schemes address a simple objective that focuses on enhancing the efficiency by either increasing the throughput or decreasing the amount of consumed power in the system.

The third approach, referred to as utility adaptive (UA) resources allocation, aims at optimizing a utility function that is formulated based on MSP objectives [43, 44, 45, 46]. This utility function can be used as a system key performance indicator; the higher the utility, the better the performance. This approach will be adopted to formulate the optimization problem in this paper.

The problem is developed as an integer linear program (ILP) with all applicable power and integrality constraints to produce the resource blocks (RBs) allocation matrix. The allocation matrix states the optimal assignment of the RBs to the user equipments (UEs) in order to achieve the problem objective. The ILP is optimally solved using Cplex©, however the optimal solution shows a high computational complexity. Thus, for practical purposes, a suboptimal algorithm using sub-gradient optimization technique is presented.

A distributed radio resources management scheme for LTE-A systems is proposed. The evolved nodeB (eNB) periodically allocate RBs to the UEs according to an allocation matrix. The eNBs calculate the allocation matrices by solving the problem to perform the optimal resources allocation in each frequency partition (FP). The concept of down-link frequency partition configurations (DFPCs) proposed by WiMAX 2 is adopted in partitioning the LTE-A frequency spectrum in our system model and we recommend the implementation of this concept in LTE-A networks.

A simulation is conducted to show the achieved utility, consumed power and spectral efficiency of the proposed scheme compared to other optimization schemes. Results for the proposed scheme indicate an enhancement in the consumed power and

the spectral efficiency while maintaining minimum rate, delay, and QoS requirements. A comparison between the optimal and suboptimal solutions is performed and the gap between them is analyzed.

This paper investigates the joint optimal spectral efficiency and power allocation problem in LTE-A multi-cell systems:

1. A resources allocation utility based frame work is proposed. The frame work introduces a novel utility function used to increase the MSPs return on investment by increasing their profits through maximizing the network spectral efficiency and decreasing their operational cost by minimizing the power consumption in the network.

2. An analysis that show the complexity of both the optimal and suboptimal solutions is developed to justify the applicability of the suboptimal solution for practical implementation.

3. A comparison for both the achievable spectral efficiency and the consumed power between the proposed frame work and other well established schemes (e.g. maximizing data rate, minimizing power consumption) is conducted to show the merit of our proposal.

The paper is organized as follows. In section II, the system model is developed. Section III is devoted to the resources allocation scheme. In section IV, the utility function is developed and the problem is formulated. Section V presents the simulation, results, and the performance of the optimal and suboptimal allocation solutions. Finally, in section VI, conclusions are drawn.

2.3 System Model

eNBs are assumed to exploit an adaptive fractional frequency reuse (AFFR) technique to allow different FPs to utilize different power levels and patterns. A maximum of four FPs is allowed to be reused in each cell to help mitigate interference and maximize the network spectral efficiency.

The down-link transmissions in a multi-cell LTE-A system is modeled in this section. The network consists of the seven identical adjacent hexagonal cells denoted as b where $(b \epsilon 1, 2, ..., 7)$. Each cell in the network contains only one centric eNB. Each cell in the network consists of three sectors and each sector employs a directional antenna and only the subset of UEs in this sector are receptive at that antenna. This multi antenna technique achieves spatial orthogonality between UEs and reduces the interference experienced by each UE in the system.

Four FPs are utilized in the network and denoted as R where $(R \epsilon 1, 2, 3, 4)$. Frequency partitioning is employed by dividing the cells into two (inner $\Re_0$ and outer $\Re_1$) regions. Each region uses a different FP to achieve better spectral efficiency and reduce the inter-cell interference (ICI) as shown in the fractional frequency reuse (reuse-3) scenario presented in Fig.2.3.1.

The maximum number of RBs allocated to each FP and the set of RBs $\{1, 2, ..., N_j\}$ allocated to each UE are denoted as N and $|N_j|$ respectively, where N is determined according to the utilized DFPC in the network. The concept of DFPCs proposed by WiMAX 2 is applied in our LTE-A system model. The number of DFPCs and their specifications depend on the total bandwidth and the fast fourier transform (FFT) size. The DFPC determines the ratio between the different FPs, thus changing the FPs count ($FPCT$) and size (FPS) according to the utilized DFPC in the network. The DFPC changes on semi static basis according to the traffic load and users distribution in the network. Each RB is denoted as i where $(i \epsilon 1, 2, ..., N)$. A DFPC update interval (T) is assumed with value $(T = 1hr)$ to determine the duration by

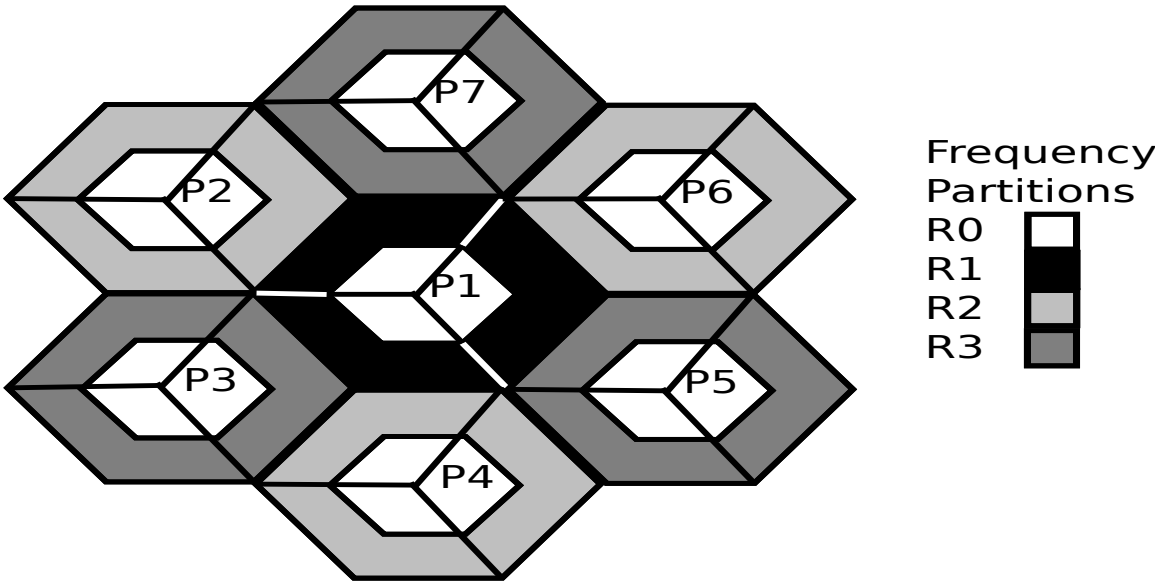

Figure 2.3.1: AFFR (reuse-3) Network Model.

which the DFPC changes. The maximum number of UEs in each cell partition is S and each UE is denoted as j where $(j \epsilon 1, 2, ..., S)$. The UEs are assumed to be uniformly distributed inside the cell, however, the distribution changes for different cells.

The system bandwidth and the bandwidth allocated to each UE are denoted as W *and* W_j respectively. The total number of utilized modulation types is M and each modulation type is denoted as k where $(k \epsilon 1, 2, ..., M)$. The type of modulation specifies the amount of data bits carried by each symbol.

An adaptive modulation and coding (AMC) technique is assumed to adjust the transmission modulation type according to the channel state conditions. When the channel utilized by a UE is in a good state, a high modulation type is utilized. However, in case of poor channel state, UEs use a low modulation type.

The maximum amount of power transmitted in cell b and in each $FP(R)$ of cell b are denoted as P_{max} and $P_{R,max}$ respectively. The amount of power consumed due to date transmission of UE j over RB i with modulation type k is denoted as P_{ijk}. A resource metric information $RM(R)$ is transmitted by each eNB for each $FP(R)$ in order to balance the load between the different FPs and is denoted as $RM(R)$.

The minimum required and achievable data rate by each UE in $FP(R)$ of cell b are denoted as $r_{min,j}$ and r_j respectively. The normalized queue delay and the maximum delay for a UE are denoted as θ_j and $\theta_{max,j}$ respectively. The QoS coefficient and the minimum QoS requirements of each UE are denoted as α_j and $\alpha_{min,j}$ respectively.

RBs are allocated in the system every Transmission Time Interval (TTI) and the assignment indicator $x_{ijk}(t)$ is the result of data transmission by UE j over RB i with modulation type k at the TTI. The data rate, bandwidth, expected spectral efficiency, normalized spectral efficiency, and achievable utility due to the allocation of RB i with modulation type K to UE j are denoted as r_{ijk}, W_{ijk}, ESE_{ijk}, NSE_{ijk} and U_{ijk} respectively.

The interference and noise experienced by UEs in the system are modeled. Small scale fading is modeled with Rayleigh distribution. Path loss is modeled as $\kappa\ d^{-\beta}$ where d is the distance between a given UE and eNB, β is the path loss exponent where $2 \leq \beta \leq 4$ and κ is a constant for a given environment. The shadowing in the system is modeled as a log-normal distribution $N\ (0,\sigma_{sh})$ with $4dB \leq \sigma_{sh} \leq 12dB$. UEs are exposed to co-channel interference that emerges from only the two neighboring sectors. However, they are isolated from the inter-cell interference because each RB is assigned exclusively to only one UE at any time.

2.4 Resources Allocation Scheme

In this section, the FPs, bandwidth, and power assignment are discussed. In addition, the spectral efficiency calculations, QoS, and delay requirements for the proposed scheme are presented.

2.4.1 Frequency partitions assignment

The DFPC is centrally controlled by a centralized network controller (CNC). The number of DFPCs and their specifications depend on the total bandwidth (W) and the FFT size utilized as shown in Table 2.4.1 [33]. DFPC determines the ratio between the different FPs, thus the $FPCT$ and the FPS change according to the utilized DFPC in the network. The eNBs must transmit the $FPCT$ to the UEs periodically in order to inform them of the frequency reuse factor utilized in the network. The different DFPCs stated in Table 2.4.1 are denoted as C_c and $N = 96$ is assumed, also the ratio between the different FPs bandwidth $\mathcal{W}_R$ is presented.

Table 2.4.1: Down-link frequency partition configurations (FFT = 2048, W = 20 MHz).

C_c	R_0:R_1:R_2:R_3	FPCT	FPS(No. of RBs)	$\mathcal{W}_0$:$\mathcal{W}_1$:$\mathcal{W}_2$:$\mathcal{W}_3$
C_1	1:0:0:0	1	96:0:0:0	20:0:0:0
C_2	0:1:1:1	3	0:32:32:32	0:6.67:6.67:6.67
C_3	1:1:1:1	4	24:24:24:24	5:5:5:5
C_4	3:1:1:1		48:16:16:16	10:3.33:3.33:3.33
C_5	5:1:1:1		60:12:12:12	12.5:2.5:2.5:2.5
C_6	9:1:1:1		72:8:8:8	15:1.67:1.67:1.67
C_7	9:5:5:5		36:20:20:20	7.5:4.17:4.17:4.17
C_8	0:1:1:0	2	0:48:48:0	0:10:10:0
C_9	1:1:1:0	3	32:32:32:0	6.67:6.67:6.67:0

The $FPCT$ specifies the number of regions in each cell b. The eNBs use $FP(R_0)$ in region $\Re_0$ and the three $FPs(R_1, R_2, R_3)$ in case of reuse 3 ,or the two $FPs(R_1, R_2)$ in case of reuse 2 in region $\Re_1$. However, in case of no reuse, the cell contains only one region $\Re_0$ that utilizes the whole frequency spectrum.

2.4.2 Spectral efficiency calculations

2.4.2.1 Normalized spectral efficiency

Spectral efficiency specifies how efficiently the limited frequency spectrum is utilized. Each UE j will have to calculate its achievable NSE_{ijk} over all RBs i and

modulation types k according to

$$NSE_{ijk} = \frac{ESE_{ijk}}{RM(R)} \tag{2.4.1}$$

where ESE_{ijk} the expected spectral efficiency is defined as

$$ESE_{ijk} = \frac{r_{ijk} * (1 - PER)}{W_{ijk}} \tag{2.4.2}$$

where PER is the UE estimated packet error rate.

The UEs in each cell report their $SINR$ measurement to the eNBs. The higher the $SINR$ estimated by an eNB over a RB, the smaller bandwidth allocated to it and the higher the type of modulation this UE can achieve. The RBs to be assigned to a UE by the eNB is selected to provide the best $SINR$ to that UE. According to Shannon's channel capacity formula, the data rate achieved by a UE j in any $FP(R)$ of cell b is determined as

$$r_{ijk} = W_{ijk} \ \log_2(1 + SINR_{ijk}) \tag{2.4.3}$$

2.4.2.2 Resource metric information

$RM(R)$ information for $FP(R)$ assists the UE to select it's preferred DFPC in order to balance the cell load between its different FPs. $RM(R)$ for reuse-3 calculated for different FPs can be expressed as

$$RM(R) = \begin{cases} 1 & R = 0 \\ 3 - RM(2) - RM(3) & R = 1 \\ 0 \leq RM(2) < 1 & R = 2 \\ 0 \leq RM(3) < 1 & R = 3 \end{cases} \tag{2.4.4}$$

2.4.3 Power assignment

The eNB has different FPs power levels $P_{R,max}$ calculated according to (2.4.5). Cell center $FP(R_0)$ is reused in the cell center with fixed power level ($P_{0,max}$). Cell edge $FPs(R_1, R_2, R_3)$ are reused in the cell edge with power boosted level ($P_{3,max}$) and in the cell center with power de-boosted levels ($P_{1,max}, P_{2,max}$). A power control factor denoted as β is used to determine the maximum power ($P_{R,max}$) in each FP R where $\beta = 0$ for non existing partitions in any DFPC [47].

$$Power\ levels\ (P_{R,max}) \begin{cases} P_{0,max} = \beta P & \beta = 1 \\ P_{1,max}, P_{2,max} = \beta P & \beta < 1 \\ P_{3,max} = \beta P & \beta > 1 \end{cases} \tag{2.4.5}$$

AFFR power patterns are employed by eNBs to manage the power adaptation. The number of AFFR power patterns available in the network is specified according to the DFPC that determine the reuse factor. Fig. 2.4.1 shows an example for different AFFR power patterns for network with reuse_3. Cell (1) in Fig. 2.3.1 is assumed to use pattern (1), while cells (3,5,7) are assumed to use pattern (2), and cells (2,4,6) are assumed to use pattern (3). The AFFR power model shown in Fig. 2.4.1 aims at using the whole spectrum in the cell center and increasing the maximum transmission power limits in the cell edge to enhance the achieved spectral efficiency while preventing ICI between neighboring cells.

The amount of consumed power P_{ijk} by the UE j according to the allocation of RB i with modulation type k is expressed as

$$P_{ijk} = 4\pi d^2 * \frac{2^{r_{ijk}/W_{ijk}}}{\rho * SINR_{ijk}} \tag{2.4.6}$$

Where $\rho = \frac{-1.5}{ln(5*BER)}$ is constant and it depends on the bit error rate (BER) [48].

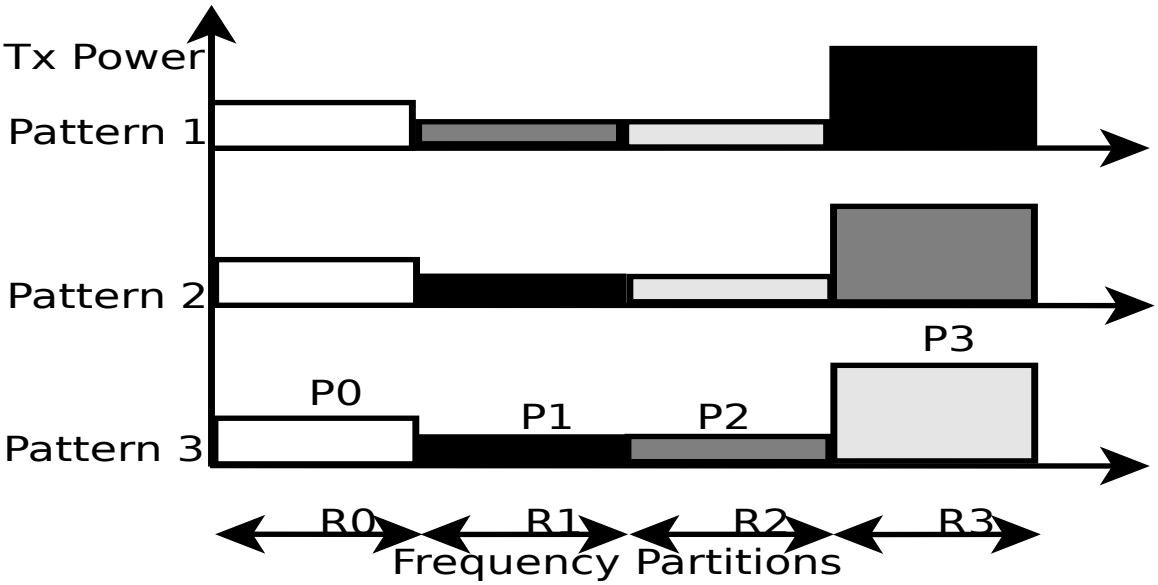

Figure 2.4.1: AFFR (reuse-3) power patterns.

2.4.4 QoS and delay requirements

The normalized queue delay θ_j experienced by UE j traffic is formulated as shown

$$\theta_j = \frac{\tau_j - \bar{\tau}_j}{\bar{\tau}_j} \tag{2.4.7}$$

Where τ_j is the delay a UE j traffic experience in the queue up to the end of the TTI. $\bar{\tau}_j$ is the average delay experienced over all UEs in the queue at the end of the TTI. The QoS coefficient α_j is a parameter that specifies the priority of the transmission of UE j in the cell b. The QoS traffic priority coefficient α_j is represented as a quantized number where $|\alpha_j| = \{1, 2, 3, 4, 5, 6, 7, 8\}$ and each number demonstrates one of eight QoS levels that map to a specific traffic type [43].

2.5 Utility and Problem Formulation

2.5.1 Utility formulation

The new proposed utility function is used a transmission priority metric to identify the best RBs to be allocated to a UE under certain modulation type. A utility-function u_{ijk} defined for each UE is expressed as

$$u_{ijk} = \frac{NSE_{ijk}}{P_{ijk}} \tag{2.5.1}$$

The total system utility U at allocation time is calculated as the summation of all the UEs utilities and defined as

$$U = \sum_{i=1}^{N} \sum_{j=0}^{S} \sum_{k=1}^{M} u_{ijk} x_{ijk} \tag{2.5.2}$$

Where $x_{ijk} \in \{\Omega_{R,b}\}$ in (2.5.2) is an assignment indicator where $x_{ijk} = 1$ if RB i is assigned to UE j with modulation type k, and $x_{ijk} = 0$ otherwise. $\Omega_{R,b}$ is the assignment matrix in any partition R of cell b at time of allocation. A null UE $j = 0$ for which $u_{i0k} = 0$ for all RBs i and modulation types k is defined. $x_{i0k} = 1$ means that RB i is not assigned to any UE at allocation time. This case indicates that there is no transmission scheduled on this RB over this allocation time slot. Each RB can only be assigned to one and only one user in the system, however, one user may be assigned many RBs.

2.5.2 Problem formulation

2.5.2.1 Optimal allocation

We formulate the resources allocation problem in a multi-cell LTE-A system to maximize (U). The problem is formulated as shown from (2.5.3) to (2.5.11) to enable

each eNB to calculate the assignment matrices ($\Omega_{R,b}$) for each $FP(R)$.

$$max \quad \sum_{i=1}^{N}\sum_{j=0}^{S}\sum_{k=1}^{M} u_{ijk}x_{ijk} \tag{2.5.3}$$

$$S.t. \quad \sum_{i=1}^{N}\sum_{j=0}^{S}\sum_{k=1}^{M} P_{ijk}x_{ijk} \leq P_{max} \tag{2.5.4}$$

$$\sum_{j=0}^{S}\sum_{k=1}^{M} P_{ijk}x_{ijk} \leq P_{R,max} \tag{2.5.5}$$

$$\forall\ (i = 1, 2, ...N) \tag{2.5.6}$$

$$\sum_{i=1}^{N}\sum_{j=0}^{S}\sum_{k=1}^{M} x_{ijk} = 1 \tag{2.5.7}$$

$$r_j \geq r_{min,j} \tag{2.5.8}$$

$$\theta_j \leq \theta_{max,j} \tag{2.5.9}$$

$$\alpha_j \geq \alpha_{min,j} \tag{2.5.10}$$

$$\forall\ j = (1,2,3,...,S)$$

$$x_{ijk} \in \{0, 1\} \tag{2.5.11}$$

Where (2.5.4) indicates that the total amount of power consumed in the cell shall not exceed the eNB maximum power. The constraint in (2.5.5) shows the maximum power available to each RB must not exceed the FP maximum power. The binary variable x_{ijk} in (2.5.7) and (2.5.11) ensures that each RB is assigned to one and only one UE. The achievable data rate in (2.5.8) is constrained by the minimum data rate required by each UE. The constraints in (2.5.9) and (2.5.10) ensure the satisfaction of the maximum delay and minimum QoS requirements respectively.

2.5.2.2 Suboptimal allocation

The high computational cost incurred to find the optimal solution of the allocation problem is considered as a limitation for the practical implementation of the proposed

scheme. To overcome this limitation, a suboptimal greedy heuristic algorithm is introduced to decrease the complexity of the problem solution.

A greedy heuristic is used to solve the problem. Greedy algorithm 2.1 shows the details of the greedy heuristic to calculate the suboptimal solution. The algorithm starts by initializing the inputs. For each RB in the system that is expected to be modulated by a UE, the UEs achievable utilities on that RB are arranged in a decreasing order. This is followed by checking the power constraints per eNB and per FP for satisfaction. The algorithm then checks for that the rate, delay, and QoS requirements constraints are satisfied. Finally, the assignment matrix, UEs utilities and amount of power consumed due to the allocation are returned by the algorithm.

Algorithm 2.1 Greedy Heuristic

Require: (N); (S); (K); (P_{max}); ($P_{R,max}$); U[i][j][k]; P[i][j][k];
  for $i \leq N; j \leq S; k \leq K$ **do**
    Sort U[i][j][k] and corrosponding P[i][j][k]; in an ascending order
  end for
  while $P_T \leq P_{max}$ **do**
    if $P_{i,T} \leq P_{R,max} \& r_j \geq r_{min} \& \theta_j \leq \theta_{max} \& \alpha_j \geq \alpha_{min}$ **then**
      U = U + U[i][j][k]
      $P_T = P_T + P[i][j][k]$
      $P_{i,T} = P_{i,T} + P[i][j][k]$ $(\forall\ (i = 1, 2, ...N))$ **return** U[i][j][k] & P[i][j][k]
      set $x[i][j][k] = 1$ in $\Omega_{R,b}$
    end if
  end while
  return $(\Omega_{R,b})$

2.5.3 Complexity analysis

The ILP formulated in (2.5.3)-(2.5.11) can be mapped as a Multiple Choice Knapsack Problem (MCKP). This mapping shows that the solution of the resources allocation optimization problem in a multi-cell system is an NP-hard. MCKP is a problem in combinatorial optimization that consist of a set of multiple classes with several items in each class and to solve the problem an item and only one item must be

chosen from each class to determine the problem objective subject to the applicable constraints.

Similarly to the MCKP, in our problem we assume that there are N classes *(C_1, C_2 , . . ., C_N)* of items to be packed in a knapsack with a maximum capacity of $P_{i,max}$ where each class represents one of the PRUs and consist of *K(S+1)* items corresponding to the *(S+1)* UEs including null allocation over the K modulation constellation. The UE j that achieve maximum utility for PRU i with modulation K is packed in the knapsack at the time of allocation. Each item $\in\{\ C_i\ \}$ has a profit u_{ijk} and a weight p_{ijk}. The purpose of the solution is to choose one item from each class such that the profit sum presented in the system utility U is maximized without having the weight sum represented in the power consumed exceeds P_{max}.

The most successful techniques for solving such problems are branch and bound algorithms that use either linear programming, Lagrangian relaxation or variation of Lagrangian relaxation for bounding purposes [49]. We used Cplex© to find the optimal solution of the optimization problem. Cplex© uses branch-and-bound search with modern algorithmic features such as cuts and heuristics to solve the ILP problems [50].

Discussing the efficiency of algorithm 1 (greedy heuristic) is essential to prove its ability for practical implementation. The computational complexity of the greedy heuristic is analyzed by analyzing the worst case running time of the algorithm.

Initially, in each ABS, the algorithm sorts and labels all U_{ijk} for all DFPCs in a descending order. This step takes time $O(RMSN\,log\,MSN)$ to sort the utility arrays and an additional time $O(RMSN)$ to label the new sorted arrays. U_{max} is calculated by iterating through the AMSs sorted utility array incrementing U_{ijk} while checking for the violation of power constraints. This step incurs a maximum time $O(RMSN)$ to get accomplished.

The worst case running time is calculated by summing up the times required to run the algorithm. Algorithm 1 needs a total $O(RMSN\,log\,MSN)$ to calculate U_{max} for all FPs. The computational complexity analysis for the worst case running time shows that Algorithm 1 is a polynomial-time algorithm.

2.6 Simulation and Results

In this section two simulation scenarios are used and results using the parameters described in Table 2.6.1. are presented. The first scenario shows a comparison between the values of optimal utility, power consumption and NSE for four different schemes. The four schemes are; maximizing the utility function (S1), maximizing the NSE (S2), maximizing the data rate (S3), and minimizing the total power (S4).

Table 2.6.1: System parameters

Description	Value
No. of RBs (N)	96
No. of UEs (S)	60
No. of modulation types (M)	8 (QPSK, 16 QAM, 64 QAM)
No. of eNBs (b)	7
No. of frequency partitions (R)	4
Max. eNB Power ($P_{b,max}$)	20 W
Bit error rate (BER) range	10^{-4} - 10^{-6}
UE queue delay range (τ_j)	150 μs - 6 ms
Channel bandwidth (BW)	10 MHz
Transmission Time Interval (TTI)	1 ms

The second scenario shows a comparison between the optimal utility (S1) and the suboptimal utility (S5). This scenario is used to show the performance of the suboptimal solution compared to the optimal solution.

The resource allocation process is repeated in the simulations each allocation time where the UE in each cell submit their bandwidth requirements to the eNBs. The eNBs first allocate the resources to satisfy the UEs minimum QoS requirements and

then allocate the rest of the resources to maximize the system utility. During the allocation each UE utility is calculated using the system parameters and the UE channel conditions ($SINR$). The number of allocation matrices produced in each allocation time depend on the number of FPs dictated by the CNC in the network. After calculating the assignment matrices ($\Omega_{R,b}$) the UEs are granted the resources.

2.6.1 Optimal allocation

Cplex© is used to solve the optimization problem and to identify the optimal solution. Simulation results generated using Java© are analyzed and discussed. The results of the maximum utility achieved by S1, S2, S3, and S4 are presented in Fig.2.6.1. The results shows that S1 achieves the best joint power and spectral efficiency compared to other schemes S2, S3, and S4. S1 and S4 achieve a high utility, however, both S2 and S3 show very poor utility due to their failure to address the power consumption minimization in their allocation scheme.

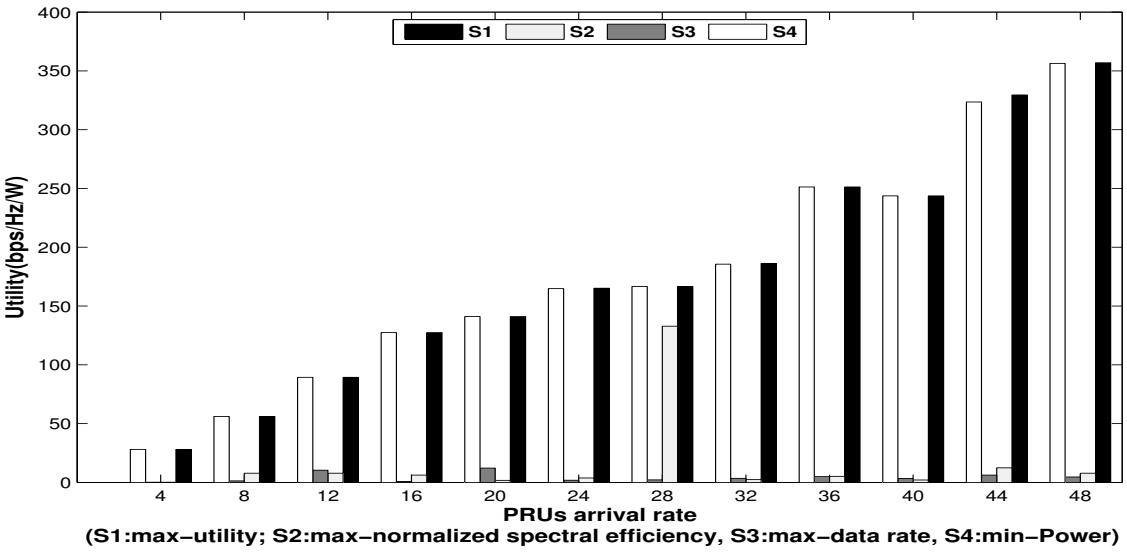

Figure 2.6.1: Maximum utility achieved for different schemes

The results in Fig.2.6.2 indicate that the proposed utility S1 almost achieve the same power efficiency as S4. The cause of the higher power efficiency achieved by S1

is that eNBs prefer to allocate resources to UEs with low power consumption. Results also show that both S2 and S3 produce a poor power efficiency as it fail to consider power consumption in the allocation process.

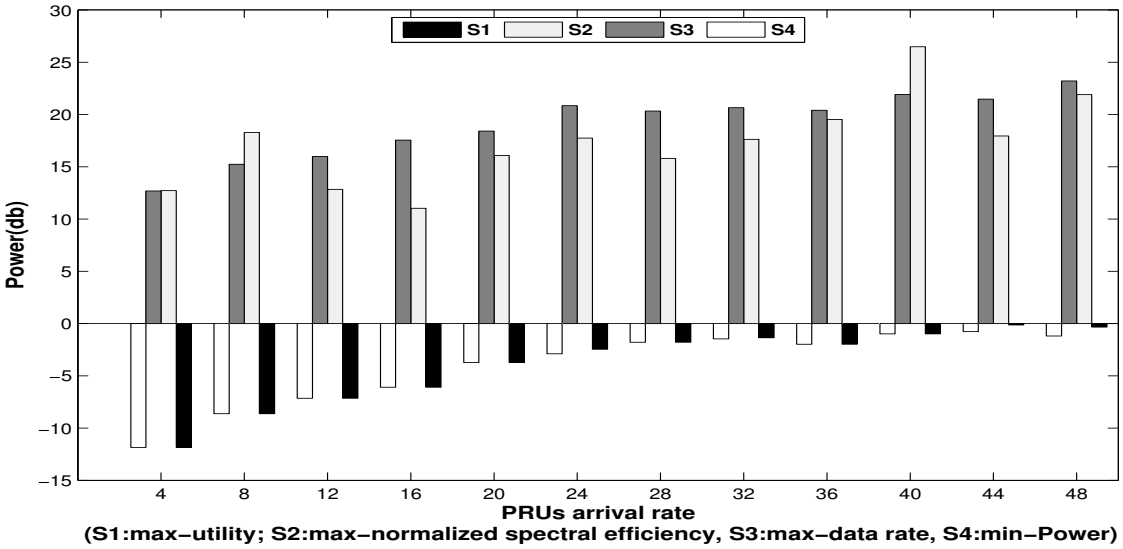

Figure 2.6.2: Power consumption for different schemes

Fig.2.6.3 demonstrates a comparison between the NSE for schemes S1, S2, S3, and S4. S1 achieved higher NSE compared to S4; this occurs due to S1 joint optimization of NSE and power. S2 achieves the maximum NSE, however, S1 achieves a NSE comparable to S2. Results show that S1 compromises between maximizing the NSE and minimizing the consumed power in order to achieve the best utility which is expected to result in the best ROI for MSPs.

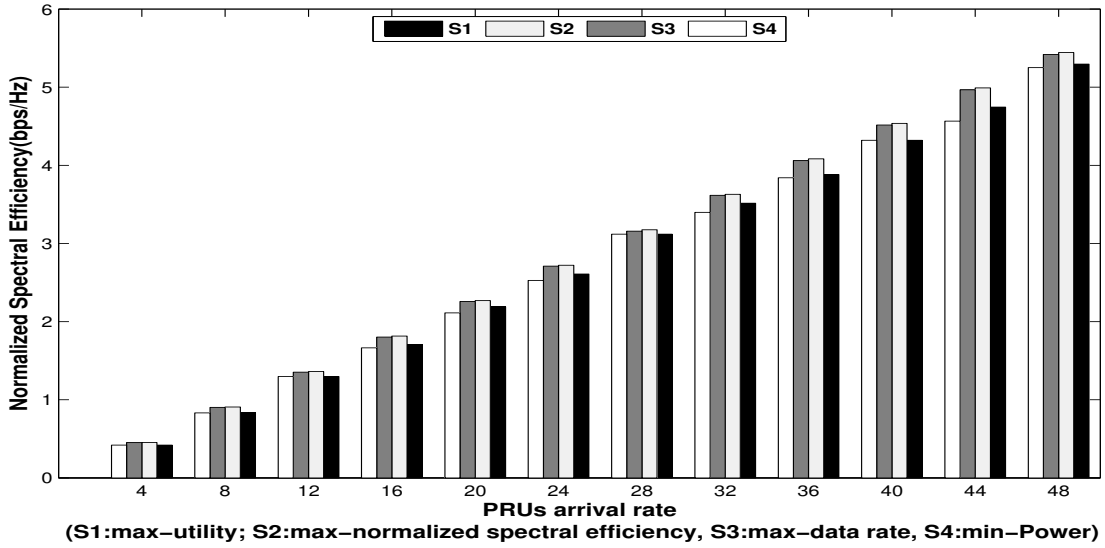

Figure 2.6.3: Normalized spectral efficiency for different schemes

The results in Figs.(2.6.4,2.6.5,2.6.6) demonstrate a comparison between the data rate, normalized queue length and traffic priority respectively between S1, S2, S3 and S4. The three figures present the minimum rate, QoS and delay requirements respectively and shows that S1 fulfills the minimum needed requirements.

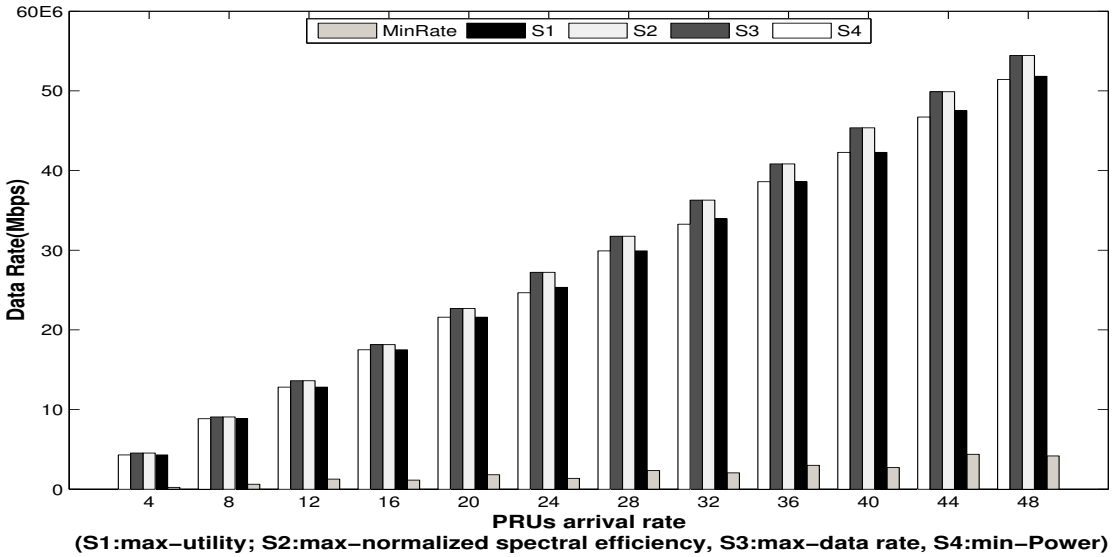

Figure 2.6.4: Data rate comparison for different maximization schemes and minimum rate requirements

Results in Fig.(2.6.4) show that S2 and S3 achieves the best data rate. S1 achieves

approximately the same data rates. The reason for the close results between S1 and S4 is due to the tendency of both schemes to allocate resources to AMSs with low power consumption which result in decreasing data rate in order to save more power while maintaining the minimum rate requirements however S1 achieves a better total average data rate than S4.

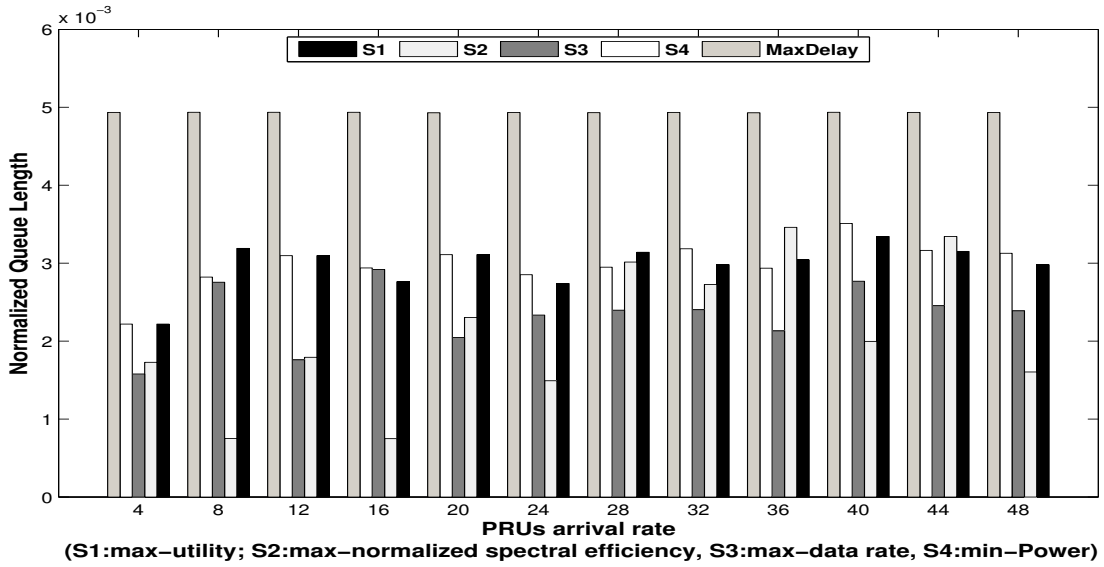

Figure 2.6.5: Normalized queue length comparison for different maximization schemes and maximum delay requirements

The normalized queue length in Fig.(2.6.5) shows that S1 is maintaining a shorter normalized queue length than that required by maximum delay requirements. Results show that S1 maintain a better total average normalized queue length than S4.

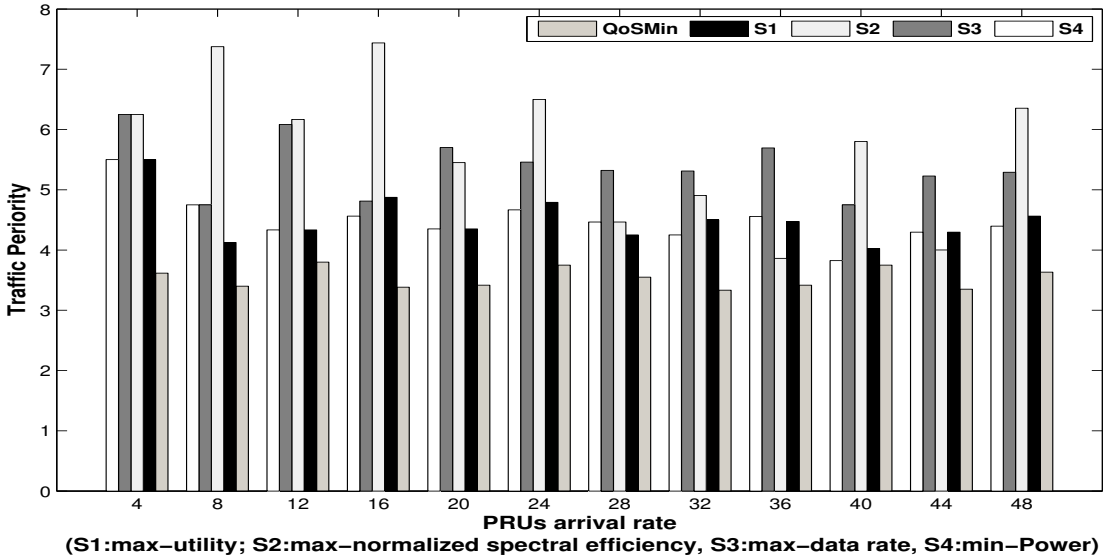

Figure 2.6.6: Traffic priority comparison for different maximization schemes and minimum QoS requirements

Fig.(2.6.6) presents a comparison between the results of the traffic priority for S1, S2 ,S3 and S4. Results show that S1 maintain a higher traffic priority results over the minimum required traffic priority and thus proof to maintain the QoS requirements. S1 shows a higher achieved total average traffic priority than S4.

2.6.2 Suboptimal allocation

The results in Fig.2.6.7 present a comparison between optimal (S1) and suboptimal (S5) solutions. The performance comparison between S1 and S5 shows an average gap of 9%. The performance gap is accepted taking into consideration the importance of employing S5 for practical implementations.

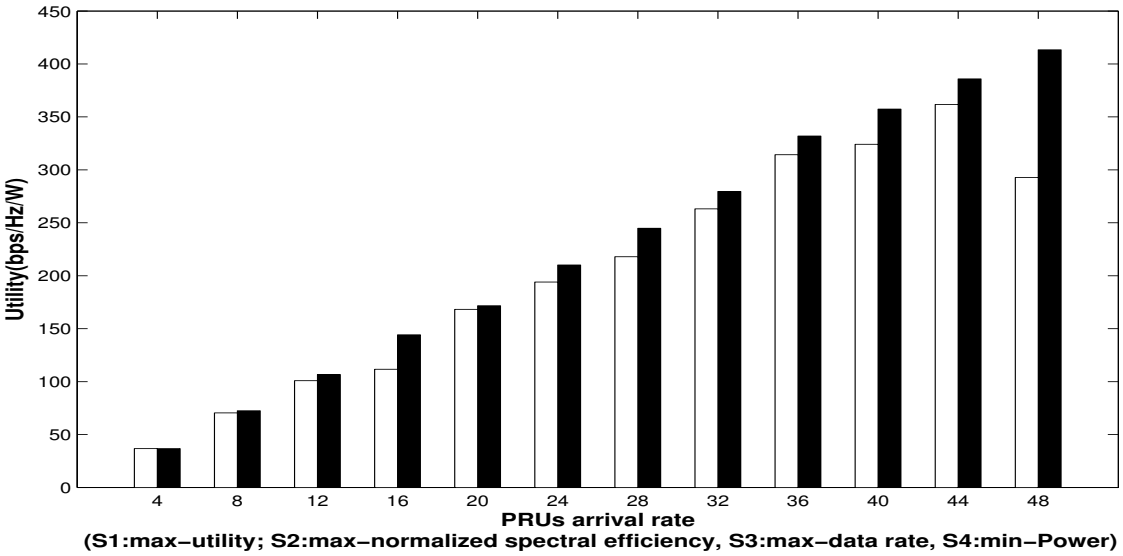

Figure 2.6.7: Maximum utility achieved for optimal and suboptimal solutions

2.7 Conclusion

In this work, we studied the resource allocation problem in LTE-Advanced (LTE-A). Adaptive fractional frequency reuse is utilized to partition each cell in the network. We proposed to utilize the down-link frequency partition configurations concept introduced by WiMAX 2 into LTE-A system to determine the number of frequency partitions in the network and the amount of resource blocks (RBs) in each cell partition.

User equipments in each cell contact the evolved nodeB (eNB) to request resources in the form of resource blocks (RBs) to satisfy their data requirements. Each eNB performs the resource allocation using a utility-based optimization scheme proposed in this paper. The designed scheme aims to maximize the mobile service provider return on investment (ROI) by jointly maximizing spectral efficiency and minimizing amount of power consumed.

An integer linear program (ILP) is formulated and the problem is solved to identify the allocation matrices in each FP that maximize the network utility. A novel utility

function is proposed to maximize the spectral efficiency and minimize the power consumption, while maintaining the QoS, rate, and delay requirements. The problem is mapped as multiple choice Knapsack problem (MCKP) which is an NP-hard problem. The optimal solution is calculated using Cplex© and a comparison with other well known schemes is conducted to show the merit of our scheme.

A polynomial time heuristic called the greedy heuristic is presented as a suboptimal solution to enhance the solution complexity and produce the assignment matrix. The results show that the gap between the optimal and suboptimal solution is within a 9% range. Finally, the proposed scheme shows through simulations a compromise between spectral efficiency and power consumed to achieve the best ROI compared to the other resources allocation schemes.

2.8 References

[32] 3GPP, *3GPP TS 36.300 version 10.8.0 Release 10, Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2*, 2012.

[33] IEEE, *Amendment to IEEE Standard for Local and metropolitan area networks Part 16 : Air Interface for Fixed and Mobile Broadband Wireless Access Systems, Advanced Air Interface*, 2010.

[34] ——, *IEEE Standard for Local and metropolitan area networks Part 16 : Air Interface for Fixed and Mobile Broadband Wireless Access Systems*, May 2009.

[35] D. Ling, Z. Lu, Y. Ju, X. Wen, and W. Zheng, "A multi-cell adaptive resource allocation scheme based on potential game for ICIC in LTE-A," *International Journal of Communication Systems*, 2013.

[36] H. Wang and W. Jia, "An optimized scheduling scheme in OFDMA WiMAX networks," *International Journal of Communication Systems*, vol. 23, pp. 23–39, 2010.

[37] R. Iyengar, K. Kar, B. Sikdar, and X. Luo, "QoS provisioning and radio resource allocation in OFDMA based WiMAX systems," ECSE Department, Rensselaer Polytechnic Institute, 2009.

[38] A. Biagioni, R. Fantacci, D. Marabissi, and D. Tarchi, "Adaptive subcarrier allocation schemes for wireless OFDMA systems in WiMAX networks," *IEEE Journal on Selected Areas in Communications*, vol. 27, pp. 217–225, 2009.

[39] N. Ksairi, P. Bianchi, P. Ciblat, and W. Hachem, "Resource allocation for downlink cellular OFDMA systems — Part I : Optimal allocation," *IEEE Transactions on Signal Processing*, vol. 58, pp. 720–734, 2010.

[40] ——, "Resource allocation for downlink cellular OFDMA systems — Part II : Practical algorithms and optimal reuse factor," *IEEE Transactions on Signal Processing*, vol. 58, pp. 735–749, 2010.

[41] N. U. Hassan and M. Assaad, "Optimal fractional frequency reuse (FFR) and resource allocation in multiuser ofdma system," in *International Conference on Information and Communication Technologies, ICICT '09*, 2009.

[42] S. Hamouda, S. Tabbane, and P. Godlewski, "Improved reuse partitioning and power control for downlink multi-cell OFDMA systems," in *BWAN06, International Workshop on Broadband Wireless Access for ubiquitous Networking*, 2006.

[43] M. Katoozian, K. Navaie, and H. Yanikomeroglu, "Utility-based adaptive radio resource allocation in OFDM wireless networks with traffic prioritization," *IEEE Transactions on wireless communications*, vol. 8, pp. 66–71, 2009.

[44] C. Wang and K.-Y. Chan, "Utility-based admission control for mobile WiMAX networks," *Wireless Networks*, vol. 19, no. 2, pp. 207–218, 2013.

[45] G. Song and Y. Li, "Cross-layer optimization for OFDM wireless networks-part I: theoretical framework," *Wireless Communications, IEEE Transactions on*, vol. 4, no. 2, pp. 614–624, 2005.

[46] ——, "Cross-layer optimization for OFDM wireless networks-part II: algorithm development," *Wireless Communications, IEEE Transactions on*, vol. 4, no. 2, pp. 625–634, 2005.

[47] T. Novlan, R. Ganti, A. Ghosh, and J. Andrews, "Analytical evaluation of fractional frequency reuse for OFDMA cellular networks," *Wireless Communications, IEEE Transactions on*, no. 99, pp. 1–12, 2011.

[48] G. Song and Y. Li, "Adaptive subcarrier and power allocation in OFDM based on maximizing utility," in *The 57th IEEE Semiannual Vehicular Technology Conference, 2003*, vol. 2, 2003, pp. 905–909.

[49] H. Kellerer, U. Pferschy, and D. Pisinger, *Knapsack Problems*, Springer, Ed. Springer, 2004.

[50] IBM, "Branch and cut in CPLEX," IBM ILOG CPLEX Optimization Studio V12.4 documentation, 2014, http://pic.dhe.ibm.com/infocenter/cosinfoc/v12r4/index.jsp?topic=2Filog.odms.ide.help2Frefcppop2Fhtml2Fbranch.htmlb.

CHAPTER 3. JOINT SPECTRAL EFFICIENCY AND POWER ALLOCATION OPTIMIZATION IN IEEE 802.16M

A paper published in *The Journal of IEEE Transactions of Mobile Computing*

Tamer R. Omar[4] [5]and J. Morris Chang[6]

3.1 Abstract

With today's limited bandwidth, high data rate services, and energy efficiency requirements, maximizing the spectral efficiency and minimizing the consumed power becomes essential. Investigating the issues impeding spectral efficiency maximization and consumed power minimization for mobile systems is crucial for solving this contemporary problem. This paper aims to optimize the utilization of the scarce mobile spectrum and the amount of power consumption in the multi-cell IEEE 802.16m

[4]Graduate student and Associate Professor, respectively, Department of Electrical and Computer Engineering, Iowa State University.
[5]Primary researcher and author.
[6]Author for correspondence.

networks. A return on investment model adopting a utility optimization technique is proposed; the model objective is to increase the revenue of the mobile service providers by maximizing the normalized spectral efficiency and decrease the operational cost by minimizing the power consumed in the network. Based on this model, we propose two phases scheme (distributed and centralized) to solve the joint spectral and power optimization problem. The problem is solved to identify the optimal distributed resources assignment and the central down-link frequency partition configuration that achieves the model objective. Simulation results show that the optimal solution significantly improves the system power consumption while maintaining the normalized spectral efficiency, yet with high computational complexity. Accordingly, an effective suboptimal solution utilizing a polynomial-time heuristic is proposed for practical implementations.

Keywords: IEEE 802.16m, normalized spectral efficiency, radio resource management, resource metric.

3.2 Introduction

The increased demand for high speed Internet access and mobile Internet services is exponentially driving the development of fourth generation (4G) technologies. 4G networks utilize both the IEEE 802.16 (WiMAX) and the third generation partnership program long term evolution (3GPP-LTE) technologies [51, 52]. Broadband wireless access based on WiMAX standard was approved on October 2010 as an international mobile telecommunication advanced (IMT-Advanced) technology [53]. Optimal radio spectrum utilization and power consumption are crucial in managing WiMAX systems resources.

Table 3.2.1: Table of Acronyms

Acronym	Description	Acronym	Description
ABS	Advanced Base station	FPS	Frequency Partition Size
AMS	Advanced Mobile Station	HNCE	Higher Network Control Entity
DFPC	Down-Link Frequency Partition Configuration	NSE	Normalized Spectral Efficiency
ESE	Expected Spectral Efficiency	PRU	Physical Resource Unit
FFR	Fractional Frequency Reuse	ROI	Return on Investment
FP	Frequency Partition	RRM	Radio Resource Management
FPCT	Frequency Partition Count	SE	Spectral Efficiency

In WiMAX networks, advanced radio resource management (RRM) schemes that effectively increase spectral efficiency (SE) or decrease power consumption perform a central task in the resource management process. WiMAX introduces a new centralized RRM mechanism called the down-link frequency partition configurations (DFPC) in its latest amendment IEEE 802.16m [54]. This paper aims to determine the optimal DFPC among the list of available DFPCs supported by WiMAX (more details in Section 3.5) that can efficiently allocate resources in order to achieve the joint maximum normalized spectral efficiency (NSE) and minimum power consumption in the network. The optimal DFPC selection is realized by employing an advanced RRM model.

The RRM model utilizes a centralized semi-static adaptive fractional frequency reuse (AFFR) radio resource allocation scheme. In this research, we aim to select the optimal DFPC through two phases. The first phase adopts a distributed approach implemented by each advanced base station (ABS) in the network to determine the optimal local DFPC in its cell. The second phase adopts a centralized approach implemented by a higher network control entity (HNCE) which calculates the optimal global DFPC to be utilized in the network using the optimal local DFPCs recom-

mended by each ABS. Fig. 3.2.1 shows a flow diagram for the RRM process model. The flow diagram presents the model inputs, processes, decision points, and outputs

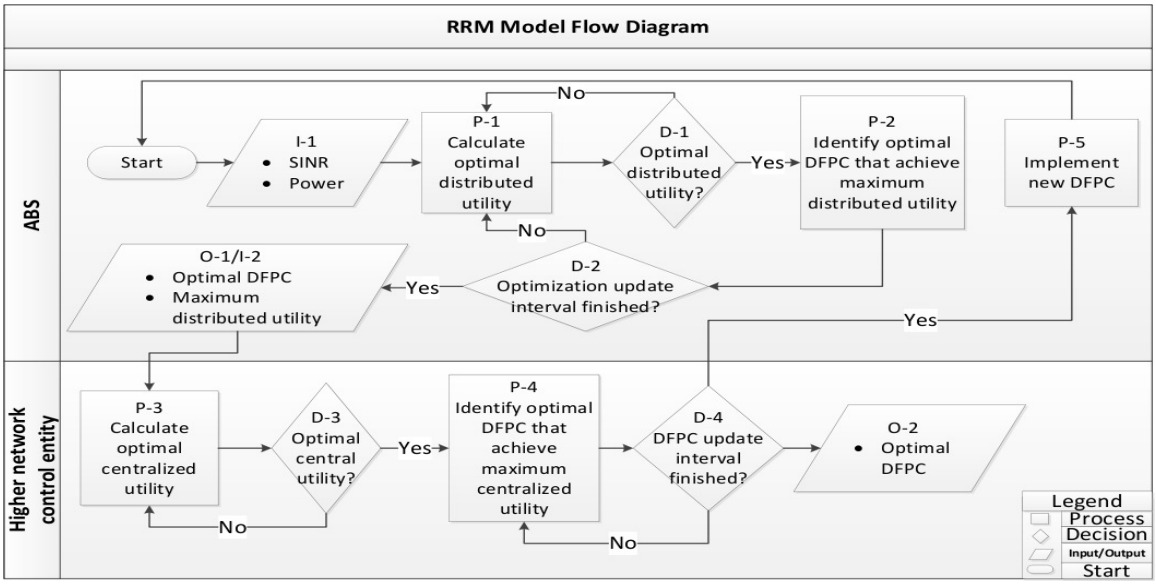

Figure 3.2.1: RRM model flow diagram

A new two phase (distributed and centralized) scheme is proposed to solve the problem of identifying the optimal DFPC to be utilized in the network. First, the distributed phase problem is formulated as an integer linear program (ILP) which aims at maximizing a newly proposed distributed utility function. The distributed utility function maps the mobile service provider's return on investment (ROI) from utilizing the network. The ROI model proposed aims to increase the mobile service provider's revenues by maximizing the NSE and decrease the cost by minimizing the power consumption. All applicable power and integrality constraints are applied to the formulated problem. Second, the central phase problem is formulated as a utility maximization problem for the average sum of the distributed utility per DFPC achieved by each cell in the network.

The problem solution presented identifies the optimal DFPC that achieves the maximum central utility in the network. A comparison between the performance of the proposed scheme and other rate adaptive resource allocation schemes is conducted (e.g [55], [56]). The results show that the proposed scheme is capable of decreasing

the amount of power consumed in the network while maintaining high values for the achievable NSE. The optimal solution demonstrates a high computation complexity. Thus, for practical implementation purposes a suboptimal greedy heuristic algorithm is proposed to reduce the complexity of the optimal solution. The computational complexity analysis for the proposed suboptimal greedy algorithm shows that it is a polynomial-time algorithm.

The paper is organized as follows. The related work is discussed in section II. The system model is introduced in section III. Section IV provides details about DFPC and the power allocation mechanism. The proposed (centralized and distributed) utilities are presented in Section V. Central and distributed phase problems formulation are discussed in section VI. The problem optimal and suboptimal solutions are presented in Section VII. Section VIII presents the simulation and discusses the results. Finally, in section IX, conclusions are drawn. The acronyms used in this paper are listed and explained in Table 3.2.1.

3.3 Related Work

RRM controls frequency partitioning, multi-connection assignment, and resource units scheduling in the network. Frequency partitions (FPs) in WiMAX divide the cell area into two regions (e.g. cell center and cell edge). Each region utilizes a part of the frequency spectrum. Centralized and distributed static, semi-static, and dynamic fractional frequency reuse (FFR) schemes are proposed in previous research studies to increase the network throughput and SE or to minimize the total power consumption.

The authors in [56] proposed two distributed allocation schemes for FFR operation. First scheme dynamically allocates power across FPs by gradual power adaptation through analysis of the optimal power allocation problem. Second scheme assigns users to achieve load balancing between FPs and performs the assignment

with minimum signaling overhead. The results show that the SE can be enhanced by using the proposed schemes under various network conditions. Using dynamic FFR, the authors in [57] use a utility function for allocating sub-carriers to users according to their geographical regions and then apply opportunistic scheduling for assigning the sub-carriers in each cell. Adaptive modulation and coding techniques are used to increase the throughput and a random access sub-band is applied to improve the fairness of the system.

Graph theory is adopted in [59] and [60] to propose AFFR schemes that improve the cell throughput and the users service rate. The authors in [59] proposed AFFR scheme per cell load conditions to enhance the conventional strict fractional frequency reuse (SFFR). The resource allocation problem is translated into a graph coloring problem where a graph is constructed to match the specific version of the utilized AFFR, followed by coloring the graph using a graph algorithm. A graph-based framework is also proposed in [60] to implement AFFR in a multi-cell network. The scheme utilized enhances the SFFR by enabling adaptive spectral sharing based on cell-load conditions. An interference graph that matches the specific realization of FFR and the network topology is constructed. The graph is colored using a heuristic algorithm. Both proposed schemes offer significant performance improvement in terms of cell throughput and service rate.

Frequency partitioning techniques that aim at mitigating interference are addressed in [61, 62, 63]. Universal frequency reuse (UFR) is presented in [61] to control mutual interference among neighboring cells. The authors assign the whole frequency to all cells and design resource allocation rules to avoid inter-cell interference (ICI). A threshold loading factor is used to maintain ICI at minimum level and increase the spectral reuse efficiency. The results show that the UFR provides high spectral reuse efficiency. AFFR with selected power boosted FPs are discussed in [62] to allow the control of maximum power limits per FP; different power patterns can be

employed to determine these limits. In [63] the authors address the joint interaction between interference management and energy utilization. An energy-efficient power optimization scheme is developed for a two-user network with ideal cooperation, then a more generic non-cooperative power optimization scheme is presented to improve the trade-off between energy efficiency and SE. The authors show through simulation in a network with limited interference that the proposed scheme improves the energy and spectral efficiency.

All the previous schemes aimed to maximize either the throughput and SE or minimize the power consumption. Moreover, the scope of all schemes was local using a distributed solution in each ABS. However, the following three important problems that are addressed in this paper were not discussed in previous work:

1. Studying the global implementation of the RRM model in the network. For interference mitigation [54] suggests that at any time, one and only one, DFPC can be implemented in the network. This condition results in a conflict between the local optimal DFPC in each cell and the global optimal DFPC in the network which increases the importance of studying the network globally.

2. Addressing the DFPCs dynamic behavior according to the network topology, load conditions, and users distribution by identifying the optimal DFPC periodical changes that aims to maximize the utilization of the network resources.

3. Using a newly approach in formulating the problem as a joint optimization of the NSE and power consumption to increase the ROI from the network.

3.4 System Model

We consider the down-link transmission adopting orthogonal frequency division multiple access (OFDMA) in a multi-cell WiMAX network. The network consists of

seven identical adjacent hexagonal cells and a central HNCE responsible for globally controlling the network resources as shown in Fig. 3.4.1. It is worse mentioning that the proposed scheme can be applied to other networks that consist of any number of cells. However, the scalability mainly depends on the used topology by the MSPs in implementing their networks.

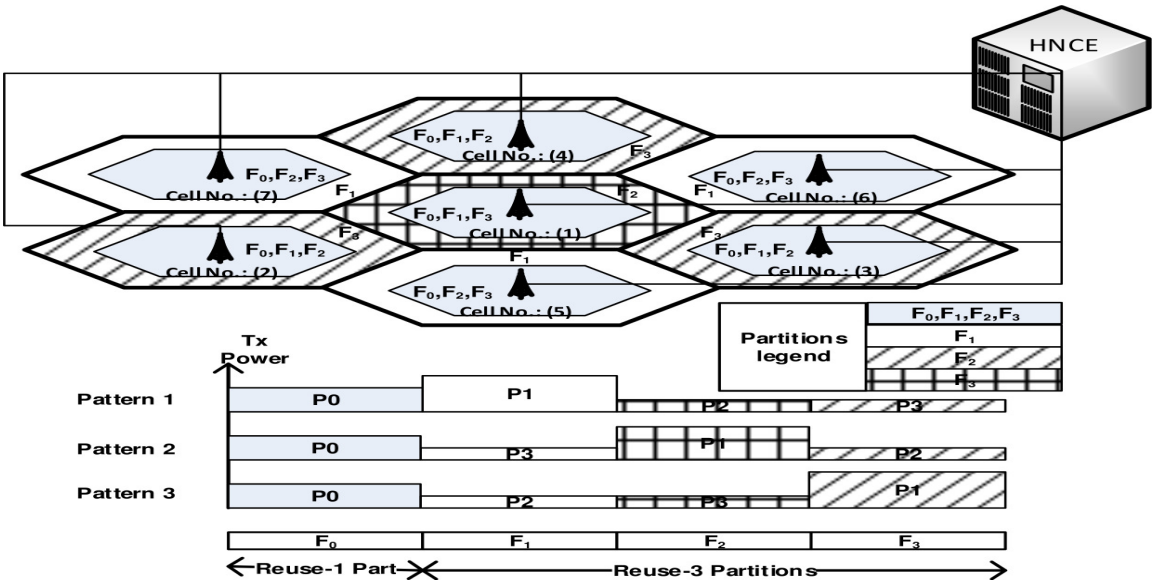

Figure 3.4.1: Network configuration

Each cell contains one centric ABS. The maximum number of cells is B and each cell is denoted by $\mathcal{L}_b$ where $b = \{1,2,...,B\}$. The maximum number of FPs in each cell is M and each FP is denoted by $\mathcal{F}_i$ where $i = \{1,2,...,M\}$. The maximum number of advanced mobile stations (AMSs) with pending traffic in each FP is denoted by S and each AMS is denoted by $\mathcal{S}_j$ where $j = \{1,2,....,S\}$. The number of AMSs differ from one cell FP to another and they are assumed to be uniformly distributed. A physical resource unit (PRU) is the basic physical unit used for resource allocation in the network. AMSs in each cell are only able to modulate PRUs in their assigned FP but not in any other FPs. The maximum number of PRUs in each FP is denoted by N and each PRU is denoted by $\mathcal{P}_k$ where $k = \{1,2,...,N\}$. For any DFPC used in the network, the number of utilized FPs is known as the frequency partition count

(FPCT) and the number of PRUs allocated to each FP is known as the frequency partition size (FPS) [54].

AFFR mechanism in WiMAX indicates a maximum number of DFPCs to be used in the network. The maximum number of utilized DFPCs in the network is $\mathbb{C}$ and each DFPC is denoted by C_c where c ={1,2,....,$\mathbb{C}$}. The maximum number of utilized DFPCs is determined according to the utilized bandwidth and the employed fast fourier transform (FFT) in the network. The bandwidth and FP bandwidth are denoted by $\mathcal{W}$ and $\mathcal{W}_i$ respectively.

3.5 DFPC and Power Allocation

This section aims at understanding the DFPC and the power allocation mechanism introduced in WiMAX. This interpretation is crucial in formulating the problem and identifying the optimal DFPC.

3.5.1 Down-link frequency partition configuration (DFPC)

WiMAX regulates a maximum of four FPs in the network. The utilized DFPC determines the FPCT in the network that ranges from one to four FPs. In case FPCT is greater than 1, each ABS in the network uses FP ($\mathcal{F}_0$) in the cell center. In case of reuse-2, each ABS uses only one of the two FPs ($\mathcal{F}_1$, $\mathcal{F}_2$) in the cell edge. In case of reuse-3, each ABS uses only one of the three FPs ($\mathcal{F}_1$, $\mathcal{F}_2$, $\mathcal{F}_3$) in the cell edge. Fig. 3.4.1 shows a reuse-3 network configuration with four FPs where FP ($\mathcal{F}_0$) is used in the cell center and FPs ($\mathcal{F}_1$, $\mathcal{F}_2$, $\mathcal{F}_3$) are used in the cell edge.

Identifying the optimal DFPC is affected by the network topology, load conditions, and users distribution. Table 3.5.1 presents a detailed description of the DFPCs that can be used in a network with the 20 MHz bandwidth and FFT Size = 2048. Other tables are offered in [54] to networks with different bandwidth and FFT sizes. Each

row in Table 3.5.1 specifies one of the DFPCs, for example a network that uses DFPC (C_4) has four FPs (FPCT = 4), $\mathcal{F}_0$ is the cells center FP and its size is three times the size of the cells edge FPs in the network (3:1:1:1). The number of PRUs (FPS) allocated to $\mathcal{F}_0$ to be distributed on the AMSs are 48 PRUs with a total bandwidth $\mathcal{W}_0$ of 10 MHz. Similarly ($\mathcal{F}_1$, $\mathcal{F}_2$, $\mathcal{F}_3$) are three equal cell edge FPs, according to DFPC (C_4) each edge FP is allocated 16 PRUs to be distributed on its AMSs with a total bandwidth of 3.33 MHz for each FP ($\mathcal{W}_1$:$\mathcal{W}_2$:$\mathcal{W}_3$).

Table 3.5.1: Down-link frequency partition configurations (FFT = 2048, $\mathcal{W}$ = 20 MHz).

C_c	$\mathcal{F}_0$:$\mathcal{F}_1$:$\mathcal{F}_2$:$\mathcal{F}_3$	FPCT	FPS(PRUs)	$\mathcal{W}_0$:$\mathcal{W}_1$:$\mathcal{W}_2$:$\mathcal{W}_3$
C_1	1:0:0:0	1	96:0:0:0	20:0:0:0
C_2	0:1:1:1	3	0:32:32:32	0:6.67:6.67:6.67
C_3	1:1:1:1	4	24:24:24:24	5:5:5:5
C_4	3:1:1:1		48:16:16:16	10:3.33:3.33:3.33
C_5	5:1:1:1		60:12:12:12	12.5:2.5:2.5:2.5
C_6	9:1:1:1		72:8:8:8	15:1.67:1.67:1.67
C_7	9:5:5:5		36:20:20:20	7.5:4.17:4.17:4.17
C_8	0:1:1:0	2	0:48:48:0	0:10:10:0
C_9	1:1:1:0	3	32:32:32:0	6.67:6.67:6.67:0

Four groups with the same FPCT are illustrated in Table 3.5.1;

1. Group 1 includes (C_1) with FPCT = 1.
2. Group 2 includes (C_8) with FPCT = 2.
3. Group 3 includes (C_2, C_9) with FPCT = 3.
4. Group 4 includes (C_3, C_4, C_5, C_6, C_7) with FPCT = 4.

For interference mitigation purposes, only one group is assumed to be used in the network at any time instance. However, the network can switch between different groups every DFPC update interval (T). Once the operating groups is chosen based on to the network topology, the optimal DFPC can be identified from within the chosen groups DFPCs. However in case of disconnected networks with multiple subsets

controlled under the same HNCE, the HNCE can choose a different DFPC group for each subset. The optimal DFPC changes on semi-static basis in order to adapt with the variations in the network load conditions and users distribution in the network.

3.5.2 Power allocation mechanism

AMSs in cell center experience good channel conditions but need to control their power transmission levels to avoid ICI. However, AMSs in the cell edge suffer from bad channel conditions that require them to boost their transmission power levels. WiMAX adopts an AFFR power allocation mechanism that aims at both preventing ICI and improving NSE.

3.5.2.1 AFFR power allocation mechanism

The power pattern presentation in Fig. 3.4.1 shows an example of the AFFR power allocation mechanism. Regular, boosted, and de-boosted transmission power levels are utilized in each cell. AFFR uses different power patterns for each group of neighbor cells; ABSs in cells (1), (2, 3, 4), and (5, 6, 7) use patterns (2), (3), and (1) respectively. ABS segment IDs are used for managing the power patterns utilization in the network by ensuring that neighbor ABSs use different power patterns [54].

Power patterns in the cell center shown in Fig. 3.4.1 has a regular power level range (P_0) for ($\mathcal{F}_0$) and power de-boosted levels range of (P_2, P_3) for ($\mathcal{F}_2$, $\mathcal{F}_3$). In the cell edge, however, the power patterns has power boosted level range (P_1) for ($\mathcal{F}_1$). Controlling the power de-boosted levels (P_2, P_3) helps avoid ICI generated from the cell center AMSs, while controlling the power boosted level (P_1) in the cell edge enhance the NSE.

3.5.2.2 Power allocation for frequency partitions

The maximum transmission power level in each cell is defined by P_{max}. FP ($\mathcal{F}_0$) is reused in the center of each cell with a fixed maximum power level ($P_{0,max}$) FPs ($\mathcal{F}_1$, $\mathcal{F}_2$, $\mathcal{F}_3$) are reused with a maximum power boosted level ($P_{1,max}$) in the cell edges and a maximum power de-boosted levels ($P_{2,max}$, $P_{3,max}$) in the cell center. In general, the maximum power $P_{i,max}$ in each FP is calculated according to (3.5.1).

$$P_{i,max} = \begin{cases} P_{0,max} = \alpha P_{max} & \alpha < 1 \\ P_{1,max} = \beta P_{max} & \beta = 1 \\ P_{2,max}, P_{3,max} = \gamma P_{max} & \gamma \lll 1 \end{cases} \tag{3.5.1}$$

where α, β, γ are the power control factors used to determine the value of $P_{i,max}$ and $\alpha, \beta, \gamma = 0$ for non-existing FP in any DFPC [58].

3.5.3 Spectral efficiency

In order for ABSs to determine the optimal local DFPC, at system entry each AMS in the cell is required by the ABS to calculate it's achievable NSE for all cell FPs. The AMSs compare the values of the NSE in all FPs and identify the FP with the maximum calculated NSE. AMSs inform the ABS by their preferred frequency partition (PFP) using the preferred frequency partition indicator (PFPI). The feedback from the AMSs support the ABS decisions to admit AMSs into different cell FPs.

3.5.3.1 SINR and rate calculations

The signal to interference plus noise ratio (SINR) threshold model proposed in [58] is adopted. SINR experienced when allocating PRU ($\mathcal{P}_k$) at cell ($\mathcal{L}_b$) FP ($\mathcal{F}_i$) to AMS ($\mathcal{S}_j$) is denoted by $SINR_{bijk}$. AMSs are classified in the cell regions according

to their SINR threshold denoted by δ_b into cell center or cell edge AMSs. AMSs with $SINR_{bijk} \geq \delta_{\mathrm{b}}$ are classified as cell center AMSs, and AMSs with $SINR_{bijk} < \delta_{\mathrm{b}}$ are classified as cell edge AMSs. Adaptive modulation and coding (AMC) is assumed, AMSs adjust their transmission constellation according to the channel state conditions; in case of good channel conditions AMSs use a high order modulation while in case of poor channel condition a low modulation order is used. The AMC constellation (e.g. QAM) is selected in the ABS by the link adaptation procedure according to the SINR exhibited by each AMS and regularly reported to the ABS. The SINR is calculated according to (3.5.2)

$$SINR_{bijk} = \frac{P_{bijk}G_{bijk}}{\sum_{v\epsilon I_l} P_{livk}G_{livk} + N_0 w} \tag{3.5.2}$$

where P_{bijk} is the power consumed for allocating PRU ($\mathcal{P}_k$) to AMS ($\mathcal{S}_j$) in cell ($\mathcal{L}_b$) FP ($\mathcal{F}_i$), P_{livk} is the power consumed for allocating PRU ($\mathcal{P}_k$) to AMS ($\mathcal{S}_v$) in all interfering cells ($\mathcal{L}_l$) FP ($\mathcal{F}_i$), where $v\epsilon I_l$ and I_l is the set of interfering AMSs to AMS ($\mathcal{S}_j$) that uses the same FP ($\mathcal{F}_i$) in all interfering cells ($\mathcal{L}_l$). N_0 is the thermal noise density and w is the sub-carriers separation. G_{bijk} is the channel gain experienced by AMS ($\mathcal{S}_j$) in cell ($\mathcal{L}_b$) FP($\mathcal{F}_i$) on PRU ($\mathcal{P}_k$), G_{livk} is the channel gain experienced by AMS ($\mathcal{S}_v$) in cell ($\mathcal{L}_l$) FP ($\mathcal{F}_i$) on PRU ($\mathcal{P}_k$). The channel gain is calculated as shown in (3.5.3)

$$G_{bijk} = 10^{\frac{-\Gamma_{bijk}(d)}{10}} \omega_{bijk}\varsigma_{bijk} \tag{3.5.3}$$

where $\Gamma_{bijk}(d)$ is the path loss at distance d. ω_{bijk} and ς_{bijk} are the shadowing and fading coefficient respectively [64].

According to the Shannon's theorem, the data rate r_{bijk} achieved by each AMS can be expressed as shown in (3.5.4)

$$r_{ijk} = w \log_2(1 + \lambda SINR_{ijk}) \tag{3.5.4}$$

where $\lambda = \frac{-1.5}{ln(5*BER)}$ and BER is the Bit Error Rate [55].

For each cell ($\mathcal{L}_b$) FP ($\mathcal{F}_i$) the achieved throughput denoted by $\mathcal{R}$ is calculated as shown in (3.5.5)

$$\mathcal{R}_{bi} = \sum_{j=1}^{S} \sum_{k=1}^{N} w \log(1 + \lambda SINR_{bijk}) \tag{3.5.5}$$

The total throughput is the sum of the achievable throughput in cell ($\mathcal{L}_b$) FP ($\mathcal{F}_i$) by all AMSs according to the PRUs allocation results from solving the optimization problem by each ABS.

3.5.3.2 Normalized spectral efficiency

The NSE calculated as shown in (3.5.6) specifies how efficiently the limited frequency spectrum is utilized.

$$NSE_{ijk} = \frac{ESE_{ijk}}{RM_i} \tag{3.5.6}$$

where ESE_{ijk} and RM_i are AMS ($\mathcal{S}_j$) expected spectral efficiency from allocating PRU ($\mathcal{P}_k$) in FP ($\mathcal{F}_i$) and the resource metric (RM) in FP($\mathcal{F}_i$) respectively.

3.5.3.3 Expected spectral efficiency

The ESE is calculated according to (3.5.7)

$$ESE_{ijk} = \frac{r_{ijk}(1 - PER)}{\mathcal{W}_i} \tag{3.5.7}$$

where r_{ijk} and $\mathcal{W}_i$ are the AMS expected data rate and FP bandwidth respectively. PER is the AMS estimated Packet Error Rate. [7] The expected value of the PER is denoted by packet error probability E_p. For a data packet length of n bits, E_p is calculated as shown in (3.5.8)

$$E_p = 1-(1-b_e)^n \tag{3.5.8}$$

where b_e is the bit error probability, b_e is equal to the expected value of the BER.

3.5.4 Resource metric information

AMSs periodically check their achievable NSE for all cell FPs in order to update the ABS by their PFP. In order to obtain more distributed control, WiMAX employs a resource metric information that impacts the calculation of the NSE. Each ABS periodically sends the RM information in each super-frame to all AMSs in the cell. AMSs utilize the RM values to calculate the NSE. The RM values selected by each ABS changes the values of the AMSs achievable NSE and impacts their selection for the PFP.

The NSE calculated by each AMS depends on the RM values indicated by the ABS. RM_i for networks with reuse-3 and reuse-2 is calculated for different FPs $\mathcal{F}_i$ according to (3.5.9) and (3.5.10) respectively [54].

$$RM_i = \begin{cases} 1 & i = 0 \\ 3 - RM_2 - RM_3 & i = 1 \\ 0 \leq RM_2 < 1 & i = 2 \\ 0 \leq RM_3 < 1 & i = 3 \end{cases} \tag{3.5.9}$$

[7] PER presents the number of incorrectly received data packets divided by the total number of received packets. A packet is declared incorrect if at least one of the packet bits is erroneous.

$$RM_i = \begin{cases} 1 & i = 0 \\ 2 - RM_2 & i = 1 \\ 0 \leq RM_2 < 1 & i = 2 \end{cases} \tag{3.5.10}$$

where RM_2 and RM_3 are the values for the resource metric for FPs $\mathcal{F}_2$ and $\mathcal{F}_3$ respectively. RM_2 and RM_3 are assumed as a random values in the allowed range in (3.5.9) and (3.5.10). The ABS transmits the resource metric values as a quantized fractional number *"y"* between zero and one. Each AMS receives and decodes the quantized resource metric to determine the RM real value [54]. The identification of the optimum values of RM_2 and RM_3 is important in balancing the load between different FPs, however, its not considered in the scope of this paper.

The following brief example is used to clarify the effect of the RM values on the calculations of the NSE. In case of reuse-3 for the lower bound $RM_2 = 0$, the calculated NSE by the AMSs in $\mathcal{F}_2$ will tend to infinity which will be a reason for the these AMSs to choose $\mathcal{F}_2$ as their PFP. However in case of upper bound $RM_2 \simeq 1$, the calculated NSE by the AMSs in $\mathcal{F}_2$ will be approximately equal to the expected calculated ESE which will prevent the AMSs from joining $\mathcal{F}_2$ if the ESE is low or at least not competitive enough to other FPs. Similar to RM_2, the values RM_3 have the same effect on the calculations of the NSE in $\mathcal{F}_3$. The changes in RM_2 and RM_3 are significant to the calculations of RM_1 as shown in (3.5.9). RM_1 values increases and the expected NSE achieved in $\mathcal{F}_1$ decreases if RM_2 and RM_3 decreases and vise versa. This results in the variation of the AMSs decision to choose $\mathcal{F}_1$ as their PFP.

3.5.5 Utility update intervals

The maximum expected central utility is calculated every DFPC update interval denoted by T while the maximum expected distributed utility is calculated every

optimization calculation interval denoted by t. There is a dependency between t and T values, the value of t shall be less than T and a suitable t interval should be selected to present the network dynamics. For example $T = 5t$ implies that the maximum central utility is calculated using five calculated samples of the maximum distributed utility during T interval.

The sampling rate is determined according to the value of the sampling factor denoted by $\tau = {}^{T}/_{t}$. The optimum value of τ is important, a larger sampling rate (e.g. 50 samples/sec, sample each super frame) increases the accuracy of the optimal solution, however, a small sampling rate (e.g. 5 samples/sec, sample each 10 super frames) decreases the load incurred on the ABSs from solving the optimization problem. Thus for a T interval of 1 hr with high sampling rate $\tau = 180,000$ sample is needed while for low sampling rate $\tau = 18,000$ sample. This trade-off between accuracy and computational load needs to be addressed before choosing the optimal sampling factor (τ).

In WiMAX, T ranges from 1 to 143,200 minutes (≈99.4 days) with a default of 1 hr [54]. T duration controls the dynamics of the DFPC utilized in the network. Moreover, t duration specifies the accuracy of the maximum central utility calculations . Suitable selection of τ increases the accuracy of the optimal solution, and decreases the load incurred on ABSs due to the iterative calculation of the optimal distributed utility.

In the proposed scheme, the optimal distributed utility and the DFPC achieving it are calculated in the ABSs. Therefore a message must be communicated between each ABS and the HNCE. This message is responsible for updating the HNCE by the values of the distributed utility and their corresponding DFPC numbers. This will be an 18-20 bytes message send by each ABS to inform the HNCE with the needed information to calculate the maximum central utility and to identify the optimal

DFPC. The communication between the ABSs and the HNCE incur an overhead which is proportional to the number of the ABSs in the network.

3.6 Utility and Problem Formulation

The two phase problem aims at identifying the the optimal DFPC to be implemented in the network. The distributed phase calculates the PRUs allocation matrix X_b and determines the optimal DFPC locally in each cell $\mathcal{L}_b$ by using a newly proposed distributed utility function. The central phase calculates the DFPCs allocation matrix X_c and identifies the optimal DFPC globally in the network by using a central utility function. The objective of these scheme is to jointly maximize the NSE and minimize the power consumption in the network.

3.6.1 Utility function formulation

The proposed scheme uses two utility functions to identify the optimal DFPC. The distributed phase utility function denoted by U_b is calculated locally by each ABS and the central phase utility function denoted as U_c is calculated globally by a HNCE for the network.

3.6.1.1 Scheme distributed phase "Utility function"

The objective of the proposed distributed utility is to increase the ROI of the MSPs' by jointly increasing the NSE and decreasing the consumed power. In order to achieve this objective, the utility is selected in a manner that minimize the calculation load incurred on the ABSs. As previously mentioned, the AMSs periodically check their achievable NSE and report it to the ABSs and also the calculations of the consumed power is a regular task performed by the AMSs. However, other forms of the utility can be adopted to achieve the same objective but with extra calculation

cost. The utility achieved by each AMS is denoted by U_{ijk}. The AMS utility is calculated as shown in (3.6.1).

$$U_{ijk} = \frac{NSE_{ijk}}{P_{ijk}} \tag{3.6.1}$$

where NSE_{ijk} and P_{ijk} are the NSE achieved and the power consumed respectively due to the allocation of PRU ($\mathcal{P}_k$) to AMS ($\mathcal{S}_j$) in FP ($\mathcal{F}_i$).

U_b is the sum of the utility achieved by all AMSs in each cell $\mathcal{L}_b$. U_b is calculated by each ABS according to (3.6.2)

$$U_b = \sum_{i=1}^{M}\sum_{j=1}^{S}\sum_{k=1}^{N} U_{ijk} \tag{3.6.2}$$

The maximum distributed utility achieved by each ABS during t interval is denoted by $U_{b,max}$ and the maximization problem that calculate $U_{b,max}$ is addressed in section 3.6.2.1.

3.6.1.2 Scheme central phase "Utility function"

The central utility denoted by U_c is the sum of $U_{b,max}$ per C_c achieved by all ABSs in the network. The HNCE calculates U_c as shown in (3.6.3)

$$U_c = \sum_{b=1}^{B} U_{b,max}/B \tag{3.6.3}$$

for each DFPC the central utility calculated by the HNCE each T interval using U_cs reported by the ABSs each t interval is denoted by $U_{c,T}$ and expressed as

$$U_{c,T} = \sum_{t=1}^{T} U_{c,t} \tag{3.6.4}$$

the maximum central utility that determine the DFPC to be implemented in the network is denoted by $U_{c,max}$ and the maximization problem that calculate $U_{c,max}$ is addressed in section 3.6.2.2.

3.6.2 RRM model

The proposed RRM model is executed using a two phase (centralized and distributed) approach; The first phase solves a distributed problem each optimization calculation interval (t) implemented locally by each ABS. The objective of the scheme distributed phase is to calculate $U_{b,max}$. The second phase solves a centralized problem each DFPC update interval (T) implemented globally by a HNCE. The objective of the scheme central phase is to calculate $U_{c,max}$ and to identify the optimal DFPC to be utilized in the network.

3.6.2.1 Scheme distributed phase "Problem formulation"

Each ABS calculates the AMSs expected achievable data rates E$[r_{ijk}]$ for all current group DFPCs. The AMSs expected SINR E$[SINR_{ijk}]$ values for all current group DFPCs are assumed to be the same as the current measured SINR values. In order to calculate the AMSs expected utility E$[U_{ijk}]$, each ABS calculates the AMSs expected E$[NSE_{ijk}]$ and the expected power consumption E$[P_{ijk}]$ for all current group DFPCs. After calculating the E$[U_{ijk}]$, the ABS solves the joint optimization problem and calculates the expected maximum distributed utility E$[U_{b,max}]$ for all the current group DFPCs. The ABS reports the maximum distributed utilities and their corresponding DFPCs to the HNCE.

The distributed phase joint optimization problem for each C_c is modeled to calculate $U_{b,max}$ as shown in (3.6.5) to (3.6.11).

$$\underset{X_b}{arg\,max} \quad \sum_{i=1}^{M}\sum_{j=1}^{S}\sum_{k=1}^{N} U_{ijk}x_{ijk} \tag{3.6.5}$$

$$s.t \quad \sum_{i=1}^{M}\sum_{j=1}^{S}\sum_{k=1}^{N} P_{ijk}x_{ijk} \leq P_{max} \tag{3.6.6}$$

$$P_{ijk}x_{ijk} \leq P_{i,max} \tag{3.6.7}$$

$$\sum_{k=1}^{N} r_{ijk} \geq r_{ijk,min} \tag{3.6.8}$$

$$\forall(j = 1, 2, ..., S) \tag{3.6.9}$$

$$\sum_{i=1}^{M}\sum_{j=1}^{S}\sum_{k=1}^{N} x_{ijk} = 1 \tag{3.6.10}$$

$$x_{ijk} = \{0, 1\} \tag{3.6.11}$$

where U_{ijk} is the utility function achieved by AMS ($\mathcal{S}_j$) using PRU ($\mathcal{P}_k$) in FP ($\mathcal{F}_i$). The maximum ABS transmission power constraint in (3.6.6) states that the total power assigned to all AMSs are limited by ABS maximum power P_{max}. The power constraint in (3.6.6) ensures that AMSs in each FP limit their maximum transmission power to $P_{i,max}$. The constraint in (3.6.8) shows that the achievable rate requirement for each AMS should guarantee the minimum rate requirement $r_{ijk,min}$. This minimum rate is required to satisfy the QoS in the system. The binary variables x_{ijk} are the assignment indicators of the assignment matrix X_b, and $x_{ijk} = 1$ if PRU ($\mathcal{P}_k$) at FP ($\mathcal{F}_i$) is assigned to AMS ($\mathcal{S}_j$) and $x_{ijk} = 0$ otherwise. The constraint on x_{ijk} in (3.6.11) and (3.6.10) ensures that each PRU is assigned to one and only one AMS. This problem's affine objective and constraint functions together with the integrality constraint construct a convex ILP optimization problem[65].

3.6.2.2 Scheme central phase "Problem formulation"

The optimal DFPC is identified by solving the U_c maximization problem. The problem modeled in (3.6.5) to (3.6.10) calculates $U_{b,max}$ for each C_c in each cell. The

U_c maximization problem aims to calculate $U_{c,max}$ and determine optimal DFPC from all possible DFPCs. Each ABS sends the values of the optimal distributed utilities each t interval with their corresponding $C_c s$ to the HNCE. The HNCE utilizes this information to calculate the $U_{c,T}$ achieved during the T interval by solving the problem modeled in (3.6.12).

$$\underset{X_c}{arg\,max} \quad U_{c,T} x_{c,T} \tag{3.6.12}$$

$$x_{c,T} = \{0, 1\} \tag{3.6.13}$$

where the binary variables $x_{c,T}$ are the assignment indicators for the assignment vector X_c, c is the configuration number and T is the allocation time. $x_{c,T} = 1$, if C_c is a possible DFPC and $x_c = 0$, otherwise. $U_{c,T}$ is the sum of the U_c calculated every interval t during T interval for each C_c as shown in (3.6.4). Every T interval, the HNCE compares the $U_{c,T}$s for all DFPCs by solving the trivial optimization problem in (3.6.12) to select the DFPC with $U_{c,max}$. The DFPC with $U_{c,max}$ is the selected optimal DFPC to be implemented in the network.

3.7 Problem Solution

3.7.1 Optimal solution

3.7.1.1 Scheme distributed phase

The distributed phase problem to be solved by each ABS is a binary ILP optimization problem. This problem is classified as NP-hard problem with high computational complexity when optimally solved. It's solved by determining $U_{b,max}$ per DFPC in each ABS.

3.7.1.2 Scheme central phase

After determining $U_{b,max}$ per DFPC in each ABS, $U_{c,T}$ is incremented in the HNCE by $U_{b,max}$ every t interval to calculate $U_{c,T}$ achieved by each DFPC. The $U_{c,T}$s for all DFPCs are then compared to determine $U_{c,max}$ and identify its corresponding DFPC to be utilized in the network.

3.7.2 Suboptimal solution

3.7.2.1 Scheme distributed phase "Greedy heuristic algorithm"

The optimal solution of the U_b maximization problem (ILP problem) is limited by its high computational complexity. This limitation appears significantly in case of short T intervals (e.g $T = 1$ min) when the problem optimal solution is required within T time frame. A suboptimal solution using a greedy algorithm is used to overcome this limitation. Algorithm 1 presents the greedy heuristic implemented by each ABS in the network .

Algorithm 1 starts every time interval t by requesting inputs and initializing variables. For all expected DFPCs the U_{ijk} matrices and their corresponding P_{ijk} matrices are arranged in a decreasing order. For each DFPC, $U_{b,max}$ is calculated for all cell FPs by iterating through the U_{ijk} sorted matrices, incrementing U_{ijk} and P_{ijk}. Each iteration, the power constraints per ABS ($P_{b,max}$) and per FP ($P_{i,max}$) are checked for satisfaction. Finally, the algorithm returns the $U_{b,max}$ and its corresponding C_c for all DFPCs.

3.7.2.2 Scheme central phase "Utility maximization algorithm"

The optimal solution of the U_c maximization problem is of low complexity; however, for practical implementation purposes we recommend algorithm 2 to be used by the HNCE to calculate $U_{c,max}$.

Algorithm 3.2 Greedy Heuristic

```
for t < T do
    Inputs     : (ℂ);(M);(S);(N);(P_{b,max});(P_{i,max});
                 ([U_{ijk}]);([P_{ijk}]);(t);(T)
    Inialize   : ([U_{b,max}] = 0);([P_b] = 0);
                 ([U_c] = 0);([U_{c,T}] = 0);([U_{c,max}] = 0)
    Sort       : ([U_{ijk}]) /* Sort ([U_{ijk}]) in descending order */
    Sort       : ([P_{ijk}]) /* Sort ([P_{ijk}]) in corrospondance to ([U_{ijk}]) */
    for c ≤ ℂ do
        for i ≤ M do
            for j ≤ S do
                for k ≤ N do
                    while P_b ≤ P_{b,max} do
                        if P_{ijk} ≤ P_{i,max} then
                            State : U_{b,max} = U_{b,max} + U_{ijk}
                            State : P_b = P_b + P_{ijk}
                        end
                    end
                end
            end
        end
    end
    Return     : (U_{b,max}) & (C_c) /* (C_c) Corrosponding to ([U_{b,max}]) */
    State      : t = t + t
end
```

Algorithm 2 starts by requesting inputs and initializing variables. Every time interval T, U_c is used to calculate $U_{c,T}$ for all DFPCs in the network. $U_{c,T}$s for all DFPCs are compared and finally, the algorithm returns $U_{c,max}$ and the optimal DFPC to be implemented in the network.

3.7.3 Complexity analysis for suboptimal solution

Discussing the efficiency of algorithm 1 (greedy heuristic) is essential to prove its ability for practical implementation. The computational complexity of the greedy heuristic is analyzed by analyzing the worst case running time of the algorithm.

Initially, in each ABS, the algorithm sorts and labels all U_{ijk} for all DFPCs in a descending order. This step takes time $O(\mathbb{C}MSN\,log\,MSN)$ to sort the utility arrays

Algorithm 3.3 Central Utility Maximization

Inputs : $(B);([U_{b,max}]);(\mathbb{C});(t);(T)$
Inialize : $([U_c]=0);([U_{c,T}]=0);([U_{c,max}]=0)$
for t $< T$ **do**
  for $c \leq \mathbb{C}$ **do**
    for $b \leq B$ **do**
      State : $U_c = \sum_{b=1}^{B}(U_{b,max})/B$
    end
    State : $U_{c,T} = U_{c,T} + U_c$
  end
  State : t = t + t
end
Return : $[U_{c,T}]$ & $[C_c]$ /* (C_c) Corresponding to $([U_{c,T}])$ */
if t $= T$ **then**
  Compare : $[U_{c,T}]$
    $\forall(C_c : c = 1, 2, ..., \mathbb{C})$
  State : $U_{c,max} = max(U_{c,T})$
end
Return : $(U_{c,max})$ & (C_c) /* (C_c) Corresponding to $([U_{c,max}])$ */

and an additional time $O(\mathbb{C}MSN)$ to label the sorted arrays. $U_{b,max}$ is calculated by iterating through the AMSs sorted utility array incrementing U_{ijk} while checking for the violation of power constraints. This step incurs a maximum time $O(\mathbb{C}MSN)$ to get accomplished.

The worst case running time is calculated by summing up the times required to run the algorithm. Algorithm 1 needs a total $O(\mathbb{C}MSN\,log\,MSN)$ to calculate $U_{b,max}$ for all DFPCs. The computational complexity analysis for the worst case running time shows that Algorithm 1 is a polynomial-time algorithm.

Algorithm 2 incur a polynomial worst case running time $O(\mathbb{C}B\tau)$ to calculate $U_{c,max}$ that identifies the optimal DFPC to be used in the network. Finally, the analysis of the worst case running time for calculating $U_{b,max}$ in algorithm 1 and $U_{c,max}$ in algorithm 2 shows that both algorithms are polynomial-time suitable for practical implementation.

The computation time for running the greedy heuristic is measured in addition to the computational complexity to show the time frame needed to execute the algorithm. The computational time is calculated on a Linux machine (x86_64) with $Intel^{(R)}Xeon^{(R)}$ CPU (X5550 @ 2.67GHz) and 94.5 GB RAM. A number of 60 AMSs are allocated 96 PRUs in both FPs of a single ABS using configuration C_3 to show the allocation time needed by the greedy heuristic to perform the allocation process. The computational time for the allocation is calculated as (389 ms) under the aforementioned conditions.

3.8 Simulation and Results

3.8.1 Optimal solution

This section presents the simulation and results of the proposed optimal solution. We run the simulation to calculate $U_{c,max}$ and identify the optimal DFPC. Cplex® is used to solve the problem and to calculate the X_{ijk} that achieves $U_{b,max}$. After determining $U_{b,max}$ per DFPC in each ABS, a Java® program is developed to increment the values of $U_{b,max}$ generated by Cplex every t interval to calculate $U_{c,T}$ achieved by each DFPC.

Three different schemes aiming at maximizing different objectives are adopted in the simulation. First scheme denoted as S_1 represents our scheme that maximize the central utility (U_c), the second scheme denoted as S_2 only maximizes the normalized spectral efficiency (NSE), and the third scheme denoted as S_3 maximizes the achieved throughput ($\mathcal{R}$). Different DFPCs are used in the simulation,the results for a network that utilizes DFPCs from group 4 (C_2 to C_6) are presented and the parameters used in the simulation are described in Table 3.8.1.

A performance comparison between our proposed scheme S_1 and optimization schemes S_2 and S_3 is presented. The performance evaluation of S_1 against S_2 meant

Table 3.8.1: System parameters

Parameter	Value
Number of Cells (B)	7
Number of FPs (M)	4
Number of AMSs (S)	60
Number of PRUs (N)	96
Number of DFPCs (c)	5 ($C_2 - C_6$)
BER	10^{-6}
ABS Maximum Power ($P_{\max}$) $[w]$	10
SINR Threshold (δ_b) $[db]$	18.5
DFPC Update Interval (T) $[hr]$	1
Optimization Calculation Interval (t) $[Sec]$	20
Bandwidth ($\mathcal{W}_i$) $[MHZ]$ $\{(\mathcal{F}_0)/(\mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3)\}$	(20){(10)/(3.33)}

to clarify the difference between them in terms of NSE and power consumption due to the consideration of the power consumption in the formulation of S_1. In addition, the performance evaluation of S_3 is intended to show a comparison between our proposed scheme S_1 and rate adaptive resource allocation schemes such as those presented in [55, 56].

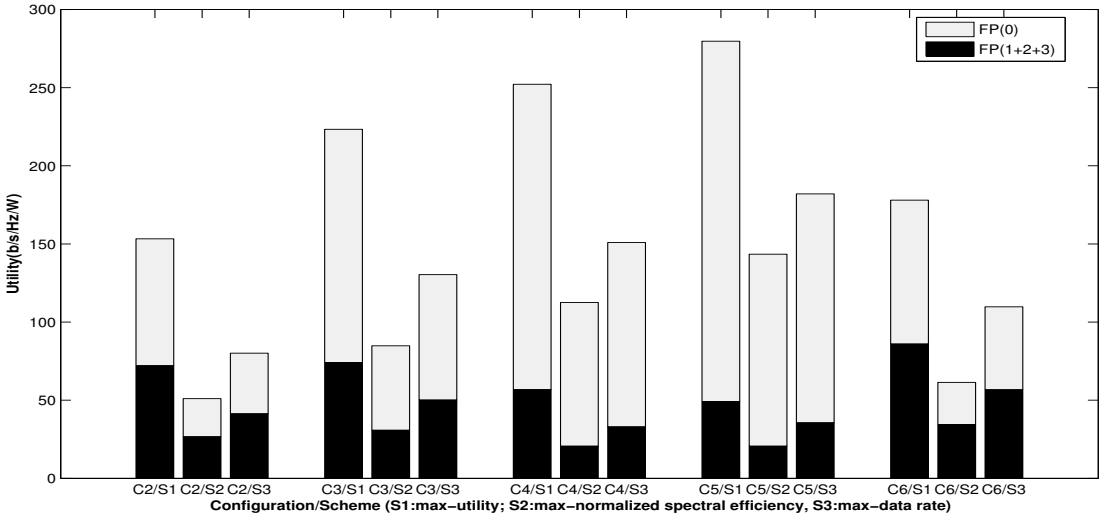

Figure 3.8.1: Comparison between U_c achieved by $S1$, $S2$, and $S3$ in cells center FP ($\mathcal{F}_0$) and cells edge FPs ($\mathcal{F}_1$,$\mathcal{F}_2$,$\mathcal{F}_3$)

The results in Fig. 3.8.1 show that $U_{c,max}$ in the network and in cells center FP ($\mathcal{F}_0$) are achieved by C_5/S_1. $U_{c,max}$ in cells edge FPs ($\mathcal{F}_1$, $\mathcal{F}_2$, $\mathcal{F}_3$) is achieved by

C_6/S_1. The previous finding indicates that, in case of high load condition in cells edge, using C_6/S_1 instead of C_5/S_1 can boost $U_{c,max}$ by using a DFPC that assigns more resources to the cells edge users. This result shows the effect of the network load conditions and users distribution on the dynamics of the chosen DFPC that achieves $U_{c,max}$.

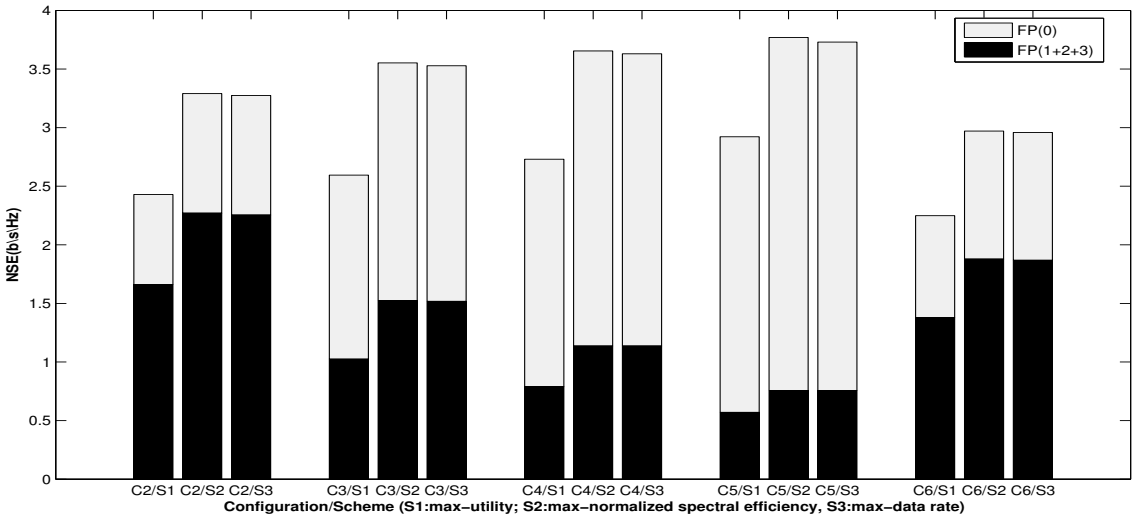

Figure 3.8.2: Comparison between $S1$, $S2$, and $S3$ maximum NSE in cells center FP ($\mathcal{F}_0$) and cells edge FPs ($\mathcal{F}_1$,$\mathcal{F}_2$,$\mathcal{F}_3$)

The results in Fig. 3.8.2 show a comparison between the NSE for group 4 DFPCs. The maximum network NSE is achieved by C_5/S_2. However, the achieved NSE by C_5/S_1 the DFPC with $U_{c,max}$ is lower than C_5/S_2 due to the influence of minimizing the power consumption considered by C_5/S_1 on maximizing the NSE. S_2 maximizes the NSE without considering the amount of power consumed by selecting users with maximum NSE without considering their power consumption requirements. The results show 1 (b/s/Hz) difference between NSE achieved by C_5/S_1 and C_5/S_2.

Fig. 3.8.2 shows that C_5/S_2 achieves the maximum NSE in cells center FP ($\mathcal{F}_0$) which is the same DFPC that achieve $U_{c,max}$. Moreover, the maximum NSE in the cell edge FPs ($\mathcal{F}_1$, $\mathcal{F}_2$, $\mathcal{F}_3$) is achieved by C_2/S_2 not C_3/S_2 the DFPC with $U_{c,max}$ in

the cells edge. This indicates that the consideration of minimizing the power consumption in the allocation process influences the DFPC identification decision.

The results in Fig. 3.8.3 shows that C_4/S_1 is the DFPC that consumes the minimum consumed power in both the network and in cells edge FPs ($\mathcal{F}_1$, $\mathcal{F}_2$, $\mathcal{F}_3$). Moreover, C_2/S_1 consumes the minimum consumed power in cells center FP ($\mathcal{F}_0$). Fig. 3.8.3 shows that C_5/S_1 the DFPC with $U_{c,max}$ consumes higher amount of power than C_4/S_1, yet C_5/S_1 shows significant difference in the minimum amount of power consumed compared to C_5/S_2 the DFPC with maximum NSE. The large amount of power consumed by C_5/S_2 is due to either high data rates achieved by cell center users or poor channel conditions exhibited by users in the cell edge.

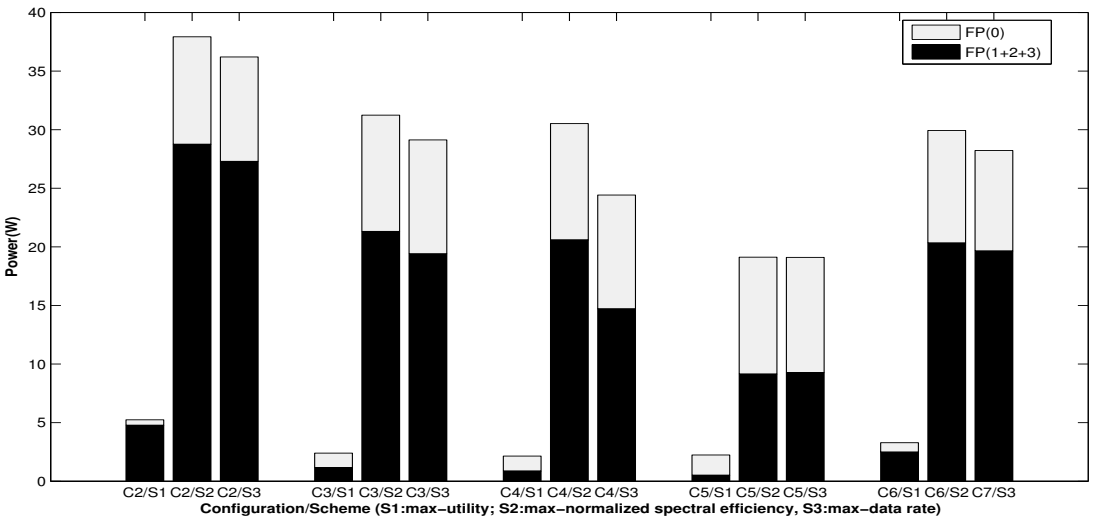

Figure 3.8.3: $S1$, $S2$, and $S3$ minimum power consumption comparison in cells center FP ($\mathcal{F}_0$) and cells edge FPs ($\mathcal{F}_1$,$\mathcal{F}_2$,$\mathcal{F}_3$)

The results in Fig. 3.8.4 show a comparison between the maximum achievable $\mathcal{R}$ for group 4 DFPCs. C_5/S_3 achieves maximum $\mathcal{R}$ in both the network and cells center FP ($\mathcal{F}_0$). Moreover, C_2/S_3 achieves the maximum $\mathcal{R}$ in the cells edge FPs ($\mathcal{F}_1$, $\mathcal{F}_2$, $\mathcal{F}_3$).

The achievable $\mathcal{R}$ of C_5/S_1 the DFPC with $U_{c,max}$ is less than C_5/S_3 due to the consideration of the power consumption that enforces S_1 to choose users with both

maximum $\mathcal{R}$ and minimum amount of power consumed. The reason for the increase in $U_{c,T}$ achieved by C_5/S_1 over other S_1 DFPCs (C_2, C_3, C_4, C_6) is due to the users good channel conditions in the cells center which cause an increase in their achievable $\mathcal{R}$ as shown in Fig. 3.8.4.

The results show that C_5/S_1 achieves $U_{c,max}$, C_5/S_2 achieves the maximum NSE, and C_5/S_3 achieves the maximum $\mathcal{R}$. These results indicate that although C_5 is the optimal DFPC to be used in the network but each allocation scheme utilizes C_5 generates a different ROI from the network.

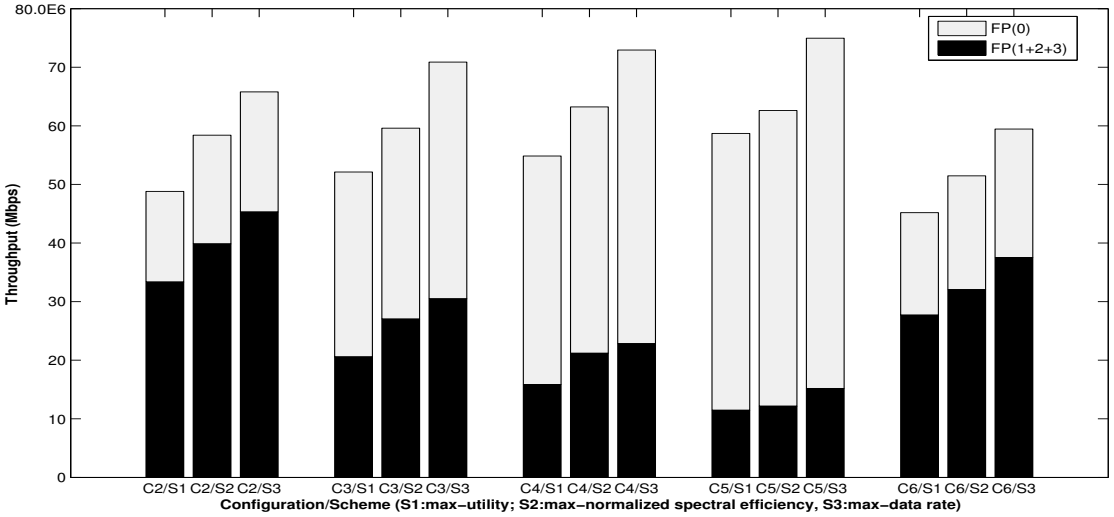

Figure 3.8.4: Comparison between $S1$, $S2$, and $S3$ maximum R in cells center FP $(\mathcal{F}_0)$ and cells edge FPs $(\mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3)$

Results in Fig. 3.8.5 are presented to show the effect of the proposed scheme S_1 on the system performance metrics (power and NSE). Results of the joint optimization using S_1 shows a trade-off between the power consumption and the spectral efficiency. This trade-off appears in the form of an increase in the amount of power consumed with the increase in the achieved NSE.

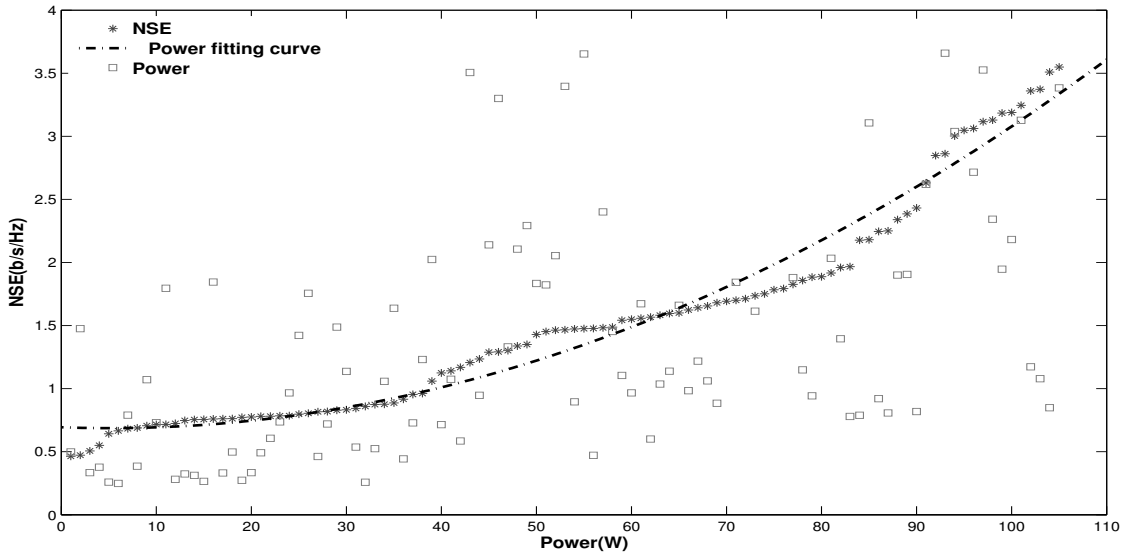

Figure 3.8.5: Trade-off between the achieved NSE and consumed power using scheme $S1$

3.8.2 Suboptimal solution

A simulation is conducted to evaluate the efficiency of the suboptimal solution. Fig. 3.8.6 shows a comparison between the maximum U_c for S_1 and the greedy heuristic denoted by S_4. The results show that the heuristic succeeded to identify the same DFPC (C_5) as the optimal DFPC. The results also show an approximate gap of 4% between S_1 and S_4.

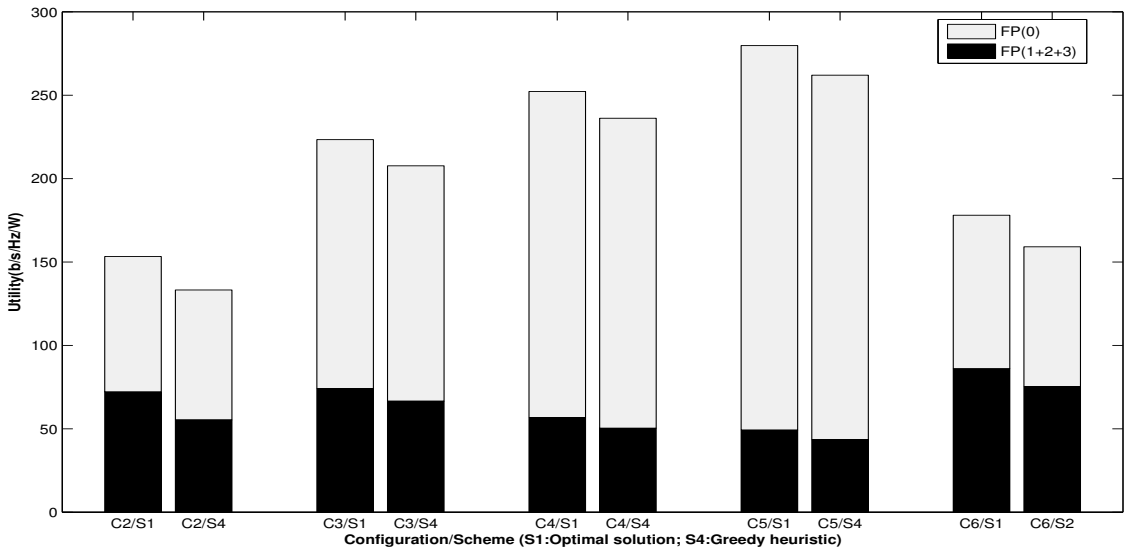

Figure 3.8.6: S_1 versus S_4: U_c maximization comparison in cells center FP ($\mathcal{F}_0$) and cells edge FPs ($\mathcal{F}_1$,$\mathcal{F}_2$,$\mathcal{F}_3$)

3.9 Conclusion

In this study, a radio resource management (RRM) model that utilizes adaptive fractional frequency reuse mechanism in WiMAX is proposed to identify the optimal down-link frequency partition configuration (DFPC) to be implemented in the network. The objective of the model is to improve the mobile service providers MSPs' return on investment by maximizing the normalized spectral efficiency (NSE) and minimizing the total power consumption in the network.

To solve the problem, we adopted a utility based optimization technique in which we proposed a central and distributed utility functions calculated by a higher network control entity and the advanced base stations (ABS) respectively. A joint optimization problem is formulated to calculate the distributed utility in each ABS and a utility maximization problem is formulated to calculate the maximum central utility in the network. The problem solution succeeded to calculate the maximum central utility and to identify the optimal DFPC to be utilized. The optimal solution though shows a high computational complexity. Thus, for practical implementation purposes

a suboptimal heuristic called the greedy heuristic is additionally proposed to solve the problem. A simulation for the proposed model shows accepted results and a comparison between the optimal and the suboptimal solutions is conducted to evaluate the suboptimal solution efficiency and investigate the gap with the optimal solution. Finally, this study proposes a RRM model that succeeded to maximize the NSE and to minimize the power consumed and we recommend our model for implementation by MSPs.

3.10 References

[51] IEEE, *IEEE Standard for Local and metropolitan area networks Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems*, May 2009.

[52] 3GPP, *TR 25.814, Physical Layer Aspects for Evolved UTRA.*, 2009.

[53] B. Harpham, "IEEE 802.16m approved as IMT-Advanced Technology," WiMAX Forum, October 2010, http://www.wimaxforum.org/news/2650.

[54] IEEE, *IEEE Standard for Local and metropolitan area networks Part 16: Air Interface for Broadband Wireless Access Systems Amendment 3: Advanced Air Interface*, 2011.

[55] S. Ali and V. Leung, "Dynamic frequency allocation in fractional frequency reused OFDMA networks," *IEEE Transactions on Wireless Communications*, vol. 8, no. 8, pp. 4286–4295, 2009.

[56] J. Kim and W. S. Jeon, "Two practical resource allocation techniques for fractional frequency reuse in IEEE 802.16m networks," in *7th International Wireless Communications and Mobile Computing Conference (IWCMC), 2011*, july 2011, pp. 261 –265.

[57] Z. Mohades, V. T. Vakili, S. M. Razavizadeh, and D. Abbasi-Moghadam, "Dynamic Fractional Frequency Reuse (DFFR) with AMC and Random Access in WiMAX System," *Wireless Personal Communications*, vol. 68, no. 4, pp. 1871–1881, Feb 2013.

[58] T. Novlan, R. Ganti, A. Ghosh, and J. Andrews, "Analytical evaluation of fractional frequency reuse for OFDMA cellular networks," *IEEE Transactions on Wireless Communications*, no. 99, pp. 1–12, 2011.

[59] R. Chang, Z. Tao, J. Zhang, and C. Kuo, "A graph approach to dynamic fractional frequency reuse (FFR) in multi-cell OFDMA networks," in *IEEE International Conference on Communications, ICC'09*, 2009, pp. 1–6.

[60] R. Y. Chang, Z. Tao, J. Zhang, and C.-C. J. Kuo, "Dynamic fractional frequency reuse (D-FFR) for multicell OFDMA networks using a graph framework," *Wireless Communications and Mobile Computing*, 2011.

[61] K. T. Kim and S. K. Oh, "A universal frequency reuse system in a mobile cellular environment," in *IEEE 65th Vehicular Technology Conference*, april 2007, pp. 2855 –2859.

[62] N. Himayat, S. Talwar, A. Rao, and R. Soni, "Interference management for 4g cellular standards [wimax/lte update]," *Communications Magazine, IEEE*, vol. 48, no. 8, pp. 86 –92, august 2010.

[63] G. Miao, N. Himayat, G. Li, A. Koc, and S. Talwar, "Interference-Aware Energy-Efficient Power Optimization," in *IEEE International Conference on Communications, 2009. ICC '09.*, June 2009, pp. 1–5.

[64] H. Lei, L. Zhang, X. Zhang, and D. Yang, "A novel multi-cell OFDMA system structure using fractional frequency reuse," in *IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications*, sept. 2007, pp. 1 –5.

[65] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004.

CHAPTER 4. DOWN-LINK SPECTRUM ALLOCATION IN 5G HETNETS

An Invited paper published in *The proceeding of the 2014 International Wireless Communication & Mobile Computing Conference (IWCMC' 2014)*

Tamer R. Omar[8] [9], Ahmed E. Kamal[10], J. Morris Chang[11]

4.1 Abstract

Fifth generation mobile systems (5G) target an Average Area Spectral Efficiency (AASE) over hundred $Gbps/km^2/user$ for future mobile systems with an Energy Dissipation (ED) per unit area similar to the current ED levels. Heterogeneous networks (HetNets) with high density of deployed small cells are currently adopted to aid in achieving the target ED and AASE by 5G. Limited spectrum availability requires efforts to manage the spectrum utilization in such dense deployments. Development of new network architectures and Radio Resources Management (RRM) schemes is important to address such challenges. The objective of this work is to propose a new architecture that consists of a Decision Support System (DSS) and a data collection

[8]Graduate student, Professor and Associate Professor, respectively, Department of Electrical and Computer Engineering, Iowa State University.
[9]Primary researcher and author.
[10]Author for correspondence.
[11]Author for correspondence.

system to dynamically manage and control the spectrum allocation process. The DSS generates spectrum allocation patterns using non-parametric estimation and statistical analysis for the collected data. A new RRM model using a Plan, Do, Control and Act (PDCA) cycle is proposed as a new self optimization module in the self organizing network framework. The PDCA model utilizes the new architecture and the allocation patterns to **dynamic**ally predict future spectrum allocation. Results show improvement in the AASE achieved using the PDCA model compared to conventional spectrum allocation.

Keywords: 5G, HetNets, Dynamic Spectrum Management, AASE.

4.2 Introduction

The Cisco visual networking index forecasts that the dependence on mobile phone and data services by mobile subscribers will increase 10-15 fold between 2012 and 2017 [66]. Major mobile service providers (MSPs), standardization bodies and different forums are exerting maximum efforts to find novel solutions that offer users high quality services which meet their expectations. The mobile and wireless communications enablers for the twenty-twenty information society (METIS) is leading the first international and large-scale research activity on 5G [67]. Efforts on 5G are supposed to produce a new standard that will replace the current 4G technologies into the next decade.

Coordinated multi-point (COMP) transmission/reception, carrier aggregation, MSPs network (e.g. spectrum) sharing, advanced multiple-input multiple-output techniques, relays and enhanced inter-cell interference coordination are the enabling technologies for 4G and act as the baseline for 5G. The authors in [68] provide an overview for these enabling technologies from the physical layer aspects while [69] explains the possible improvements and the associated challenges for those enabling technologies.

Rate adaptive, margin adaptive, and utility based dynamic spectrum allocation (DSAn) and physical resources allocation (PRAn) schemes are suggested in previous work to improve spectrum utilization [70]. However, Heterogeneous Networks (HetNets) which consist of conventional macro cells overlaid by small cells are currently proposed as the most feasible method to improve spectrum utilization specially that spectral efficiency has reached its limit. DSAn and PRAn are implemented using rate adaptive and COMP schemes in [71, 72] to maximize the system throughput. Geometric programming is used in [71] to optimize users minimum rate under variable bandwidth and variable power. Resource partitioning is proposed in [72] to improve the performance of HetNets using a two cell selection criterion, one based on maximizing the received signal strength with a potential bias and the other based on maximizing the product of signal to interference plus noise ratio (SINR) and bias.

Self organizing networks (SON) for HetNets are discussed in [73] with a main goal of reducing capital and operational expenditure that increase due to the increased number of network parameters to be monitored. Also other different autonomic functionalities like auto configuration, self optimization, diagnosis, and self healing are evaluated.

Traffic Geo-Location service is the service of identifying the real-time geographic location and the traffic load of user equipments (UEs). The benefits and technical issues of traffic Geo-Location service is explored in [74]. Geo-Location service is discussed as a tool that aid MSPs to deliver an optimized environment to deploy macro cells (MCs) to small cells (SCs) off loading solutions.

According to the best of our knowledge none of the previous approaches presents a module in the SON framework to operate as a decision support system (DSS) that helps automate the optimization process of the PRAn and DSAn. Thus, this paper develops a new architecture and a supporting SON module to optimally distribute and utilize the available resources and frequency spectrum in order to maximize the

benefit of the available limited spectrum. A traffic Geo-location resources allocation database (Geo-LRADB) is proposed to be populated using Geo-Location service to store the geographical information of the users allocated resource blocks (RBs).

The goal of this paper is to address the DSAn and PRAn problem using a new allocation approach

1. A new four step Radio Resource Management (RRM) model, referred to as the PDCA (Plan-Do-Check-Act), is proposed. The model aims at identifying the optimal DSAn and PRAn in terms of MSPs return on investment (ROI) by enhancing the average area spectral efficiency in HetNets.

2. A new network architecture is proposed that consists of two components

   - A data collection system (DCS) infrastructure that includes a Geo-LRADB is introduced to the distributed network control and management system (NCMS) in the network architecture whose purpose is to aid as a traffic data repository.
   - A DSS that applies the PDCA model is introduced as a new self optimization module within the SON framework. The model represents the core functionality of the DSS and utilizes the data collected by the distributed NCMS to predict and recommend the optimal DSAn to be implemented in the network.

3. A non-parametric estimation analysis and a stochastic linear programming with recourse is formulated and adopted to enhance the spectrum allocation by utilizing the previous history for predicting future allocations.

The remainder of this paper is organized as follows. Network and system model is introduced in Section II. Section III provides details about the RRM mechanism in HetNets. The RRM module proposed for DSAn self-optimization in the SON

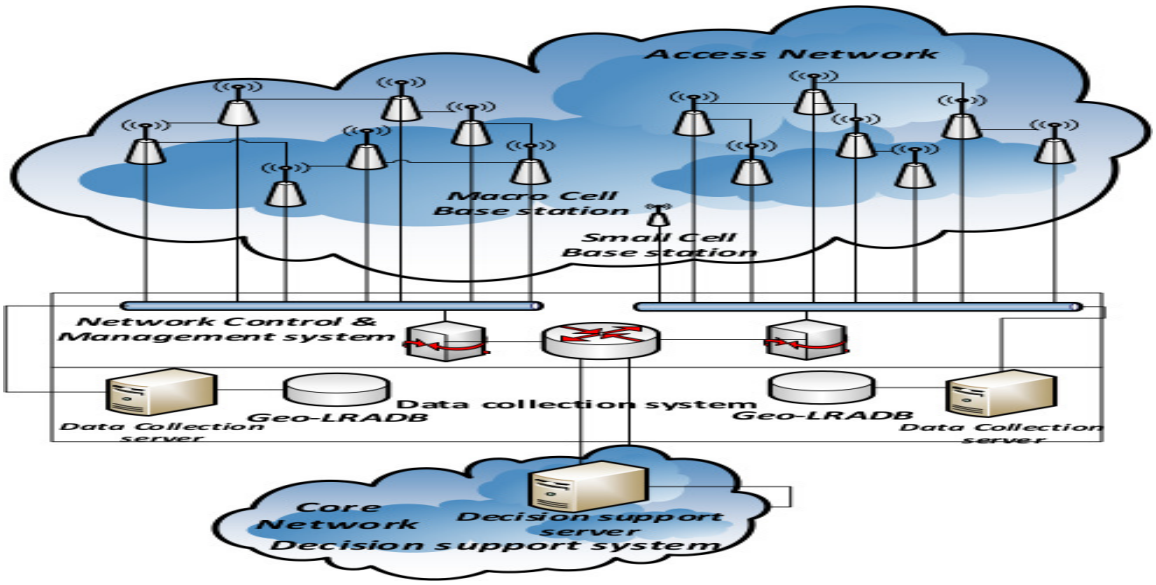

Figure 4.3.1: Network architecture

framework is presented and the optimization problem is formulated in Section IV. Simulation results are presented and discussed in Section V. Finally, in Section VI conclusions are drawn.

4.3 Network and System Model

This section introduces the proposed HetNet architecture, system model, and the power allocation model.

4.3.1 Network architecture

The proposed architecture adds a DSS and a DCS to the current network architecture as shown in Fig. 4.3.1.

- A core network (CN) central DSS consists of a decision support server (DSSrvr) and DSAn optimization methodology is proposed. The DSS is recommended for implementation as a radio resources self optimization module in the SON

frame work. The DSS is responsible for globally planning and controlling the DSAn in the access network (AN).

- A DCS contains a group of data collection servers (DCSrvrs) and a Geo-LRADBs replicated in the NCMS and distributed over the AN to collect all necessary data required by the radio resources allocation process.

The data collected by the DCS is analyzed by the DSS to centrally manage the dynamic spectrum allocation self optimization in the HetNet, as described below :

4.3.1.1 Decision support system

The DSS is a central management system using a DSSrvr to perform the needed functions. The DSS, in addition to the dynamic control and management of the DSAn self optimization, can be used in optimally controlling and managing multiple parameters in the AN, and recommending deployment plans and strategies to the NCMS. Examples of the tasks that can be managed by the DSS are ;

- Dynamically controlling the SINR threshold between different frequency partitions (FPs).
- Updating COMP deployment plan for Macro Cell-Base Stations (MC-BSs) and the deployment plan of the small cell-base stations (SC-BSs) inside the MCs.
- Dynamically changing the MC-BS beam-forming strategies.
- Managing the adaptive frequency reuse power patterns,
- Performing resource allocation trending analysis.

4.3.1.2 Data collection system

The DCSrvrs and a Geo-LRDBs proposed in the DCS within the NCMS perform several tasks;

- Data collection: the real time physical resource allocation in the AN is collected.
- Data aging: the old data that is no longer needed is aged and deleted according to the aging rules.
- System maintenance: the Geo-LRADB relational database management system is maintained.

4.3.2 System model

We consider the down-link transmission in a multi-cell HetNet. Fig. 4.3.2 shows an example network which consists of seven identical adjacent hexagonal MC-BSs. Five external SC-BSs ($ESCs$) are deployed outside of the AN coverage to extend coverage. Three internal SC-BSs ($ISCs$) are used inside of the AN coverage for capacity increase. The maximum number of FPs in each MC in the AN is M and each FP is denoted by $\mathcal{F}_{mc,i}$, where $i = \{1,2,...,M\}$.

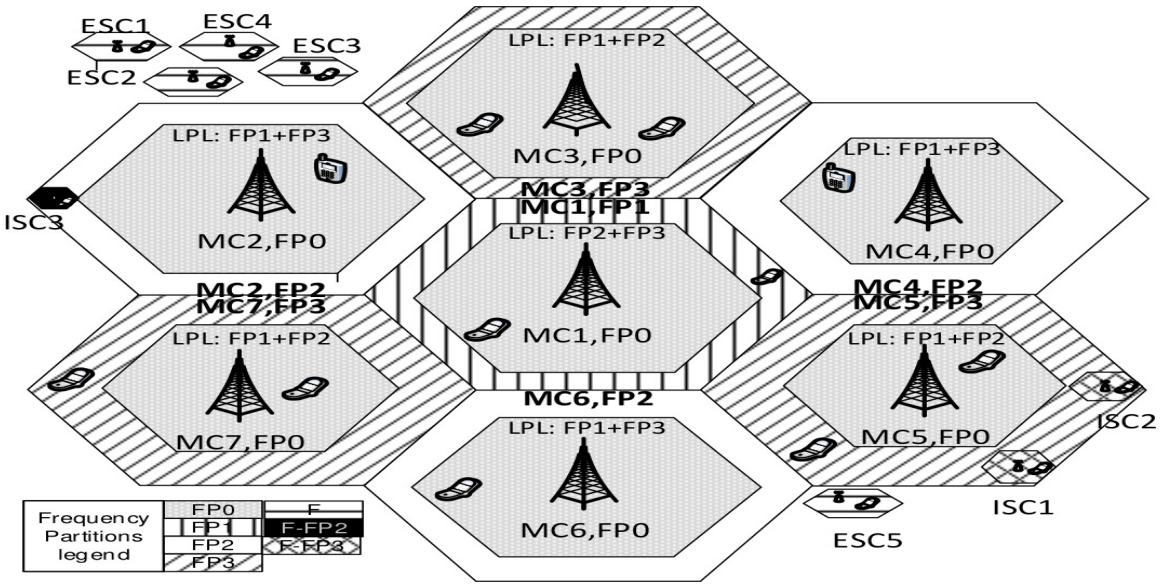

Figure 4.3.2: System model

4.3.3 Power allocation model

4.3.3.1 Power patterns

The idea of using power patterns is introduced by WiMAX [75] and can be extended to 5G networks to aid as a transmission power adaptation technique. Power patterns are used to specify the maximum allowed transmission power level in each FP. Different patterns have different maximum power levels allocated according to the area served by the FP (cell center FP, cell edge FP or SC FP). Fig. 4.3.3 shows the different power patterns used by the MC-BSs, the max power level used in each FP in the MC and the power patterns used by the SC-BSs for our system model that partition the utilized spectrum using reuse-3 partitioning as shown in Fig. 4.3.2: For example MC_1 uses pattern 1 where P_0 is used in the cell center for FP $\mathcal{F}_0$, and Low Power Level (LPL) P_2 is used for FP $(\mathcal{F}_2, \mathcal{F}_3)$ to avoid Inter-Cell Interference (ICI). P_1 is used in the cell edge for FP $\mathcal{F}_1$ to enhance the spectrum utilization by improving the UEs transmission rates. $MC_{2,4,6}$ use pattern 2, and $MC_{3,5,7}$ use pattern 3 to specify the maximum power levels in different FPs. $ESC_{1,2,3,4,5}$ use pattern 4, $ISC_{1,2}$ implemented in MC_5 edge use pattern 5 where UEs use LPL P_2 in all the spectrum $\mathcal{F}$ except in FP $\mathcal{F}_3$ to avoid interference with UEs in MC_5 edge. ISC_3 use pattern 6 to specify the maximum power level in the SCs.

4.3.3.2 Power levels

Fig. 4.3.3 maps the adaptive power allocation used by MC-BSs in the different FPs and the SC-BSs of the system model shown in Fig. 4.3.2. FP $\mathcal{F}_0$ is reused in the center of each MC with a fixed maximum power level $P_{0,max}$. FPs ($\mathcal{F}_1$, $\mathcal{F}_2$, $\mathcal{F}_3$) are reused with a maximum power boosted level $P_{1,max}$ in the MC edges and a maximum LPL $P_{2,max}$ in the cell center. All spectrum $\mathcal{F}$ is reused in $ESC_{1,2,3,4,5}$ with a maximum LPL $P_{3,max}$. FP $(\mathcal{F} - \mathcal{F}_3)$ is reused in SCs $ISC_{1,2}$ and FP $(\mathcal{F} - \mathcal{F}_2)$ is

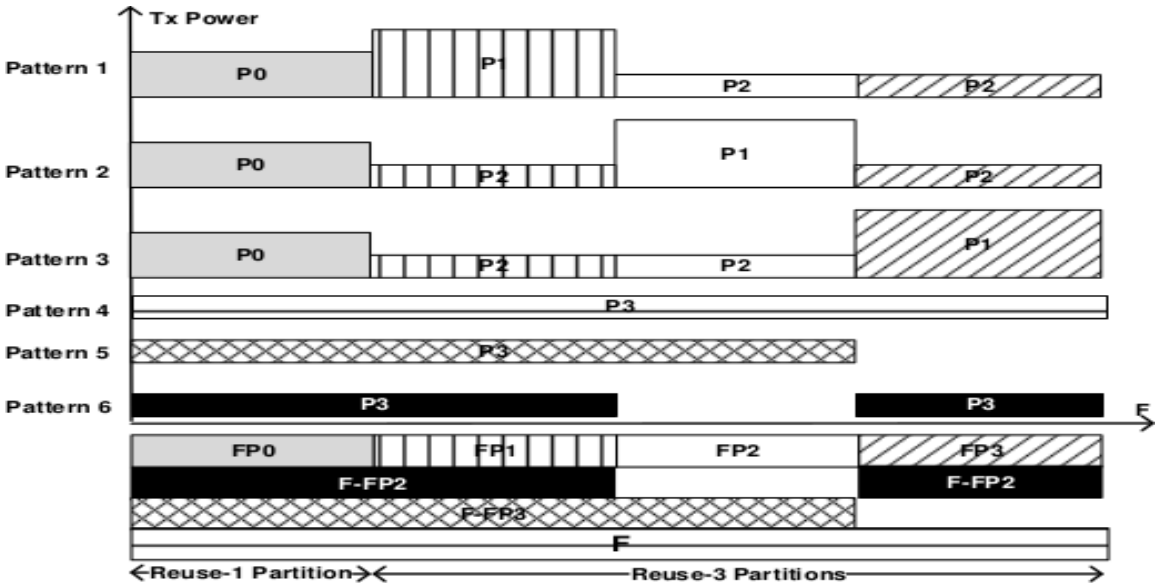

Figure 4.3.3: Power Patterns

used in SC ISC_3 with a LPL $P_{3,max}$.

The utilized Adaptive Fractional Frequency Reuse (AFFR) mechanism decides the maximum power levels and power patterns to be used in the AN. Adaptive power allocation improves the energy dissipation by adapting the UE transmission power levels according to their locations. Thus, the MC-BS and SC-BS use specific power patterns to aid in mitigating ICI between different FPs in the AN.

4.4 Radio Resource Management

DSAn and PRAn are two extremely important RRM functions that allocate the amount of spectrum to each FP and the amount of RBs to each UE, respectively. This section introduces the down-link frequency partition configuration (DFPC) basics, frequency allocation hierarchy in a HetNet, the DSAn utility based optimization problem formulation and the Geo-LRADB structure.

4.4.1 Down-link frequency partition configuration

AFFR is a promising mechanism adopted by 4G technology to enhance the spectral efficiency in the AN by controlling the frequency partition size (FPS), frequency partition count (FPCT), and UEs maximum transmission power limits.

The maximum number of utilized DFPCs in the AN is $\mathbb{C}$ and each DFPC is denoted by C_c , where $c \epsilon \{1, 2, ..., \mathbb{C}\}$. The maximum number of utilized DFPCs is determined according to the utilized bandwidth ($\mathcal{W}$) and the employed Fast Fourier Transform (FFT) in the AN.

A list of the DFPCs that can be dynamically implemented in the AN according to the user requirements and the network load conditions are presented in [75]. Each DFPC specifies the number of FPs in terms of their FPS and FPCT.

4.4.2 Dynamic spectrum allocation

DSAn is calculated centrally by the DSS and implemented globally in the AN to control the distribution of the RBs between MC-BSs cell regions and SC-BSs existing in a HetNet. DSAn aims at performing the optimal spectrum allocation by identifying the best DFPC to be used in the AN.

4.4.2.1 Distributed utility formulation

A number of utility based frameworks for scheduling RBs to UEs in the AN are proposed in previous work. Allocation schemes are presented to calculate the allocation matrices of scheduling the RB k to UE j for each FP i in MC-BS b in order to optimize a utility function U_{ijk}. The distributed maximum utility denoted by $U_{b,max}$ is calculated by maximizing the summation of the aforementioned utility ($max \sum_i \sum_j \sum_k U_{ijk} x_{ijk}$) for all UEs in each MC.

4.4.2.2 Central utility formulation

The proposed central utility function is used to identify the optimal DFPC. The central utility function is denoted by $U_{c,max}$ and is calculated globally by the DSS. $U_{c,max}$ is the sum of $U_{b,max}$ achieved by all MC-BSs and SC-BSs in the AN. The DSS calculates $U_{c,max}$ per configuration C_c as $U_{c,max} = \sum_{b=1}^{B} U_{b,max}$ and uses it in identifying the DFPC to be implemented in the AN.

4.4.2.3 DSAn problem formulation

The DCS stores the real time allocation parameters used to achieve $U_{b,max}$ in the AN at the Geo-LRADB. The expected central utility for all DFPCs are calculated by the DSS periodically each time interval (T) using $U_{b,max}$, where T is the DFPC update interval. The maximum central utility $U_{c,max,Actual}$ that is achieved due to the implementation of a specific DFPC in the AN during a T interval is denoted as U_{DSS}. The calculation of $U_{c,max,Calculated}$ for all DFPCs requires access to a significant amount of information. One way to reduce the amount of information is by limiting the calculations of $U_{c,max,Calculated}$ to only the DFPCs that can possibly be implemented in the AN. An example of the possible DFPCs is the group of DFPCs with a same FPCT.

The DCS calculates the UEs' expected achievable data rates $r_{ijk,calculated}$ for all DFPCs. The UEs' expected SINR values $SINR_{ijk,calculated}$ for all DFPCs are calculated using the reported values, $SINR_{ijk,Actual}$, by the UEs. In order to calculate the UEs expected utility $U_{ijk,calculated}$, DCS calculates the UEs expected normalized spectral efficiency $NSE_{ijk,calculated}$ as

$$NSE_{ijk} = \frac{r_{ijk}(1 - PER)}{RM(i) * \mathcal{W}_i}$$

where PER and $RM(i)$ are the packet error rate and the resource metric, respectively.

The DCS also calculates the expected power consumption $P_{ijk,calculated}$ is calculated for all DFPCs as

$$P_{ijk} = A * \frac{2^{r_{ijk}/\mathcal{W}_j}}{\rho * SINR_{ijk}}$$

where A is the area calculated from the Geo-location information collected in the Geo-LRADB, $\rho = \frac{-1.5}{ln(5*BER)}$ is constant and BER is bit error rate. After calculating the expected utility $U_{ijk,calculated}$ as

$$U_{ijk,calculated} = \frac{NSE_{ijk,calculated}}{P_{ijk,calculated}}$$

the DCS calculates the expected maximum distributed utility $U_{b,max,calculated}$ for all current group DFPCs.

The optimal DFPC is identified by aggregating $U_{b,max,calculated}$ for all MC-BSs and SC-BSs per DPFC and comparing them to choose the DFPC with $U_{c,max,calculated}$. The DSS aided by the information collected in the DCS calculates the $U_{c,max,calculated}$ achieved during the most recent T interval by solving the problem below:

$$max \quad \sum_{c=1}^{C} U_{c,max,Calculated} x_c \tag{4.4.1}$$

$$s.t \quad \sum_{c=1}^{\mathbb{C}} x_c = 1 \tag{4.4.2}$$

$$x_c = \{0, 1\} \tag{4.4.3}$$

where X_c is the assignment vector and the binary variable $x_c \in X_c$ is the assignment indicator. $x_c = 1$, if C_c is the assigned DFPC, and $x_c = 0$ otherwise. In this trivial optimization problem, every T interval the DSS compares the $U_{c,max,Calculated}$ for all DFPCs to select the DFPC with the maximum $U_{c,max,calculated}$ for implementation in the next T interval.

4.4.3 Geo-location resource allocation database

The Geo-LRADB is a relational database that we proposed in this paper as a repository for the data needed in the optimization of frequency spectrum and the identification of the DFPC by the DSS.

A relational database model is updated in the DCS to collect the RRM data from the MC-BS. The collected data is used to calculate the expected distributed and central utilities in future T intervals.

4.5 Radio Resource Management in Self Organizing Networks

A new RRM model is proposed in this section to act as the core methodology used by the DSS for managing and controlling the radio resources in the network. The new RRM model denoted as the PDCA model is suggested for continuous improvement of the spectrum dynamic allocation. The DSS is recommended to be added as a self optimization module in the SON framework to automate the dynamic spectrum allocation process. A spectrum allocation methodology that relies on the statistical analysis of the collected data and a problem formulation using stochastic programming with recourse is proposed.

4.5.1 Plan-Do-Check-Act model

The PDCA model virtually calculates the DSAn and PRAn allocation for all C_c to recommend the DFPC changes in the AN that achieve $U_{c,max}$. The PDCA model learning cycle has four repeatable steps; 1) Plan, 2) Do, 3) Check, and 4) Act. The functions of each of the four steps are:

- Plan: Understand the gap between the real time central utility $U_{c,max,Actual}$ calculated due to the current DFPC and the expected $U_{c,max,Calculated}$ calculated for other DFPCs.

- Do: Using the SON resources allocation self optimization module, implement the optimal DFPC to be used in the AN according to the data collected in the Act phase.

- Check: Observe the change on the achieved distributed utility $U_{b,max,Actual}$ after applying the updated DFPC .

- Act: Collect the new data by the DCS in the Geo-LRADB and communicate it to the DSS.

4.5.2 Dynamic spectrum allocation

The analysis of the DSAn dynamics during the PDCA cycle is important to build allocation patterns which are used in a stochastic program to estimate the DFPC to be implemented in future intervals

4.5.2.1 Patterns analysis

A transition probability matrix (stochastic matrix) is used to obtain allocation patterns from the data collected at $Table_3$ and $Table_4$ in the Geo-LRADB. The utilization of the stochastic matrix helps in building a DSAn pattern based on the empirical data collected in the DCSrvrs and analyzed by the DSS module implemented in the SON framework. This allocation pattern can be used for future application of the DFPC to be implemented in HetNets.

The continuous variations in the network load encountered due to changes in users' distributions and traffic requirements in each cell has an important effect on the DSAn process. The DFPC needs to change periodically every DFPC update interval (T)

in order to address variations in the achieved utility $U_{c,max,Actual}$ due to network load dynamics.

Using empirical data, we construct the time homogeneous transition probability matrix, P^{T_{n-1},T_n} of size ($\mathbb{C}x\mathbb{C}$), with elements $p_{ij}^{T_{n-1},T_n}$ expressing the transition from DFPC i at current allocation time T_{n-1} to DFPC j at future allocation time T_n defined as

$$p_{ij}^{T_{n-1},T_n} = p\{U_j^{T_n} \text{ is optimal at time n } | U_i^{T_{n-1}} \text{ was optimal at time n-1}) \tag{4.5.1}$$

for any c_i which is the DFPC achieving the optimal central utility $U_{c,max,Actual}$ at allocation time T_{n-1} and c_j which is the DFPC achieving the maximum central utility $U_{c,max,Calculated}$ at allocation time T_n. DFPC is always determined after finding the UEs locations from the Geo-location database. The transition probability matrix is recalculated every allocation time to consider the last transition in calculating the probability matrix. A non-parametric estimate (histogram model) is used to present the distribution of the empirical data utilized in calculating the probability distribution $p\{U_j^{T_n}|U_i^{T_{n-1}})$.

4.5.2.2 Problem formulation

A stochastic integer linear program with recourse is used to calculate the maximum utility $U_{c,max,Calculated}$. At stage n where ($n \geq 1$) the values of $p_{ij}^{T_{n-1},T_n}$ of random variable ξ in the stochastic matrix P^{T_{n-1},T_n} as well as the previous actual utilities $(U_{c,max,Actual}^{T_0}, ..., U_{c,max,Actual}^{T_t}, ..., U_{c,max,Actual}^{T_{n-1}})$ can be calculated using the DSS historical data. The recourse in the stochastic program is used in order to represent the expected deficiency resulting in network utilization U_{DSS} at time T_n due to the choice of a a certain DFPC according to the calculated utility $U_{c,max,Calculated}$. This penalty is enforced to account for the DSS learning from previous decisions in identifying the

Table 4.5.1: Model variables

T_n	Allocation Time
$U_{c,max,calc}$	Max utility calculated by the DSS
$U_{c,max,Act}$	Actual utility collected by the DCS
$p_{ij}^{T_{n-1},T_n}$	Transition probability matrix
$[y_j]^+$	Penalty due to previous allocation
$x_c^{T_n}$	Allocation Matrix

future DFPC recommendations. The penalty incurred is determined after the observation of the empirical data and calculating the possible difference in the calculated central utility $U_{c,max,Calculated}$, which makes the x_c, a first stage decision variable.

The stochastic integer linear program with recourse costs at time T_n is modeled as

$$max \sum_{c=1}^{\mathbb{C}} U_{c,max,Calculated}^{T_n} x_c^{T_n} - \sum_{j=1}^{\mathbb{C}} \sum_{c=1}^{\mathbb{C}} P_{ij}^{T_{n-1},T_n} [y_j]^+ x_c^{T_n} \tag{4.5.2}$$

$$\sum_{c=1}^{\mathbb{C}} x_c^{T_n} = 1 \tag{4.5.3}$$

$$x_c^{T_n} = \{0, 1\} \tag{4.5.4}$$

equation (4.5.5) evaluates the penalty $[y_j]^+$as

$$[y_j]^+ s,t \quad \begin{cases} y_j = (U_{j,max,Calc}^{T_{n-1}} - U_{c,max,Act}^{T_{n-1}}) & \forall\, y_j \geqslant 0 \\ y_j = 0 & otherwise \end{cases} \tag{4.5.5}$$

Table 4.5.1 shows the key variables of the model. The solution of (4.5.2) to (4.5.4) identifies the DFPC that achieves the expected maximum utility in the AN in the next T interval.

4.6 Simulation and Results

This section presents a comparison between the two DSAn schemes proposed, first scheme using the equations (4.4.1)-(4.4.3) of section III and second scheme using (4.5.2)-(4.5.4) of section IV. We used this schemes to calculate $U_{c,max,Calculated}$ and identify the optimal DFPC in the AN.

The first scheme is a memory-less scheme using integer linear programing and the second scheme is memory aided using stochastic programming with recourse to identify the optimal DFPC to be implemented in the AN and the schemes are denoted by S_1 and S_2 respectively. Simulation for DFPCs with four FPs (C_3 to C_7) is conducted and results are presented using simulation parameters in Table 4.6.1.

Table 4.6.1: System parameters

Parameter	Value
Number of Cells (B)	7(MC) - 5(ESC) - 3(ISC)
Number of FPs (M)	4
Number of UEs in MC	60
Number of UEs in each SC	20
Number of RBs	96
Number of DFPC update interval (T)	360
Number of DFPCs (c)	5 ($C_3 - C_7$)
MC-BS Maximum Power ($P_{\max}$) $[w]$	25
SINR Threshold (δ_b) $[db]$	18.5
Bandwidth ($\mathcal{W}_i$) $[MHZ]$	20

A comparison between the DFPC allocated by scheme S_1 and scheme S_2 is presented in Fig. 4.6.1. The results show that S_1 and S_2 have different allocation decisions, the decisions taken by S_2 are affected by the history of the previous allocations collected in the DCSrvrs.

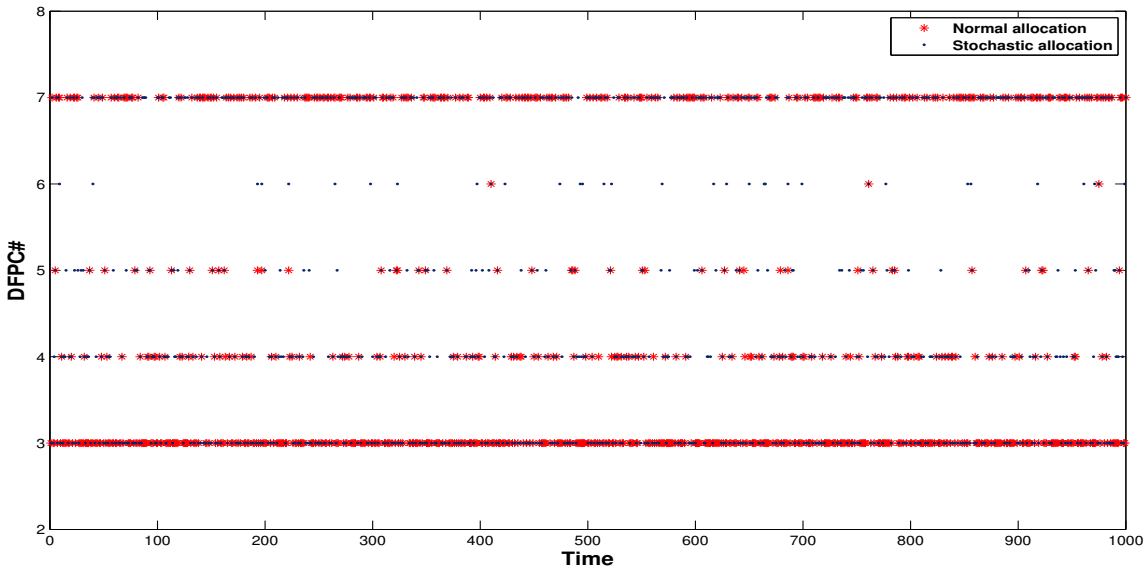

Figure 4.6.1: DFPC: Comparison between $S1$ and $S2$.

The network operation is simulated by conducting two consecutive simulation runs. In the first run the DFPC transition probability matrix, P^{T_{n-1},T_n} is created using the normal allocation model S_1 for a sufficient number of iterations to capture the UEs dynamics in the network. In the second run the DFPC transition probability matrix is used in the stochastic allocation model S_2 to calculate the penalty incurred due to previous allocations. The expected DFPCs are then calculated by solving the optimization problems for both models. The expected DFPCs recommended by the DSS are used to operate the network and the central utility $U_{c,max,actual}$ is measured each DFPC update interval to evaluate the network utilization.

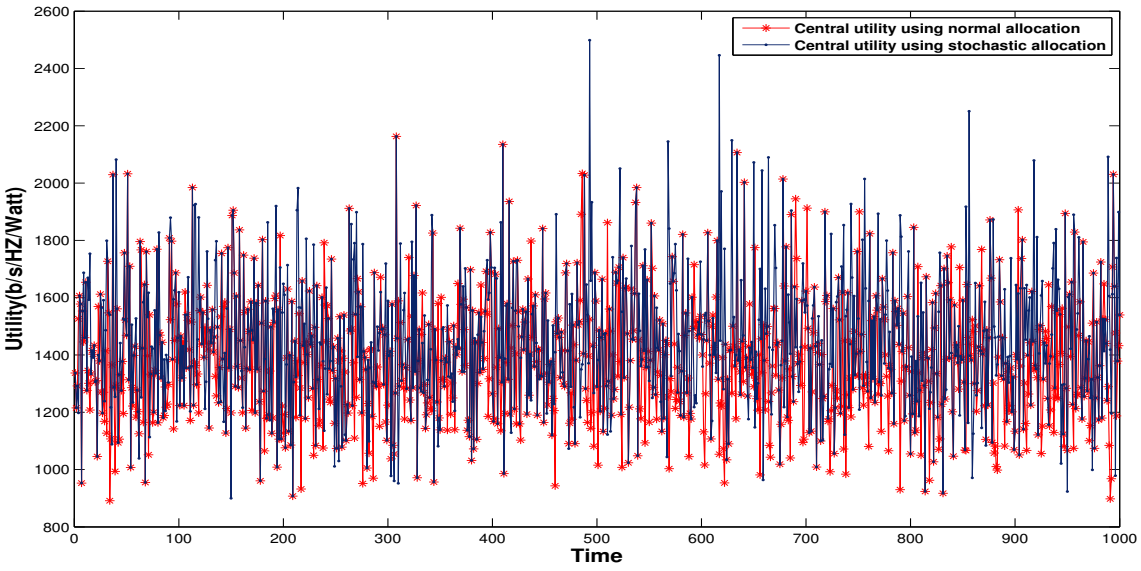

Figure 4.6.2: Maximum Central Utility: Comparison between $S1$ and $S2$.

The results in Fig. 4.6.2 show a comparison between the measured maximum central utility $U_{c,max,actual}$ for S_1 and S_2. Results show that the allocation of the DFPC using S_2 is achieving better central utility $U_{c,max,actual}$ than that achieved according to S_1. The allocation of the DFPC using S_1 outperforms 10.3 % of the time while S_2 outperforms 33.2 % and both models generate the same DFPC allocation 56.5% of the time. Thus the results verified that S_2 outperforms S_1 for 22.8% of the time. Also the average central utility $U_{c,max,Actual}$ using S_2 is calculated as 1536.5 b/s/Hz/watt while for S_1 it sums up to 1416.1 b/s/Hz/watt with a significant difference of 120.4 b/s/Hz/watt between the two allocation models..

4.7 Conclusion

The mobile and wireless communications enablers for the twenty-twenty information Society (METIS) in their research activities on 5G systems are targeting an average area spectral efficiency (AASE) over hundred Gbps/km^2/user. Heterogeneous networks (HetNets) that consist of both macro and small cells are currently viewed as the most capable solution to satisfy the expected future AASE requirements.

This work presents a new network architecture for the next generation (5G) of wireless communication systems. The architecture aims at building a self organizing network (SON) module in the core network. The module acts as a decision support system (DSS) supported by a Geo-location data collection infrastructure for radio resources utilization. The new network architecture using the DSS is expected to manage the spectrum utilization by forecasting the best down-link frequency partition configuration (DFPC) to be utilized in the access network (AN).

Finally a plan-do-control-act (PDCA) model for radio resource management is proposed to be used by the DSS in the SON module. The PDCA model aims at continuous improvement of the spectrum allocation process. A statistical analysis for the empirical data collected in a proposed data collection system (DCS) is performed by the DSS. Utilizing the DSS and the collected data, an optimization problem using stochastic linear programing with recourse and a discrete probability distribution modeled as a histogram is formulated to identify the best DFPC to be implemented in the AN. Simulations is conducted and results for the proposed scheme show that identifying future DFPC by learning from old allocations and using stochastic allocation with recourse improves the spectrum utilization efficiency over using a memoryless allocation method.

4.8 References

[66] "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2012-2017," white paper, Cisco, Febr. 2013."

[67] Mobile and wireless communications Enablers for the Twenty-twenty Information Society (METIS), *Intermediate description of the spectrum needs and usage principles, Deliverable D5.1,* , 2013.

[68] T.-T. Tran, Y. Shin, and O.-S. Shin, "Overview of enabling technologies for 3GPP LTE-advanced," *EURASIP Journal on Wireless Communications and Networking*, 2012.

[69] I. F. Akyildiz, D. M. Gutierrez-Estevez, and E. C. Reyes, "The evolution to 4G cellular systems: LTE-Advanced," *Physical Communication*, vol. 3, no. 4, pp. 217 – 244, 2010.

[70] D. Ling, Z. Lu, Y. Ju, X. Wen, and W. Zheng, "A multi-cell adaptive resource allocation scheme based on potential game for ICIC in LTE-A," *International Journal Communication Systems*, 2013.

[71] B. Krasniqi and C. Mecklenbrauker, "Maximization of the Minimum Rate by Geometric Programming for Multiple Users in Partial Frequency Reuse Cellular Networks," in *IEEE Vehicular Technology Conference (VTC Fall)*, 2011, pp. 1–5.

[72] K. Balachandran, J. Kang, K. Karakayali, and K. Rege, "Cell Selection with Downlink Resource Partitioning in Heterogeneous Networks," in *IEEE International Conference on Communications Workshops (ICC)*, 2011, pp. 1–6.

[73] O. Østerbø and O. Grøndalen, "Benefits of Self-Organizing Networks (SON) for Mobile Operators," *Journal of Computer Networks and Communications*, vol. 2012, no. 862527, p. 16, 2012.

[74] O. Guéret, "Exploring Traffic Geo-Location as a Key Strategy to Offload Macro Networks with Small Cells Solutions," Alcatel-Lucent, 2011, http://webform.alcatel-lucent.com/res/alu/survey/alu2CustomForm.jsp?cw=alu2CorpDocDownload&LMSG_CABINET=Docs_and_Resource_Ctr&LMSG_CONTENT_FILE=White_Papers/Alcatel-Lucent_GeoLocation.pdf.

[75] IEEE, *IEEE Standard for Local and metropolitan area networks Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems*, May 2009.

CHAPTER 5. FAULT-TOLERANT PLANNING OF RELAY STATION LOCATIONS IN WIMAX NETWORKS

A paper ready to be submitted to *The Journal of IEEE Transactions of Mobile Computing*

Tamer R. Omar[12], Zakhia Abichar [13], Ahmed E. Kamal[14], J. Morris Chang[15].

5.1 Abstract

WiMAX uses Relay Stations (RS) to extend the networks coverage and increase the spectrum efficiency. However, the standard does not specify how to plan the locations of the RSs within the network. Several papers introduced strategies for planning the locations of RSs in the WiMAX architecture. However, placement of RSs in WiMAX networks such that an RS failure will not interrupt the service, hence

[12]Graduate student, Primary researcher and author, Department of Electrical and Computer Engineering, Iowa State University, USA.

[13]Author for correspondence, Lecturer, Department of Electrical Engineering and Computer Science University of Central Florida, USA.

[14]Author for correspondence, Professor, Department of Electrical and Computer Engineering, Iowa State University, USA.

[15]Author for correspondence, Associate Professor, Department of Electrical and Computer Engineering, Iowa State University, USA.

making the network fault-tolerant, is an important design and planning problem that has not been addressed in the literature. In this paper, we address this problem, and present an Integer Linear Program (ILP) formulation for planning RS locations with fault-tolerance. We allow one RS to fail while keeping the provided service at a designated level (defined in terms of throughput to users). We consider the problem of RS locations planning using both out-band and in-band modes and an interference model is presented to consider the in-band mode and to address the effect of interference on RSs placement planning. We present numerical results that show how our model can be used to plan the positions of RSs. We also incorporate the existence of obstacles in the planning, such as large structures or natural formations, that might happen in real life. To the best of our knowledge, this is the first work that addresses planning the RS locations in WiMAX in a fault-tolerant manner.

Keywords: WiMAX, IEEE 802.16j, Relay stations, Network architecture and design, Network topology, Fault tolerance.

5.2 Introduction

After the success and wide deployment of Wireless Local Area Networks (WLAN) [76], the area of wireless networks has witnessed the standardization process for broadband wireless access networks. The IEEE 802.16 standards provide last-mile connectivity and they have been touted to fill several needs: last-mile end user access, initial deployment of infrastructure in unwired areas and providing access to mobile users.

The networks running on the IEEE 802.16 standards are known as WiMAX (Worldwide Interoperability for Microwave Access) [77, 78]. The main stations in a WiMAX network are the Base Station (BS) and the Subscriber Stations (SS), or alternatively called Mobile Stations (MS) when the device is mobile.

The IEEE 802.16j amendment introduced the use of Relay Stations (RSs) in WiMAX networks. The goal of using RSs is to support the connectivity between the BS, on one side, and the SSs and MSs, on the other side. The RS can extend the range of a BS. For example, there could be users that are out of reach of the BS and cannot connect to the network. With the placement of an RS between the user and the BS, the user would be able to connect; hence, the range of the BS is extended. The RS can be also used to enhance the capacity of the BS. For example, even if all the users are in range with the BS, placing one or more RSs in the cell allows higher data rates and enhancing the cell's capacity as a result. An introduction to the 802.16j standard is provided in [79, 80]. The 802.16j standard, however, doesn't specify how the RSs should be placed in the network. This issue remains open to research. It is the goal of this paper to devise a technique for planning the RS locations with fault tolerance to avoid failures in a WiMAX network.

5.2.1 Motivation

In this paper, we consider the problem of placing several RSs to support a BS in order to extend the coverage, improve the rate, and at the same time provide a resilient operation to the WiMAX network. In real life, the users expect a reliable service. In real life many businesses rely on the Internet connection in order to be able to function. If the network is planned with no fault tolerance, an RS failure might result in disconnecting some users.

We consider the case in which at most one RS might fail at a certain time. It could be any RS among the used RSs. With an adequate level of service, the RS should be fixed before another RS fails. This assumption (allowing only one RS to fail) also allows keeping the cost of the system reasonable.

In order to provide fault tolerance, we define for each user a full bit rate and a backup bit rate. When there is no failure among the RSs, the users receive service at

the full bit rate. However, in the case of a failure, we consider that offering service to affected users at a reduced rate is better than no service at all. Thus, in the case of an RS failure, the users receive service at the backup rate. This definition also allows users who primarily depend on the Internet for business to have a backup rate that can be made equal to the full rate. Thus, these users will function without service degradation even in the case of an RS failure.

The input to our problem is the location of the BS and the location of the users (SSs and MSs) and their respective demands represented by the bit rate. Users are represented by Test Points (TP). For example, if there are several houses located nearby of each other with demands of 2, 3 and 5 Mbps, they could be represented by a TP (located in a centric point to the houses) with a demand equal to the total SS demands of 10 Mbps.

The planning solution we present in this paper aims at placing the RSs in the network to achieve several goals. The specific goals are:

1. All the service area should be covered with connection to the network. The area is defined through the TPs; thus by providing connectivity between all the TPs and the BS, the service area will be covered.

2. The throughput demand of all the TPs should be satisfied. There should be a connection between the TP and the BS with a flow equal to the predefined demand of the corresponding TP.

3. The number of RSs placed by our solution should be minimized in order to reduce the equipment, installation and operation cost.

4. In case an RS in the network fails the network should continue to operate and provide service to the TPs at a predefined level of service, which we call the backup service rate. Thus, our planning method provides fault-tolerance and resilience to single RS failures.

5.2.2 Contribution

In a previous contribution [81], we provided a planning solution that satisfies only the first three goals listed above. There was no fault-tolerance provision in our initial approach. Thus, if an RS fails, there was no guarantee that the level of service provided would be adequate to the subscribers. Hence, in this paper, we extend our approach to incorporate fault tolerance and ensure that users are served adequately in the case of an RS failure.

Automation of fault tolerance detection and correction process can be handled through Self Organized Networks (SON) frame work. Cost efficient and high quality network operation using SON to optimize different access networks radio resources is discussed in [82]. An overview about the inter-working between different SON functionalities and some exemplary use cases are presented. Self healing is one of the main functions of the SON frame work that can be utilized to achieve a fault tolerant network. Self healing mechanism through cooperative clusters is proposed in [83] to deploy and manage the increasing number of small-cell networks. Resources utilization performance in both normal and failure modes of a small-cell network is evaluated and authors show that their proposed mechanism outperform other conventional mechanisms.

There are several papers in the literature that address the problem of placing RSs in the WiMAX network. These are reviewed in the next section. However, to the best of our knowledge, there is no work on planning the locations of RSs in WiMAX networks with fault-tolerance. There have been some approaches on placing relays in a fault-tolerant manner in other types of networks such as wireless sensor networks.

We formulate the RS planning problem using a Mixed Integer Linear Program (MILP). We present numerical results by solving our model with CPLEX v12.2. We believe that solving the model directly to obtain results is a valid approach since a

limited number of RSs will be installed in real-life to serve one BS and also planning is not a real-time operation.

The problem is solved and the allocation is made using both out-band and in-band modes. To address the in-band mode, an interference model is introduced and the maximum link rates is calculated taken the interference into consideration. Since the interference model results in a non-linear formulation of the problem. We mapped the formulation to a binary linear formulation by expanding the state space, hence avoiding non linearity.

We present numerical results that show how our model finds the number and locations of RSs. Our model also specifies all the links that are used and gives the rate on each link. Also, for every RS that is used in the main topology (used when no RS is in a failure condition), the model gives a corresponding backup topology in case this RS fails.

The rest of this paper is organized as follows. Section II presents the related work and Section III presents the network model. The optimization model with fault-tolerance is provided in Section IV which includes both out-band and in-band operational modes. Numerical results are given in Section V, and the conclusions are given in Section VI.

5.3 Related Work

This section presents earlier work in the literature that is related to the problem of planning the RS locations with fault tolerance.

5.3.1 Planning RS location in WiMAX networks

Our previous work in [81] presents a model for planning the RS locations in a WiMAX networks. However, in the previous work, there was no guarantee of service

if an RS fails since fault-tolerance was not considered. In this paper, we extend our model to provide resilience to relay features.

A clustering algorithm based on uniform cluster concepts is proposed in [84] to select the BS and RS locations from candidate positions depending on the traffic demands. The authors introduce a scheme that makes adaptive decision for selecting the deployment sites of the BS and RS. The simulation results show that the scheme achieves a high performance for the network throughput and coverage ratio. However the work did not address the effectiveness of the scheme in case of RS failures.

A planning model is presented in [85] to find the locations of BSs and RSs in the network. The model is formulated as an optimization problem using integer programming. In this model, there is at most one RS between the SS/MS and the BS. A maximum of two hops is allowed. Since the standard does not have a limit on the number of hops going through the RSs, this assumption may impose unneeded restrictions.

The same authors of the work above present an extension of their work in [86]. In this paper, they consider a large coverage area which increases the computation time of the model to reduce the computation time, they divide the area into clusters and apply the approach above to every cluster. Then, the cases on the boundaries of the clusters are solved to find the overall solution. This paper similarly limits the number of hops to two.

In [87], an RS placement model is presented. This work is based on cooperative transmission between the source node and the relay node to provide a better signal to the destination node. They consider the decode-and-forward scheme and the compress-and-forward scheme for cooperative transmission. This model is different from our work since it considers the placement of a single RS to serve multiple SSs. The authors present another RS placement solution in [88], where the cooperative transmission paradigm is used in multi-hop relaying for the purpose of range exten-

sion. Also, the same authors present in [89, 90] an RS placement solution that uses the cooperative transmission technique for the purpose of capacity enhancement.

In [91], a model is presented to find the locations of RSs that extend the range of a BS in a WiMAX network. This work defines preset topologies and finds the RS placement for these topologies; in comparison, our model in this paper and in [81] can work with any topology. This work also considers RS location planning for sector-based topology. Each sector uses a frequency that is different from adjacent sectors in order to reduce the interference.

In [92], the problem of joint BS and RS deployment is considered and an optimization model is presented. Due to the large size of the problem, the model takes a long time to solve. Thus, the authors also present an efficient heuristic algorithm to find the problem sub-optimal solution.

In [93], the problem of RS placement in the WiMAX network is considered. The location of the RSs and the bandwidth allocation to users are found. This work assumes that users' demands could change due to fluctuations in traffic demand and due to mobility. Thus, the optimization of the RS locations is found on a long-term basis and the bandwidth allocation to users is found on a short-term basis.

The authors of [94] consider using relays for the purpose of capacity enhancements as follows. There is a BS, an area that can be totally covered by the BS and a given number of relays. This work decides where to place the relays to maximize the system capacity.

In [95], the following paradigm is considered for the placement of RSs in WiMAX networks. The number and locations of BSs are given. The goal of the problem is to place RSs that use the transparent mode. In this mode, the RSs do not transmit control information; the control information are only transmitted by the BS. The RSs are thus in range of the BS and the goal of the RS placement is capacity enhancement.

Other approaches used in BSs and RSs placement in IEEE 802.16j is presented in [96, 97] with the goal to enhance the overall network capacity.

5.3.2 Planning relay locations with fault-tolerance

There are approaches in the literature that provide relay location planning with fault-tolerance. But these approaches have been designed for wireless sensor networks and not for WiMAX networks.

In [98], an Integer Linear Program (ILP) model is presented for placing relays in sensor networks to provide fault-tolerance in case some nodes fail. The main issue was connectivity, regardless of bandwidth requirements, which implies that all relay nodes may be operational all the time. The same authors present an extension of their work in [99], which takes into consideration the routing strategy in order to reduce the battery consumption.

Other approaches on fault-tolerant placement of relay nodes are given in [100, 101, 102, 103, 104, 105, 106, 107].

The approaches presented in this section consider a multitude of issues and configurations for the RS use in WiMAX networks. There are multiple approaches on planning RS placement in WiMAX networks. There are also approaches that plan the locations of relay nodes with fault-tolerance, although not in WiMAX networks. To the best of our knowledge, there is no approach that provides the planning of RS locations in WiMAX network with fault-tolerance. Moreover, approaches introduced for wireless sensor networks do not consider downgrading the offered service due to failures. Hence, our paper is the first to propose such an approach.

5.4 Network Model

This section presents the network model that we consider in this paper.

5.4.1 Relay station non-transparent mode

The IEEE 802.16j Standard defines two modes of RS operation: a transparent mode, and a non-transparent mode.

In the transparent mode, the users (SSs and MSs) are unaware of the presence of an RS. The RS does not transmit control information (such as down-link map and up-link map). These are transmitted by the BS. Thus, all the SSs and MSs are within range of the BS. However, the RSs are used in the transparent mode for the purpose of capacity enhancement.

In the non-transparent mode, the RS transmits control information to the SSs it serves. Multi-hop routes are allowed in the non-transparent mode. The goal for using non-transparent RSs is to extend the range of the network and to also enhance the capacity.

5.4.2 Duplexing mode

When multi-hop RSs are used, the frames of two stations that are in range should be duplexed either in the time domain or in the frequency domain in order to avoid interference.

The IEEE 802.16j Standard [77] allows using different frequencies for RSs serving the same BS. Thus, we make the assumption that the RSs duplex their transmission in Frequency Division Duplex (FDD) mode. For example, on a two-hop route BS - RS - SS, we can have a transmission of rate r on the BS-RS hop and another transmission of the same rate on the RS-SS hop. This happens if the two hops are using different frequencies. With Time Division Duplex (TDD), the two hops will

alternate in transmission using the same frequency channel. However, each hop will have a larger bandwidth since the bandwidth is not divided anymore. We use FDD for simplicity, but our model can be modified easily to work with TDD.

5.4.3 Link capacity

Our model allocates a rate on each link that is used in the produced topology. The allocated rate on a link is bounded by the maximum capacity of the link. The maximum capacity of a wireless link can be modeled with the Shannon-Hartley equation as given in [108]. It is given by the equation: $C = B.log_2(1 + SINR)$, where C is the capacity in bit/sec, B is the channel bandwidth in Hz.

The SINR is the signal to interference plus noise ratio that can be calculated as $SINR = S/[N_0 + I]$, where S is the received signal strength on link i, N_0 is the noise density and I is the signal strength received from all interferes j on link i. The capacity changes with the distance since the $SINR$ degrades when the distance increases. The SINR on a link i can be expressed as

$$SINR_i = \frac{p_i}{(d_{ij})^{\alpha}\left[N_0 + \sum_{j \neq i} p_i/(d_{ij})^{\alpha}\right]} > \beta$$

where p_i is the signal received power, d_{ij} is the Euclidean distance between the transmitter and receiver, $\alpha > 2$ is the path loss exponent and β is the antenna gain.

Other factors also affect the link capacity such as the coding and modulation schemes. When a high SINR is measured on the link, coding and modulation schemes with high rates are used. However, when the SINR is low, robust coding and modulation schemes are preferred to limit the Bit-Error-Rate (BER), although they provide low data rates. Table 5.4.1 shows the achievable bit rates for the OFDMA physical layer as given in the standard [109]. QPSK is more robust but achieves a small rate. On the other hand, 64-QAM is less robust but achieves a high rate.

Table 5.4.1: OFDMA rates (Mbps) for various modulation schemes using 7 (Mhz) bandwidth

QPSK	QPSK	16-QAM	16-QAM	64-QAM	64-QAM
1/2	3/4	1/2	3/4	2/3	3/4
5.82	8.73	11.64	17.45	23.27	26.18

The factors that affect a link's capacity can be combined in an equation. For a link from node i to node j, the maximum rate is: $m_{i;j} = \Gamma(SNR_{i;j},\ \beta,\ Cod,\ Mod)$, where Γ is the upper-bound on the BER, Cod is the coding scheme and Mod is the modulation scheme. Γ is the function that maps all the three parameters to the maximum rate.

Any definition of the function Γ can work with our model. However, for simplicity, we assume that the maximum rate changes with distance. In real-life scenarios, there is usually a field survey which precedes the network deployment [110, 111, 112]. The link rates are selected based on the links characteristics such as, the SINR (Signal-to-Interference-plus- Noise-Ratio), fading, the specifics of the terrain and interference with other wireless systems.

5.4.4 Definition of fault-tolerance

The planning model we present in this paper allows the failure of an RS without interrupting service to the users, albeit at a reduced bit rate, hence tolerating equipment failure.

We assume that only one RS will fail at a given time. This is a reasonable assumption since usually in the time it takes the RS to be repaired, there is a very small probability that another RS will fail. This is true since the number of RSs supporting a BS cannot be very large. This assumption will keep the cost of RSs small, since tolerating the failure of two or more RSs at the same time requires installing many extra RSs. This might not be a cost effective approach.

For every set of customers, represented by a Test Point (TP), a tuple $\{r_i,\ rb_i\}$ defines the requested service rates. When all the RSs are operational, the full rate or a TP i, given by r_i is provided. However, when there is an RS failure, a reduced rate which is the backup rate, rb_i is provided. We have $rb_i \leq r_i$. Users who request the same service rate even in the case of an RS failure will have $rb_i = r_i$.

5.5 Optimization Model

This section presents the optimization model for the RS planning problem with fault-tolerance in both out-band and in-band modes. The model takes as input (1) the possible sites where an RS can be installed, (2) the location of the Test Points (TP) that represent the users' traffic, (3) the rates (full and reduced) in Mbps of each TP; the full rate is provided when all the RSs are operational and the reduced rate is guaranteed when there is an RS failure. (4) Finally, the model takes as an input the maximum rate on any link: BS-RS, BS-TP, RS-RS and RS-TP, which depends on the link characteristics such as distance, SINR and bandwidth.

The output of our model is the full-rate (main) topology and the reduced-rate (backup) topologies. The full-rate topology is defined by the number of RSs used, their positions, the links used, the rate on each link and, finally, the connection node for each TP (either the BS or an RS). Each of the backup topologies corresponds to a failure in one of the RSs used in the main topology. For example, if the main topology uses RS 1, RS 3 and RS 8, then there will be three backup topologies that are used in the case any of these RSs fail.

For any TP, TP_i, its full rate is designated by r_i and its reduced rate is designated by rb_i (rb stands for backup rate), which is the minimum acceptable rate in the case of failure.

Let $R = \{RS_0, ..., RS_{N-1}\}$ be the set of candidate sites for RS with cardinality $|R| = N$. Similarly, let $T = \{TP_0, ..., TP_{M-1}\}$ be the set of Test Points (TP) that represent the user traffic with cardinality $|T| = M$.

5.5.1 Decision variables

The following decision variables define the full-rate topology.

$$dR_i = \begin{cases} 1; & \text{an RS is deployed in site R}S_i \\ 0; & \text{otherwise (i } \epsilon \text{ R)} \end{cases}$$

$$dBR_i = \begin{cases} 1; & \text{a link is used between the BS and R}S_i \\ 0; & \text{otherwise(i } \epsilon \text{ R)} \end{cases}$$

$$dBT_i = \begin{cases} 1; & \text{T}P_i \text{ is assigned to the BS} \\ 0; & \text{otherwise (i } \epsilon \text{ T)} \end{cases}$$

$$dRR_{ij} = \begin{cases} 1; & \text{a link is used between the R}S_i \text{ and R}S_j \\ 0; & \text{otherwise (i, j } \epsilon \text{ R)} \end{cases}$$

$$dRT_{ij} = \begin{cases} 1; & \text{T}P_j \text{ is assigned to the R}S_i \\ 0; & \text{otherwise (i } \epsilon \text{ R, j } \epsilon \text{ T)} \end{cases}$$

We also define variables that are similar to the above in order to specify the backup topologies. These variables are: $dR_i^k, dBR_i^k, dBT_i^k, dRR_{ij}^k$and dRT_{ij}^k. The term k indicates the RS that has failed. For example, when $k = 3$, these variables define the backup topology that is used when RS 3 fails.

We also define decision variables that designate the assigned flow (in Mbps) on each link. While the previous variables were binary, the flow variables take integer values. We use integer values since the rate used on a link is usually one of a subset of the supported rates by the physical (PHY) layer. Thus, the rate is selected from a given set of rates.

In the full-rate topology, the variables fBR_i, and fRR_{ij} designate the flow on the links from BS-to-RS, and RS-to-RS, respectively, where i and fBR_i are indexes of RSs ($i,\ j \epsilon R$).

Similarly, in the backup topology, the variables fBR_i^k and fRR_{ij}^k designate the flow on the links from BS-to-RS and RS-to-RS, respectively, where $i,\ j$ and k are indexes of RSs ($i,\ j,\ k \epsilon R$).

5.5.2 Topology constraints

The following constraints define the topology of the RS domain. They ensure that when a link is used in the solution,the two end nodes of the link exist (i.e., the RSs are selected). They also ensure that a TP is connected either directly to the BS or to only one RS; we use this condition for simplicity of problem formulation and for simplicity of network operation.

First, when there is a link between the BS and RS i, there should be an RS deployed at site RS i. This is ensured by the following constraints in the full-rate and the backup topologies.

$$dBR_i \leq dR_i \qquad \forall\ i \epsilon R \tag{5.5.1}$$

$$dBR_i^k \leq dR_i^k,\ i \neq k, \quad \forall\ i,\ k \epsilon R \tag{5.5.2}$$

When there is a link between RS i and RS j, two RSs should be installed at sites RS i and RS j. This is ensured by the following constraints

$$dRR_{ij} \leq \frac{dR_i + dR_j}{2} \forall\ i,\ j \epsilon R \tag{5.5.3}$$

$$dRR_i^k \leq \frac{dR_i^k + dR_j^k}{2},\ i \neq k,\ j \neq k, \forall\ i,\ j,\ k \epsilon R \tag{5.5.4}$$

When there is a link between RS i and TP j, an RS should be deployed at site RS i. This is ensured by the following constraints.

$$dRT_{ij} \leq dR_i \qquad \forall\, i \,\epsilon\, R,\ \forall\, j \,\epsilon\, T \tag{5.5.5}$$

$$dRT^k_{ij} \leq dR^k_i,\ i \neq k,\ j \neq k, \quad \forall\, i,\ k \,\epsilon\, R,\ \forall\, j \,\epsilon\, T \tag{5.5.6}$$

The following constraints send all the traffic of a TP either through a direct link with the BS or through a single TP.

$$dBT_i + \textstyle\sum_{j\,\epsilon\,R} dRT_{ji} = 1 \qquad \forall\, i \,\epsilon\, T \tag{5.5.7}$$

$$dBT^k_i + \textstyle\sum_{j\,\epsilon\,R,\, j\neq k} dRT^k_{ji} = 1 \quad \forall\, i \,\epsilon\, T \tag{5.5.8}$$

5.5.3 Flow constraints

The flow constraints ensure that the amount of data that is transported is balanced and sufficient for the demands of all the TPs.

5.5.3.1 Flow balance at the BS:

In the main topology, the total traffic going out of the BS should be equal to the sum of the full rates, r_i, of all the TPs. This condition is ensured by the following equation.

$$\sum_{i\,\epsilon\,R} fBR_i.dBR_i + \sum_{j\,\epsilon\,T,\, mBT_j \geq r_j} r_j.dBT_j = \sum_{j\,\epsilon\,T} r_j \tag{5.5.9}$$

At a backup topology, the rate provided to TP_i is greater than or equal to rb_i. Then, this condition is used.

$$\sum_{i \,\epsilon\, R,\, i \neq k} fBR_i^k.dBR_i^k + \sum_{j \,\epsilon\, T,\, mBT_j \geq rb_j} rb_j.dBT_j^k = \sum_{j \,\epsilon\, T} rb_j \tag{5.5.10}$$

We are interested in keeping the system linear. Thus, we use the following transformation and substitute in Eq. (9).

$$X_i = fBR_i.dBR_i$$

where X_i is an auxiliary variable.

Eq. (9) therefore becomes:

$$\sum_{i \,\epsilon\, R} X_i + \sum_{j \,\epsilon\, T,\, mBT_j \geq r_j} r_j.dBT_j \;\; = \;\; \sum_{j \,\epsilon\, T} r_j \tag{5.5.11}$$

X_i can be evaluated using the following set of linear constraints, where Q is a large number such that $Q > max(fBR_i),\ \forall\, i \,\epsilon\, R$.

$$X_i \geq Q.dBR_i - Q + fBR_i \quad \forall\, i \,\epsilon\, R \tag{5.5.12}$$

$$X_i \leq fBR_i \quad \forall\, i \,\epsilon\, R \tag{5.5.13}$$

$$X_i \geq 0 \quad \forall\, i \,\epsilon\, R \tag{5.5.14}$$

$$X_i \leq Q.dBR_i \quad \forall\, i \,\epsilon\, R \tag{5.5.15}$$

Similarly, we use the following transformation for Eq. (10).

$$X_i^k = fBR_i^k.dBR_i^k,\ i \neq k \quad \forall\, i,\ k \,\epsilon\, R \tag{5.5.16}$$

Then, Eq. (10) becomes:

$$\sum_{i \epsilon R,\, i \neq k} X_i^k + \sum_{j \epsilon T,\, mBT_j \geq rb_j} rb_j.dBT_j^k = \sum_{j \epsilon T} rb_j \tag{5.5.17}$$

X_i^k is evaluated like X_i was evaluated in Eq. (12) to Eq. (15).

5.5.3.2 Flow balance at an RS:

At an RS, the amount of traffic that is coming from the BS and from upstream RSs is equal to the amount of traffic that is going to downstream RSs and to TPs that are directly connected to the RS. This is ensured by the following constraint.

$$fBR_i.dBR_i + \sum_{j \epsilon R} fRR_{ji}.dRR_{ji} = \sum_{j \epsilon R} fRR_{ij}.dRR_{ij} + \sum_{y \epsilon T,\, mBT_{iy} \geq r_y} r_y.dRT_{iy} \forall\, i \epsilon R \tag{5.5.18}$$

The equation above is made linear by using the transform $Z_{ij} = fRR_{ij}.dRR_{ij}$ and becomes:

$$X_i + \sum_{j \epsilon R} Z_{ji} = \sum_{j \epsilon R} Z_{ij} + \sum_{y \epsilon T,\, mBT_{iy} \geq r_y} r_y.dRT_{iy} \quad \forall\, i \epsilon R \tag{5.5.19}$$

Z_{ij} is evaluated like X_i was evaluated in Eq. (12) to Eq. (15).

For the backup topologies, the flow conservation at the RS is ensured by the following equation.

$$fBR_i^k.dBR_i^k + \sum_{j\,\epsilon\,R,\,j\neq k} fRR_{ji}^k.dRR_{ji}^k =$$

$$\sum_{j\,\epsilon\,R,\,j\neq k} fRR_{ij}^k.dRR_{ij}^k + \sum_{y\,\epsilon\,T,\,mBT_{iy}\geq rb_y} rb_y.dRT_{iy}^b$$

$$i \neq k,\ \forall\, i,\ k\,\epsilon\,R \qquad (5.5.20)$$

The equation above is made linear by using the transform $Z_{ij}^k = fRR_{ij}^k.dRR_{ij}^k$ and it becomes:

$$X_i^k + \sum_{j\,\epsilon\,R,\,j\neq k} Z_{ji}^k = \sum_{j\,\epsilon\,R,\,j\neq k} Z_{ij}^k +$$

$$\sum_{y\,\epsilon\,T,\,mBT_{iy}\geq rb_y} rb_y.dRT_{iy}^b \qquad i \neq k,\ \forall\, i,\ k\,\epsilon\,R \qquad (5.5.21)$$

Z_{ij}^k is evaluated similar to how X_i was evaluated in Eq. (12) to Eq. (15).

5.5.3.3 Flow balance at a TP:

In the main topology, the amount of traffic between the BS or the RS and the TP should be equal to the full-rate, r_i of the TP. This is ensured by the following constraint.

$$\sum_{i=i,\,mBT_i\geq r_i}^{i} r_i.dBT_i + \sum_{j\,\epsilon\,R,\,mRT_{ji}\geq r_i} r_i.dRT_{ji} = r_i$$

$$\forall\, i\,\epsilon\,T \qquad (5.5.22)$$

For the backup topologies, the amount of traffic at the TP should be equal to rb_i. This is ensured by the following constraint.

$$\sum_{i=i,\, mBT_i \geq rb_i}^{i} rb_i.dBT_i^k + \sum_{j\,\epsilon\, R,\, j\neq k,\, mRT_{ji} \geq rb_i} rb_i.dRT_{ji} = rb_i$$

$$\forall\, i\,\epsilon\, T\ ,\ k\,\epsilon\, R \tag{5.5.23}$$

5.5.4 Constraint on the link capacity

The model that we have so far assumes the out-band mode, and in this case the maximum capacity of link BR_i, RR_i, BT_i and RT_i are mBR_i, mRR_i, mBT_i and mRT_i respectively.

To accommodate the in-band mode, the capacity on link i depends on the activity of other links j, which may interfere with link i. We will need to set the capacity on link i such that it corresponds to the capacity that is subject to interference with other active links in the system. Rather than using a non-linear formulation, we use a linear formulation that comes on the cost of an expanded space state. This is done by using a binary linear formulation in which the capacity of link i is defined as follows:

- Assume that the maximum number of other links which can interfere with link i are a_i, and the setof such links is $L_i^x = l_1, l_2, ..., l_{a_i}$ where $x\epsilon\,\{BR,\ RR,\ BT,\ RT\}$.
- The capacity of link i given that links in the n^{th} subset $F_n^x \subseteq L_i^x$ are active, including the empty subset, is given by $c_{F_n^x}^x$, and is precomputed.

In order to find out which combination of links are active, a binary variable I_j^x is defined as being equal to 1 if link j is active. The number of active links in each F_n^x is evaluated as follows:

$$A_{F_n^x}^x = \sum_{j\,\epsilon\, F_n^x} I_j^x \quad \forall\, F_n^x \subseteq L_i^x \tag{5.5.24}$$

Then, to find the combination that has all of its member links active. We define a binary variables $B_{F_n^x}^x$, which will be equal 1 only if all links in the subset F_n^x are

active. $Bx_{F_n^x}$ can be evaluated using the following constraints

$$Bx_{F_n^x} \geq Ax_{F_n^x} - |F_n^x| + 1 \tag{5.5.25}$$

$$Bx_{F_n^x} \leq \frac{Ax_{F_n^x}+\delta}{|F_n^x|+\delta} \tag{5.5.26}$$

where δ is a small number. The addition of δ to both the numerator and dominator is to include the case of empty subset, in which case both $|F_n^x|$ and $Ax_{F_n^x}$ are zeros.

Therefore, the maximum capacity of a link can be evaluated as the minimum for all cases in which $Bx_{F_n^x} = 1$. The upper bounds of the rate on this link ($mBR^i_{F_n^{BR}}$, $mRR^{ij}_{F_n^{RR}}$, $mBT^i_{F_n^{BT}}$, $mRT^{ij}_{F_n^{RT}}$) for main topology and ($mBR^{ik}_{F_n^{BR}}$, $mRR^{ijk}_{F_n^{RR}}$,$mBT^{ik}_{F_n^{BT}}$, $mRT^{ijk}_{F_n^{RT}}$) for backup topologies are input parameters to the problem and are calculated as follows:

$$mBR^i_{F_n^{BR}} = cBR^i_{F_n^{BR}} BBR^i_{F_n^{BR}} + (1 - BBR^i_{F_n^{BR}})M \tag{5.5.27}$$

$$mRR^{ij}_{F_n^{RR}} = cRR^{ij}_{F_n^{RR}} BRR^{ij}_{F_n^{RR}} + (1 - BRR^{ij}_{F_n^{RR}})M \tag{5.5.28}$$

$$mBT^i_{F_n^{BT}} = cBT^i_{F_n^{BT}} BBT^i_{F_n^{BT}} + (1 - BBT^i_{F_n^{BT}})M \tag{5.5.29}$$

$$mRT^{ij}_{F_n^{RT}} = cRT^{ij}_{F_n^{RT}} BRT^{ij}_{F_n^{RT}} + (1 - BRT^{ij}_{F_n^{RT}})M \tag{5.5.30}$$

$$mBR^{ik}_{F_n^{BR}} = cBR^{ik}_{F_n^{BR}} BBR^{ik}_{F_n^{BR}} + (1 - BBR^{ik}_{F_n^{BR}})M \tag{5.5.31}$$

$$mRR^{ijk}_{F_n^{RR}} = cRR^{ijk}_{F_n^{RR}} BRR^{ijk}_{F_n^{RR}} + (1 - BRR^{ijk}_{F_n^{RR}})M \tag{5.5.32}$$

$$mBT^{ik}_{F_n^{BT}} = cBT^{ik}_{F_n^{BT}} BBT^{ik}_{F_n^{BT}} + (1 - BBT^{ik}_{F_n^{BT}})M \tag{5.5.33}$$

$$mRT^{ijk}_{F_n^{RT}} = cRT^{ijk}_{F_n^{RT}} BRR^{ijk}_{F_n^{RT}} + (1 - BRT^{ijk}_{F_n^{RT}})M \tag{5.5.34}$$

where M is a very large number and the maximum flow that can be transmitted on a link i is limited by the power transmitted, the link distance and the coding and modulation schemes. The maximum rate that can be assigned on the link from the BS to RS i, fBR_i, is limited by $mBR^i_{F_n^{BR}}$, where $mBR^i_{F_n^{BR}}$ is the maximum rate on

this link. A similar notation is used for all the other links and the constraints that ensure the upper bound are the following

$$SC * fBR_i \leq nBR^i_{F^{BR}_n} mBR^i_{F^{BR}_n} \forall\, F^{BR}_n \subseteq L^{BR}_i \tag{5.5.35}$$

$$SC * fRR_{ij} \leq nRR^{ij}_{F^{RR}_n} mRR^{ij}_{F^{RR}_n} \forall\, F^{RR}_n \subseteq L^{RR}_{ij} \tag{5.5.36}$$

$$SC * fBR^k_i \leq nBR^{ik}_{F^{BT}_n} mBR^{ik}_{F^{BR}_n} \forall\, F^{BR}_n \subseteq L^{BR}_{ik} \tag{5.5.37}$$

$$SC * fRR^k_{ij} \leq nRR^{ijk}_{F^{RR}_n} mRR^{ijk}_{F^{RR}_n} \forall\, F^{RR}_n \subseteq L^{RR}_{ijk} \tag{5.5.38}$$

$$SC * fBT_i \leq nBT^i_{F^{BT}_n} mBT^i_{F^{BT}_n} \forall\, F^{BT}_n \subseteq L^{BT}_i \tag{5.5.39}$$

$$SC * fRT_{ij} \leq nRT^{ij}_{F^{RT}_n} mRT^{ij}_{F^{RT}_n} \forall\, F^{RT}_n \subseteq L^{RT}_{ij} \tag{5.5.40}$$

$$SC * fBT^k_i \leq nBT^{ik}_{F^{BT}_n} mBT^{ik}_{F^{BR}_n} \forall\, F^{BT}_n \subseteq L^{BT}_{ik} \tag{5.5.41}$$

$$SC * fRT^k_{ij} \leq nRT^{ijk}_{F^{RT}_n} mRT^{ijk}_{F^{RR}_n} \forall\, F^{RT}_n \subseteq L^{RT}_{ijk} \tag{5.5.42}$$

$$i \neq k,\ j \neq k,\ i,j,k\, \epsilon\, R$$

where the variable $nBR^i_{F^{BR}_n} = [0,\ 1,\ ...,\ SC]$ corresponding to the number of sub-carriers allocated to any RS out of a total of S sub-carriers which is an input parameter, where SC is assumed as an integral power of 2 (e.g.,$SC = 512$)[16]. The constraint in eq.(5.5.35) is a nonlinear constrain with $nBR^i_{F^{BR}_n}$ a discrete variable and $mBR^i_{F^{BR}_n}$ a continuous variable, the following equations are used to transform it to a linear form. A new variable $vBR^i_{F^{BR}_n} = nBR^i_{F^{BR}_n} * mBR^i_{F^{BR}_n}$ is defined and a binary expansion for $nBR^i_{F^{BR}_n}$ is performed as

$$nBR^i_{F^{BR}_n} = \sum_{r=0}^{R-1} 2^r dnBR^{ir}_{F^{BR}_n} \forall\, F^{BR}_n \subseteq L^{BR}_i,\ i\, \epsilon\, R \tag{5.5.43}$$

$$R = log_2(SC) \tag{5.5.44}$$

$$dnBR^{ir}_{F^{BR}_n} \epsilon\, [0,1] \tag{5.5.45}$$

[16] with out loss of generality and to reduce the model complexity, we consider SC to be a power of 2

Then $vBR^{i}_{F^{BR}_{n}}$ can be rewritten as

$$vBR^{i}_{F^{BR}_{n}} = \sum_{r=0}^{R-1} 2^{r} wBR^{ir}_{F^{BR}_{n}} \forall\ F^{BR}_{n} \subseteq L^{BR}_{i},\ i, r\ \epsilon\ R \tag{5.5.46}$$

where

$$wBR^{ir}_{F^{BR}_{n}} = dnBR^{ir}_{F^{BR}_{n}} * mBR^{i}_{F^{BR}_{n}} \forall\ F^{BR}_{n} \subseteq L^{BR}_{i},\ i, r\ \epsilon\ R \tag{5.5.47}$$

Now the constraint in eq.(5.5.35) is converted into nonlinear constraint but with $dnBR^{i}_{F^{BR}_{n}}$ a discrete variable and $mBR^{i}_{F^{BR}_{n}}$ a continuous variable which can be linearized using the same approach used in eq.(5.5.12-5.5.15)

$$\begin{aligned}
wBR^{ir}_{F^{BR}_{n}} &\geq Q.dnBR^{ir}_{F^{BR}_{n}} - Q + mBR^{i}_{F^{BR}_{n}} && (5.5.48)\\
wBR^{ir}_{F^{BR}_{n}} &\leq mBR^{i}_{F^{BR}_{n}} && (5.5.49)\\
wBR^{ir}_{F^{BR}_{n}} &\geq 0 && (5.5.50)\\
wBR^{ir}_{F^{BR}_{n}} &\leq Q.mBR^{i}_{F^{BR}_{n}} && (5.5.51)\\
\forall\ F^{BR}_{n} &\subseteq L^{BR}_{i},\ i\ \epsilon\ R
\end{aligned}$$

similarly the same conversion is used for variables ($vBT^{i}_{F^{BT}_{n}}$, $vRR^{ij}_{F^{RR}_{n}}$, $vRT^{ij}_{F^{RT}_{n}}$) and the backup topologies variables ($vBR^{ik}_{F^{BT}_{n}}$, $vBT^{ik}_{F^{BT}_{n}}$, $vRR^{ijk}_{F^{RR}_{n}}$, $vRT^{ijk}_{F^{RT}_{n}}$). Also, when an RS fails, the rate on all the links incident on it is zero.

5.5.5 Objective function

The primary objective of our solution is to minimize the total number of RSs used. This will minimize the cost of RS installation. We define the term S_i which indicate if an RS is installed at site RS i either in the main topology or in any other topology.

$$S_i \geq dR_i \qquad i \,\epsilon\, R \tag{5.5.52}$$

$$S_i \geq dR_i^k \qquad i \neq k,\ i,\ k \,\epsilon\, R \tag{5.5.53}$$

To minimize the total number of RSs that are installed, $\sum S_i\ i\ \epsilon\ R$ should be minimized.

We also aim at reducing the number of RSs used in backup topologies. Minimizing the number of RSs used in every topology allows us to remove lengthy paths. For example, if a TP can connect to the BS by going through one RS only, it is better not to use two RSs for this TP. Thus, minimizing the number of RSs in the backup topologies will make a TP use the minimum necessary number of RSs it needs to connect to the BS.

The term T_k designates the number of RSs used in the backup topology when RS k fails. We have the following constraint.

$$T_i \geq \sum_{i\,\epsilon\,R,\, i\neq k} dR_i^k \forall\ k \,\epsilon\, R \tag{5.5.54}$$

For a similar reason as above, we aim to minimize the number of RSs that is used in the main topology, designated by the term V. We have the constraint:

$$V \geq \sum_{i\,\epsilon\,R} dR_i \tag{5.5.55}$$

The term Obj combines the terms above. The main term from the above is $\sum_{i\,\epsilon\,R} S_i$ since it gives the number of RSs that should be installed. It should be given a higher weight than the other terms. The maximum value of $\sum_{k\,\epsilon\,R} T_k$ is N^2 and the maximum value of V is N. Thus, we give the weight $N^2 + N$ to the term with S_i.

$$Obj = (N^2 + N).\sum_{i \epsilon R} S_i + \sum_{k \epsilon R} T_k + V \tag{5.5.56}$$

Then, the objective function is:

$$Minimize\ Obj \tag{5.5.57}$$

5.6 Numerical Results

This section shows the numerical results. Planning problems are presented and solved with our model. Firstly, we show examples of planning without fault-tolerance. In these examples, when an RS fails, there is no guarantee of service to the TPs. Secondly, we show examples with fault-tolerance, where service will be guaranteed even in case of failure of an RS. By using CPLEX v12.2, which runs on a modern multi-core machine, we obtained solution times for realistic scenarios that are under 10 seconds. Thus, the computation time is reasonable for practical cases.

5.6.1 Planning RS locations for out-band mode

For out-band mode the interference between the links are not considered and the maximum rate that can be assigned on the link from the BS to RS i, fBR_i, is limited by $fBR_i \leq mBR_i$. Similarly for all the other links the constraints that ensure the upper bound are calculated according to

$$fBR_i \leq mBR_i \quad i \epsilon R$$

$$fRR_{ij} \leq mRR_{ij} \quad i,\ j \epsilon R$$

$$fBR_i^k \leq mBR_i \quad i \neq k,\ i,\ k \epsilon R$$

$$fRR_{ij}^k \leq mRR_{ij} \quad i \neq k,\ j \neq k,\ i,\ j,\ k \epsilon R$$

5.6.1.1 Planning without Fault-Tolerance

This section presents the initial results of planning the RS locations without fault-tolerance. In this section, where no fault-tolerance is considered, the variables and constraints in the model that are used for fault-tolerance are omitted. These are all the variables that have an index k, which are: dR_i^k , dBR_i^k , dBT_i^k , dRR_i^k , fBR_i^k , fBR_i^k and fRR_{ij}^k.

Also, the objective function will change. In the case where we do not have fault-tolerance, the objective is simply to minimize the total number of RSs used. Then, the objective function is:

$$min \sum_{i \, \epsilon \, R} dR_i \tag{5.6.1}$$

This section also shows how our model considers the existence of obstacles in the planning field. We consider two types of obstacles. A lake-type obstacle does not allow installing RSs on it. However, it allows a wireless link to cross the lake. The second one is the mountain-type obstacle, which could be a steep mountain or a tall structure. This type of obstacle also does not allow installing an RS on it, and it does not allow a wireless link to go through it.

Theoretical Model

The following is an RS planning example that uses our solution. The problem is shown in Fig. 5.6.1(a). The planning area is made discrete by the use of a square grid. The BS location is on the top line of the grid as shown in the figure. We select this setting since we consider that the BS is at the edge of the connected area. The area below the BS does not have the connection and we plan to connect this area through the BS. Without loss of generality, we can use any topology with our model.

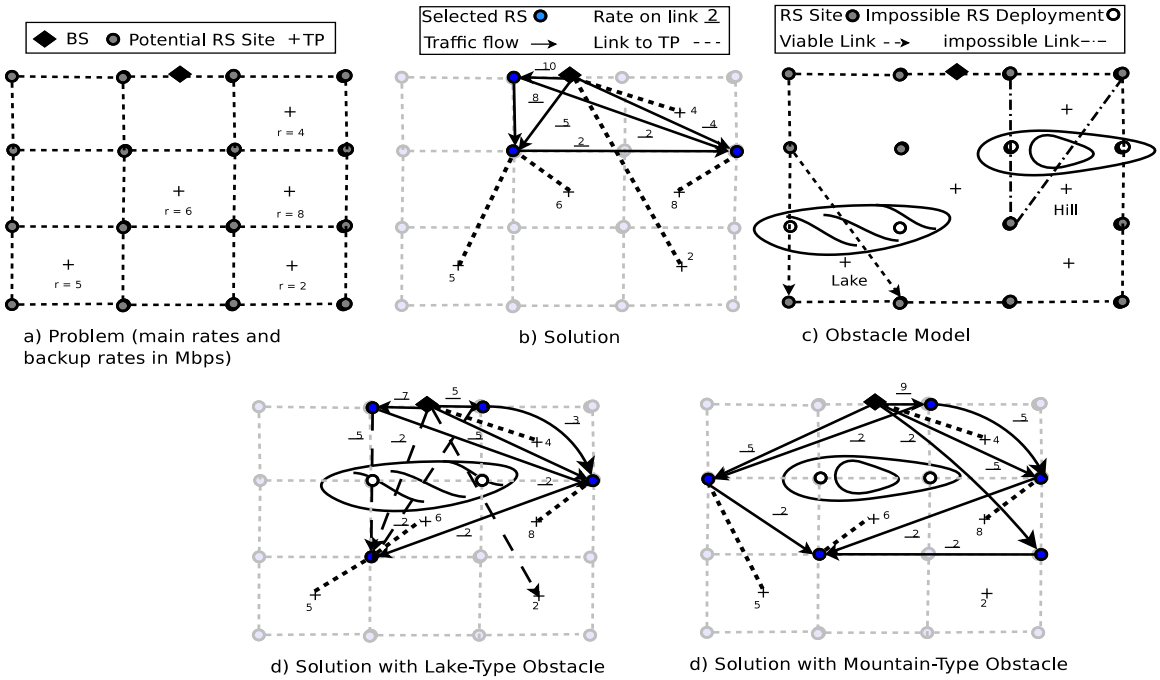

Figure 5.6.1: Planning RS Locations without Fault-Tolerance for Out-Band Model

The potential sites for an RS are the corners of a grid square. In the 4-by-4 grid, the RS sites are numbered 0 to 15, as shown in the figure. The possible site of a TP is in the center of a square. The TPs are numbered 0 to 9. In the figure, the TP numbers are 2, 4, 5, 6 and 8. The number shown in the figure next to each TP is its traffic demand in Mbps.

The maximum rate on the links is shown in Table 5.6.1. The distance unit is the side length of a square in the grid. The table shows the feasible rate for the corresponding distance interval (per the model in Section **??**). The solution to this planning example is shown in Fig. 5.6.1(b). The shaded RS sites are the ones that have been selected. Three RSs are needed for this problem, which are RS1, RS5 and RS7. The solid line links are the BS-RS and RS-RS links. The dashed lines are the BS-TP and RS-TP links. The underlined numbers are the link rates allocated by the solution. The rates of the dotted links are equal to the corresponding TPs' rates. The arrows on the links show the flow of traffic in the down-link to facilitate interpreting the results. However, the traffic may go in the up-link or down-link direction.

Table 5.6.1: LINK RATES Out-Band Mode

Distance (unit)	Link Rate (Mbps)
if distance <= 1	rate = 10
else if distance <= 2	rate = 5
else if distance <= 3	rate = 2
else if distance <= 4	rate = 1
else	rate = 0

We note the following observations from this example:

- The distance from the TP to the BS doesn't necessarily indicate a direct or relayed connection. For example, the TP with demand of 2 Mbps is the farthest from the BS. However, its demand is relatively low which can be satisfied by a single link. On the other hand, TPs which are closer to the BS have higher demands, and require the use of relay stations.

- Secondly, the RS-to-RS links help in reducing the number of relays. In our example, there is more traffic to the right of the BS (4+8+2 = 14) than the left (5) and the middle (6). Thus, in the solution, the diagonal and horizontal links, both with rate of 2 Mbps, between RSs (1,7) and (5,7), respectively, relay the traffic from the right side to the less congested left side. If this was not the case, more relays would be needed on the right side.

Obstacle Model In this paper, we considered the topological conditions such as the existence of obstacles in the coverage area that might occur in real-life. For example, there might be a lake in the coverage area that would prohibit the installation of an RS in the lake. However, the wireless link is able to go over the lake. Alternatively, there might be a small hill or a large building in the coverage area. A hill with a rough terrain would prohibit installing an RS on it and it would also block a wireless

link from traversing it. Our planning solution can accommodate both types of these obstacles.

In this part, we consider the existence of obstacles in the planning area. In real life, obstacles could be natural such as a hill, a forest or a lake, or man-made structures such as buildings, water towers or others. In this subsection, we consider two cases of obstacles as shown in Fig. 5.6.1(c). First, there might be lakes or other similar obstacles in the area. This type of obstacles does not allow deploying an RS in it. However, two adjacent RSs can transmit on a link that goes over this type of obstacle, as shown in the figure. In the model, the RS location over this obstacle is canceled by setting the maximum rate on such RS to null. In the model, these are the parameters for when RS_i is canceled: $m_i^{BR} = 0$ as the rate to the BS, $m_{ij}^{RR} = 0\ , \forall j\ \epsilon\ R$ as the rate to other RSs, and $m_{ik}^{RT} = 0\ , \forall k\ \epsilon\ T$ as the rate to all TPs.

The other type of obstacles that we consider doesn't allow a wireless link to penetrate it. For example, it could be a small hill in the area. We also consider that it might be rough terrain and it would not allow deploying an RS on it. This type is also shown in Fig. 5.6.1(c). The same conditions that were for the first type apply here to cancel the use of RSs that coincide with the obstacle. In addition, we need to cancel the links that traverse the obstacle too. So for every link that is blocked by the obstacle from the BS to RS_i, from the BS to TP_j or from RS_i to RS_k and from RS_i to TP_j , we have $m_i^{BR} = 0$, $m_j^{BT} = 0$, $m_{ik}^{RR} = 0$, $m_{ij}^{RT} = 0$.

Planning with Lake-Type Obstacles: Now we show how the planning is solved by our model with obstacles in the planning area. We take the planning case in Fig. 5.6.1(a) and we insert obstacles in it, once a lake and once a hill. Then we compare the planning outcome due to the effect of the obstacles.

First, we start by inserting a lake in the planning area as shown in Fig. 5.6.1(c). The lake covers the RS locations of RS5 and RS6. Hence, it is no longer possible to

deploy RSs at these sites. We note the following observations in comparison to Fig. 5.6.1(b).

- The number of RSs has changed from 3 to 4. This shows that an obstacle might increase the number of RSs. Previously, RS1, RS5 and RS7 were used. The obstacle makes RS5 no longer available. In the new planning, the RSs used are RSs 1, 2, 7 and 9.

- As we designated in the model, the links can traverse a lake-type obstacle. Hence, in the planning result, many links traversed the lake. Notably, from RS9, there are 3 links that go over the lake to RS1, the BS and RS2.

- Finally, the links that were direct from TPs to BS remain unchanged since the lake-type obstacle doesn't interfere with these links.

Planning with Mountain-Type Obstacles: In the following case, we insert a mountain-type obstacle in the planning area. This type of obstacle, as the lake-type, cancels the overlapping RS position. In addition, it cancels the links traversing it. Thus, the mountain-type obstacle is more restrictive than the lake-type. The result of the planning is shown in Fig. 5.6.1(e). The result shows that here, 5 RSs are now needed to support the TPs. This is intuitive since this obstacle has more restrictions than in the previous case, which required 4 RS, and compared to the initial case with no obstacles, which required 3 RSs.

We note the following observations:

- The network used links that are dispersed around the mountain, since it is not possible to go through it.

- Notice for TP4 (with rate of 2 Mbps), it was able to communicate directly with the BS in the case with no obstacles and in the case with lake obstacle. Now, it cannot do this anymore, because of the mountain, and it goes through RS2.

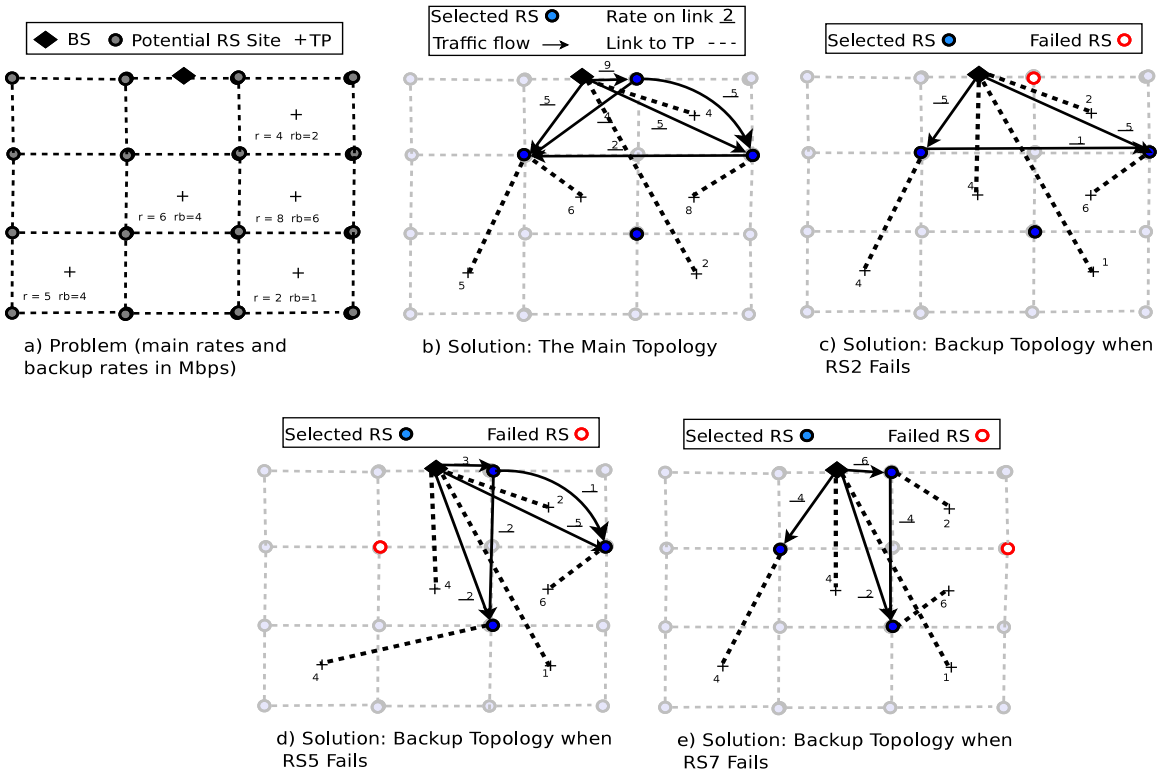

Figure 5.6.2: Planning RS Locations with Fault-Tolerance for Out-Band Model

- Finally, we saw that more restrictions in the planning area on the RS locations and the viable links will likely require more RSs to satisfy the TPs.

5.6.1.2 Planning with Fault-Tolerance

In this part, we present planning results with fault-tolerance. The problem input is shown in Fig. 5.6.2(a) and it has the same TP locations and rates as the example in Fig. 5.6.1(a). In this case, there are also backup rates for each TP, which are smaller than or equal to the the main rate.

Fig. 5.6.2(b) shows the main topology which supports the main rates of the TPs. This topology, similar to that in Fig. 5.6.1(b), supports the same normal operation rates, and also uses three RSs. However, unlike the topology in Fig. 5.6.1(b), it uses RS 2, RS 5 and RS 7. Moreover, it also requires the installation of an additional relay

at site RS 10. Even though RS 10 is not used in the main topology, it is required in case one of the three used RSs fails.

In Fig. 5.6.2(b), the TPs with rates of 4 and 2 Mbps connect directly to the BS since their direct link can support the required rate. This is similar to Fig. 5.6.1(b). Each of the other TPs connects to the RS that is closest to it.

Fig. 5.6.2(c) shows the backup topology that is used when RS 2 fails. In this topology, the backup rates are supported, which are smaller than the main rates in this example. Due to the lower rates, now three TPs are able to have a direct connection to the BS (compared to two in the main topology). The other two TPs connect through RSs. In this topology, RS 10 is also not used since RS 5 and RS 7 are able to support the TPs' demands, which makes the recovery from RS 2 failure faster, since RS 10 need not be used.

The topology in Fig. 5.6.2(d) is the case when RS 5 fails. In this case, three RSs are needed to support the TPs. Notice that in the previous topology, RS 5 was strategically located between the BS and the TP in the low-left corner. Since RS 5 has failed, there is no RS that can satisfy this TP. Thus, two RSs are used to connect this TP.

In Fig. 5.6.2(e), the topology that is used when RS 7 fails is shown. Now the TP in the low-left corner is able to connect via RS 5. However, the TP with demand 6 Mbps, which was 9 previously relying on RS 7 cannot connect with only one RS. Then, RS 2 and RS 10 convey the traffic of this TP in this case. However, the TP connects only to RS 10, and RS 10 connects to both the BS and RS 2 in order to receive the data. Finally, when RS 10 fails, we can continue to use the main topology as in Fig. 5.6.2(b) since this topology doesn't use RS 10.

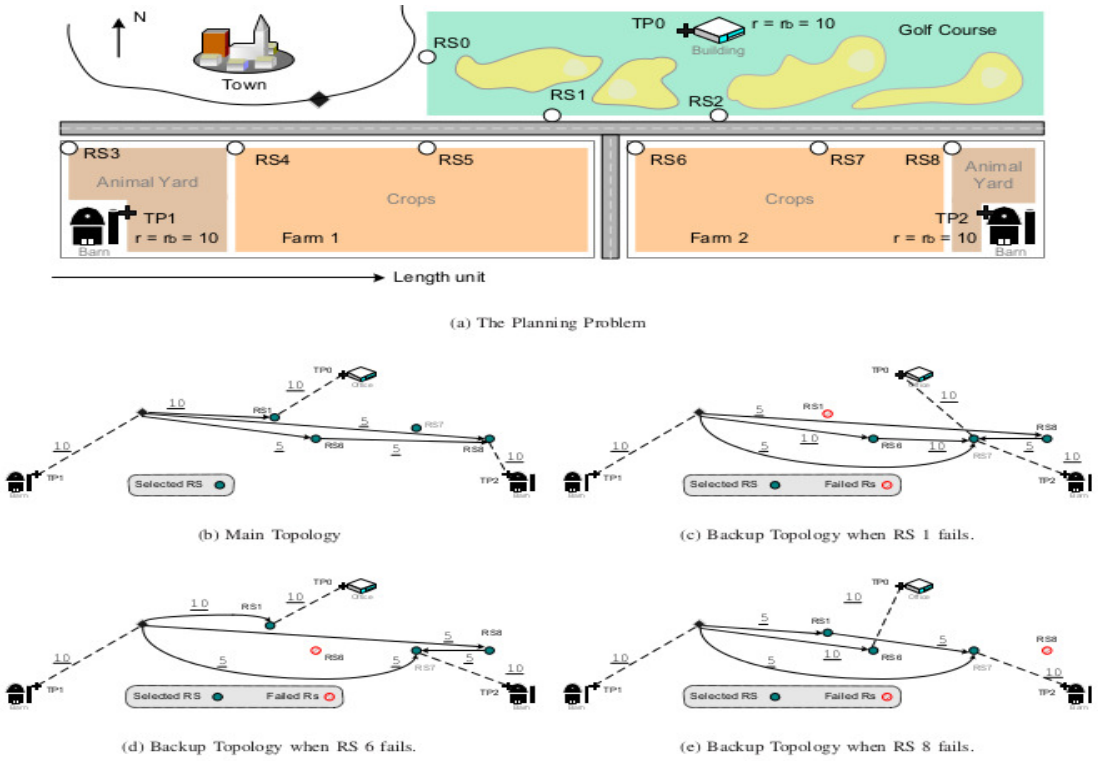

Figure 5.6.3: Planning of RS Locations in a Realistic Scenario. Fault-Tolerance is Provided for Out-Band Model

5.6.1.3 A Realistic Planing Case

In this part, we consider a case where we need to plan the locations of RSs. This scenario is similar to ones that could happen in real life where our solution can be used. This planning case is shown in Fig. 5.6.3(a).

On the top-left corner of the figure, there is a town that has Internet connectivity. However, there is no connectivity at the areas outside of the city limits. There is a BS at the edge of the town, marked by the black diamond.

The goal of this planning case is to connect the three properties adjacent to the town through the BS. They are a golf course east of the town, a farm south of the town and another farm southeast of the town. Each of the three entities has a single building: the main building at the golf course and the barns at the farms. In each

property, the TP is at the corresponding building. Each of the TP has a regular demand rate of $r_i = 10$ Mbps. In case of an RS failure, it is required to receive the same level of service, i.e., no drop in rate should be experienced. Therefore, $rb_i = 10$ Mbps.

Each property has provided possible locations for RSs to be deployed at their site. There are three RS sites at each property marked RS 0 through RS 8 in Fig. 5.6.3(a). In the layout of the properties, there are large areas that cannot have an RS, such as in the middle of the golf course or in the middle of the crop field. Therefore, the RS sites are located at the edges of fields. Our model can work with any layout of the potential RS sites. Once the X- and Y-coordinates of the RS sites, BS and TPs are known, our model generates the optimal number and locations of the RSs.

In this example, the data rates on a link are determined according to Table II. The length unit is shown in Fig. 5.6.3(a).

Fig. 5.6.3(b) shows the main topology. It turns out that 4 RSs are needed for this planning case. They are RS 1, RS 6, RS 7 and RS 8.

In the main topology, RS 7 is not used. In the main topology, TP 1 connects directly to the BS since it is at proximity and is able to obtain 10 Mbps through a direct link. TP 0, which is a bit farther, requires two hops, which go through RS 1. TP 2 is even further than TP 0. Therefore, it connects to the TP through RS 6 and RS 8.

In Fig. 5.6.3(c) the topology when RS 1 fails is shown. TP 1 doesn't have any change and still has a direct connection to the BS. But TP 2 connects to RS 7 even tough RS 8 is closer to it. This happens since we restrict the TP to connect to only one RS (or to the BS). RS 7 is closer to the BS and therefore is able to obtain a higher data rate than RS 8 can do. As a result, TP 2 is linked to RS 7.

In Fig. 5.6.3(d), a similar observation happens for TP 2 which connects through RS 7 even though RS 8 is closer to it than RS 7. In this topology, TP 0 uses a single

relay, RS 1, to reach the BS. And, finally, TP 1 also uses a direct connection to the BS.

Fig. 5.6.3(e) shows the topology when RS 8 fails. Here, TP 0 uses a two-hop connection through RS 6. TP 2 takes its traffic from RS 7.

5.6.2 Planning RS locations for in-band model

For in-band mode the interference between different links is taken into consideration and the maximum rate that can be assigned on any link is calculated according to the interference model listed in section (5.5.4). In the following planning case we are trying to present the effect of the interference consideration on the allocation of RSs using the planning with fault tolerance model and compare it with out-band model.

5.6.2.1 Planning with fault-tolerance

Planning results for the in-band with fault-tolerance model are presented to show the interference effect on the planning process. Interference is modeled such that the transmission from each BS to another RS or TP is interfered by transmission from other RSs within 1 unit grid, similarly the transmission from each RS to an RS or TB is interfered by the RSs within the 1-by-1 grid distance. Fig. 5.6.4(a) shows the locations of the TPs and RSs, in this in-band case eight RSs and eight TPs are used. For the case of in-band mode the different maximum link rates are calculated according to eq.(5.5.27 - 5.5.42) and the $SINR$ is calculated according to eq.(4.4.1). The main and backup rate requirement for each TP are shown in Fig. 5.6.4(a), where backup rates (rb_i) are smaller than the the main topology required rates (r_i). The The parameters used in the simulation are shown in Table 5.6.2.

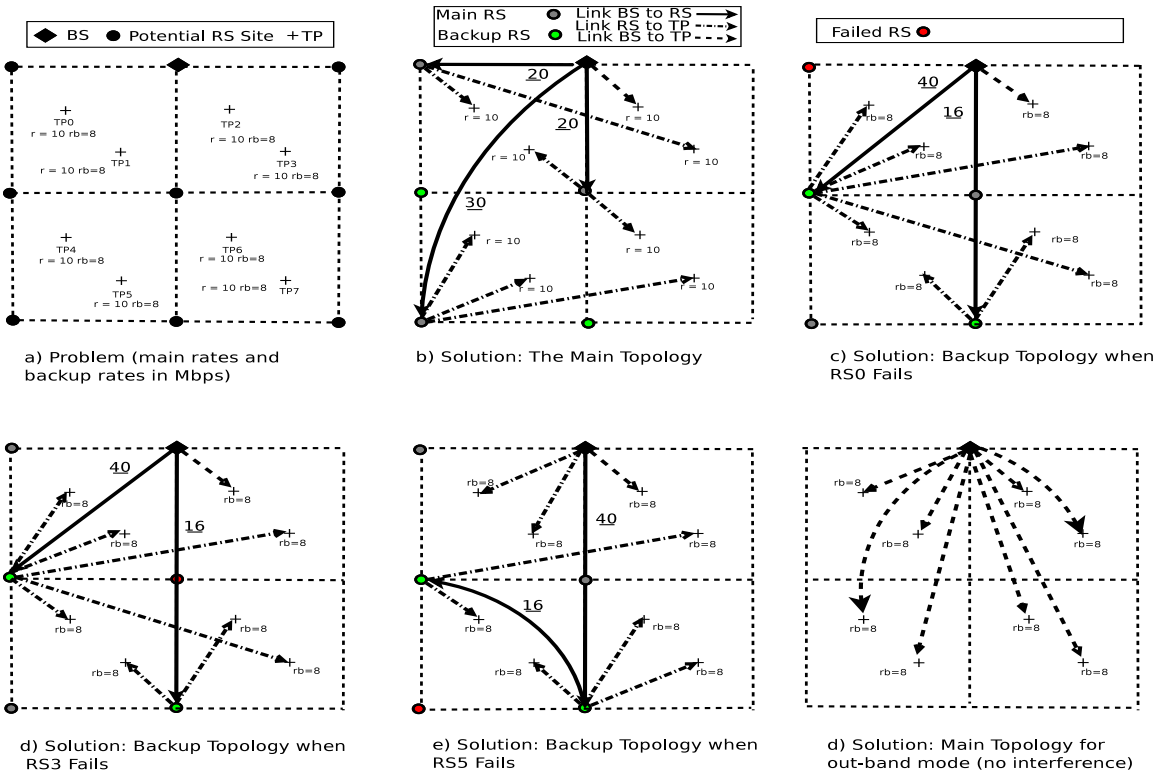

Figure 5.6.4: Planning RS Locations with Fault-Tolerance in In-Band Model

Table 5.6.2: System parameters

Description	Value
Band Width	5 MHZ
Transmitter Power	46 dBm / 39.81 W
path loss exponent (α)	2
Receiver Noise	-104 dBm
Coverage area	12 KM * 12 KM

Fig. 5.6.4(b) shows the main topology which supports the main rates to the TPs. This topology supports the equal normal operational rates for all TPs (10 Mbps), and uses three RS (RS 0, RS 3, RS 5) to satisfy all the TPs rate requirements. In Fig. 5.6.4(c,d,e) the backup topology for the in-band mode is presented, in case of any RS failure in the main topology, the network will implement the backup solution using two RSs (RS 2, RS 6). Both RSs will operate to support the backup rates (8 Mbps) to the TPs. The solution shows that the BS supports TP 2, RS 2 supports TPs (0, 1, 3, 4,7) and RS 6 supports TPs (5, 6) in both backup plans for (RS 0 and RS 3). However in backup plan for RS 5, TPs (0, 1, 2) are supported by the

BS. RS 6 supports TPs (5, 6, 7) and RS 2 supports TPs (3, 4). The increase in the number of RSs and the diversity of their locations and which TPs they serve are due to the consideration of the interference caused by the in-band mode. Results in Fig. 5.6.2(f) for the main topology without interference considerations shows that the TPs rate requirements are all satisfied by direct links from the BS. The reason is that the BS to TPs links maximum rates are capable of delivering the TPs rate requirements without any relaying. This comparison shows that modeling the problem with interference constraints requires RSs implementation, but in case of ignoring the interference in the model no relaying is required, in order to support the TPs with same rate requirements. The comparison clearly shows the importance of considering the interference effect in planning the RSs placement.

5.7 Conclusions

In this paper, we considered the problem of planning the RS locations in the WiMAX network in a fault-tolerant manner. To the best of our knowledge, this is the first work that provides fault-tolerance in planning RS locations in WiMAX. We provided a Mixed-Integer Linear Program (MILP) that formulates the planning problem. The allocation problem is studied in both the out-bound and in-bound relaying modes. We solved the problem with CPLEX and obtained numerical results that show how our model produces the main topology and the backup topologies of a network. Finally, we considered the existence of obstacles in the planning field, such as a large structure or a natural obstacle. We showed how our model can deal with these obstacles and plan the network around them effectively.

5.8 References

[76] Z. Abichar and J. Chang, "A medium access control scheme for wireless lans with constant-time contention," *Mobile Computing, IEEE Transactions on*, vol. 10, no. 2, pp. 191–204, Feb 2011.

[77] IEEE Std 802.16j-2009 (Amendment to IEEE Std 802.16-2009), *IEEE Standard for Local and metropolitan area networks Part 16: Air Interface for Broadband Wireless Access Systems Amendment 1: Multiple Relay Specification*, june 2009.

[78] IEEE Std 802.16m-2011 (Amendment to IEEE Std 802.16-2009), *IEEE Standard for Local and metropolitan area networks Part 16: Air Interface for Broadband Wireless Access Systems Amendment 3: Advanced Air Interface*, 2011.

[79] S. Peters and R. Heath, "The future of WiMAX: Multihop relaying with IEEE 802.16j," *Communications Magazine, IEEE*, vol. 47, no. 1, pp. 104–111, January 2009.

[80] Y. Yang, H. Hu, J. Xu, and G. Mao, "Relay technologies for WiMax and LTE-advanced mobile systems," *Communications Magazine, IEEE*, vol. 47, no. 10, pp. 100–105, October 2009.

[81] Z. Abichar, A. Kamal, and J. Chang, "Planning of Relay Station Locations in IEEE 802.16 (WiMAX) Networks," in *Wireless Communications and Networking Conference (WCNC), 2010 IEEE*, April 2010, pp. 1–6.

[82] J. Belschner, P. Arnold, H. Eckhardt, E. Kuhn, E. Patouni, A. Kousaridas, N. Alonistioti, A. Saatsakis, K. Tsagkaris, and P. Demestichas, "Optimisation of Radio Access Network Operation Introducing Self-x Functions: Use Cases, Algorithms, Expected Efficiency Gains," in *Vehicular Technology Conference, 2009. VTC Spring 2009. IEEE 69th*, April 2009, pp. 1–5.

[83] E. Chu, I. Bang, S. H. Kim, and D. K. Sung, "Self-organizing and self-healing mechanisms in cooperative small-cell networks," in *Personal Indoor and Mobile Radio Communications (PIMRC), 2013 IEEE 24th International Symposium on*, Sept 2013, pp. 1576–1581.

[84] J.-Y. Chang and Y.-S. Lin, "A clustering deployment scheme for base stations and relay stations in multi-hop relay networks," in *Journal Computers and Electrical Engineering*, vol. 40, no. 2, 2014, pp. 407 – 420.

[85] Y. Yu, S. Murphy, and L. Murphy, "Planning Base Station and Relay Station Locations in IEEE 802.16j Multi-Hop Relay Networks," in *Consumer Communications and Networking Conference, 2008. CCNC 2008. 5th IEEE*, Jan 2008, pp. 922–926.

[86] ——, "A Clustering Approach to Planning Base Station and Relay Station Locations in IEEE 802.16j Multi-Hop Relay Networks," in *Communications, 2008. ICC '08. IEEE International Conference on*, May 2008, pp. 2586–2591.

[87] B. Lin, P.-H. Ho, L.-L. Xie, and X. Shen, "Optimal Relay Station Placement in IEEE 802.16J Networks," in *Proceedings of the 2007 International Conference on Wireless Communications and Mobile Computing*, ser. IWCMC '07. New York, NY, USA: ACM, 2007, pp. 25–30.

[88] B. Lin and P.-H. Ho, "Dimensioning and location planning of broadband wireless networks under multi-level cooperative relaying," *Wireless Communications, IEEE Transactions on*, vol. 8, no. 11, pp. 5682–5691, November 2009.

[89] B. Lin, M. Mehrjoo, P.-H. Ho, L.-L. Xie, and X. Shen, "Capacity Enhancement with Relay Station Placement in Wireless Cooperative Networks," in *Wireless Communications and Networking Conference, 2009. WCNC 2009. IEEE*, April 2009, pp. 1–6.

[90] B. Lin, P.-H. Ho, L.-L. Xie, X. Shen, and J. Tapolcai, "Optimal Relay Station Placement in Broadband Wireless Access Networks," *Mobile Computing, IEEE Transactions on*, vol. 9, no. 2, pp. 259–269, Feb 2010.

[91] S.-J. Kim, S.-Y. Kim, H.-W. Lee, S.-W. Ryu, H.-W. Lee, and C.-H. Cho, "Multi-Hop Relay Based Coverage Extension in the IEEE802.16j Based Mobile WiMAX Systems," in *Networked Computing and Advanced Information Management, 2008. NCM '08. Fourth International Conference on*, vol. 1, Sept 2008, pp. 516–522.

[92] H.-C. Lu and W. Liao, "Joint Base Station and Relay Station Placement for IEEE 802.16j Networks," in *Global Telecommunications Conference, 2009. GLOBECOM 2009. IEEE*, Nov 2009, pp. 1–5.

[93] D. Niyato, E. Hossain, D. I. Kim, and Z. Han, "Relay-centric radio resource management and network planning in IEEE 802.16j mobile multihop relay networks," *Wireless Communications, IEEE Transactions on*, vol. 8, no. 12, pp. 6115–6125, December 2009.

[94] C.-Y. Chang, C.-Y. Chang, M.-H. Li, and C.-H. Chang, "A Novel Relay Placement Mechanism for Capacity Enhancement in IEEE 802.16j WiMAX Networks," in *Communications, 2009. ICC '09. IEEE International Conference on*, June 2009, pp. 1–5.

[95] Y. Yu, S. Murphy, and L. Murphy, "Interference Aware Relay Station Location Planning for IEEE 802.16J Mobile Multi-hop Relay Network," in *Proceedings of the 4th ACM Workshop on Performance Monitoring and Measurement of Heterogeneous Wireless and Wired Networks*, ser. PM2HW2N '09. New York, NY, USA: ACM, 2009, pp. 201–208.

[96] ——, "Planning Base Station and Relay Station Locations for IEEE 802.16j Network with Capacity Constraints," in *Consumer Communications and Networking Conference (CCNC), 2010 7th IEEE*, Jan 2010, pp. 1–5.

[97] C.-Y. Chang and M.-H. Li, "A placement mechanism for relay stations in 802.16j WiMAX networks," *Wireless Networks*, vol. 20, no. 2, pp. 227–243, 2014.

[98] A. Bari, A. Jaekel, and S. Bandyopadhyay, "Optimal Placement of Relay Nodes in Two-Tiered, Fault Tolerant Sensor Networks," in *Computers and Communications, 2007. ISCC 2007. 12th IEEE Symposium on*, July 2007, pp. 159–164.

[99] A. Bari, Y. Xu, and A. Jaekel, "Integrated Placement and Routing of Relay Nodes for Fault-Tolerant Hierarchical Sensor Networks," in *Computer Communications and Networks, 2008. ICCCN '08. Proceedings of 17th International Conference on*, Aug 2008, pp. 1–6.

[100] S. Misra, S. D. Hong, G. Xue, and J. Tang, "Constrained Relay Node Placement in Wireless Sensor Networks: Formulation and Approximations," *Networking, IEEE/ACM Transactions on*, vol. 18, no. 2, pp. 434–447, April 2010.

[101] J. Bredin, E. Demaine, M. Hajiaghayi, and D. Rus, "Deploying Sensor Networks With Guaranteed Fault Tolerance," *Networking, IEEE/ACM Transactions on*, vol. 18, no. 1, pp. 216–228, Feb 2010.

[102] X. Han, X. Cao, E. Lloyd, and C.-C. Shen, "Fault-Tolerant Relay Node Placement in Heterogeneous Wireless Sensor Networks," in *INFOCOM 2007. 26th IEEE International Conference on Computer Communications. IEEE*, May 2007, pp. 1667–1675.

[103] W. Zhang, G. Xue, and S. Misra, "Fault-Tolerant Relay Node Placement in Wireless Sensor Networks: Problems and Algorithms," in *INFOCOM 2007.*

26th IEEE International Conference on Computer Communications. IEEE, May 2007, pp. 1649–1657.

[104] P. Meena, D. Gurjar, A. Singh, and S. Verma, "Optimal positioning of base station in wireless sensor networks: A survey," in *Intelligent Computing, Networking, and Informatics*, ser. Advances in Intelligent Systems and Computing, D. P. Mohapatra and S. Patnaik, Eds. Springer India, 2014, vol. 243, pp. 1135–1143.

[105] A. Samson Arun Raj, K. Ramalakshmi, and C. Priyadharsini, "A Survey on Classification of Fault Tolerance Techniques Available in Wireless Sensor Network," in *International Journal of Engineering Research and Technology*, vol. 3, no. 1, 2014.

[106] M. Nivedita and G. Raja, "Efficient relay station placement strategy for broadband wireless networks - 4G," in *International Conference on Recent Trends In Information Technology (ICRTIT)*, April 2012, pp. 282–286.

[107] H. Wang, X. Yin, C. Chen, and X. Wang, "DPRP: Dual-path relay placement in WiMAX mesh networks," in *IEEE Wireless Communications and Networking Conference (WCNC)*, April 2013, pp. 1597–1602.

[108] W. Stallings, *Wireless Communications and Networks*. Prentice Hall, 2001.

[109] IEEE Std 802.16-2004 (Revision of IEEE Std 802.16-2001), *IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Fixed Broadband Wireless Access Systems*, 2004.

[110] T. Tsourakis and K. Voudouris, "WiMax network planning and system's performance evaluation," in *Wireless Communications and Networking Conference, 2007.WCNC 2007. IEEE*, March 2007, pp. 1948–1953.

[111] M. Molina-Garcia and J. Alonso, "Planning and Sizing Tool for WiMAX Networks," in *Radio and Wireless Symposium, 2007 IEEE*, Jan 2007, pp. 403–406.

[112] J. Garcia-Fragoso and G. Galvan-Tejada, "Cell planning based on the WiMax standard for home access: a practical case," in *Electrical and Electronics Engineering, 2005 2nd International Conference on*, Sept 2005, pp. 89–92.

CHAPTER 6. GENERAL CONCLUSIONS

6.1 General Discussion

This thesis is organized in a paper based format where each chapter include a manuscript published or prepared to be submitted into a scholarly journals or a conference proceedings. The thesis consists of four chapters, each chapter presents one of the papers included in this thesis and a general introduction and conclusion. In this thesis the problems of physical resources allocation, spectrum allocation, network architecture and fault tolerance cell planning is addressed. Novel solutions for the problems is proposed to enhance the network performance in terms of physical and spectrum resources allocation. New architecture is proposed to alleviate the allocation process and a fault tolerance cell planning is suggested to guarantee the provided service level. Each problem is formulated to achieve the required objectives with all subjected constraints and different feasibility programing methods such as integer linear programming (ILP), mixed ILP, stochastic programming are utilized. The optimal solutions for the aforementioned problems are calculated using Cplex®. Due to the hight complexity nature of this category of optimization problems, in cases where applicable a suboptimal solution is suggested for practical implementation purposes. A simulation is conducted to alleviate the comparison of the proposed solutions with the state of the art solutions existing in the literature. The proposed solutions shows

an adequate performance that can help mobile service providers increase their return on investment.

6.2 Recommendations for Future Research

The work in this thesis addressed different aspects of the physical and spectrum resources allocation that can be extended by future research work. The problem of controlling the resource metric information to enhance the allocation process is recommended to achieve more efficient physical and spectrum resources allocation in different frequency partitions. Another area of improvement is the consideration of longer period of historical data in the stochastic program using the proposed network architecture and the decision support system to enhance the learning process from empirical data collected in the data collection system. Other learning algorithms are also beneficial to explore and compared to accelerate the allocation process. Finally the work of fault tolerance in cell planning using relay stations in WiMax networks is recommended to be extended to small cells HetNets with the involvement of self organizing networks in a self healing framework as these are the most recent technologies that gain more research interest.

ACKNOWLEDGEMENTS

I would like to express my gratitude to my academic advisor, Dr. Morris Chang and Dr. Ahmed Kamal, for all there valuable advice, guidance, and support during my doctoral program. There continued support was always a drive for me both on the technical and the professional levels. Special thanks are also extended to the distinguished faculty members, who served on my thesis committee—Dr. Lu Ruan, Dr. Zhao Zhang , and Dr. Sang Kim, for all their guidance and constructive feedback. I would like to extend my sincere thanks to my family—my wife for her continuous support and understanding all through the years of my Ph.D. work, my mother, for being my driving force, my brother, for his continuous encouragement, and my aunt for her support during the days of this journey. This dissertation would not have been possible without my strong family's support and pride in this accomplishment. I would also like to extend my gratitude to my dearest friend, Eng. Tarek Elkayat, for his valuable technical and moral support during my Ph.D. years.