Resource Allocation and End-to-End Quality of Service for Cellular Communications Systems in Congested and Contested Environments

Mohammad Ghorbanzadeh

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Electrical Engineering

Thomas C. Clancy, Chair
Luiz A. DaSilva
Robert W. McGwier
Harpreet S. Dhillon
Alireza Haghighat

October 23, 2015
Arlington, Virginia

Keywords: Resource Allocation, Radio Resource Blocks, Discrete Optimization, Quality of Service (QoS), Secure Auctions, LTE, Convex Optimization, Radar Spectrum Sharing, Channel Conditions.

©Copyright 2015, Mohammad Ghorbanzadeh

Resource Allocation and End-to-End Quality of Service for Cellular Communications Systems in Congested and Contested Environments

Mohammad Ghorbanzadeh

ABSTRACT

This research addresses the concept of radio resource allocation for cellular communications systems operating in congested and contested environments with an emphasis on end-to-end quality of service (QoS). The radio resource allocation is cast under a proportional fairness formulation which translates to a convex optimization problem. Moreover, the resource allocation scheme considers subscription-based and traffic differentiation in order to meet the QoS requirements of the applications running on the user equipment in the system. The devised resource allocation scheme is realized through a centralized and a distributed architecture and solution algorithms for the aforementioned architectures is derived and implemented in the mobile devices and the base stations. The sensitivity of the resource allocation scheme to the temporal dynamics of the quantity of the users in the system is investigated. Furthermore, the sensitivity of the resource allocation scheme to the temporal dynamics in the application usage percentages is accounted for. In addition, a transmission overhead of the centralized and distributed architectures for the resource allocation schemes is performed. Furthermore, the resource allocation scheme is modified to account for a possible additive bandwidth done through spectrum sharing in congested and contested environments, in particular spectrally coexistent radar systems. The radar-spectrum additive portion is devised in a way to ensure fairness of the allocation, high bandwidth utilization, and interference avoidance. In order to justify the aforesaid modification, the interference from radar systems into the Long Term Evolution (LTE) as the predominant 4G technology is studies to confirm the possibility of the spectrum sharing. The preceding interference analysis contains a detailed simulation of radar systems, propagation path loss models, and a third generation partnership project compliant LTE system. The propagation models are Free Space Path Loss (FSPL) and Irregular Terrain Model (ITM). The LTE systems under consideration are macro cell, outdoor small cells, and indoor small cells. Furthermore, the resource allocation under channel consideration is formalized such that the resources are allocated under a congested environment and based on the quality of channel the users have in the network as well as the quality of service requirements of the applications running on the mobile devices.

Dedication

To the glory of God who loved the world so much that gave his only Son that whosoever believes into Him shall not perish, but have everlasting life.

To my wife Diane for all her sacrifices, unconditional love, and consistent care.

Acknowledgments

I would like to thank to my PhD advisor, Dr. Charles T. Clancy, for supporting me during these past four years. Charles is someone you will instantly be fond of. He's the funniest advisor and one of the smartest people I know. I hope that I could be as lively, enthusiastic, and energetic as Charles and to someday be able to command an audience as well as he can. Charles has been supportive and has given me the freedom to pursue various projects without objection. He has also provided insightful discussions about the research. I am also very grateful to Charles for his scientific advice and knowledge and many constructive discussions and suggestions. He is my primary resource for getting my science questions answered and was instrumental in helping me crank out this thesis in a short time. I also have to thank the members of my PhD committee, Dr. Luiz A. DaSilva , Dr. Robert W. McGwier, Dr. Harpreet S. Dhillon, and Dr. Alireza Haghighat for their helpful advice and suggestions in general. I will forever be thankful to Dr. Ahmed Abdelhadi, while he was a research professor at Virginia Tech. He provided great discussion many times during my graduate school career. He was and remains a fantastic role model for a scientist, mentor, and teacher. I also thank Dr. Eugene Visotsky and Mr. Prakash Moorut of Nokia Networks for providing insights in various parts of this thesis. I wish all these people the best in their future life and career endeavors. I specially thank my wife Diane for her immense sacrifice during the term of my graduate school and for her encouragement, without which I would have not been able to finish this thesis. I thank my mom, Azam, my dad, Rahim, my brother, Babak, and my aunt, Eshrat who have always been great encouragement for me during my graduate school years. My hard-working parents have sacrificed their lives for my brother and myself and provided unconditional love and care. I love them so much, and I would not have made it this far without them. The best outcome from these past four years is finding my best friend, soulmate, and wife. I married the best person out there for me. There are no words to convey how much I love her. Diane has been a true and excellent supporter and has unconditionally loved me during my good and bad times. she has been non-judgmental of me and instrumental in instilling confidence. She has faith in me and my intellect even when I felt like digging hole and crawling into one because I didnt have faith in myself. These past several years have not been an easy ride, both academically and personally. I truly thank Diane for sticking by my side, even when I was irritable and depressed. I feel that what we both learned a lot about life and strengthened our commitment and determination to each other and to live life to the fullest.

Contents

Dedication **iii**

Acknowledgments **iv**

Contents **v**

List of Figures **x**

List of Tables **xxi**

1 Quality of Service in Communications Networks **1**

1.1 Current Trends in End-to-End QoS in Cellular Networks 5

1.1.1 QoS in Wired IP Networks 6

1.1.2 QoS in Wireless Networks 7

1.1.2.1 Link Layer QoS 7

1.1.2.2 Network Layer QoS 9

1.1.2.3 Application Layer QoS 10

1.1.2.4 Hardware QoS 10

1.1.2.5 Cross Layer QoS 11

1.2 Spectrum as a Congested and Contested Environment 13

1.3 Modern-Day Networks Quality of Service Challenges 15

1.4 Problem Statement 20

1.5 Contributions 25

1.6 Chapters Organization 27

2 Utility Functions, Optimization, and Resource Allocation **30**

2.1 Application Utility Functions 31

2.2 Resource Allocation and its Formulations 37

2.2.1 Max-Min Resource Allocation 38

2.2.2 Proportional Fairness 39

2.2.3 Frank Kelly Algorithm 40

2.3 Previous Studies on Resource Allocation 41

2.4 Chapter Summary 44

3 A Centralized Architecture for Resource Allocation **45**

3.0.1 Contributions 46

3.1 Resource Allocation Optimization 47

3.1.1 System Model 48

3.2 Centralized Architecture 54

3.3 Existence of Global Optimal Solution 54

3.4 Centralized Solution Algorithm 55

3.5 Centralized Architecture Simulation 57

3.5.1 Rate Allocation and Bids for $10 \leq R \leq 200$ 61

3.6 Centralized Architecture Real-World Implementation 62

3.6.1 Implementation Results 64

3.7 Chapter Summary 70

4 A Distributed Architecture for Resource Allocation **72**

4.0.1 Contributions 73

4.1 Distributed Architecture 74

4.1.0.1 EURA Optimization Problem 75

4.1.0.2 IURA Optimization Problem 76

4.2 Distributed Architecture's Global Optimal Solution 77

4.2.1 EURA Global Optimal Solution 77

4.2.2 IURA Global Optimal Solution 78

4.3 Distributed Solution Algorithm 79

4.3.1 EURA Algorithm 79
4.3.1.1 EURA Convergence Analysis 82
4.3.1.2 EURA Robust Algorithm 86
4.3.2 IURA Algorithm 88
4.4 Mathematical Equivalence 88
4.5 Distributed Architecture Simulation 94
4.5.1 Rate Allocation and Bids for $10 \leq R \leq 200$ 98
4.5.2 Pricing for $10 \leq R \leq 200$ 103
4.6 Benchmark 104
4.7 Chapter Summary 106

5 Traffic and Sensitivity Analysis of Resource Allocation Architectures 108
5.0.1 Contributions 109
5.1 Traffic/Sensitivity Analysis under UE Quantity Dynamics 111
5.1.0.1 Distributed Architecture with Rebidding 116
5.1.0.2 Distributed Architecture without Rebidding 119
5.1.0.3 Centralized Architecture 122
5.2 Traffic/Sensitivity Analysis under Application Usage Dynamics 123
5.2.0.4 Distributed Architecture with Rebidding 124
5.2.0.5 Distributed Architecture without Rebidding 126
5.2.0.6 Centralized Architecture 129
5.3 Computational Complexity Considerations 130
5.4 Traffic Analysis Simulation 130
5.4.1 Transmission Overhead Analysis 131
5.4.2 Sensitivity to Application Usage Changes 132
5.4.3 Sensitivity to UE Quantity Changes 136
5.5 Chapter Summary 138

6 Radio Resource Block Allocation Architecture 141
6.1 Introduction 141
6.2 Discrete Optimization and Continuous Optimization Relaxation 145

6.2.1 Resource Block Optimization Formulation 148
6.2.2 Continuous Rate Allocation Relaxed Optimization Problem 149
6.3 Resource Block Allocation Optimization Solution 153
6.4 Resource Block Allocation Simulation Results 157
6.4.1 Algorithm Convergence for R =50 159
6.4.2 Algorithm Convergence for $50 \leq R \leq 100$ 162
6.4.3 Resource Block Allocation 165
6.5 Chapter Summary 169

7 Practical Challenges in Resource Allocation 171
7.1 Radio Resource Allocation in Congested/Contested Environments 171
7.1.1 Contributions 173
7.2 Architecture for Fair Allocation 174
7.2.1 Congested Environment Radio Resource Allocation Optimization . . 176
7.2.2 Solution Algorithm 178
7.3 Practical Examples for Congested Environments and Resource Assignment . 181
7.4 Radar Systems and Interference into LTE 184
7.4.1 Radar Simulation 184
7.4.2 Macro cell, Micro cell, and Indoor LTE Simulation 187
7.4.3 Propagation Model Simulation 190
7.4.4 Radar to Macro Cell LTE UL Interference 192
7.4.5 Radar to Outdoor Small Cell LTE UL Interference 193
7.4.6 Radar to Indoor Small Cell LTE UL Interference 195
7.5 Resource Allocation Simulation under Radar Interference 197
7.6 Chapter Summary 200

8 Resource Allocation, LTE, and Channel Effect 203
8.0.1 Contributions 204
8.0.2 Organization 205
8.1 Channel Quality, Modulation, and Coding 205
8.2 Radio Resource Management 208

8.3 Resource Allocation Efficacy and Channel Conditions 210
8.4 Channel-Aware Distributed Resource Allocation Formulation 212
8.4.1 Determining ϵ 214
8.4.2 Determining K 216
8.5 Global Solution Existence 217
8.5.1 Solution for Channel-Aware EURA Optimization 220
8.5.2 IURA Global Optimal Solution 223
8.6 Simulation Results 225
8.6.1 Large Scale Network Planning 228
8.7 Chapter Summary 312

9 Conclusions **315**
9.1 Findings 315
9.2 Future Trajectory 320

References **322**

Index **335**

List of Figures

1.1 Fixed, mobile, and mobile broadband subscribers' growth: The mobile and mobile broadband subscriber's quantity have observed a dramatic growth from 2008 to 2014 [1]. 2

1.2 Cross Layer QoS 12

1.3 NTIA set radar-comm. exclusion zones reaching inland to a large extent. Such separations preclude the viability of the 3.5 GHz band for mobile broadband and excludes catering communications services to more than 40% of the U.S. population [2]. 15

1.4 Thesis Organization. 29

2.1 Utility function $U(x)$ for delay-tolerant applications vs. rate allocation x. . . 32

2.2 Utility function $U(x)$ for hard real-time applications vs. rate allocation x. . . 33

2.3 Utility function $U(x)$ for soft real-time applications vs. rate allocation x. . . 34

2.4 Delay-tolerant and real-time applications have respectively logarithmic and sigmoidal utility functions U_{ij} plotted against the application-assigned rates r_{ij}. 37

3.1 System Model: Single cell, within the cellular network, with an eNB covering $M = 6$ UEs each with simultaneously running delay-tolerant and relay-time applications represented by logarithmic and sigmoidal utility functions respectively. 49

3.2 Centralized Algorithm: Resources are allocated to the applications running on the UEs in a monolithic stage, in which UEs transmit their application utility parameters to their eNB, which calculates the optimal application rates and transmit them to the germane UEs. 56

3.3 The system contains 6 UEs, each concurrently running a delay-tolerant and real-time application with respective identically colored logarithmic and sigmoidal utility functions U_{ij} vs. the application-assigned rates r_{ij} plots in Figure 3.3(a). Utility slope curvature functions, the first derivative of the application utility natural logarithms S_{ij} with respect to the application rates r_{ij} are illustrated in Figure 3.3(b) where identical colors relate to the applications on one UE. 58

3.4 Figure 3.4(a) plots the aggregated utilities, multiplications of the usage-percentage-powered application utility functions $V_i(r_i)$ vs. the UE rates r_i, where $i \in \{1, ..., 6\}$. Figure 3.4(b) illustrates the aggregate slope curvature functions , first derivative of the the aggregate utility natural logarithms $S_i(r_i)$. Furthermore, decay function-induced robustness effect is depicted; As we can see, the lack of decay functions yields in the system instability revealed in the shadow price oscillation. 60

3.5 Figure 3.5 depicts the optimal rates allocated to the UEs by the distributed scheme vs. BS resources. No user is dropped as no assignment is zero. 62

3.6 Figure 3.6 depicts the optimal application rates r_{ij} vs. the eNB rate R. Applications running on an UE are identically colored. As we can see, real-time applications are initially allocated more resources as opposed to the delay-tolerant ones due to their urgent need for resources. 63

3.7 Implementation system model contains UEs, WiFi-connected to the Internet, run delay-tolerant and real-time applications whose optimal rates are assigned by the resource block (RB) entity on which the centralized resource allocation is instrumented. The RB is installed on the router which shapes the traffic based on the rates allocated by the resource allocation scheme. 64

3.8 Real-world network architecture contains two VMs "Guest 1" and "Guest 2" which respectively host the router and resource allocation method implemented on the the resource block (RB) entity. Furthermore, an office-Internet virtual switch for external devices, three phone-network virtual switches, and a network management virtual switch are created on a two physical two PCI-ported IBM server. The Phones and WiFi access point (AP) have their private network with IP address 192.168.2.1. 66

3.9 Wireshark Throughput Analysis without applying the resource allocation to the network in Figure 3.8 with the parameters in Table 3.3 under an overall bandwidth constraint $R = 1$ Mbps and no traffic shaping. Black/Green/Red curve shows the network/YouTube 1/HTTP 1 throughput. HTTP 1 download completed in 1200 s with YouTube 1 incurring multiple buffering, adversely affecting the QoE as HTTP 1 (red curve) uses the entire available capacity (hits the black curve). 68

3.10 Wireshark throughput analysis with applying the resource allocation to the network in Figure 3.8 with the parameters in Table 3.3 under the bandwidth constraint $R = 1$ Mbps and shaping. Black/Green/Red curve shows the network/YouTube 1/HTTP 1 throughput. HTTP 1 download completed in 2650 s with no YouTube 1 buffering occurrences. Occasionally, one application rate zeros and the other application utilizes the maximum bandwidth to achieve a consistent allocation to elevate users' QoE. 69

4.1 System Model: Single cell, within the cellular network, with an eNB covering $M = 6$ UEs each with simultaneously running delay-tolerant and relay-time applications represented by logarithmic and sigmoidal utility functions respectively. 75

4.2 The centralized resource allocation assigns application after receiving the request from each application. The problem can simply be decomposed to network and device optimization where the BS only assigns UE rates, and the UE assigns application rates. 89

4.3 The system contains 6 UEs, each concurrently running a delay-tolerant and real-time application with respective identically colored logarithmic and sigmoidal utility functions U_{ij} vs. the application-assigned rates r_{ij} plots in Figure 4.3(a). Utility slope curvature functions, the first derivative of the application utility natural logarithms S_{ij} with respect to the application rates r_{ij} are illustrated in Figure 4.3(b) where identical colors relate to the applications on one UE. 96

4.4 Figure 4.4(a) plots the aggregate utilities, multiplications of the usage-percentage-powered application utility functions $V_i(r_i)$ vs. the UE rates r_i, where $i \in \{1, ..., 6\}$. Figure 4.4(b) illustrates the aggregated slope curvature functions , first derivative of the the aggregated utility natural logarithms $S_i(r_i)$. Furthermore, decay function-induced robustness effect is depicted; As we can see, the lack of decay functions yields in the system instability revealed in the shadow price oscillation. 97

4.5 UE Optimal Rates: This figure depicts the optimal rates allocated to the UEs by the distributed scheme vs. eNB resources. No user is dropped as no assignment is zero. 99

4.6 UE Bids: This figure illustrates the UE bids for acquiring the resources vs. the eNB rate. The applications requiring more resources bid higher. When bandwidth is scarce, applications needing more resources bid significantly higher than the others. The plots reveal that the higher bid are tantamount to receiving more resources. 100

4.7 Application Optimal Rates: This figure depicts the optimal application rates r_{ij} vs. the eNB rate R. Applications running on an UE are identically colored. As we can see, real-time applications are initially allocated more resources as opposed to the delay-tolerant ones due to their urgent need for resources. . . 101

4.8 Application Bids: This figure illustrates the applications bids in the UEs. The real-time applications bid higher when the resources are scarce, while the opulence of eNB resources escalates the application rates and reduces the UE bids. 102

4.9 Shadow Price p vs. eNB Resources R: Availability of more eNB resources reduces the shadow price. 104

4.10 Comparison of the resources allocated by our proposed method and sigmoidal-fitted logarithm. 106

5.1 System Model: Single cell of the cellular network with a base station (BS) covering $M = 6$ UEs simultaneously running delay-tolerant and relay-time applications represented by logarithmic and sigmoidal utility functions respectively. 111

5.2 Transmission Overhead vs. BS Rate R: The centralized resource allocation scheme incurs a drastically less transmission overhead as opposed to the distributed approach, as it requires less message exchanges before it converges to the optimal rates. 132

5.3 The UEs allocated rates r_i and pledged bids w_i vary in time as the application usage percentages change. Figures 5.3(a) and 5.3(b) respectively depict the UE rate and bid changes as the applications usages vary in time. The time procession reduces the rate and bid jumps due to the convergence. 135

5.4 The shadow price varies with temporal changes of the application usage percentages. The more usage, the higher the shadow price. 136

5.5 Transmission Overhead: The plots compares the centralized and distributed algorithms with respect to their transmission overhead when application usage percentages change. The centralized algorithm is independent of the termination threshold δ, whereas the smaller thresholds causes more iterations of the algorithm leading to escalated transmission overheads. 136

5.6 Comparisons between the centralized and distributed architectures divergence from optimal solutions when a new users enters the cellular network. This figure depicts the rate error r_i when UE6 enters the network. 138

5.7 UE Bids with Temporally Changing Applications Usages: This figure shows the bid errors when UE6 enters the network and depicts that the rates/bids remain optimal for the centralized architecture. 139

5.8 UE Bids with Temporally Changing Applications Usages: This figure portrays the shadow price error for the distributed and centralized architectures, where the latter introduces no errors. 139

6.1 Global Mobile Subscriptions between Q1 - Q2 2013 (Adopted from [3]): Lower (Upper) histograms depict the total subscription quantity (subscription augmentation quantity). Such enormous escalations of the number of subscriptions make dynamic resource allocation indispensable. 143

6.2 Monthly mobile subscribers' traffic has undergone drastic increases between 2007-2012 (Adopted from [3]). Data, as opposed to voice, has the lion's share of the generated traffic increase, so resource allocation strategies with an attention to the type of data traffic is desirable. 144

6.3 Exemplar System Utility Function: Two UE system with rates r_1 and r_2 where $U(r_1, r_2) = U(r_1)U(r_2)$ is the system utility function. As annotated, the minimum in the plot happens at $X = r_2 = 54.2$ and $Y = r_1 = 64.3$, so it is the optimum rate allocation, and the value of the utility function is $U(r_1, r_2) = 20$. 154

6.4 Exemplar Boundary Points: The blue point on r_1 - r_2 plane represents the continuous rate assigned to r_1 and r_2 $(r_1, r_2) = (54.2, 64.3)$. The red point $(r_1, r_2) = (54, 64)$, green point $(r_1, r_2) = (54, 65)$, yellow point $(r_1, r_2) = (55, 64)$, and cyan point $(r_1, r_2) = (55, 65)$ represent the discrete rates possible and the closest to the blue point. 155

6.5 Experiment Set-up: 6 UEs are under coverage of an eNB, where UEs 1, 2, and 3 (4, 5, and 6) are running real-time (delay tolerant) applications represented by sigmoidal (logarithmic) utility functions. The goal is for eNB to allocate resource blocks to UEs. 158

6.6 Utility Functions for the Discrete Optimization Problem: Each curve represents utility function for one of the UEs in the experiment, where 3 are delay tolerant applications corresponding to logarithmic utility functions and 3 are real-time applications corresponding to sigmoidal utility functions. The horizontal axis represents the resource blocks of size one unit bandwidth. . . 159

6.7 Utility Functions for the Continuous Optimization Problem: The horizontal axis represents continuous valued rates, and utilities as continuous functions of the rates. Each curve represents utility function for one of the UEs, where 3 are delay tolerant applications corresponding to logarithmic utility functions and 3 are real-time applications corresponding to sigmoidal utility functions. 160

6.8 Continuous Rate Allocations vs. Iterations for the Relaxed Problem: Red, green, and blue curves are for VoIP, standard streaming video, and HDV applications in UEs 1, 2, and 3, and cyan, purple, and yellow curves are the rates allocated to FTP applications in UEs 4, 5, and 6. Higher rates are allocated to real-time applications (UEs 1, 2, and 3) vis-a-vis delay tolerant ones (UEs 4, 5, and 6). 161

6.9 Continuous Allocations Bids vs. Iterations for the Relaxed Problem: Red, green, and blue curves are bids for UEs running VoIP, standard streaming video, and HDV applications that is UEs 1, 2, and 3, and cyan, purple, and yellow curves are the bids performed by FTP applications in UEs 4, 5, and 6. Real-time applications (UEs 1, 2, and 3) initially bid higher so that more rates are allocated to them as opposed to the delay tolerant ones (UEs 4, 5, and 6). 162

6.10 Shadow Price vs. Iterations for the Relaxed Problem: As algorithm calculates continuous rates and shadow prices, the value of the shadow price stabilizes in the steady state. These values are sent to all the UEs by the eNB in each iteration of the algorithms. 163

6.11 Continuous Rate Allocations vs. eNB Bandwidth for the Relaxed Problem: Steady state continuous allocated rates to the UEs increase as the eNB has more available resources. Red, green, and blue curves are bids for UEs running VoIP, standard streaming video, and HDV applications that is UEs 1, 2, and 3, and cyan, purple, and yellow curves are the bids performed by FTP applications in UEs 4, 5, and 6. 164

6.12 UE Bids vs. eNB Bandwidth for the Relaxed Problem: As more resources become available to the eNB, applications bid lower to have bandwidths allocated to them. Red, green, and blue curves are bids for UEs running VoIP, standard streaming video, and HDV applications that is UEs 1, 2, and 3, and cyan, purple, and yellow curves are the bids performed by FTP applications in UEs 4, 5, and 6. 164

6.13 Shadow Price vs. eNB Bandwidth for the Relaxed Problem: As more resources become available for the eNB, the shadow price decreases. This is in line with the fact that applications bid lower once resources are opulent. 165

6.14 Resource Block Allocation versus eNB Bandwidth for the Discrete Optimization Problem: Red, green, and blue curves are bids for UEs running VoIP, standard streaming video, and HDV applications that is UEs 1, 2, and 3, and cyan, purple, and yellow curves are the bids performed by FTP applications in UEs 4, 5, and 6. Other discrete allocations are also feasible. 166

6.15 Computational Complexity: Red curve shows the semi-log computational complexity as a function of the number of UEs once we do not exclude non-violating sequence possibilities from the proposed methodology. The blue curve is the semi-log computational complexity once we have applied the technique to exclude the sequences which violate the constraint for the continuous relaxed problem. 168

7.1 FCC spectrum allocation map defines the spectrum assignments in the US. . 173

7.2 Cellular communications system: Cell sector colors indicate frequency bands (f_1, f_2, f_3) and identically colored sectors imply frequency reuse whose topological pattern minimizes the inter-cell interference. The radar and pink sectors operations in the same frequency band f_3 create interferes. 175

7.3 Three sigmoidal (purple, green, red) and logarithmic (yellow, pink, cyan) utilitiy functions for real-time and delay tolerant applications. 176

7.4 Resource Allocation Stage Diagram: This illustration only depicts one stage of the allocation whereas the assignment is performed once over the radar spectrum (stage 1) then over the carrier spectrum where radar is not operating (stage 2). The procedure in both stages is similar with the difference that the stage 2 includes a rate shift to include the stage 1-allotted bandwidth from radars spectrum (adopted from [4]). 182

7.5 Radar-WiMAX exclusion zones (yellow curves) reach 577 km, where no 3.5 GHz communications system can currently be deployed [2]. This omits major US cities. 184

7.6 Radar simulation scenario subsumes a radar next to the 3.5 GHz LTE. . . . 186

7.7 Radar pulses radiated on the LTE during the antenna dwell time. 186

7.8 Radar Antenna Pattern [2]: The black curve represents the pattern after the first side lobe is 14.4 dB below the main lobe. The back lobe is constant at -50 dB. 187

7.9 Composite antenna pattern for the macro cell LTE BSs for 3-sector cells from equation (3) [5]. 188

7.10 : LTE macro cell layout simulation. 189

7.11 LoS FSPL in green and NLoS ITM loss in blue and red for macro and small cells respectively represent the radar signal degradation in dB vs. distance in km. 192

7.12 Normalized mean UE throughput for LTE macro cell and cochannel radar. Macro cell LTE operates within NTIA exclusion zones, and at 50 km the throughput loss is less than 30%. 194

7.13 SINR for the LTE system in the UL direction when a cochannel radar at the center of the LTE band affects the BSs. SINR recover between the radar pulses and drops during the time the BS is hit by a radar pulse. This is promising as only certain symbols are affected in an LTE subframe. 194

7.14 UE mean throughput for out-of-band interference from radar 50 km away from the macro cell LTE system. The throughput loss is drastically alleviated. . . 195

7.15 Normalized UE mean throughput for outdoor small cells and a cochannel radar. At 50 km, there is 55% throughput loss for the small cells. 195

7.16 Normalized UE mean throughput for indoor small cells and a cochannel radar. 196

7.17 Cellular communications system: Three sector cells are covered by BSs, controlled by an MME unit. The cell phone callers indicate frequency bands such that sectors with the same colored cell phones are reusing identical bands. Such a topological reuse pattern is aimed at minimizing the inter-cell interference. 198

7.18 Rate allocation to the UEs when a radar interfering with sectors 3 of cells is in the vicinity of the cellular communications system. The allocated rates are initially allocated to non-interfering sectors from the radar operating frequency bands and no bandwidth is allocated to the interfering sectors. In the second stage of the allocation, the rates start allocating to the interfering sectors only in order to insure fairness. Once certain bandwidth is allocated to the sector, the remainder of the bandwidths can be allocated to all the sectors. 199

7.19 Rate allocation to the UEs when no radar is in the vicinity of the cellular communications system. The allocated rates are similar since the pattern of the applications in the cells are alike. 199

8.1 Link Adaptation System: LTEs physical layer with Turbo coding and modulation modules [6]. 206

8.2 LTE Radio Resource Block [6]. 208

8.3 LTE Frame Structure [7]. 209

8.4 System Model for Resource Allocation. 211

8.5 System Model and channel for Resource Allocation. 212

8.6 Parameter ϵ vs SNR. 216

8.7 The system contains 9 UEs, each concurrently running a real-time application with respective identically colored sigmoidal utility functions. The lower SNR UEs are allocated more resources so that their throughput meet their QoS requirements. 226

8.8 The system contains 9 UEs, each concurrently running two real-time application with respective identically colored sigmoidal utility functions. The lower SNR UEs are allocated more resources so that their throughput meet their QoS requirements. 227

8.9 The system contains 9 UEs, each concurrently running a real-time application with respective identically colored sigmoidal utility functions. The lower SNR UEs are allocated more resources so that their throughput meet their QoS requirements. 228

8.10 The system contains 9 UEs, each concurrently running a real-time application with respective identically colored sigmoidal utility functions. The lower SNR UEs are allocated more resources so that their throughput meet their QoS requirements. 229

8.11 10 km $\times$ 10 km area covering Falls Church and Annandale in Northern Virginia, where our cell planning and resource allocation occurs. 230

8.12 Ues are randomly distributed in the area in Figure 8.11, and since they choose the serving base station according to the strongest downlink signal that they receive from all base stations, they are distributed distinctively amongst the base stations as in this figure. The numbers on each pie piece shows the percentage of the UEs of 500 total UEs that belong to a particular cell. 232

8.13 Ues are randomly distributed in the area in Figure 8.11, and since they choose the serving base station according to the strongest downlink signal that they receive from all base stations, they are distributed distinctively amongst the base stations as in this figure. The numbers on each pie piece shows the percentage of the UEs of 500 total UEs that belong to a particular cell. 233

8.14 The system contains 31 UEs, each concurrently running a real-time application. 235

8.15 UE Bids pledged to eNB 1 236

8.16 The system contains 31 UEs, each concurrently running a real-time application. 238

8.17 The system contains 23 UEs, each concurrently running a real-time application. 239

8.18 UE Bids pledged to eNB 2 240

8.19 The system contains 23 UEs, each concurrently running a real-time application. 242

8.20 The system contains 22 UEs, each concurrently running a real-time application. 243

8.21 UE Bids pledged to eNB 3 244

8.22 The system contains 22 UEs, each concurrently running a real-time application. 246

8.23 The system contains 28 UEs, each concurrently running a real-time application. 247

8.24 UE Bids pledged to eNB 4 248

8.25 The system contains 28 UEs, each concurrently running a real-time application. 250

8.26 The system contains 14 UEs, each concurrently running a real-time application. 251

8.27 UE Bids pledged to eNB 5 252

8.28 The system contains 14 UEs, each concurrently running a real-time application. 254

8.29 The system contains 20 UEs, each concurrently running a real-time application. 255

8.30 UE Bids pledged to eNB 6 256

8.31 The system contains 20 UEs, each concurrently running a real-time application. 258

8.32 The system contains 32 UEs, each concurrently running a real-time application. 259

8.33 UE Bids pledged to eNB 7 260

8.34 The system contains 32 UEs, each concurrently running a real-time application. 262

8.35 The system contains 27 UEs, each concurrently running a real-time application. 263

8.36 UE Bids pledged to eNB 8 264
8.37 The system contains 27 UEs, each concurrently running a real-time application. 266
8.38 The system contains 26 UEs, each concurrently running a real-time application. 267
8.39 UE Bids pledged to eNB 9 268
8.40 The system contains 26 UEs, each concurrently running a real-time application. 270
8.41 The system contains 14 UEs, each concurrently running a real-time application. 271
8.42 UE Bids pledged to eNB 10 272
8.43 The system contains 14 UEs, each concurrently running a real-time application. 274
8.44 The system contains 24 UEs, each concurrently running a real-time application. 275
8.45 UE Bids pledged to eNB 11 276
8.46 The system contains 24 UEs, each concurrently running a real-time application. 278
8.47 The system contains 25 UEs, each concurrently running a real-time application. 279
8.48 UE Bids pledged to eNB 12 280
8.49 The system contains 25 UEs, each concurrently running a real-time application. 282
8.50 The system contains 21 UEs, each concurrently running a real-time application. 283
8.51 UE Bids pledged to eNB 13 284
8.52 The system contains 21 UEs, each concurrently running a real-time application. 286
8.53 The system contains 25 UEs, each concurrently running a real-time application. 287
8.54 UE Bids pledged to eNB 14 288
8.55 The system contains 25 UEs, each concurrently running a real-time application. 290
8.56 The system contains 31 UEs, each concurrently running a real-time application. 291
8.57 UE Bids pledged to eNB 15 292
8.58 The system contains 31 UEs, each concurrently running a real-time application. 294
8.59 The system contains 14 UEs, each concurrently running a real-time application. 295
8.60 UE Bids pledged to eNB 16 296
8.61 The system contains 14 UEs, each concurrently running a real-time application. 298
8.62 The system contains 18 UEs, each concurrently running a real-time application. 299
8.63 UE Bids pledged to eNB 17 300
8.64 The system contains 18 UEs, each concurrently running a real-time application. 302
8.65 The system contains 30 UEs, each concurrently running a real-time application. 303

8.66 UE Bids pledged to eNB 18 304
8.67 The system contains 30 UEs, each concurrently running a real-time application. 306
8.68 The system contains 23 UEs, each concurrently running a real-time application. 307
8.69 UE Bids pledged to eNB 19 308
8.70 The system contains 23 UEs, each concurrently running a real-time application. 309
8.71 The system contains 32 UEs, each concurrently running a real-time application. 311
8.72 UE Bids pledged to eNB 20 312
8.73 The system contains 32 UEs, each concurrently running a real-time application. 313

List of Tables

2.1 Applications Utility Function Parameters 36

3.1 Applications Utility Parameters - Centralized Architecture 59

3.2 Network throughput for the network in Figure 3.8 under neither bandwidth constraints nor traffic shaping. 67

3.3 Smart phone Applications under $R = 1$ Mbps Network Constraint with/without Traffic Shaping 68

3.4 Bandwidth Consumption and Download Time. 70

4.1 Applications Utility Parameters - Distributed Architecture 98

5.1 Applications Utility Function Parameters - Traffic Analysis 131

5.2 Application Status Differentiation 133

6.1 An Example of the Continuous Rate Allocations versus the Resource Block Assignments for eNB Rate 100 and 50: The former yields in 8 possible discrete rate allocations whilst the latter generates 1 discrete rate allocation alone. . 167

7.1 Radar parameters from [2], except those marked with * where we used typical parameters for medium-to-large shipborne S-band radars [8]. 185

7.2 Macro Cell LTE Parameters for Our Simulations (adopted from 3GPP [5]). . 190

7.3 ITM Parameters 196

7.4 Application Utility Functions of the Cellular Network. 202

8.1 LTE DL Link Adaptation 207

8.2 LTE DL Link Adaptation 210

8.3 Characterizing epsilon for different channels. (M: Modulation, CR: Coding Rate,SE: Spectral Efficiency, T: Transmit Mode [9]) 215

8.4 Characterizing K for different channels. (BW: Bandwidth, TB: Transmit Bandwidth, TREPF: Total Resource Element per Frame, TREPS: Total Resource Element per Second, OREPF: Overhead Resource Element per Second, OREPS: Overhead Resource Element per Second, DREPS: Data Resource Element per Second). 217

Chapter 1

Quality of Service in Communications Networks

The number of mobile broadband users and the volume of the traffic generated by them increase incessantly every year. Based on a report by the Infonetics Research [1] , as depicted in Figure 1.1, the number of the mobile subscribers grew from 5.1 billion in 2010 to 6.5 billion by the end of 2014, while the number of the mobile broadband subscriptions escalated from 548.9 million in 2010 to 1.5 billion by late 2014, whereas the quantity of the fixed broadband subscribers do not observe any dramatic growth. The disproportional augmentative trend in the mobile broadband subscriptions vs. fixed and mobile subscriptions results from the global adoption of the data hungry mobile broadband smart devices that are bandwidth killers and cause grave concerns for Mobile Network Operators (MNOs) and infrastructure vendors. As such, there has been a perpetual demand for more bandwidth allocation to the mobile broadband services.

However, communications networks have finite resources such as the radio spectrum and transport backhaul which are expensive and are shared between numerous users running many services. Thus, there has been several initiatives by the government to share its

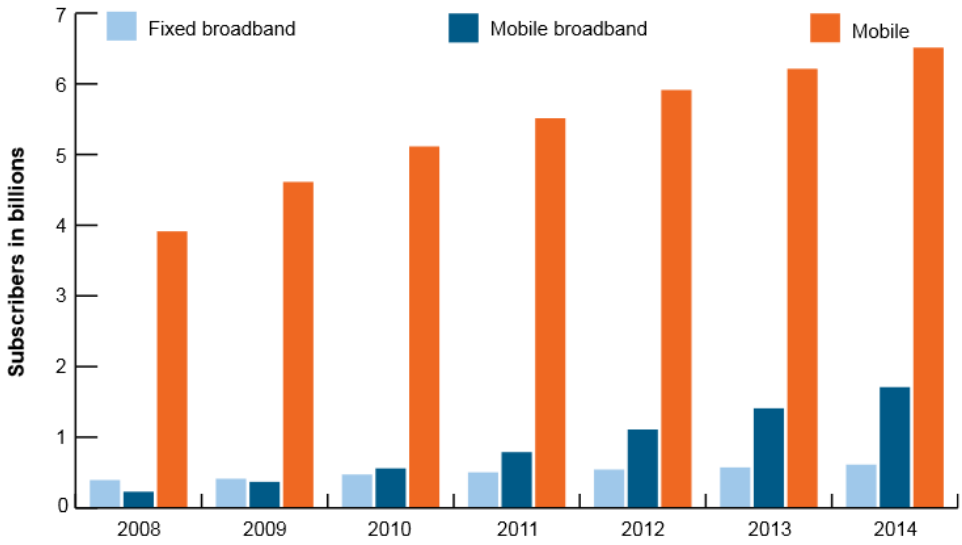


Figure 1.1: Fixed, mobile, and mobile broadband subscribers' growth: The mobile and mobile broadband subscriber's quantity have observed a dramatic growth from 2008 to 2014 [1].

bandwidth with the mobile broadband services; this in turn causes significant feasibility issues with regard to the sustenance of the incumbents operation as well as the entrants performance in the face of the coexistence which have to be incorporated into the operation of the *au courant* communications systems.

On the other hand, while many vendors have bluntly suggested capacity-additive approaches to cater to the mobile broadband traffic growth, such proposals can be premature. According to Nokia [10], the annual data volume originating from the mobile broadband users has undergone a $10,000\%$ increment to 23 Exabyte by the end of 2015. Therefore, capacity is only one issue, and unique mobility properties such as intermittent connectivity and location shifting, coupled with the the proliferation of smart devices and inexpensive mobile data tariffs, compounded by the emergence and prevalence of the quadruple play, added by the widespread deployment the fourth generation (4G) and the emergence of the fifth generation (5G) cellular communications networks, which substitute the conventional fixed-line connectivity, create numerous challenges in achieving a good end-to-end network performance, a satisfactory user experience, and a profitable operator revenue. Hence, communications

systems serving the mobile broadband aught to be far more intelligent in handling the traffic growth.

First, the voluminous mobile broadband traffic incorporates the traditional multi-play services such as Internet Protocol (IP) Television (IPTV), voice-over-IP (VoIP), and video conferencing which have fairly well-known quality of service (QoS) requirements from a network operation perspective. Hence, such a hybrid mobile broadband traffic should be treated with regard to the QoS requirements of the multi-play services which generated the traffic. Moreover, the evolution of the cellular networks into an all-IP paradigm in 4G and 5G introduces new challenges for the traditionally circuit-switched services such as telephony which require the operators to guarantee a minimum level of performance.

In addition, the escalating uptake of the cloud- and web-based services necessitates new QoS-related requirements for the network performance. In particular, there is a need to differentiate the mobile subscribers. In accordance with Zokem [11], less than 10% of the mobile broadcast users subscribed to a relatively small number of monthly minutes generate the bulk of the traffic due to the nature of the services that they are using. Such highly temporal usage patterns of the applications generating the mobile broadband traffic should also be taken in to account for intelligent cellular networks addressing the complexities of the present-day mobile Internet traffic. As a case in point, a voluminous amount of signaling traffic is generated by the smart phone applications such as social networking and instant messaging since they intermittently connect to the network in the background hundreds to thousands times a day. This situation will be adversely exacerbated by the advent of mobile gaming applications which not only produce an excessive interactive signaling over extensive time durations, but also conspicuously add to the volume of the data traffic over the network. For instance, Zynga servers carry about 0.82 petabytes of data on a daily basis [12] and shifting them to smart devices will push the data traffic higher than any network planners ever conceived.

An apparent takeaway from the discussions above in the context of the mobile Internet traffic is severe congestion occurrences on the MNO's Radio Access Networks (RANs) and backhaul as the main bottlenecks of an all-IP cellular communications system. The RAN becomes a bottleneck as the radio spectrum is scarce and the users' traffic transfer to and from the base stations in the uplink (UL) and downlink (DL) directions respectively. This traffic includes signaling spikes killing the bandwidth as well as the application-generated traffic which has various QoS requirements and encompasses temporal dynamics based on the user's focus on the device.

Besides, all the mobile broadband traffic aggregates into the backhaul network, whose congestion is exacerbated by the fact that it subsumes mainly legacy networks, which adversely impact the QoS-requirements of the hybrid traffic by introducing excessive delays, jitter, and packet loss. Such performance parameters should be incorporated into present-day cellular system in order to ensure the QoS-requirements within the backhaul which directly impacts the performance of the applications, generating the mobile Internet traffic, and thereby, the users' quality of experience (QoE) which is their perception of the cellular communications networks end-to-end performance and speaks directly to the subscriber churn and operator revenue.

And last but not the least, while the aforementioned discussions with regard to over-provisioning the cellular system through more bandwidth, traffic and subscriber differential treatments to meet the QoS requirements in the RAN, and proper treatment of the traffic in the backhaul are essential to elevating the QoE within the cellular communications network ecosystem, there is an equally important pressing need to devise promising data-centric business models to properly address increasing the bandwidth by maximizing the spectrum incumbents' revenue and ensuring the fairness to the entrants.

To recap, while extending the capacity of the mobile broadband services is a must in view of the ever-increasing smart device-generated traffic, the MNOs should optimize their

network efficiency via QoS treatments which improves the QoE in the network and maximizes the utilization of the resources. Introducing efficient and dynamic traffic management and resource allocation techniques can improve users QoE, which reduces the subscriber churn, meanwhile it proves a business case for operators to make an effort in leveraging QoS-related technologies in their networks. The QoS mechanisms, implemented in the RAN and the core, must span from the core network to the RAN in order to preclude congestion occurrences in any portion of the network, inasmuch as a localized congested section might adversely affect the overall network performance and degrade the QoE.

The remainder of this chapter is organized as follows. Section 1.1 provides with some information on current trends for QoS in cellular communications systems. It looks at QoS in wired IP and wireless networks and discusses various approaches that have been adopted for provisioning QoS in these networks. Section 1.2 discusses the complexities associated with the spectrum as the blood of wireless communications systems which make it a congested and contested environment. Section 1.4 presents the problem statement in the current article. Finally, section1.6 presents the organization of this entire document.

Next, section 1.1 provides with a high level discussion of the current methods for provisioning QoS in networks.

1.1 Current Trends in End-to-End QoS in Cellular Networks

In this section, we attempt to conduct a literature survey on the current research in the scope of the QoS in cellular communications networks. In particular, we consider certain QoS solutions for the all-IP wired as well as for the wireless networks.

1.1.1 QoS in Wired IP Networks

The ways to provide with QoS in the IP networks is through Integrated Services (IntServ) [13] and Differentiated Services (DiffServ) [14]. These methods bestow QoS mechanisms as scheduling, routing, and shaping on the routers, as depicted in Figure. IntServ provides with end-to-end QoS guarantees through reserving resources per flow in the nodes along the path by means of Resource Reservation Protocol (RSVP) protocol [15]. The RSVP identifies a session using the destination IP address, port number, and the transport layer protocol for IP version 4 (IPv4) or the destination IP address and flow label for IP version 6 (IPv6). Since reservation needs a policing process, a QoS broker communicates the policy decisions to the routers which may enforce the policies. The RSVP protocol consists of a path message, originating from source, and a reservation message, originating from the destination. The main advantage of this method is provisioning excellent guarantees; however, it has severe scalability issues specially with regard to maintaining per-flow operations in the routers.

In contrast, DiffServ is relatively scalable QoS method, but it can not provide any QoS guarantees. DiffServ relies on the idea of classifying packets using the differentiated code point (DSCP) field of IP header [16] which defines the three priority classes Best Effort (BE), Assured Forwarding (AF), and Expedited Forwarding (EF). BE provides no guarantees what so ever and its DSCP is 000000, while AF subdivides into four independent classes each with three drop precedence values for the queue. EF whose DSCP is 101110 provides a very small drop probability, latency, and jitter such that it can be considered a virtual leased line (VLL). EF packets in DiffServ-enabled routers undergo short queues and are quickly serviced. However, DiffServ networks require access control mechanisms in their Access Routers (ARs) so that only authorized users can inject such high priority packets into the network.

Besides QoS-enabled routers, a QoS broker [17] can control the network and manage the resources. For the sake of scalability, this entity can be repeated and the network can

be hierarchical [3-14] so that a broker manages the resources in a part of the network. For example, for a satellite network which provides communication services to the user equipment (UE), the client can switch between the satellite and terrestrial wireless networks. When the UEs support a dual access, it is important to select the appropriate network for each application based on the available resources and the application type and such a selection can be terminal or network initiated. For the multicast and broadcast services, a terminal-initiated approach seems natural since it would be hard for the network operator to select the optimum interface per individual for a large number of users.

1.1.2 QoS in Wireless Networks

The challenges that have been introduced by the widespread deployment of the smart devices in cellular communications systems (chapter 2) have led to an upsurge in the research works which target QoS in the RAN. In addition, Long Term Evolution (LTE) [18] has become the paramount 4G technology widely deployed all over the globe. Hence, many of the current studies of RAN QoS are tailored to the LTE infrastructure. Such works usually focus on a specific layer of the protocol stack including Physical (PHY), link layer, network, or application layer.

1.1.2.1 Link Layer QoS

Larmo et. al. [19] focused on the PHY and link layer design of the LTE. They indicated that deploying the single-carrier frequency division multiple access (SC-OFDMA) [20] in the UL direction improves (reduces) the peak-to-average power ratio (PAPR) in the LTE system which increases the spectral efficiency for the cell-edge users or increases the cell radius. The authors also mentioned that leveraging 12 subcarriers and 180 kHz channels in the radio resource blocks (RRBs) of the LTE generates very narrow frames. Therefore, users UL data

transmissions can be scheduled depending on the current channel quality.

Next, Piro et. al. [21] developed a framework for the QoS at the link level. Their approach used the discrete control theory to create a maximum throughput scheduler in the DL direction for the real-time multimedia services in the LTE networks. Next, Marchese et. al. [22] used a link layer QoS method to adapt the bandwidth, which is going to be allocated to a buffer, for conveying the heterogeneous traffic. The authors in [23] proposed a traffic segmentation approach to quantify the QoS at a given QoE in terms of the spectral efficiency, cost, and resource over-provisioning. In their work, the core network classified incoming packets based on a deep packet inspection (DPI) and marked the packet header using the DSCP. Then, the packets were scheduled using queue weights.

Next, Ali et. al. [24] proposed a two-level scheduler based on the game theory. The first level distributed RRBs among the classes with different QoS requirements; then, a delay-based air interface scheduler in the second level fulfilled delay requirements defined for the LTE classes. Then, a cooperative game between various service class flows was done and a Lagrangian formulation was used to find an associated Pareto optimality [25]. The delay based scheduler would check each user's packet delays within its respective service class and make scheduling decisions in the DL direction based on the current channel conditions. Next, the authors in [26] suggested a mathematical framework based on the supply function bidding in microeconomics theory [27]. Their method enabled users to opportunistically compromise the efficiently based on their demands such that each user is satisfied with a lesser QoS whilst users social welfare is maximized. Then, Monghal et. al. [22] introduced a decoupled time-frequency orthogonal frequency division multiple access (OFDMA) [28] packet scheduler for the LTE to control the throughput fairness amongst the users.

1.1.2.2 Network Layer QoS

Some of QoS works in the cellular systems rely on the network layer. For example, Lee [29] studied QoS-based routing in a multi-hop wireless network with an eye on energy efficiency. He investigated the effect that various routing methods produced from an energy efficiency perspective. Specifically, he demonstrated his method accompanies trade-offs among the end-to-end delay, throughput, and energy efficiency and suggested that the shortest path routing is a good candidate once the delay, bandwidth, and energy are considered jointly. Next, Alasti et. al. [30] explained about the QoS provisioning built within the Worldwide Interchangeability for Microwave Access (WiMAX) [31] and LTE, as two then-competing 4G technologies. They discussed service flows and bearers as QoS mechanisms aiming at the network layer. Similarly, Ekstorm [32] surveyed the QoS provisioning for traffic separation within the 3rd generation partnership project (3GPP) standardization [9] which ultimately led to those of the LTE, i.e. the bearers and service flows.

Then, Mota [33] studied the LTE backhaul QoS and showed that the backhaul creates a bottleneck. He suggested using QoS Class Identifiers (QCI)-DSCP as well as the policing, scheduling, and shaping of the flows to provide with QoS over the backhaul. On the other hand, the authors in [34] investigated the QoS at the network layer from a policy management perspective which emphasized on the traffic and subscriber differentiation. Moreover, they looked at the operator revenue by creating business models for the QoS and suggested a gradual adoption of the policy management approach, which initially concentrated on reducing the congestion and ultimately on applying the policies per subscriber and even per data flow. Next, Li [35] investigated the QoS, traffic management, and resource management in Universal Mobile Terrestrial System (UMTS) [9] and suggested a distributed infrastructure in the user equipment and base stations (BSs) to provide with end-to-end QoS in UMTS networks. He suggested using traffic shaping in the UEs to delay the traffic if necessary. He also applied traffic shaping at the BSs to deal with congested cells. Then, Gorbil et.

al. [36] proposed a network layer QoS solution to supporting real-time traffic in multi-hop wireless networks. They used a hybrid routing protocol to enable QoS traffic support for real-time traffic via an intelligent path selection at source nodes and based on the required QoS parameters. They dynamically adapted the paths as the network topology changed.

1.1.2.3 Application Layer QoS

Some other research works concentrated on declaring the QoS requirements through machine understandable languages. For instance, Tournier et. al. [37] suggested a component-based QoS architecture using fractional components and considering main QoS concepts. Also, Ahmed et. al. [38] proposed a semantic end-to-end QoS model for smart phones using the ontologies. This model addressed basic QoS properties related to the service environment, application services, and user level and leveraged a Web Service Quality Model (WSQM) - an XML-based standardization of expressing the QoS between services and customers - to declare the QoS taxonomy. In essence, their model addressed main elements acting in the service environment (infrastructure, services, and users) and their dynamic nature along with the possibility of incorporating a domain-specific QoS into the modeling. They developed ontologies which could formulate robust QoS descriptions that combined the rich semantics of the QoS ontology with the accuracy of the QoS languages.

1.1.2.4 Hardware QoS

Some other works concentrated on hardware issues, in particular the battery life, and their impact on QoS. Jung et. al. [39] worked on reducing the energy consumption of the devices and on increasing the successful interactions between the smart phones and the application servers via creating a surrogate of the client in the cloud. Their approach relied on a thick client in the cloud to enhance the inquiry throughput to elevate the QoS for real-time applications. In their work, a thin client on the cellphone orders the thick client

by means of tokens to take charge of heavy load tasks such as communications with a bursty server of a real-time application. Next, the authors in [40] focused on signaling issues in 3G and 4G and the abrupt burstiness that can be produced by users who simultaneously update their phones and how this affects the battery life.

1.1.2.5 Cross Layer QoS

A system efficiency is required to target a mass market in wireless networks, as the QoS requested by a user does not care about the resource utilization and QoS becomes a conflicting concept. However, QoS can be addressed efficiently through a cross-layer design. In this situation, the cross-layer information will be exchanged from the higher layers to the lower ones within the wireless technology protocol stack or vice versa. While in the Open Systems Interconnection (OSI) model [41], nonadjacent layers can only communicate through the intermediate levels, a cross-layer approach can help exchange the information between the layers which are not adjacent to each other. Some cross-layer proposals suggest a global coordinator [42,43] while others recommend a Medium Access Control (MAC)-centric cross-layer air interface, illustrated in Figure 1.2. The global coordinator gathers the information from the miscellaneous layers and places them in a shared memory. Global coordinator, which can be in the application or MAC layer, can also be in application layer or MAC layer.

In a cross-layer approach, the layers have specific requirements. As a case in point, for the PHY, the radio channel should be consistently estimated and the signal strength and the bit error rate (BER) estimations should be available to the implement modulation and coding and to select suitable formats at the link layer. For the network layer, mobility should be considered so that the link layer would prioritize the users during handover phases. Moreover, mapping IP QoS features to the link layer radio resource management (RRM) is essential. For transport layer, RRM methods should be modified to treat the Transport

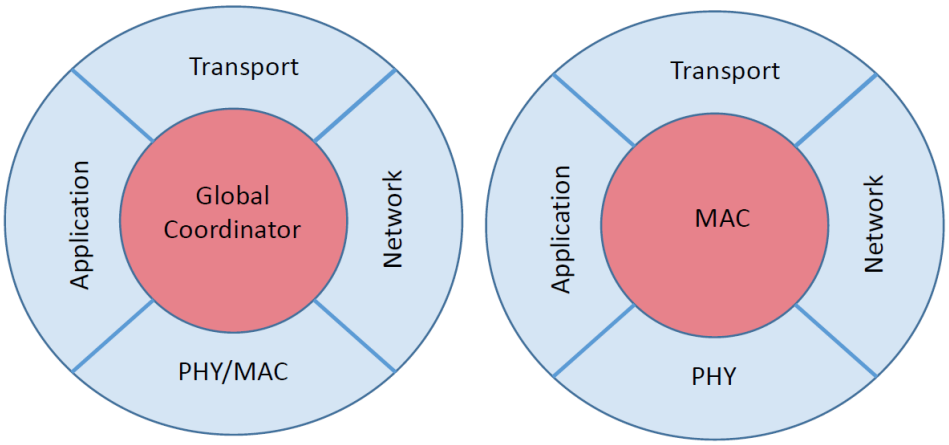


Figure 1.2: Cross Layer QoS

Control Protocol (TCP) vs. Universal Datagram Protocol (UDP) [41] and broadcast vs. multicast traffic appropriately. Application layer must consider traffic types and monitoring actions performed jointly with the link layer functions to provision with prioritization.

It is noteworthy that this method is protocol-dependent and if we used an Asynchronous Transfer Mode (ATM) network [41], the cross-layer QoS requirements would be different. For the MAC sublayer, the most important thing is the bandwidth allocation which can be a constraint, for example for satellite links in the UL direction shared among the users. Another example for the MAC sublayer is in the context of the ATM networks which need a lookup table including Virtual Path Indicator (VPI) and Virtual Circuit Indicator (VCI) to support the QoS for various ATM service categories. The authors in [44] modified the ATM protocol stack to realize the cross-layer QoS concept. At the network layer, Call Admission Control (CAC) [45] is of high consequence. The CAC algorithms perform a procedure during the call setup if a new connection can be accepted without violating existing commitments. The CAC methods should be mapped properly into the link layer RRM protocols.

It should be mentioned that most research works so far concentrate only on very specific aspects of the QoS such as the streaming video or on how to perform the CAC fairly. However, an end-to-end QoS investigation has not been addressed for large scale MNO networks. It

would be interesting to develop an an end-to-end QoS architecture which could consider the QoS for large scale networks spanning from the RAN and the core network. Such a distributed QoS architecture requires a traffic conditioning at the core network and the RAN. Furthermore, a devised QoS architecture should be dynamic enough to be responsive to the quickly fluctuating channel conditions. Albeit research works has consistently indicated the high potential of the backhaul for congestion, controlling the QoS over the backhaul for the UL/DL traffic has not been considered yet.

The QoS-related mechanism that we deploy in the network ultimately deals with traffic or bandwidth to carry the traffic. The bandwidth would be allocated to handling the traffic and is taken from the spectrum assigned for mobile broadband. In section 1.2, we talk about the spectrum and how it is congested by its users.

1.2 Spectrum as a Congested and Contested Environment

Mobile broadband networks will face a tremendous increase in data traffic volumes over the next 20 years. In order to meet this need, large amounts of spectrum will be a key prerequisite for any RAN evolution. To satisfy this demand, MNOs will need new spectrum allocations [10]. However, the created demand for more bandwidth far exceeds the available commercial spectrum and has spurred the Federal Communications Commission (FCC) to consider spectrum sharing, which is an elegant solution to utilizing shareable spectrum bands efficiently.

In spite its attractiveness, spectrum sharing is challenging as incumbent systems need to be shielded from harmful interference from the entrant systems and vice versa. As a case in point, in response to the Presidents Council of Advisers on Science and Technology

(PCAST)s report [46] on realizing the full potential of the government-held spectrum, Federal Communications Commissions (FCC) issued a Notice of Proposed Rulemaking (NPRM) [47] that suggested designating 115 MHz of bandwidth in the 1675 - 1710, 1755 - 1780, 3550 - 3650 (*aka* the 3.5 GHz band), 4200-4220, and 4380 - 4400 MHz bands for mobile broadband. Soon after the release of the NPRM, the National Telecommunications and Information Administration (NTIA) conducted a Fast Track Evaluation [2] to study the interference between the band incumbents and WiMAX as the paramount mobile broadband system at the time.

For the case of the 3.5 GHz band, containing the largest segment (100 MHz) from the released 115 MHz spectrum, the Fast Track Evaluation identified S-band radars [8] and satellite systems as the band incumbents and suggested establishing exclusion zones reaching 577 km inland as depicted in Figure 1.3. The results indicated that no 3.5 GHz cellular communications system should be deployed in the exclusion zones which subsumed large metropolitan cities in the coasts of the the United States (US). On the other hand, over 55% of the US population lives within 50 miles from the shoreline [48], and the inability to cater for this huge market which includes metropolises like New York City and Los Angeles ensues severe financial caveats for the MNOs. Thus, any efforts to judiciously decrease the exclusion zones is of great interest for the MNOs, a desirable step toward realizing 3.5 GHz radar and mobile broadband network coexistence, and an inspiration for the spectrum sharing in the other bands. However, the aforementioned exclusion zones were developed from the link budget analysis of radars and WiMAX systems, whereas only a precise simulation of the nuances of radar and WiMAX systems, as opposed to the simplistic link budget analyses, can lead to relevant exclusion zones.

Henceforth, mobile broadband spectrum as the blood of the wireless communications system will be a congested environment constricted by the notion of spectrum sharing, i.e. by the band incumbent from the perspective of the band entrant and vice versa. On the

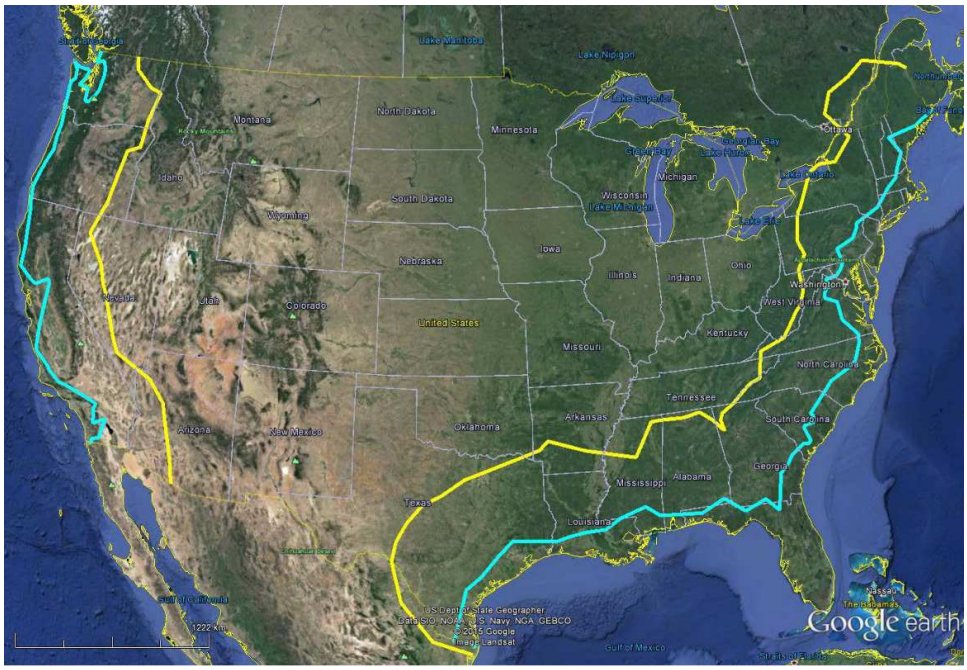


Figure 1.3: NTIA set radar-comm. exclusion zones reaching inland to a large extent. Such separations preclude the viability of the 3.5 GHz band for mobile broadband and excludes catering communications services to more than 40% of the U.S. population [2].

other hand, mobile broadband spectrum is becomes a contested environment in the realm of modern-day communications networks inasmuch as an empty band, providing resources to either an incumbent or entrant system, can be contested for when both the incumbent and entrant operate in the vicinity of one another.

1.3 Modern-Day Networks Quality of Service Challenges

Because of the fact that the incessant demand for more data in modern-day mobile broadband networks grows far beyond the spectrum licensed for commercial wireless communications, there is an urgency to augment the spectral resources designated for mobile broadband. While assigning more spectrum to mobile broadband services is highly desirable, the cur-

rent noncommercial under-utilization of the spectrum arises a new horizon which reuses the spectrum to cater to the mobile broadband devour for bandwidth. This fascinating notion of spectrum re-utilization is further strengthened by the perpetual bandwidth solicitation from the future mobile broadband networks outfitted with interactive heavy gaming and multimedia applications which can push the data demands exceedingly beyond what network planners can ever envisage. Realization of this idea has been proposed by spectrum-governing agencies like FCC through various measures such as spectrum sharing and spectrum auctioning which aim at provisioning more resources for mobile communications.

On the other hand, while assigning more spectral pieces and reusing under-utilized spectrum for mobile broadband services can alleviate the hunger of present-day communications networks for data to some extent, the evolution of networks' smart devices towards more complexity by letting them host a wide variety of applications to enhance a large gamut of daily tasks further diversifies the traffic types that run on the network. Such traffic diversity is tightly bound to the applications performance and the users perception of the network operation, and it can lead to very stringent requirements for the traffic in order to meet the QoS.

In light of the aforesaid traffic diversity and QoS constraints, even over-generously adding to the mobile broadband networks' bandwidth via spectrum auctions and sharing does not eliminate a need for developing sophisticated radio resource allocation entities as integral parts of intelligent cellular communications systems. Having a complex radio resource allocation scheme handle temporal data demands and QoS requirements of modern communications networks inundated with smart devices can fine-grain the procedure to disseminate the scarce valuable spectrum efficiently. Such resource allocation schemes should be able to account for a wide range of QoS-related issues.

First, the devised radio resource allocation methods should be spectrally efficient in disseminating the available resources so that neither the resources are underutilized nor they

are wasted. The spectral efficiency necessitates that every small portion of the available spectrum is utilized such that all the UEs in the RAN receive their required resources based on their traffic demands. On the other hand, excessive resources should not be assigned to the UEs which could get by much less bandwidth to meet their data volume demands.

On the other hand, modern mobile broadband networks are replete with smart devices which can concurrently run many applications to enhance daily tasks. The applications range a wide variety of categories and aim at fulfilling a large gamut of duties from financial services to educational sessions to administrative works to entertainment programs. Therefore, the applications generate various traffic types with miscellaneous QoS requirements which should be met in order for the applications to perform properly. For instance, the tolerance of the applications to the delays in the network should be considered. Therefore, a resource allocation scheme which is to be deployed in modern cellular communications networks should account for the type of the running applications which generate a hybrid traffic with stringent QoS requirements, whose fulfillment elevates the QoE in the network.

Furthermore, the dynamic nature of the present-day cellular communications systems should be taken into the equation in the design of the radio resource allocation methods. As an illustration, UEs can be highly mobile and move from one area of the RAN to another and such a dynamism should be incorporated into the radio resource allocation approaches for the au courant cellular networks.

Another dynamism prevalent in modern cellular communications systems inundated with the smart devices is the temporal changes in that occur in the usage percentage of several applications running on a smart device. An intelligent resource allocation scheme which distributes the resources to the smart devices, which run their applications with temporal usage changes, should account for such dynamics in the network in order to efficiently allocate resources in such a time varying system.

Moreover, users should be differentiated in present-day cellular communications systems.

As highly prioritized users operate in the same cellular communications system as the public users, there should be mechanisms to treat the users in a differential manner. Such high priority users included public safety responders and national security/emergency preparedness subscribers whose traffic should be handled with the highest possible priority. In addition, users can have a variety of subscriptions with the MNOs, e.g. prepaid vs. post-paid subscribers. Besides, subscriptions to third party services can create a heavy burden on the network (for example a smart device subscribed to the Netflix). Such subscription-related concerns should be included in the novel radio resource allocation approaches in order to ensure fairness in the network and create revenue for the operators.

While such sophisticated spectrum allocation schemes might be derived through a precise mathematical modeling of all the aforementioned issues, the devised resource allocation methods should be tailored to the bandwidth assignment units deployed in present-day cellular systems, i.e. 4G and 5G. Such cellular technologies rely on a discrete assignment of the spectrum to the UEs, and purely theoretical radio resource allocation techniques may lack the capacity to be applicable to such cellular communications technologies.

Additionally, any radio resource allocation scheme should be computationally efficient. Considering the current volume and the outgrowth of the mobile Internet traffic as well as gigantic amounts of signaling traffic generated by the mobile broadband devices, radio resource allocation schemes should be able to assign the resources with a reasonable or possibly minimum amount of transmission overhead. Also, how the temporal dynamism of present-day networks from the perspective of the changes induced by the varying number of users in the system and varying application usage percentage in the smart devices would affect the radio resource allocation performance and signaling is important. An excessive signaling can be a limiting factor in realizing a resource allocation technique regardless of the amount of nuances included in its structure.

Another issue to consider for the radio resource allocation methods, devised for present-day

cellular systems and aimed at provisioning QoS in the network, is accommodating bandwidth augmentative novelties such as spectrum sharing. In the light of what was discussed so far, portions of the government-held spectrum has been released for mobile broadband purposes. Inasmuch as the majority of the currently-released bands are shared with radar systems, novel resource allocation schemes which account for the spectrum sharing with the band incumbents (for example, radars and satellites) can sustain as a part of the intelligent future-minded cellular communications networks in the long run.

Even though the creation of sophisticated yet pragmatic radio resource allocation schemes, which consider the traffic type and dynamic, user subscription, and spectrum sharing, is an effective step toward development of intelligent cellular communications systems which are capable of meeting the data volume demands and application QoS requirements of the au courant wireless networks, the generated traffic might pass through the core network. Since the core network can contain legacy networks, the voluminous mobile broadband data traffic, treated well in the RAN by an efficient resource management and allocation, can suffer severely in the legacy core network not equipped with sophisticated resource allocation and management schemes. It is noteworthy that while LTE has an all IP core network, the traffic generated by the smart devices may have to go through legacy networks and there is no way to provision an end-to-end QoS over the entire backhaul networks.

On the other hand, the outrageous perpetual increase in the traffic volume generated in mobile broadband networks causes the backhaul to be a problematic bottleneck in present-day cellular networks. Furthermore, the QoS requirements of the generated traffic will observe grim chances of being fulfilled over the entire backhaul network; this severely adversely impacts the performance of the applications running on the smart devices and degrades the QoE in the cellular network. Such a deterioration of the users' experience, which appears in the form of lengthy delays or excessive losses for the traffic, leads to the subscriber churn.

In order to provision for the delays that the traffic might incur in the backhaul, it would be

helpful to create prototypes that would model the procrastination that the backhaul network, containing legacy portions, can impose on the data traffic originated from or intended for the mobile broadband devices. Such a delay-based modeling provides an insight into how the QoS performance of the traffic generated in present-day networks may be undermined.

Ultimately, adding further spectrum to the bandwidth available for mobile broadband services should be pragmatically facilitated for the involved parties, i.e. the band incumbents and potential entrants. Spectrum is a scarce expensive resource which can be an object of high remuneration for the incumbents willing to share it with the entrants. Due to the fact the communications ecosystem alters spatially and temporally at a fast pace, such changes should be accommodated to allow for an efficient utilization of the spectrum. As a case in point, a radar system which operates in the 3.5 GHz band might be spatially closeby an in-band cellular communications system in short time periods. During other times when the radar is not in the vicinity of the cellular system, the spectrum can be better utilized by being released to mobile broadband services. Such a better utilization of the bandwidth can be achieved by devising auctioning systems to lease the spectrum to an entrant system while maximizing the monetary revenue for the band incumbent. On the other hand, such auction systems should be secure to ensure the fairness of the auctions for the potential entrants to the band. In addition, the auction privacy would be a concern for the potential entrants to preclude virtual price inflation.

1.4 Problem Statement

Because of the fact that the incessant demand for more data in modern-day mobile broadband networks grows far beyond the spectrum licensed for commercial wireless communications, there is an urgency to augment the spectral resources designated for mobile broadband. While assigning more spectrum to mobile broadband services is highly desirable, the current

noncommercial under-utilization of the spectrum arises a new horizon which reuses the spectrum to cater to the mobile broadband's devour for bandwidth. This fascinating notion of spectrum re-utilization is further strengthened by the perpetual bandwidth solicitation from the future mobile broadband networks outfitted with interactive heavy gaming and multimedia applications which can push the data demands exceedingly beyond what network planners can ever envisage. Realization of this idea has been proposed by spectrum-governing agencies like FCC through various measures such as spectrum sharing and spectrum auctioning which aim at provisioning more resources for mobile communications.

On the other hand, while assigning more spectral pieces and reusing under-utilized spectrum for mobile broadband services can alleviate the hunger of present-day communications networks for data to some extent, the evolution of networks' smart devices towards more complexity by letting them host a wide variety of applications to enhance a large gamut of daily tasks further diversifies the traffic types that run on the network. Such traffic diversity is tightly bound to the applications performance and the users perception of the network operation, and it can lead to very stringent requirements for the traffic in order to meet the QoS.

In light of the aforesaid traffic diversity and QoS constraints, even over-generously adding to the mobile broadband networks' bandwidth via spectrum auctions and sharing does not eliminate a need for developing sophisticated radio resource allocation entities as integral parts of intelligent cellular communications systems. Having a complex radio resource allocation scheme handle temporal data demands and QoS requirements of modern communications networks inundated with smart devices can fine-grain the procedure to disseminate the scarce valuable spectrum efficiently. Such resource allocation schemes should be able to account for a wide range of QoS-related issues.

First, the devised radio resource allocation methods should be spectrally efficient in disseminating the available resources so that neither the resources are underutilized nor they

are wasted. The spectral efficiency necessitates that every small portion of the available spectrum is utilized such that all the UEs in the RAN receive their required resources based on their traffic demands. On the other hand, excessive resources should not be assigned to the UEs which could get by much less bandwidth to meet their data volume demands.

On the other hand, modern mobile broadband networks are replete with smart devices which can concurrently run many applications to enhance daily tasks. The applications range a wide variety of categories and aim at fulfilling a large gamut of duties from financial services to educational sessions to administrative works to entertainment programs. Therefore, the applications generate various traffic types with miscellaneous QoS requirements which should be met in order for the applications to perform properly. For instance, the tolerance of the applications to the delays in the network should be considered. Therefore, a resource allocation scheme which is to be deployed in modern cellular communications networks should account for the type of the running applications which generate a hybrid traffic with stringent QoS requirements, whose fulfillment elevates the QoE in the network.

Furthermore, the dynamic nature of the present-day cellular communications systems should be taken into the equation in the design of the radio resource allocation methods. As an illustration, UEs can be highly mobile and move from one area of the RAN to another and such a dynamism should be incorporated into the radio resource allocation approaches for the au courant cellular networks.

Another dynamism prevalent in modern cellular communications systems inundated with the smart devices is the temporal changes in that occur in the usage percentage of several applications running on a smart device. An intelligent resource allocation scheme which distributes the resources to the smart devices, which run their applications with temporal usage changes, should account for such dynamics in the network in order to efficiently allocate resources in such a time varying system.

Moreover, users should be differentiated in present-day cellular communications systems.

As highly prioritized users operate in the same cellular communications system as the public users, there should be mechanisms to treat the users in a differential manner. Such high priority users included public safety responders and national security/emergency preparedness subscribers whose traffic should be handled with the highest possible priority. In addition, users can have a variety of subscriptions with the MNOs, e.g. prepaid vs. post-paid subscribers. Besides, subscriptions to third party services can create a heavy burden on the network (for example a smart device subscribed to the Netflix). Such subscription-related concerns should be included in the novel radio resource allocation approaches in order to ensure fairness in the network and create revenue for the operators.

While such sophisticated spectrum allocation schemes might be derived through a precise mathematical modeling of all the aforementioned issues, the devised resource allocation methods should be tailored to the bandwidth assignment units deployed in present-day cellular systems, i.e. 4G and 5G. Such cellular technologies rely on a discrete assignment of the spectrum to the UEs, and purely theoretical radio resource allocation techniques may lack the capacity to be applicable to such cellular communications technologies.

Additionally, any radio resource allocation scheme should be computationally efficient. Considering the current volume and the outgrowth of the mobile Internet traffic as well as gigantic amounts of signaling traffic generated by the mobile broadband devices, radio resource allocation schemes should be able to assign the resources with a reasonable or possibly minimum amount of transmission overhead. Also, how the temporal dynamism of present-day networks from the perspective of the changes induced by the varying number of users in the system and varying application usage percentage in the smart devices would affect the radio resource allocation performance and signaling is important. An excessive signaling can be a limiting factor in realizing a resource allocation technique regardless of the amount of nuances included in its structure.

Another issue to consider for the radio resource allocation methods, devised for present-day

cellular systems and aimed at provisioning QoS in the network, is accommodating bandwidth augmentative novelties such as spectrum sharing. In the light of what was discussed so far, portions of the government-held spectrum has been released for mobile broadband purposes. Inasmuch as the majority of the currently-released bands are shared with radar systems, novel resource allocation schemes which account for the spectrum sharing with the band incumbents (for example, radars and satellites) can sustain as a part of the intelligent future-minded cellular communications networks in the long run.

Even though the creation of sophisticated yet pragmatic radio resource allocation schemes, which consider the traffic type and dynamic, user subscription, and spectrum sharing, is an effective step toward development of intelligent cellular communications systems which are capable of meeting the data volume demands and application QoS requirements of the au courant wireless networks, the generated traffic might pass through the core network. Since the core network can contain legacy networks, the voluminous mobile broadband data traffic, treated well in the RAN by an efficient resource management and allocation, can suffer severely in the legacy core network not equipped with sophisticated resource allocation and management schemes. It is noteworthy that while LTE has an all IP core network, the traffic generated by the smart devices may have to go through legacy networks and there is no way to provision an end-to-end QoS over the entire backhaul networks.

On the other hand, the outrageous perpetual increase in the traffic volume generated in mobile broadband networks causes the backhaul to be a problematic bottleneck in present-day cellular networks. Furthermore, the QoS requirements of the generated traffic will observe grim chances of being fulfilled over the entire backhaul network; this severely adversely impacts the performance of the applications running on the smart devices and degrades the QoE in the cellular network. Such a deterioration of the users' experience, which appears in the form of lengthy delays or excessive losses for the traffic, leads to the subscriber churn.

In order to provision for the delays that the traffic might incur in the backhaul, it would be

helpful to create prototypes that would model the procrastination that the backhaul network, containing legacy portions, can impose on the data traffic originated from or intended for the mobile broadband devices. Such a delay-based modeling provides an insight into how the QoS performance of the traffic generated in present-day networks may be undermined.

Ultimately, adding further spectrum to the bandwidth available for mobile broadband services should be pragmatically facilitated for the involved parties, i.e. the band incumbents and potential entrants. Spectrum is a scarce expensive resource which can be an object of high remuneration for the incumbents willing to share it with the entrants. Due to the fact the communications ecosystem alters spatially and temporally at a fast pace, such changes should be accommodated to allow for an efficient utilization of the spectrum. As a case in point, a radar system which operates in the 3.5 GHz band might be spatially closeby an in-band cellular communications system in short time periods. During other times when the radar is not in the vicinity of the cellular system, the spectrum can be better utilized by being released to mobile broadband services. Such a better utilization of the bandwidth can be achieved by devising auctioning systems to lease the spectrum to an entrant system while maximizing the monetary revenue for the band incumbent. On the other hand, such auction systems should be secure to ensure the fairness of the auctions for the potential entrants to the band. In addition, the auction privacy would be a concern for the potential entrants to preclude virtual price inflation.

1.5 Contributions

The contributions of this research study is as follows. We develop a radio resource allocation scheme, based on a proportional fairness formulation and hybrid realtime and delay-tolerant traffic, for cellular communications systems, and instrument the radio resource allocation scheme with the capability to differentiate the users to include subscription-based

prioritization into the resource allocation mechanism. Moreover, we equip the radio resource allocation method with a capability to differentiate the applications based on the QoS modeling of the applications via utility functions as well as with an ability to consider temporal changes of application usages for the UEs present in the system. Most importantly, we prove that the developed radio resource allocation scheme is convex and has a tractable global solution; therefore, the rate allocations achieved through our resource allocation scheme are optimal. Such proof migrates the problem of hybrid traffic resource allocation from an NP - hard problem to one solvable with a polynomial complexity. Besides, we show that the developed scheme refrains from dropping any users by allowing for a nonzero allocation in all times. Furthermore, we construct a centralized architecture for the radio resource allocation scheme and obtain solution algorithms leading to solution in a single set of message exchanges between UEs and their BS, which directly allocates application rates. Not only we implement the centralized approach on a real-world network, but also we decompose the application resource allocation into a simpler distributed architecture, containing network and device optimizations, and create solution algorithms to give the optimal rates in a series if messages exchanged between the applications and their host UEs and between UEs and their covering the BS. In addition, we prove that mathematical equivalence of the two architectures. Additionally, we analyze the transmission overhead of the devised centralized and distributed architectures as well as the sensitivity of the methods to temporal dynamics occurred in the number of UEs or application usages in the system. The convergence of the shadow price for both methodologies are analyzed mathematically, and variations of solution algorithms to provide with all-time convergence is provided. The radio resource allocation scheme is instrumented with spectrum-additive measures such as spectrum sharing considerations. We also expand the devised scheme to account for a discrete semi-optimal resource block allocation as opposed to continuous optimal rates so that the resource allocation procedure becomes pragmatic. Furthermore, We implement the resource allocation scheme in a real-world network and show that applying the algorithms elevates the QoE in the network.

Ultimately, to make the resource allocation more realistic, we factor in the effect of the wireless channel on the resource allocation, so that the resource allocation method considers both the application QoS requirements in terms of the bit rate as well as the quality of the channel.

1.6 Chapters Organization

The rest of this proposal is organized as follows. Chapter 2 discusses the background information needed to understand this thesis. Specially, it looks at the concept of utility functions and their role in describing QoS of traffic in wireless networks. Chapter 3 presents the centralized architecture for the proposed resource allocation scheme and provides with solution algorithms thereof. Furthermore, it implements the proposed centralized resource allocation architecture of chapter 3 on a real-world network and shows that applying the mechanism elevates the QoE in the network. Chapter 4 develops a distributed architecture for the resource allocation scheme and solutions thereof. Furthermore, it provides with a variation of the distributed architecture to add robustness into the resource allocation scheme. Chapter 5 provides with a traffic analysis of the proposed centralized and discrete architectures for the resource allocation. It further investigates the sensitivity of the proposed architectures to the dynamics incurred in the UE quantity and application usage. Chapter 6 extends the proposed resource allocation framework to account for radio resource block allocations and provides with an efficient mechanism to map theoretical continuous rates to discrete rates. Chapter 7 investigates the congested nature of the spectrum introduced by the band incumbents. It further investigates the interference effect from radar systems into LTE to see whether operation is feasible at all within such a congested/contested environment. It also extends the proposed resource allocation architecture to allow for allocating resources from portions of spectrum shared with radar system without causing much interference into the radar and vice versa, when a spectrally-coexistent radar is in the vicinity of the com-

munications system. Chapter 8 makes the resource allocation more realistic by accounting for channel conditions and includes the channel quality as a function of signal-to-noise ratio into the resource allocation. And, finally, chapter 9 concludes this thesis work and provides with suggestions for potential future trajectories from the work presented in the context of this thesis. This procedure is also depicted in Figure 1.4.

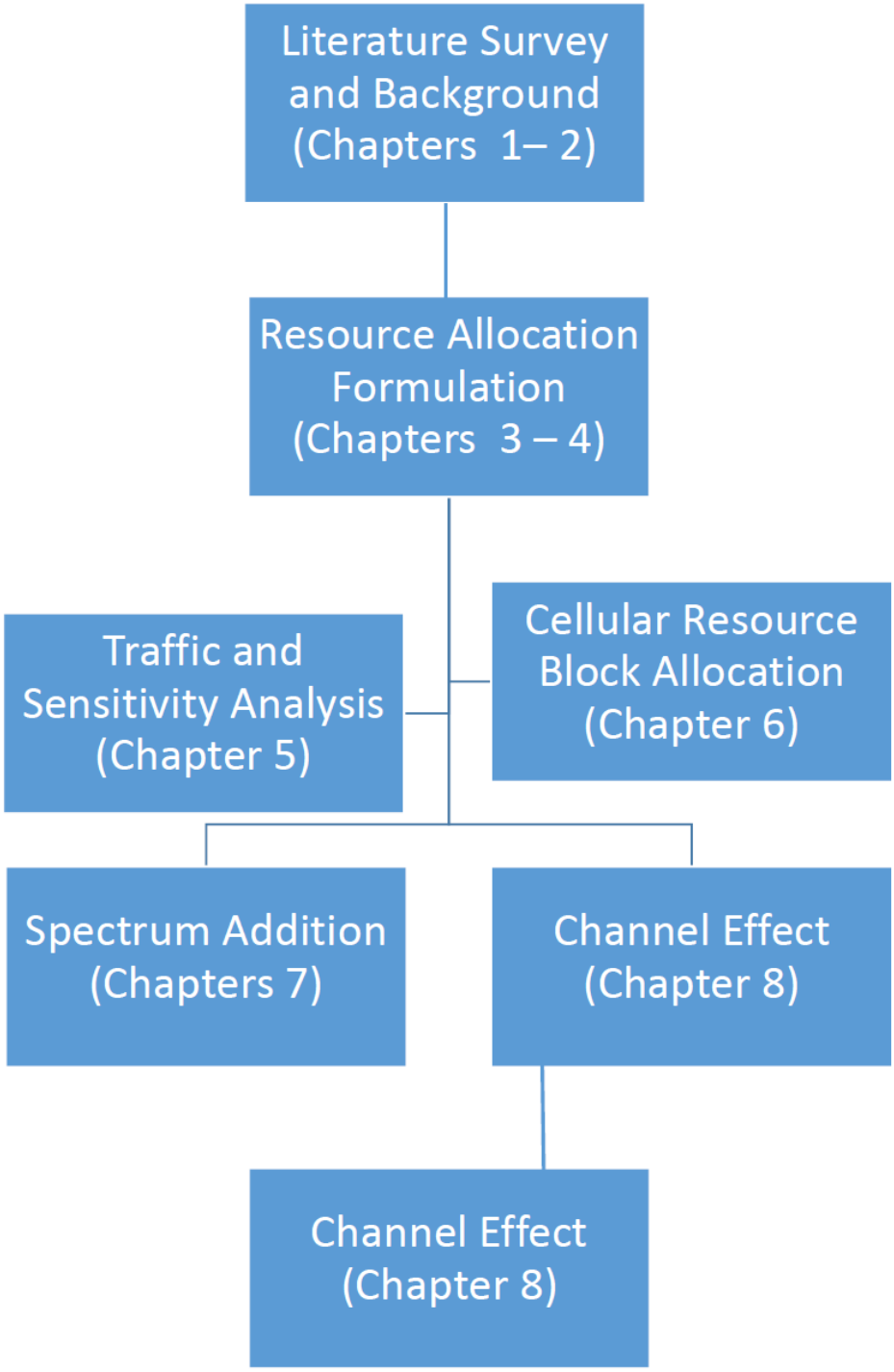


Figure 1.4: Thesis Organization.

Chapter 2

Utility Functions, Optimization, and Resource Allocation

The evolving landscape of radio resource allocation in the world of wireless communications has taken numerous directions relying on miscellaneous methods from linear algebra, machine learning, queuing theory, and so forth. A major portion of radio resource allocation research works which consider QoS have been formulated as distinctive optimization problems in the realm of linear algebra and leverage utility functions to define user requirements; Henceforth, a basic understanding of the utility functions and optimization is essential to understanding the radio resource allocation mechanisms introduced in this document. Therefore, this chapter presents the background information on application utility and optimization formulations for resource allocation. Furthermore, previous research studies on radio resource allocation is presented in section 2.3. In this chapter:

- We present the preliminary information on application utility functions as a QoS measure.
- We present the popular resource allocation formulations in the realm of cellular com-

munications systems.

- we present the basics about the Frank Kelly Algorithm as an efficient solution to convex optimization problems.
- we provide with a literature survey on topical research papers on the subject of resource allocation in mobile broadband systems.

In particular, section 2.1 explains the concept of application utility functions; section 2.2 talks about resource allocation formulations and presents two popular max-min and proportional fairness techniques germane to resource allocation; and, section 2.3 provides a literature survey about resource allocation in modern cellular networks. Next, section 2.1 presents the concept of application utility functions central to our resource allocation study.

2.1 Application Utility Functions

Utility function have been used in a wide variety of research works to model some representative characteristic of the system. For instance, [49, 50] leveraged utility functions to model the modulation schemes in a power allocation problem. In order to have mobile broadband users enjoy utilizing a cellular network, resource allocation to applications running on the smart devices operating in the network should consider QoS requirements for the traffic generated by the applications. The extent to which the QoS requirements of traffic generated by an application are fulfilled can be described by the utility function for that application, generally referred to as application utility function which maps a feasible rate allocation to a utility level which is the QoS fulfillment percentage for the application.

There are little evidence about the precise shape for application utility functions [51–54], and we can conjecture about their qualitative properties. The elastic traffic generated by traditional applications such as file transfer protocol (FTP) and simple mail transfer protocol

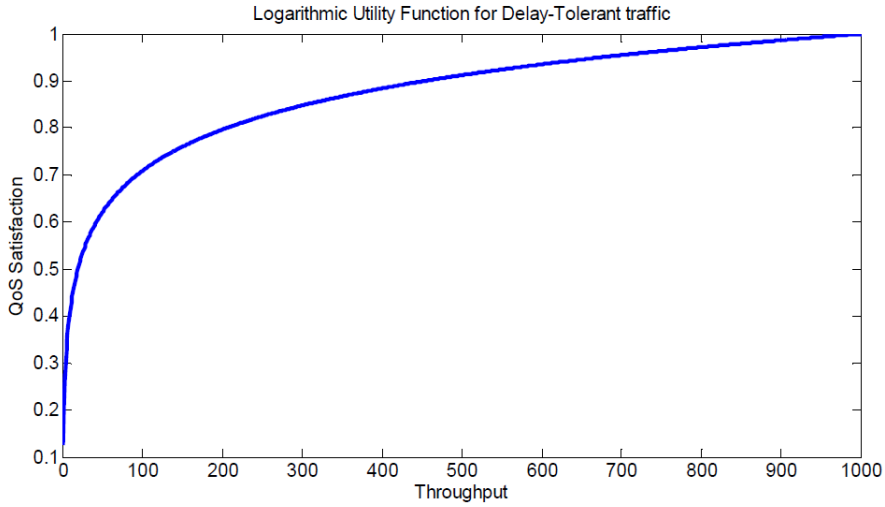


Figure 2.1: Utility function $U(x)$ for delay-tolerant applications vs. rate allocation x.

(SMTP) is characterized by its ability to have its rate adapted in presence of congestion and to tolerate delays [55–58]. Applications which generate such an elastic traffic are referred to as delay-tolerant applications. On an intuitive level, the QoS fulfillment for delay tolerant applications seems to have a decreasing marginal return for an increasing rate allocation.

The application utility functions for the elastic traffic generated by delay-tolerant applications look something like that in Figure 2.1. Here, we can observe a diminishing return in the application utility function as the assigned rate is increased. The application utility function is convex everywhere

Some other applications such as telephony and link emulation generate an inelastic traffic which requires its data to arrive within a given delay bound even though it does not care if the data arrive earlier; On the other hand, the application performs really poorly if the data arrives later than the delay bound. Such inelastic-traffic-generating applications, referred to as real-time applications, are mainly those which expect circuit-switched services and require a minimum bandwidth before meeting an acceptable performance. The application utility function for the inelastic traffic generated by real-time applications look something like that in Figure 2.2. As we can observe from this figure, so long as the delay bounds are being met

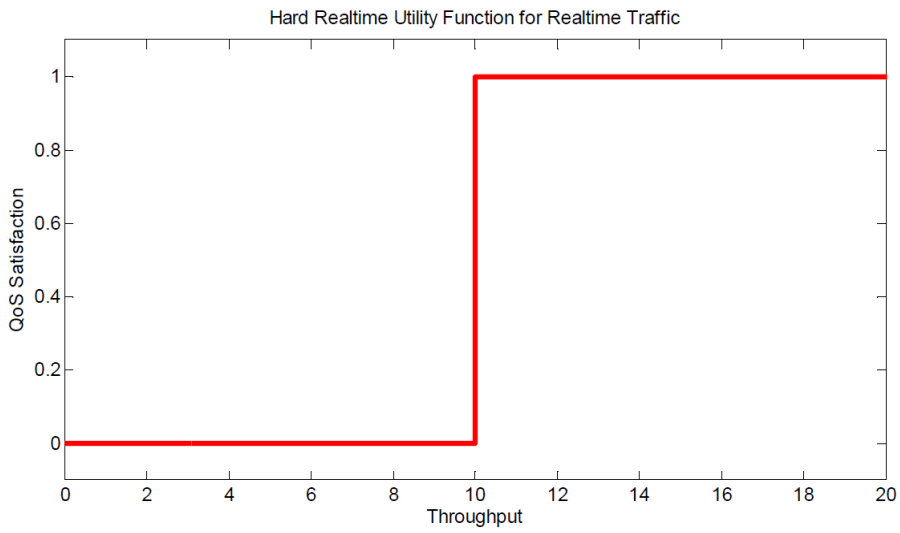


Figure 2.2: Utility function $U(x)$ for hard real-time applications vs. rate allocation x.

the the performance is almost constant, while as soon as the bandwidth share drops below what is needed to meet the required delay bounds, the performance falls sharply to zero.

While the video and audio applications have been designed with hard requirements, they can be mostly implemented to be rather tolerant of occasional violations of the delay bounds and dropped packets. The intrinsic bandwidth requirements for the real-time applications are because the inelastic traffic data generation rate is independent of the congestion over the network. Therefore, the performance of the traffic degrades severely as the bandwidth share for the application generating the traffic becomes less than the intrinsic generation rate for the traffic. Such soft real-time applications can have utility functions like Figure 2.3.

Here, the performance satisfaction is almost constant after meeting the delay bounds while it drops near zero when the bandwidth share for the application falls below what is needed to meet the delay bounds. In contrast with hard real-time applications, the performance drop of the delay-adaptive real-time application is not as sharp as the hard real-time application in Figure 2.2, while the shape is very similar. As we can see, the shape is convex in the neighborhood around zero and is concave after the minimum bandwidth allocation for the

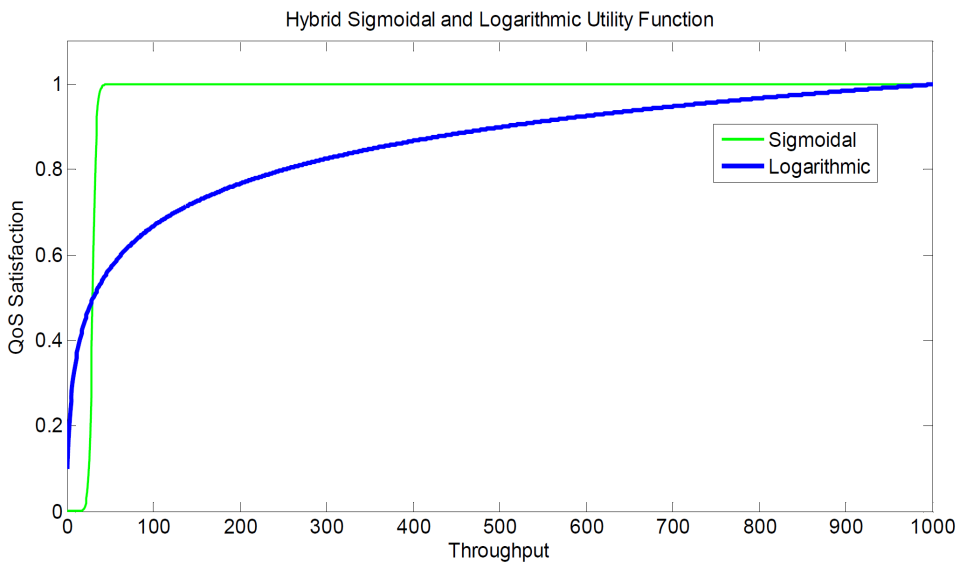


Figure 2.3: Utility function $U(x)$ for soft real-time applications vs. rate allocation x.

application is realized.

From now on, unless explicitly stated, this document refers to an application performance satisfaction vs. a function of its allocated rates as application utility function and denotes it $U(r)$ for the allocated rate r. Application utility functions have the following properties [55, 59, 60].

- $U(0) = 0$ and $U(r)$ is an increasing function r .
- $U(r)$ is twice differentiable in r and bounded above.

The first statement of the former property implies the nonnegativity of the utility functions which is expected since they represent application performance satisfaction percentage, whereas its second statement reveals that the more assigned rate, the higher the application performance satisfaction. On the flip side, the latter property indicates the continuity of the utility functions. Hybrid traffic consists of elastic and inelastic traffic streams sprung from respectively delay-tolerant and real-time applications whose utilities are conducively modeled by correspondingly normalized logarithmic and sigmoidal utility functions in equations

(2.1) and (2.2) in that order [61, 62].

$$U(r) = c\left(\frac{1}{1+e^{-a(r-b)}} - d\right) \tag{2.1}$$

Here, $c = \frac{1+e^{ab}}{e^{ab}}$ and $d = \frac{1}{1+e^{ab}}$. It can be easily verified that $U(0) = 0$ and $U(\infty) = 1$, where the former is one of the previously mentioned utility function properties and the latter indicates that an infinite resource assignment ensues 100% satisfaction. Furthermore, it is easily derivable that the inflection point of equation (2.1) occurs at $r = r^{\text{inf}} = b$, where the superscript "inf" stands for infliction. We can do this by twice differentiating $U(r)$ with respect to r and setting the second derivative equation equal to zero, i.e. $\frac{\partial^2 U}{\partial r^2} = 0 \rightarrow r = b$.

$$U(r) = \frac{\log(1+kr)}{\log(1+kr^{\max})} \tag{2.2}$$

Here, $r^{\max}$ is the maximum rate at which the application QoS is satisfied in full (100% application utility) and k is the utility function increase with augmenting the allocated rate r. It can be easily checked that $U(0) = 0$ and $U(r^{\max}) = 1$, where the former is again the basic property of the utility functions and the latter implies that a 100% QoS satisfaction occurs at $r = r^{max}$. Moreover, the inflection point of normalized logarithmic function is at $r = r^{\text{inf}} = 0$.

For the sake of illustration, the application utility functions with the parameters according to Table 2.1 are considered. Here, the sigmoidal application utility function with parameters $a = 5$, $b = 10$ approximates a step function at rate $r = 5$ and is a good model for VoIP, while parameters $a = 3$, $b = 15$ is an approximation of a real-time application with an inflection point at rate $r = 15$ and is conducive to modeling standard definition video streaming, whereas parameters $a = 1$, $b = 25$ is an estimation of another real-time application with the inflection point $r = 25$ and is appropriate for the high definition video streaming. Moreover,

Table 2.1: Applications Utility Function Parameters

Applications Utilities Parameters	
UE1 App1	Sigmoid $a = 5,\ b = 5$
UE2 App1	Sigmoid $a = 4,\ b = 10$
UE3 App1	Sigmoid $a = 3,\ b = 15$
UE4 App1	Sigmoid $a = 2,\ b = 20$
UE5 App1	Sigmoid $a = 1,\ b = 25$
UE6 App1	Sigmoid $a = 0.5,\ b = 30$
UE1 App2	Logarithmic $k = 15,\ r^{\max} = 100$
UE2 App2	Logarithmic $k = 12,\ r^{\max} = 100$
UE3 App2	Logarithmic $k = 9,\ r^{\max} = 100$
UE4 App2	Logarithmic $k = 6,\ r^{\max} = 100$
UE5 App2	Logarithmic $k = 3,\ r^{\max} = 100$
UE6 App2	Logarithmic $k = 1,\ r^{\max} = 100$

the logarithmic application utility functions with $r^{\max} = 100$ and distinct k_i parameters estimate delay-tolerant FTP applications .

The plots of the application utility functions in Table 2.1 are shown in Figure 2.4. We can observe that the real-time applications require a minimum rate, i.e. the inflection point, after which the application QoS is fulfilled to a large extent. On the other hand, the logarithmic application utility function is provided with some QoS even at low rates which is suitable for the delay-tolerant nature of the applications. Furthermore, as we can observe from Figure 2.4, in compliance with the aforementioned mathematical properties for the application utility functions, the plotted utility functions are strictly increasing continuous functions and are zero valued at zero rates.

Because the sigmoidal utility functions gain a slight QoS satisfaction only after the allocated rates surpass the inflection points of the utilities whereas the logarithmic ones obtain

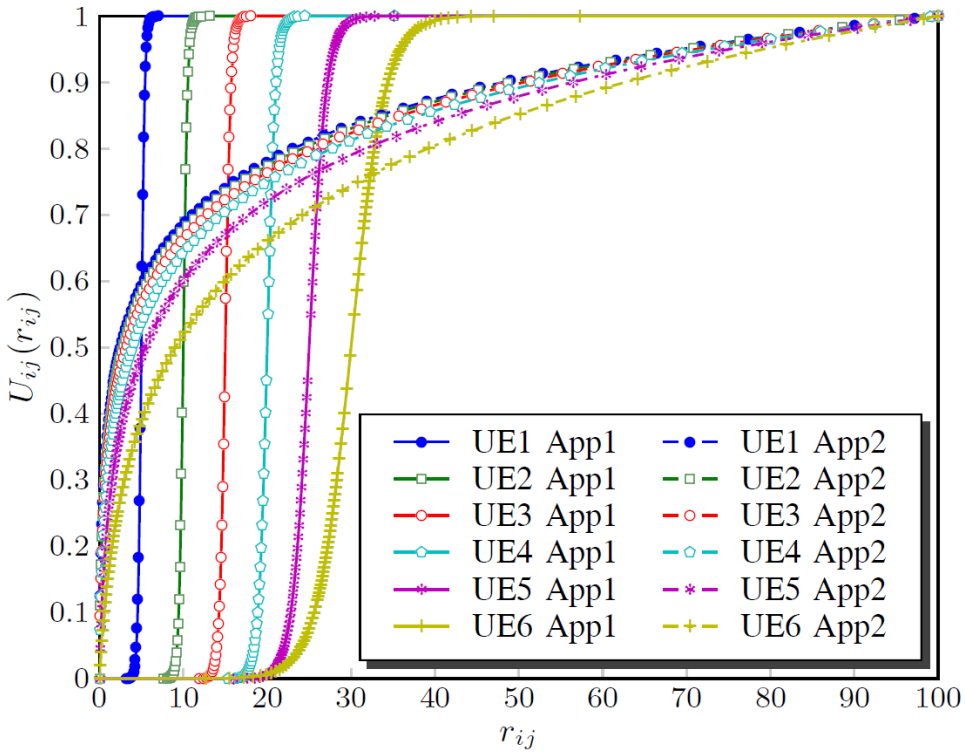


Figure 2.4: Delay-tolerant and real-time applications have respectively logarithmic and sigmoidal utility functions U_{ij} plotted against the application-assigned rates r_{ij}.

some QoS fulfillment even for a minuscule assigned bandwidth, such behaviors make sigmoidal and logarithmic utility functions suitable for modeling real-time and delay-tolerant applications respectively, and the germane mathematical analyses appear in ([61,63–65]) in nuance.

Next, section 2.2 explains some points regarding various factors that affect resource allocation formulation.

2.2 Resource Allocation and its Formulations

In mathematics literature dealing with utility functions and rate allocation, a variety of formulations have been considered for the problem of resource allocation and numerous solution methods for the formulations have been proposed. Amongst various formulations,

proportional fairness [66–68] and max-min fairness [69–72] have received a drastic attention as they can lead to optimal solutions [25,73,74]. Furthermore, the author in [75,76] defines a solution to be either Pareto optimal, or Pareto inefficient, or infeasible solutions. The latest indicate that an allocation is not possible based on the available resources and the demand from the network. On the other hand, [75] defines Pareto inefficiency as an assignment of resources which do not allocate all resources and it defines Pareto optimality as allocating all of available resources. In our resource allocation work in chapters 3 and 4, we will be using methods that lead to Pareto optimal solutions.

2.2.1 Max-Min Resource Allocation

A feasible allocation of resources achieves max-min fairness if any attempt toward increasing an assignment to one of the entities lead to decreasing the assignment in another entity which do not possess more resources than the entity for which we increase the resources. Therefore, this method achieves the highest utility with the lowest levels of utility [75,77,78]. Leveraging max-min fairness can shape the traffic, say as opposed to a First-Come First-Serve (FCFS) multiplexing, by not letting a heavy flow consisting of large packages preclude serving other flows in the network. For the utility function $U(r)$, where r is the allocated resource amount, max-min fairness can be formalized as equation (2.3).

$$r = argmax_r min_r U(r) \tag{2.3}$$

Water filling [79–81]can be used to solve the max-min fairness problems. The authors in [69] proved that a max-min fairness policy have difficulty dealing with bottlenecks in the network. Next, we look at the proportional fairness policy for resource allocation.

2.2.2 Proportional Fairness

A feasible allocation of resources achieves proportional fairness if it maximizes the system's overall utility while provisioning a minimal service to all system entities in need of resources. This is done by assigning each flow a rate inversely proportional to its resource consumption [82, 83]. For the utility function $U_i(r_i)$, where r_i is the allocated resource amount to the ith user, proportional fairness can be formalized as equation (2.4).

$$r_i = argmax_{r_i} \prod_{i=1}^{N} U_i(r_i) \tag{2.4}$$

In accordance with the properties of application utility functions in section 2.1, $U_i(r_i = 0) = 0$ which zeros the entire system utility , i.e. $\prod_{i=1}^{N} U_i(r_i)$. Hence, no user will be allocated a zero rate under this policy. Various methods for solving proportional fairness models have been proposed, and the most important of them are the Frank Kelly algorithm [84] and weighted fair queuing (WFQ) ([85–87]). Frank Kelly algorithm is an iterative procedure lets users bid for resources until the algorithm reaches an optimal rate allocation based upon the utility functions used in the formulation and the shadow price (amount of consumed resources per data bit) [84]. On the other hand, proportional fairness can be achieved by setting the inverse of the shadow price as the weights of the WFQ. We will be using a proportional fairness formulation for our radio resource allocation work presented in chapters 3 and 4. However, we would tailor the formulation to include UE priorities and application temporal usage percentages in addition to the application utility function constituents which represent the traffic.

2.2.3 Frank Kelly Algorithm

Frank Kelly algorithm as a seminal work to obtain proportional fairness was introduced in [84]. The authors in [84] also proved that their method causes Pareto optimality of a resource allocation solution that is obtained for a proportional fairness formulation. According to [84], the procedure starts by users sending their bids w_i to a resource allocation manager entity which sets the shadow price as sum of the bids averaged on the entire resources R available for the manager, i.e. $p = \frac{\sum_i^N w_i}{R}$. The ratio of the bids to the shadow price, $r_i = \frac{w_i}{p}$, derives the rate assignments. Then, users check whether the rate are optimal by solving $r_{i,textopt} = \arg\max_{r_i}\Big(U_i(r_i) - pr_i\Big)$ and if $r_i \neq r_{i,\text{opt}}$, they send new bids $w_i = r_{i,\text{opt}}p$ to the resource allocation entity and the procedure continues until a convergence occurs, that is the derivative of the utility function equals the shadow price $\frac{\partial U_i}{\partial r_i}|_{r_i=r_{i,\text{opt}}} = p$ ([84], [75]). This routine is summarized in Algorithm 1.

Algorithm 1 Frank Kelly Algorithm

Send initial bid $w_i(n = 1)$ to the resource allocation managing entity.
loop
  Calculate shadow price $p(n) = \frac{\sum_i^N w_n}{R}$.
  Receive shadow price $p(n)$ from the resource allocation managing entity.
  Calculate allocated rate $r_i = \frac{w_i(n)}{p(n)}$.
  Solve $r_{i,\text{opt}} = \arg\max_{r_i}\Big(U_i(r_i) - p(n)r_i\Big)$.
  if $r_i \neq r_{i,\text{opt}}$ **then**
    Calculate $w_i = r_{i,\text{opt}}p$.
    Send the bid $w_i(n)$ to the resource allocation managing entity.
  end if
end loop

We will be using a method based on the Frank Kelly algorithm to solve the proportional fairness resource allocation formulation that will be developed in chapters 3 and 4. Next, section 2.3 presents the previous studies about radio resource allocation.

2.3 Previous Studies on Resource Allocation

The resource allocation optimization research area has received a significant attention since the seminal network utility maximization study in [84] which allocated user rates through a utility proportional fairness maximization solved by the Lagrange multipliers [88]. Soon after, an iterative algorithm relying on the duality of the aforementioned resource allocation problem was proposed [89–91]. Whilst the traffic in these early research works had an elastic nature common for wired communication systems and approximated by concave utility functions, the advent and prevalence of high-speed wireless networks have entailed an increased utilization of real-time applications whose utility functions grow non-concave [55]. For instance, the utility of a VoIP can be represented as a step function whose utility is zero before a certain threshold rate and achieves 100% for rates larger than the threshold. Another example is a video streaming application whose utility can be approximated as a sigmoidal function convex (concave) for rates below (above) its inflection point. As such, the methods presented in ([84, 89]) incur the proceeding drawbacks: (a) Reaching optimal solutions for solely concave utility functions, they are inapplicable to the drastically escalating inelastic traffic volume of au courant networks; (b) Neither priority do they render to real-time applications with stringent QoS requirements, nor they reserve any attention for the application statuses, nor they look after subscribers' varied importance pivotal from a business standpoint.

Later, the authors in ([61, 92, 93]) presented distributed rate allocation algorithms for multi-class service offerings based on concave and sigmoidal utility functions representing applications. Despite closely approximating optimal solution, involved methods dropped users to maximize the system utility, so they could not guarantee a minimal QoS. An effort by the authors in ([94–96]) proposed a utility proportional fairness resource allocation, for users of a single-carrier communication network, cast as a convex problem with logarithmic

and sigmoidal utility functions respectively modeling delay-tolerant and real-time applications. Although their schemes prioritized the real-time applications over the delay-tolerant ones, they neither contemplated the application status or user differentiation concepts, nor regarded the hybrid traffic prevailing in modern networks.

In [75], the author considered a weighted aggregation of logarithmic and sigmoidal utilities approximated to the nearest concave utility function via a minimum mean-squared error measure inside UEs. The approximate utility function solved the rate allocation optimization through a variation of the conventional distributed resource allocation approach in [84] such that rate assignments essentially estimated optimal ones. However, the rate were only approximations and no consideration was given to user or application priorities. This work was extended by Shajaiah et. al. ([97, 98]) to allow for the application of the resource allocation in a multi-carrier network in public safety.

The authors in ([99, 100]) and ([101–103]) considered a similar multicarrier optimal resource allocation aware of the subscriber priorities. However, no attention was rendered to the temporal changes in the application usage or UE quantities. In [104], the authors adopted a non-convex optimization formulation to maximize the system utility in wireless networks consisting of applications with logarithmic and sigmoidal utility functions. A distributed process was employed to obtain the rates under a zero duality gap; but, the algorithm did not converge for a positive duality gap leading to compounding a heuristic to ensure the network stability.

In other studies, the authors of [105, 106] created a utility max-min fairness resource allocation for the hybrid traffic sharing a single path in a communications network. Similarly, [63,107,108] presented a utility proportional fairness optimization for the high signal-to-interference-plus-noise ratio (SINR) wireless networks using a utility max-min architecture, contrasted against the traditional proportional fairness algorithms [109–112] and provided a closed-form solution that refrained from network oscillations. However, neither methods

cared for any traffic or user priorities in assigning the spectrum. In ([113–115]), the authors developed a utility proportional fairness resource block allocation in wireless networks as an integer optimization problem. They initially obtained the continuous optimal rates and then took on a boundary mapping technique to extracted a pool of valid resource blocks tantamount to inferred optimal continuous rates, albeit neither hybrid traffic, nor application status, nor user importance was taken into the equation.

In a similar work [116–118], the authors organized a utility proportional fairness optimization which allocated optimal UE rates in a cellular infrastructure coexistent with radars by leveraging the Lagrange multipliers. Finally, [119–122] presented a subcarrier allocation in orthogonal frequency division multiplexed systems concentrating on delay constrained data and used network delay models [123–125] for the subcarrier assignment. And last but not the least, [126] developed a location/time/context-aware source allocation in cellular networks; however, they did not consider the temporal changes in the application usage percentage, the number of UEs, or subscribers' priority.

The authors of [2] investigated the WiMAX-radar mutual interference and concluded that large geographic separations between the two systems are required, precluding WiMAX deployability in the coastline. Cotton et. al. [127–132] performed tests, using a shipborne radar in San Diego littoral waters, measuring the temporal band occupancy and found that the 3.5 GHz spectrum is not often occupied by radar transmissions, underlining the potential of the germane band for spectrum sharing. Lackpour et. al. [133] suggested a general spectrum sharing scheme based on time, space, frequency, and system-level modifications, inconducive to real-world implementation. Sanders et. al. [134] performed experiments with RF hardlines to observe the interference effects from radar waveforms into a 3.5 GHz LTE base station (BS). They observed the throughput loss and block error rate (BLER) for the LTE system in the UL direction; however, their results were varied as some waveforms did not have any effect on the LTE while others undermined the performance drastically. Neither

did they consider any propagation models, nor did they perform any simulation of realistic radar or LTE system. Furthermore, Bjornson has extensively written on resource allocation [135–143], and a lot of theoretical contributions are provided by [93, 144–153, 153–155].

2.4 Chapter Summary

In this chapter, we introduced the concept of application utility functions and their relationship to satisfying the QoS requirement of various delay tolerant and real-time applications. We presented appropriate mathematical models for the application utility functions of real-time and delay-tolerant traffic as respectively sigmoidal and logarithmic utility functions. Furthermore, we discussed about popular resource allocation frameworks in modern resource allocation studies in cellular systems and introduced max-min and proportional fairness optimizations. Moreover, we presented the current efficient solution algorithms for the aforementioned resource allocation frameworks and presented some preliminary information on Frank Kelly algorithm as an efficient solution for the proportional fairness optimization. And, we performed a literature survey on topical resource allocation works in the realm of cellular communications systems.

Chapter 3

A Centralized Architecture for Resource Allocation

Mobile broadband services have been falling afoul of a perennially upsurged demand for radio resources during recent years. This upswing owes to the gigantic boom in mobile service subscribers' quantity as well as to the outgrowth of their generated traffic volume [3]. On the other hand, the migration of cellular network providers from offering a single service such as the Internet access to a multi-service framework, like multimedia telephony and mobile-TV [156], along with the emergence and prevalence of smartphones hosting simultaneously running delay-tolerant and real-time applications with distinctive quality of service (QoS) requirements [157, 158] arise an urgency to dynamically provisioning various bit rates to the application traffic so as to elevate users' quality of experience (QoE) tightly bound to the subscriber churn [156]. As such, incorporating service differentiation mechanisms into resource allocation methods is a matter of high consequence.

Inasmuch as applications' temporal usage percentage directly impacts the generated traffic volume and nature, e.g. the traffic elasticity, including the usage percentage as an application status differentiation in resource allocation schemes is worthwhile. Besides, cellular

network providers capability to adopt a subscription-based differentiation [156], wherein miscellaneous clients of an identical service receive differentiated subscription-based treatments (corporate vs. private, post-paid vs. pre-paid, and privileged vs. roaming users), can fine-tune resource allocation approaches. Henceforth, resource allocation modi operandi can accommodate diverse exigencies of present-day wireless networks conveying the hybrid traffic by accounting for all the aforementioned issues. Nonetheless, the majority of resource allocation proposals fizzle to address the aforesaid concerns collectively.

This chapter puts forward a novel convex utility proportional fairness maximization formulation for an optimal resource allocation in wireless networks and is outfitted with the subscriber, application status, and service differentiations parameterized respectively as user equipment (UE) subscription weights, application status weights, and application utility functions. The weights are supplied by network providers so that a foreground-running application such as a voice call attains a higher application status weight than do the background-running ones, e.g. an automatic application update process. Mobile subscribers of the system under our consideration can concurrently run multiple applications with their utility functions and statuses depending on the generated traffic nature and instantaneous usage percentage, respectively.

Casting the service differentiation under a utility proportional fairness policy prioritizes the real-time traffic over the delay-tolerant one, conducive to fulfilling QoS requirements. In addition to solving the formalized optimization problem analytically, we develop a centralized architecture for the resource allocation which assigns hybrid application rates in a monolithic stage transacted in the cellular network provider side of the communications system.

3.0.1 Contributions

In this chapter:

- We present a utility proportional fairness formulation for resource allocation in cellular communications system.
- We ornament the proposed resource allocation with mechanisms to differentiate traffic based on QoS requirements, applications temporal usages, and UE priorities.
- We develop a centralized architecture for the proposed resource allocation framework.
- We show that the devised centralized architecture is a convex optimization.
- We provide with solution algorithms for the centralized architecture which assigns application rates by the base station (BS) in a single stage.
- We provide with simulations to show the application of the proposed resource allocation in a cellular network simulation scenario.

The remainder of this chapter is organized as follows. Section 3.1 presents the optimization formulation for the radio resource allocation work in this document. Section 3.2 presents a centralized resource allocation for the utility proportional fairness resource allocation framework that is developed in this chapter. Section 3.3 proves that the devised centralized resource allocation is convex. Section 3.4 provides with solution algorithms for the devised centralized resource allocation architecture. Section 3.5 portrays simulations for the centralized resource allocation in this chapter. And, section 3.7 summarizes this chapter.

Next, section 3.1 concocts the system model for the rate allocation problem proposed in this work.

3.1 Resource Allocation Optimization

The objective is to determine optimal rates that hybrid-traffic-carrying cellular communications systems should be allocating to their UE applications so as to dynamically ensure as

such: 1) Real-time applications are rendered priority over delay-tolerant ones. 2) no user is dropped 3) Applications temporal usage is accounted for. 4) Subscription-based treatments is honored. We assume each UE contains multiple simultaneously running real-time and delay-tolerant applications, mathematically represented by sigmoidal and logarithmic utility functions as shown in the next chapter.

3.1.1 System Model

To present the system, with no loss of generality, we concentrate on a cellular network's single cell, which subsumes a BS covering M UEs (here $M = 6$) depicted in Figure 3.1, where each UE concurrently runs delay-tolerant and real-time applications represented respectively by the logarithmic and sigmoidal utility functions in section 2.1. The rate assigned by the eNB to the ith UE is denoted as r_i and the UE's aggregate utility function is shown as $V_i(r_i)$, which we relate it to the UE application utilities accordingly to the equation (3.1) below.

$$V_i(r_i) = \prod_{j=1}^{N_i} U_{ij}^{\alpha_{ij}}(r_{ij}) \tag{3.1}$$

Here, r_{ij}, $U_{ij}(r_{ij})$, and α_{ij} respectively represent the rate allocation, application utility function, and application usage percentage of the jth application running on the ith UE. Hence, we can write $\sum_{j=1}^{N_i} \alpha_{ij} = 1$ and $r_i = \sum_{j=1}^{N_i} r_{ij}$, where, for the ith UE, N_i is the number of coevally running applications and r_i presents the bandwidth allotment by the eNB. The former of the afore-written equations states the fact that the addition of the ith UE's application usage percentages proves 100% usage percentage, and the letter one implies that the the ith UE rate is the augmentation of all its N_i applications resources assignments.

Before proceeding further, the following lemma is valuable and will be used in later in this chapter for the centralized architecture that we will be developing in section 3.2.

Figure 3.1: System Model: Single cell, within the cellular network, with an eNB covering $M = 6$ UEs each with simultaneously running delay-tolerant and relay-time applications represented by logarithmic and sigmoidal utility functions respectively.

Lemma 3.1.1. *The aggregate utility natural logarithm* $\log(V_i(r_i))$ *is strictly concave.*

Proof. From equation (3.1), we can write $\log V_i(r_i) = \sum_{j=1}^{N_i} \alpha_{ij} \log U_{ij}(r_{ij})$ where $U_{ij}(r_{ij}) > 0$ in accordance with section 3.1 utility function properties. Also, logarithmic utilities (Utility concavity stems out $U'_{ij}(r_{ij}) = \frac{dU_{ij}(r_{ij})}{dr_{ij}} > 0$ and $U''_i(r_{ij}) = \frac{d^2U_{ij}(r_{ij})}{dr_{ij}^2} < 0$, resulting in $\frac{d\log(U_{ij}(r_{ij}))}{dr_{ij}} = \frac{U'_{ij}(r_{ij})}{U_{ij}(r_{ij})} > 0$ due to $U_{ij}(r_{ij}) > 0$ and $U'_{ij}(r_{ij}) > 0$ and in $\frac{d^2\log(U_{ij}(r_{ij}))}{dr_{ij}^2} = \frac{U''_{ij}(r_{ij})U_{ij}(r_{ij}) - U'^2_{ij}(r_{ij})}{U^2_{ij}(r_{ij})} < 0$ due to $U''_{ij}(r_{ij}) < 0$. Thus, the logarithmic utility natural logarithm is strictly concave. On the flip side, for a sigmoidal utility $U_{ij}(r_{ij})$ with $0 < r_{ij} < R$, we have the following inequalities amongst which the first owes to the sigmoidal function' s continuity and $0 \leq U_{ij}(r_{ij}) < 1$ and the rest are utter algebraic manipulation of the first one.

$$
\begin{aligned}
0 < c_{ij}\left(\frac{1}{1+e^{-a_{ij}(r_{ij}-b_{ij})}} - d_{ij}\right) &< 1 \\
d_{ij} < \frac{1}{1+e^{-a_{ij}(r_{ij}-b_{ij})}} &< \frac{1+c_{ij}d_{ij}}{c_{ij}} \\
\frac{1}{d_{ij}} > 1+e^{-a_{ij}(r_{ij}-b_{ij})} &> \frac{c_{ij}}{1+c_{ij}d_{ij}} \\
0 < 1 - d_{ij}(1+e^{-a_{ij}(r_{ij}-b_{ij})}) &< \frac{1}{1+c_{ij}d_{ij}}
\end{aligned}
$$

For $0 < r_{ij} < R$, we have the following inequalities, of which the first results from first additive's denominator positivity in addition to the formerly derived statement $0 < 1 -$

$d_{ij}(1 + e^{-a_{ij}(r_{ij}-bij)}) < \frac{1}{1+c_{ij}d_{ij}}$ as well as other constituents' positivity and the last one is verifiable by investigating its terms algebraically. Hence, the sigmoidal utility natural logarithm is strictly concave. As such, the applications utility functions $U_{ij}(r_{ij}) > 0$ of the system model (equation (3.1)) have strictly concave natural logarithms, meaning that the aggregated utility $\log V_i(r_i) = \sum_{j=1}^{N_i} \alpha_{ij} \log U_{ij}(r_{ij})$ is strictly concave.

$$\begin{aligned}
\frac{d}{dr_{ij}} \log U_{ij}(r_{ij}) =& \frac{a_{ij} d_{ij} e^{-a_{ij}(r_{ij}-b_{ij})}}{1 - d_{ij}(1 + e^{-a_{ij}(r_{ij}-b_{ij})})} \\
&+ \frac{a_{ij} e^{-a_{ij}(r_{ij}-b_{ij})}}{(1 + e^{-a_{ij}(r_{ij}-b_{ij})})} > 0 \\
\frac{d^2}{dr_{ij}^2} \log U_{ij}(r_{ij}) =& \frac{-a_{ij}^2 d_{ij} e^{-a_{ij}(r_{ij}-b_{ij})}}{c_{ij}\left(1 - d_{ij}(1 + e^{-a(r_{ij}-b_{ij})})\right)^2} \\
&+ \frac{-a_{ij}^2 e^{-a_{ij}(r_{ij}-b_{ij})}}{(1 + e^{-a_{ij}(r_{ij}-b_{ij})})^2} < 0
\end{aligned} \tag{3.2}$$

□

Now that we have proved the concavity of aggregate utility functions in our system model, the following lemma is also interesting and will be used later in this chapter.

Lemma 3.1.2. *The aggregate utility function $V_i(r_i)$, the slope curvature function $\frac{\partial \log V_i(r_i)}{\partial r_i}$ has its inflection point at $r_i = r_{ij}^s \approx r_{ij}^{inf}$ for j^{th} application utility function U_{ij} and is convex for $r_{ij} > \max_j r_{ij}^s$.*

Proof. For the ith UE aggregate utility $V_i(r_i)$, let $S_i(r_i) = \frac{\partial \log V_i(r_i)}{\partial r_i}$ be the aggregate utility slope curvature function, let $S_{ij}(r_{ij}) = \frac{\partial \log U_{ij}(r_{ij})}{\partial r_{ij}}$ be the jth application utility slope curvature function, and let N_i^S be the number of sigmoidal utility functions. Taking the logarithm and derivative of both sides of equation (3.1) leads to equation (3.3).

$$
\begin{aligned}
S_i(r_i) &= \frac{\partial \log V_i(r_i)}{\partial r_i} = \frac{\partial}{\partial r_i} \sum_{j=1}^{N_i} \alpha_{ij} \log U_{ij}(r_{ij}) \\
&= \sum_{j=1}^{N_i} \alpha_{ij} \frac{\partial \log U_{ij}(r_{ij})}{\partial r_{ij}} = \sum_{j=1}^{N_i} \alpha_{ij} S_{ij}(r_{ij}) \\
&= \sum_{j=1}^{N_i^S} \alpha_{ij} S_{ij}(r_{ij}) + \sum_{j=N_i^S+1}^{N_i} \alpha_{ij} S_{ij}(r_{ij})
\end{aligned} \tag{3.3}
$$

Taking the 1th and 2nd derivatives of equation (3.3), we write:

$$
\begin{aligned}
\frac{\partial S_i}{\partial r_i} = \sum_{j=1}^{N_i^S} \{ & \frac{-\alpha_{ij} a_{ij}^2 d_{ij} e^{-a_{ij}(r_{ij}-b_{ij})}}{c_{ij}\left(1 - d_{ij}(1 + e^{-a_{ij}(r_{ij}-b_{ij})})\right)^2} \\
& + \frac{\alpha_{ij} a_{ij}^2 e^{-a_{ij}(r_{ij}-b_{ij})}}{\left(1 + e^{-a_{ij}(r_{ij}-b_{ij})}\right)^2} \} \\
& - \sum_{j=N_i^S+1}^{N_i} \{ \frac{\alpha_{ij} k_{ij}^2}{(1 + k_{ij} r^{max}) \log(1 + k_{ij} r_{ij})^2} \}
\end{aligned} \tag{3.4}
$$

$$
\begin{aligned}
\frac{\partial^2 S_i}{\partial r_i^2} = \sum_{j=1}^{N_i^S} \{ & \frac{d_{ij} e^{-a_{ij}(r_{ij}-b_{ij})}(1 - d_{ij}(1 - e^{-a_{ij}(r_{ij}-b_{ij})}))}{c_{ij}\left(1 - d_{ij}(1 + e^{-a_{ij}(r_{ij}-b_{ij})})\right)^3} \\
& + \frac{e^{-a_{ij}(r_{ij}-b_{ij})}(1 - e^{-a_{ij}(r_{ij}-b_{ij})})}{\left(1 + e^{-a_{ij}(r_{ij}-b_{ij})}\right)^3} \} \times a_{ij}^3 \alpha_{ij} \\
& - \sum_{j=N_i^S+1}^{N_i} \{ \frac{\alpha_{ij} k_{ij}^2 (\log(1 + k_{ij} r_{ij}) - 1)}{(1 + k_{ij} r_{ij})^2 \log^2(1 + k_{ij} r_{ij})} \}
\end{aligned} \tag{3.5}
$$

It is easy to show that $\forall\, r_i,\ \frac{\partial S_i}{\partial r_i} < 0$. If we denote the 1th term of equation (3.4) as well as the 2nd and the 3rd terms of equation (3.5) as respectively S_i^1, S_i^2, and S_i^3, we can conclude equation (3.6) for which the properties in equation (3.7) are considerable.

$$\begin{cases} S_i^1 = \frac{\alpha_{ij} a_{ij}^3 e^{a_{ij} b_{ij}} (e^{a_{ij} b_{ij}} + e^{-a_{ij}(r_{ij} - b_{ij})})}{(e^{a_{ij} b_{ij}} - e^{-a_{ij}(r_{ij} - b_{ij})})^3} \\ S_i^2 = \frac{a_{ij}^3 \alpha_{ij} e^{-a_{ij}(r_{ij} - b_{ij})} (1 - e^{-a_{ij}(r_{ij} - b_{ij})})}{\left(1 + e^{-a_{ij}(r_{ij} - b_{ij})}\right)^3} \\ S_i^3 = \frac{\alpha_{ij} k_{ij}^2 (\log(1 + k_{ij} r_{ij}) - 1)}{(1 + k_{ij} r_{ij})^2 \log^2(1 + k_{ij} r_{ij})} \end{cases} \tag{3.6}$$

From equation (3.7), we observe that the slope curvature function S_i has the inflection point $r_i = r_{ij}^s \approx b_{ij} = r_{ij}^{\inf}$ and changes from a convex function in the vicinity of the origin to a concave function before the inflection point at $r_{ij} = r_{ij}^s$ to a convex function after the inflection point.

$$\begin{cases} \lim_{r_i \to 0} S_i^1 = \infty, \\ \lim_{r_i \to b_{ij}} S_i^1 = 0 \text{ for } b_{ij} \gg \frac{1}{a_{ij}} \ \forall \ j \\ S_i^2(b_{ij}) = 0 \\ S_i^2(r_{ij} > b_{ij}) > 0 \\ S_i^2(r_{ij} < b_{ij}) < 0 \\ S_i^3(r_{ij} > 0) > 0 \end{cases} \tag{3.7}$$

□

Moreover, the following lemma proves that the logarithms of aggregate and application utility functions are invertible and the inverse functions are strictly decreasing. This properties will later be used in this chapter.

Lemma 3.1.3. *The aggregate and application utility slope curvature functions* $S_i(r_i) = \frac{\partial \log V_i(r_i)}{r_i}$ *and* $S_{ij}(r_{ij}) = \frac{\partial \log U_{ij}(r_{ij})}{r_{ij}}$ *are invertible and their inverse functions* $r_i = S_i^{-1}(.)$ *and* $r_{ij} = S_{ij}^{-1}(.)$ *are strictly decreasing.*

Proof. The concavity of the logarithmic application utility function $U_i j$ yields in $U'_{ij}(r_{ij}) =$

$\frac{\partial U_{ij}(r_{ij})}{\partial r_{ij}} > 0$ and $U''_{ij}(r_{ij}) = \frac{\partial^2 U_{ij}(r_{ij})}{\partial r_{ij}^2} < 0$, and lemma 3.1.1 stems out $S_{ij}(r_{ij}) = \frac{\partial \log(U_{ij}(r_{ij}))}{\partial r_{ij}} = \frac{U'_{ij}(r_{ij})}{U_{ij}(r_{ij})} > 0$ and $\frac{\partial S_{ij}(r_{ij})}{\partial r_{ij}} = \frac{U''_{ij}(r_{ij})U_{ij}(r_{ij}) - U'^2_{ij}(r_{ij})}{U^2_{ij}(r_{ij})} < 0$. Also, for the application utility function, we have $U_{ij}(r_{ij}) > 0$, $U_{ij}(r_{ij})$ is increasing, and it is twice differentiable with respect to r_{ij} (section 3.1). Therefore, $S_{ij}(r_{ij})$ of the logarithmic application utility function is strictly decreasing. From equation (3.2), for the sigmoidal application utility function $U_{ij}(r_{ij})$ where $0 < r_{ij} < R$, we can write the inequalities below, giving that $S_{ij}(r_{ij})$ of the sigmoidal application utility function is strictly decreasing.

$$S_{ij}(r_{ij}) > 0, \frac{\partial}{\partial r_{ij}} S_{ij}(r_{ij}) < 0 \tag{3.8}$$

Combining equation (3.3) and inequalities 3.8 yields in inequalities (3.9). Henceforth, $S_{ij}(r_{ij})$ and $S_i(r_i)$ of all application utility functions in section 3.1 are strictly decreasing functions; thereby, the slope curvature functions $S_{ij}(r_{ij})$ and $S_i(r_i)$ are invertible and the inverse functions are strictly decreasing.

$$\begin{aligned} S_i(r_i) &= \sum_{j=1}^{N_i^S} \alpha_{ij} S_{ij}(r_{ij}) + \sum_{j=N_i^S+1}^{N_i} \alpha_{ij} S_{ij}(r_{ij}) > 0 \\ \frac{\partial S_i(r_i)}{\partial r_i} &= \sum_{j=1}^{N_i^S} \alpha_{ij} \frac{\partial S_{ij}(r_{ij})}{\partial r_{ij}} + \sum_{j=N_i^S+1}^{N_i} \alpha_{ij} \frac{\partial S_{ij}(r_{ij})}{\partial r_{ij}} < 0 \end{aligned} \tag{3.9}$$

□

Next, we develop a centralized architecture in section 3.2 to disseminate resources to the applications of UEs with the aggregated utility as in equation (3.1).

3.2 Centralized Architecture

We consider rate allocation optimization problem that assigns the application resources directly by the eNB in a singular stage. The basic formulation is illustrated in equation (3.10).

$$\begin{aligned} \max_{\mathbf{r}} \quad & \prod_{i=1}^{M} \left(\prod_{j=1}^{N_i} U_{ij}^{\alpha_{ij}}(r_{ij}) \right)^{\beta_i} \\ \text{subject to} \quad & \sum_{i=1}^{M} \sum_{j=1}^{N_i} r_{ij} \leq R, \\ & r_{ij} \geq 0, \quad i = 1, 2, ..., M, \quad j = 1, 2, ..., N_i \end{aligned} \tag{3.10}$$

Here, for M UEs covered by an eNB, $\mathbf{r} = [r_1, r_2, ..., r_M]$ is the UE allocated rate vector, R is the maximum available resources at the eNB, and β_i is a subscription-dependent weight for the ith UE. Section 3.3 proves the convexity and tractable optimal solvability of the aforementioned centralized architecture.

3.3 Existence of Global Optimal Solution

The centralized architecture's objective function $\prod_{i=1}^{M} \left(\prod_{j=1}^{N_i} U_{ij}^{\alpha_{ij}}(r_{ij}) \right)^{\beta_i}$ corresponds to $\sum_{i=1}^{M} \beta_i \sum_{j=1}^{N_i} \alpha_{ij} \log U_{ij}(r_{ij})$, reformulating equation (3.10) as equation 3.11, referred to as the log-centralized problem, for which corollary 3.3.1 is conceivable.

$$\begin{aligned} \max_{\mathbf{r}} \quad & \sum_{i=1}^{M} \beta_i \sum_{j=1}^{N_i} \alpha_{ij} \log U_{ij}(r_{ij}) \\ \text{subject to} \quad & \sum_{i=1}^{M} \sum_{j=1}^{N_i} r_{ij} \leq R, \\ & r_{ij} \geq 0, \quad i = 1, 2, ..., M, \; j = 1, 2, ..., N_i \end{aligned} \tag{3.11}$$

Corollary 3.3.1 below, we substantiate the existence a global optimal solution for the centralized architecture in equation (3.10).

Corollary 3.3.1. *The centralized architecture in equation (3.10) is convex and has a unique tractable global optimal solution.*

Proof. Substantiating lemma 3.1.1 was concomitant with proving the application utility natural logarithm concavity, which entails the the log-centralized optimization (equation (3.11)) convexity [88], ensuing the convexity of its tantamount centralized optimization (equation (3.10)) and existence of a tractable global optimal solution [88]. □

3.4 Centralized Solution Algorithm

Similarly to [84, 89], we deploy the duality for convex optimization problems to solve them efficiently. What proceeds is such an application of the duality to the centralized resource allocation optimization. The solution process for the centralized architecture in equation (3.10) consists of UE and eNB parts shown in respectively Algorithms 2 and 3, whose executions (Figure 3.2) start by UEs transmitting their application utility parameters to the eNB, which in turn solves the entire optimization by allotting the bandwidth to the applications in an optimum fashion. The rates, solutions to equation (3.3), are the values r_{ij} which solve the equation $\frac{\partial \log U_{ij}(r_{ij})}{\partial r_{ij}} = p(n)$ and are the intersection of the time varying shadow price, horizontal line $y = p(n)$, with the curve $y = \frac{\partial \log U_{ij}(r_{ij})}{\partial r_{ij}}$ geometrically.

Algorithm 2 UE Centralized Algorithm

loop
Send application utility parameters $\{a_{ij}, b_{ij}, \alpha_{ij}, k_{ij}, r_{ij}^{\max}\}$ to eNB.
Receive rates $r_i^{\text{opt}} = \{r_{i1}^{\text{opt}}, r_{i2}^{\text{opt}}, ..., r_{iN_i}^{\text{opt}}\}$ from eNB.
Allocate rate r_{ij}^{opt} internally to j^{th} applications.
end loop

Algorithm 3 eNB Centralized Algorithm

loop
Receive application utility parameters $\{a_{ij}, b_{ij}, \alpha_{ij}, k_{ij}, r_{ij}^{\max}\}$ from UEs.
Solve $\mathbf{r} = \arg\max_{\mathbf{r}} \sum_{i=1}^{M} \beta_i \sum_{j=1}^{N_i} \alpha_{ij} \log U_{ij}(r_{ij}) - p(\sum_{i=1}^{M} \sum_{j=1}^{N_i} r_{ij} - R)$. {where $\mathbf{r} = \{r_1, r_2, ..., r_M\}$ and $r_i = \{r_{i1}, r_{i2}, ..., r_{iN_i}\}$}
Send $r_i = \{r_{i1}, r_{i2}, ..., r_{iN_i}\}$ to i^{th} UE.
end loop

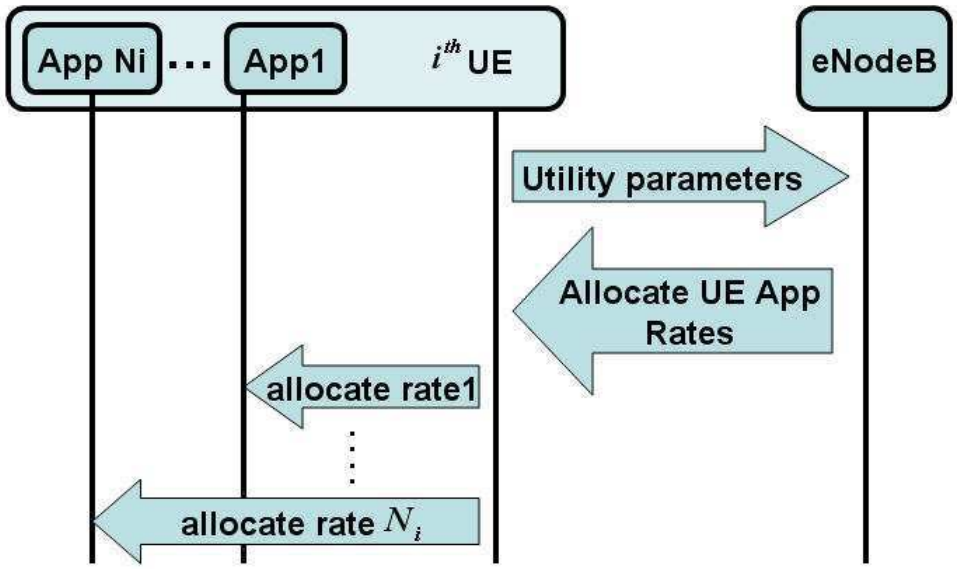


Figure 3.2: Centralized Algorithm: Resources are allocated to the applications running on the UEs in a monolithic stage, in which UEs transmit their application utility parameters to their eNB, which calculates the optimal application rates and transmit them to the germane UEs.

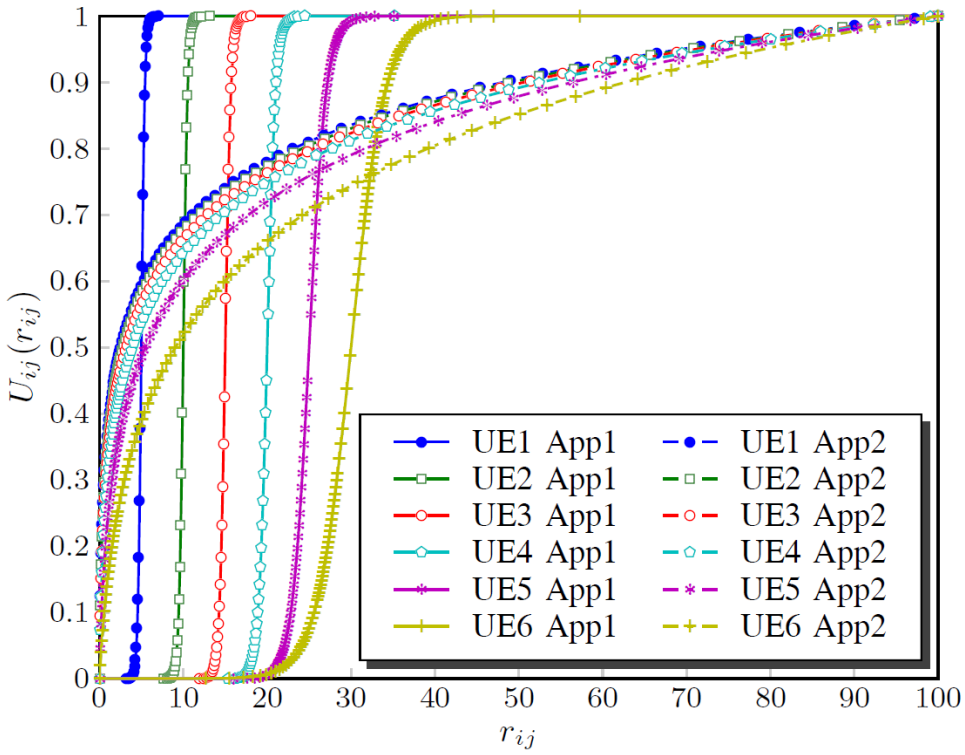
Next, section 3.5 provides with simulations for the centralized architecture developed in this chapter.

3.5 Centralized Architecture Simulation

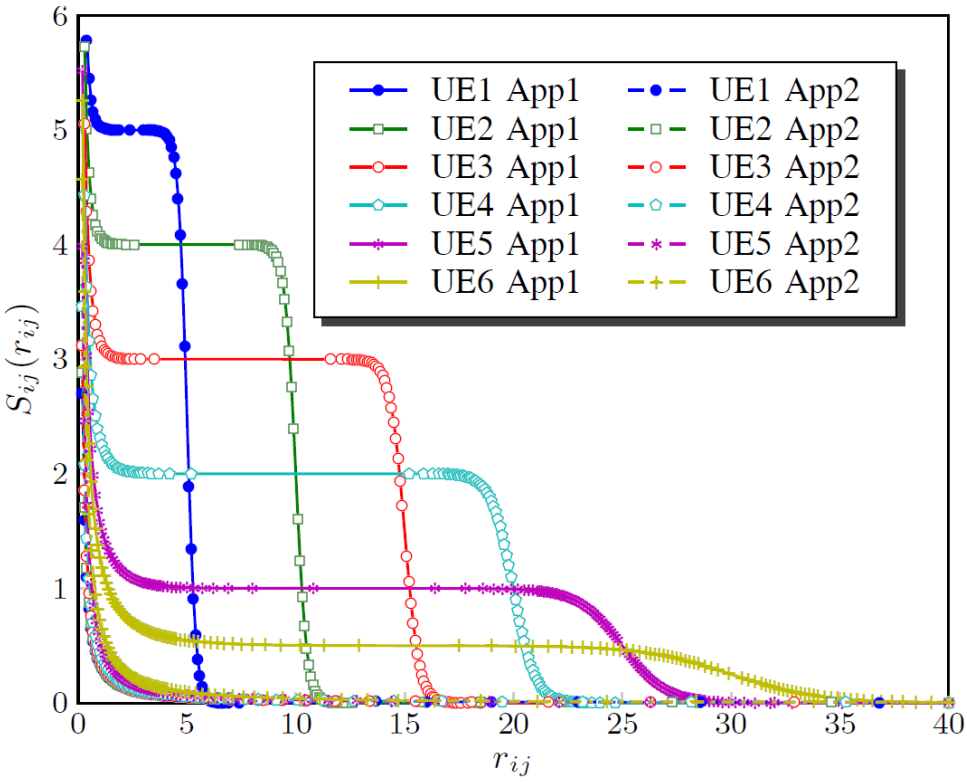
A cell with $M = 6$ UEs and an eNB, depicted in Figure 3.1, is considered and each UE concurrently runs a delay-tolerant and a real-time application with respectively logarithmic and sigmoidal utility functions with parameters in Table 3.1. The sigmoidal utility with parameters $a = 5$, $b = 10$ approximates a step function at rate $r = 5$ and is a good model for VoIP, while parameters $a = 3$, $b = 15$ is an approximation of a real-time application with an inflection point at rate $r = 15$ and is conducive to modeling standard definition video streaming, whereas parameters $a = 1$, $b = 25$ is an estimation of another real-time application with the inflection point $r = 25$ and is appropriate for the high definition video streaming. Moreover, the logarithmic utilities with $r^{\max} = 100$ and distinct k_i parameters estimate delay-tolerant FTP applications.

The plots of the utility functions in Table 3.1 are shown in Figure 3.3(a), from which we can observe that the real-time applications require a minimum rate, i.e. the inflection point, after which the application QoS is fulfilled to a large extent. On the other hand, the logarithmic utility is provided with some QoS even at low rates suitable for the delay-tolerant nature of the applications. Furthermore, as we can observe from Figure 3.3(a), in compliance with the properties mentioned in section 3.1 the utility functions are strictly increasing continuous functions, zero valued at zero rates. Furthermore, the first derivative of the utility functions natural logarithm, $S_ij(r_ij)$, are shown in Figure 3.3(b), which reflects the positivity and decreasing nature of the first derivative. This is in line with lemma 3.1.3.

Then, the centralized rate assignment procedure (Algorithms 2 and 3) were applied to the aforesaid logarithmic and sigmoidal utility functions using MATLAB. To account for the applications usage percentage, we set the application status weight vector in equation (3.1) as $\boldsymbol{\alpha} = \{\alpha_{11}, \alpha_{21}, \alpha_{31}, \alpha_{41}, \alpha_{51}, \alpha_{61}, \alpha_{12}, \alpha_{22}, \alpha_{32}, \alpha_{42}, \alpha_{52}, \alpha_{62}\}$ where α_{ij} represents the status weight of the jth application of the ith UE. It is noteworthy that the addition of application



(a) Application Utility Functions - Centralized Architecture



(b) Utility Slope Curvature Functions - Centralized Architecture

Figure 3.3: The system contains 6 UEs, each concurrently running a delay-tolerant and real-time application with respective identically colored logarithmic and sigmoidal utility functions U_{ij} vs. the application-assigned rates r_{ij} plots in Figure 3.3(a). Utility slope curvature functions, the first derivative of the application utility natural logarithms S_{ij} with respect to the application rates r_{ij} are illustrated in Figure 3.3(b) where identical colors relate to the applications on one UE.

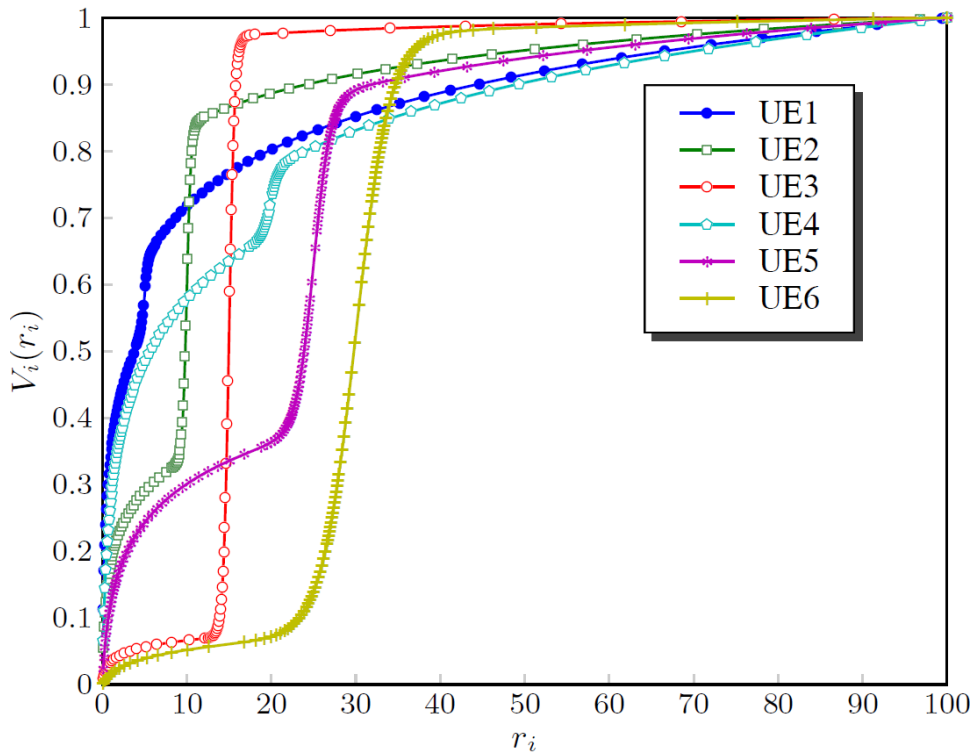
Table 3.1: Applications Utility Parameters - Centralized Architecture

Applications Utilities Parameters	
UE1 App1	Sigmoid $a = 5,\ b = 5$
UE2 App1	Sigmoid $a = 4,\ b = 10$
UE3 App1	Sigmoid $a = 3,\ b = 15$
UE4 App1	Sigmoid $a = 2,\ b = 20$
UE5 App1	Sigmoid $a = 1,\ b = 25$
UE6 App1	Sigmoid $a = 0.5,\ b = 30$
UE1 App2	Logarithmic $k = 15,\ r^{\max} = 100$
UE2 App2	Logarithmic $k = 12,\ r^{\max} = 100$
UE3 App2	Logarithmic $k = 9,\ r^{\max} = 100$
UE4 App2	Logarithmic $k = 6,\ r^{\max} = 100$
UE5 App2	Logarithmic $k = 3,\ r^{\max} = 100$
UE6 App2	Logarithmic $k = 1,\ r^{\max} = 100$

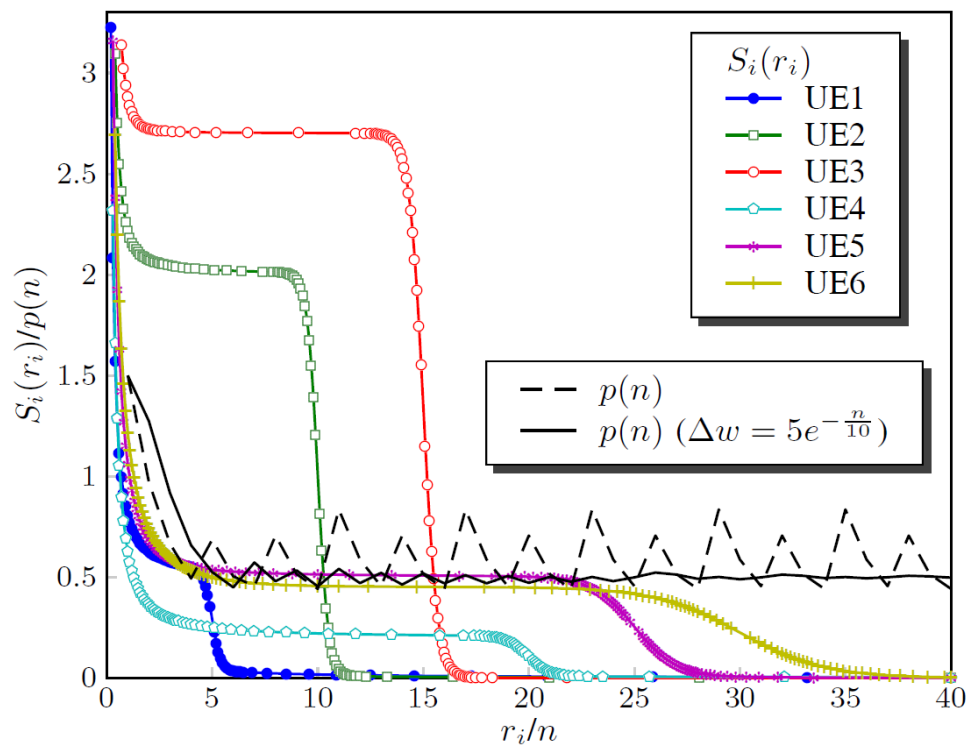
usage percentages per UE is unity, i.e. $\alpha_{i1} + \alpha_{i2} = 1$.

In addition, the aggregate utility functions $V_i(r_i)$ for $i \in \{1, ..., 6\}$ are depicted in Figure 3.4(a) and the first derivative of their natural logarithm, $S_i(r_i)$ for $i \in \{1, ..., 6\}$, are illustrated in Figure 3.4(b). As we can see, in compliance with lemma 3.1.2, the slope curvature functions inflection points occur at the application utility functions' inflection points. Furthermore, in line with lemma 3.1.3, the slope curvature functions are strictly decreasing.

Next section investigates bids and rate allocations for the UEs and applications in our system under varying eNB resource availabilities and the centralized architecture.



(a) Aggregate Utility Functions - Centralized Architecture



(b) Aggregate Slope Curvature Functions - Centralized Architecture

Figure 3.4: Figure 3.4(a) plots the aggregated utilities, multiplications of the usage-percentage-powered application utility functions $V_i(r_i)$ vs. the UE rates r_i, where $i \in \{1, ..., 6\}$. Figure 3.4(b) illustrates the aggregate slope curvature functions , first derivative of the the aggregate utility natural logarithms $S_i(r_i)$. Furthermore, decay function-induced robustness effect is depicted; As we can see, the lack of decay functions yields in the system instability revealed in the shadow price oscillation.

3.5.1 Rate Allocation and Bids for $10 \leq R \leq 200$

In the following simulations, we set the termination threshold $\delta = 10^{-4}$ and the eNB rate R to sweep from 10 to 200 with a step size of 5 bandwidth units. Besides, the application status weights is considered to be $\boldsymbol{\alpha} = \{0.1, 0.5, 0.9, 0.1, 0.5, 0.9, 0.9, 0.5, 0.1, 0.9, 0.5, 0.1\}$. It is worth mentioning that the addition of usage percentages per UE is unity, e.g. adding the $1st$ and the $6th$ components of the set (which are indeed the usage percentages for both of applications running on UE1), we get $0.1 + 0.9 = 1$, and so forth. For the centralized resource allocation (Algorithms 2 and 3), application assigned rates are depicted in Figure 3.5 with the changes in the eNB available resources R.

As we can observe, initially all the UEs are allocated some rates which is owing to the fact that they all subsume real-time applications in need of immediate rate allocations before any QoS is met. For instance, UE2 has a real-time streaming video application, which requires a bandwidth assignment right away. First of all, we see that the more resources become available at the eNB, the higher rates are assigned to the UEs. On the other hand, the dearth of the resources (small R) causes those UEs which have applications with higher bit rate requirements to bid higher in order to gain resources. For instance, since UE2 includes a real-time streaming video application, its urgent need for bandwidth allocation causes its initial higher bid for the resources, which is responded by its fast allocation portrayed in the Figure 3.5.

Then, the centralized algorithm allocates rates to their applications as illustrated in Figures 3.6. In Figure 3.6, we show the allocated applications rates $\{r_{ij} | i \in \{1, ..., 6\} \wedge j \in \{1, 2\}\}$ under changing eNB rate R. As we can observe, initially more resources are allocated to the real-time applications since these have more stringent QoS requirements.

However, when the total eNB rate exceeds the inflection point rates sum $\sum b_{ij}$ of all real-time applications incumbent in the system, eNB can allot more resources to the delay-

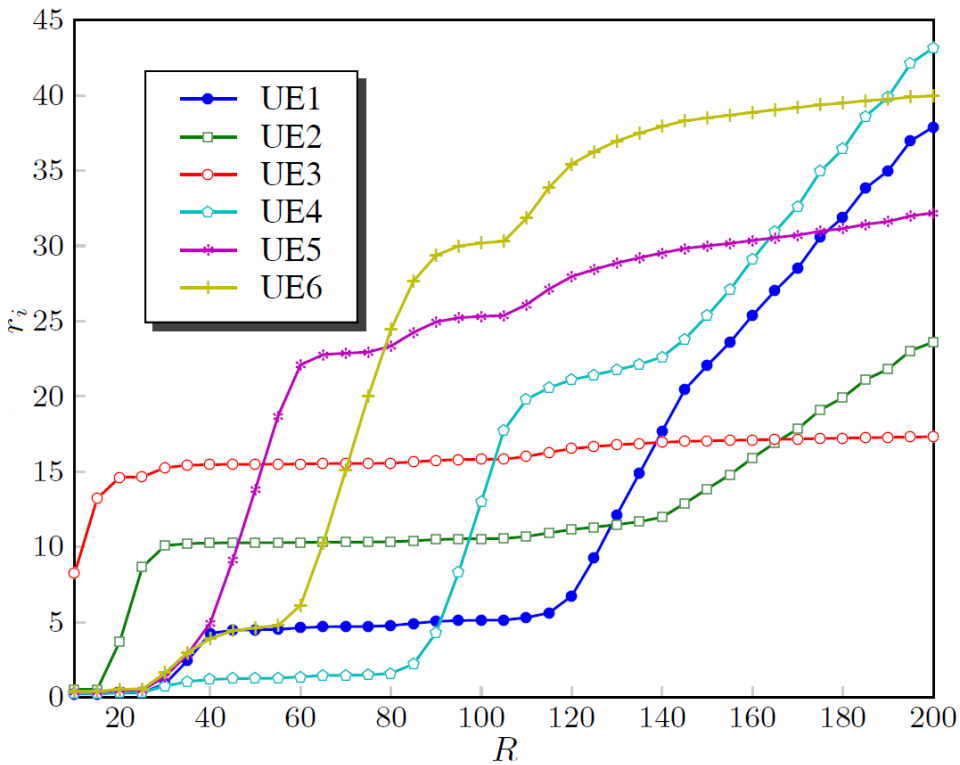


Figure 3.5: Figure 3.5 depicts the optimal rates allocated to the UEs by the distributed scheme vs. BS resources. No user is dropped as no assignment is zero.

tolerant applications with ease of mind. This behavior is observed with the rate increase and bid value plummet that take place after the eNB rate surpasses the inflection points sum, i.e. $R = \sum b_{ij} = 105$, in Figure 3.6.

3.6 Centralized Architecture Real-World Implementation

We developed a single-carrier radio resource allocation formulation cast under a utility proportional fairness framework in this chapter. We proved the optimality of the resource allocation method formulation and incorporated the traffic nature, subscriber type, and application usage percentages in the formulation. Now, we implement the centralized resource

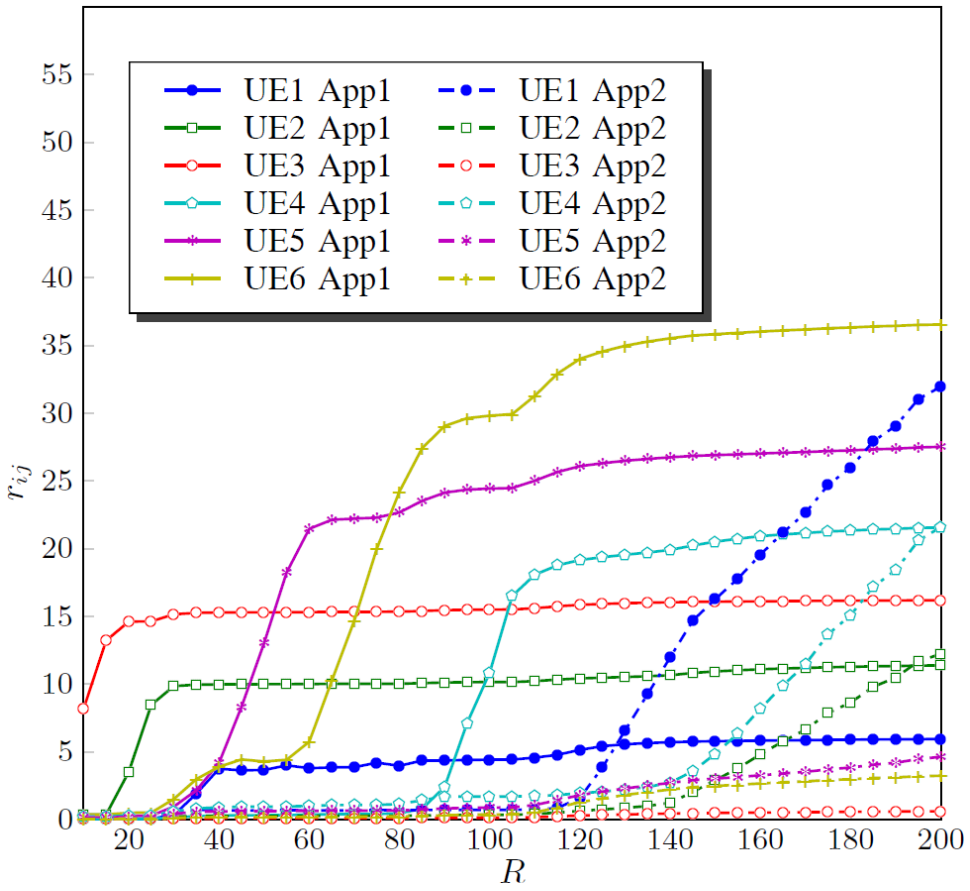


Figure 3.6: Figure 3.6 depicts the optimal application rates r_{ij} vs. the eNB rate R. Applications running on an UE are identically colored. As we can see, real-time applications are initially allocated more resources as opposed to the delay-tolerant ones due to their urgent need for resources.

allocation architecture that we developed in this chapter in a real-world network and show that applying the resource allocation scheme improves the users' QoE in the network. Since we will prove the mathematical equivalence of the centralized and distributed architectures in chapter 4, leveraging the centralized architecture for resource allocation does not affect the allocated rates within the real-world network implementation. Here,

- We implement the centralized resource allocation architecture that we developed in this chapter on a real-world network.
- We show that applying the aforementioned centralized architecture on an enforcement router improves the QoE by eliminating real-time traffic buffering.

- We show that applying the aforementioned centralized architecture reduces the total resource consumption in the network and thereby it decreases the operation costs.

3.6.1 Implementation Results

In this section, we implement the centralized resource allocation architecture in chapter 3 on a real-world network scenario depicted in Figure 3.7. The system subsumes UEs which are connected through a WiFi access point (AP) [159] to the Internet. The resource allocation is implemented on a resource broker (RB) logical entity installed on a router which shapes the traffic generated by the UEs and received by the AP based on the application rates assigned by the resource allocation scheme implemented on the RB unit.

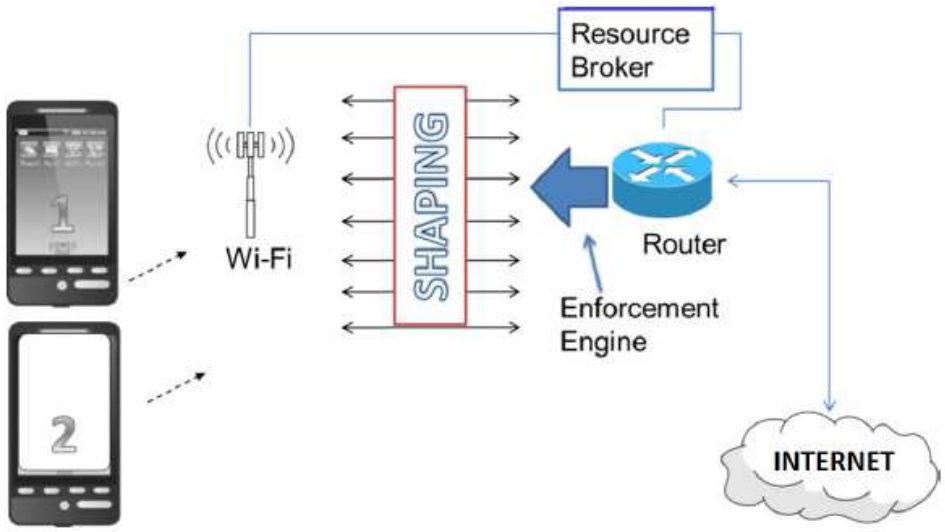


Figure 3.7: Implementation system model contains UEs, WiFi-connected to the Internet, run delay-tolerant and real-time applications whose optimal rates are assigned by the resource block (RB) entity on which the centralized resource allocation is instrumented. The RB is installed on the router which shapes the traffic based on the rates allocated by the resource allocation scheme.

To implement the scenario in Figure 3.7 on a real-world network, we leverage a personal computer (PC) to configure the network in a distributed manner to decrease the processing

load through a virtual machine (VM) architecture [160] in Figure 3.8. Here, we have used a single-socket IBM x3250 M4 server [161] with two physical and two Peripheral Component Interconnect (PCI) - enabled ports [162] to create 2 three-interfaced VMs. One VM hosts the RB entity and the other forms a router including an enforcement engine to manage rate assignments via an onboard router traffic control, and the other VM is a dedicated file server. The 2 VMs are annotated as "Guest 1" and "Guest 2" in Figure 3.8, where the router and RB sit on "Guest 1" and "Guest 2", respectively. Besides, we create 3 virtual switches intended for the phone network, for office-Internet-connected external devices, and for network maintenance/ operation issues.

The smart phones and their AP are placed on their own private Ethernet network [163] with an IP address 192.168.2.1 [41]. The router gateway settings enable connections to the office network (Intranet) and Internet. The smart phones run YouTube and hyper text transfer protocol (HTTP) download applications. The traffic generated by the real-time/delay-tolerant YouTube/HTTP applications is inelastic/elastic, and we apply the centralized resource allocation architecture in this chapter to obtain application rates (throughput). The users' QoE is reflected by YouTube traffic buffering occurrences and HTTP traffic incomplete downloads. The dearth of the aforesaid events indicates an acceptable QoE for the application users and implies the effectiveness of the resource allocation that we have introduced in this chapter.

It is noteworthy that for the test platform in Figure 3.8, the small number of phones working in a high throughput WiFi network is an object of concern as it provides too ideal of a data transfer environment to be able to appropriately illustrate the benefits that may emerge from the traffic shaping that our resource allocation scheme renders. To observe the traffic shaping effects, we should impose a higher load on the network. We do this simply by restricting the overall network bandwidth to 1 Mega bits per second (Mbps). In order to make a comparison between the scenarios with and without the resource allocation

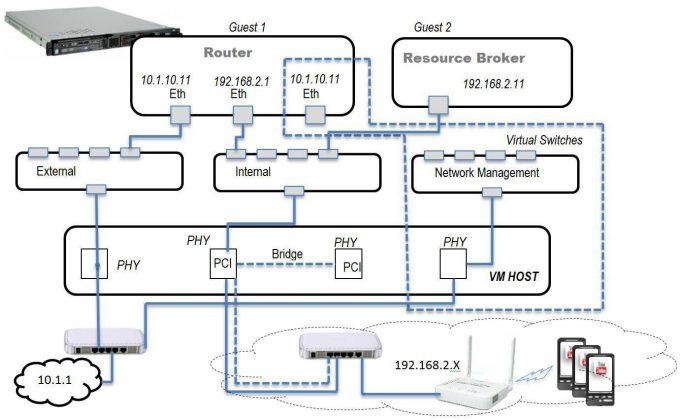


Figure 3.8: Real-world network architecture contains two VMs "Guest 1" and "Guest 2" which respectively host the router and resource allocation method implemented on the the resource block (RB) entity. Furthermore, an office-Internet virtual switch for external devices, three phone-network virtual switches, and a network management virtual switch are created on a two physical two PCI-ported IBM server. The Phones and WiFi access point (AP) have their private network with IP address 192.168.2.1.

scheme, we first apply the resource allocation method to the network in Figure 3.8 under no bandwidth constraints; then, we introduce an overall network constraint with no traffic shaping, and ultimately contrast non-shaped throughput and QoE observation to a situation where the RB entity operates under a 1 Mbps network constraint.

The YouTube and HTTP applications throughput under neither network constraints nor the resource allocation method application is shown in Table 3.2, where a low/high rate YouTube application "YouTube 1"/"YouTube 2" and a small/large file download application "HTTP 1"/"HTTP 2" (obtained from content providers of the VM-created file server) run on three UEs in the network. The speed of the network under the absence of applications is measured at 32 Mbps. For instance, the 1st scenario in the next table indicates that two phones run YouTube 1 and the average bandwidth consumption is $R_{avg} = 3.492$ Mbps, while the single-UE 3rd scenario incurs a much lower bandwidth consumption $R_{avg} = 0.951$ Mbps due to the low rate YouTube 1 application. On the other hand, the bandwidth consumption significantly increased to $R_{avg} = 2.11$ Mbps due to the single-UE high rate YouTube 2 application in the 4th scenario, and a similar rate rise is apparent in the transition from the 5th

Table 3.2: Network throughput for the network in Figure 3.8 under neither bandwidth constraints nor traffic shaping.

Scenario	Phone 1	Phone 2	Phone 3	R_{avg} (Mbps)
1	YouTube 1	YouTube 2	-	3.492
2	YouTube 1	HTTP 1	-	4.050
3	YouTube 1	-	-	0.951
4	YouTube 2	-	-	2.11
5	HTTP 1	HTTP 1	-	4.262
6	HTTP 1	HTTP 2	-	24.866
7	HTTP 1	HTTP 1	YouTube 1	13.747
Network throughput with neither application nor rate Constraints: 32 Mbps				

scenario to the 6th one where the high rate HTTP 2 replaces the low rate HTTP 1 at the 2nd phone amounting to 24.866 Mbps bandwidth consumption. The last scenario is concomitant with a lower rate $R_{avg} = 13.747$ Mbps as opposed to the 6th scenario ($R_{avg} = 24.866$ Mbps) in spite of augmenting a YouTube 1 application to phone 3 in the latter case. This can be explained by the need to transfer more bits which happen over a longer time interval vis-a-vis that of the 6th case; this slashes down the throughput relative to the 6th configuration.

To see the effect of the resource allocation scheme, implemented in the RB entity, we throttle the overall network bandwidth to $R = 1$ Mbps. Without loss of generality, we focus on deploying only two phones in the network in Figure 3.8 with IP addresses and a YouTube 1 and HTTP 1 application as illustrated in Table 3.3. Running the experiment with no resource allocation applied, we observe that YouTube 1 incurred multiple buffering instances indicating a poor QoE from its user's perspective. Furthermore, the overall average bandwidth usage was 0.963 Mbps and HTTP 1 download completed in 1200 seconds (s) (next table). Using WireShark [164, 165], we obtain the traces in Figure 3.9, where both of the applications sharing the total 1 Mbps bandwidth annotated on the black curve alternate in

Table 3.3: Smart phone Applications under $R = 1$ Mbps Network Constraint with/without Traffic Shaping .

Phone	IP	Traffic Type	Application
UE 1	192.168.2.57	Inelastic	YouTube 1
UE 2	192.168.2.98	Elastic	HTTP 1

bursty transmission intervals. In particular, at certain times, the HTTP 1 application (red curve) utilizes the entire available bandwidth, shown by the red curve reaching the black curve, which simultaneously zeros the YouTube 1 throughput illustrated by the green curve hitting the abscissa (time axis). This behavior adversely affects the QoE for the UE 1.

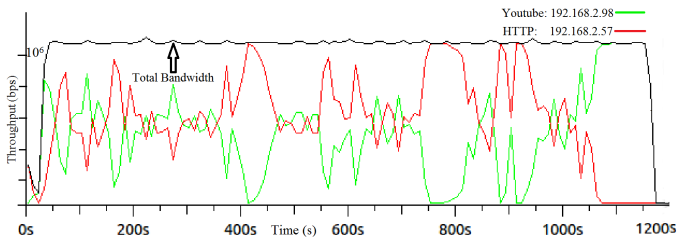


Figure 3.9: Wireshark Throughput Analysis without applying the resource allocation to the network in Figure 3.8 with the parameters in Table 3.3 under an overall bandwidth constraint $R = 1$ Mbps and no traffic shaping. Black/Green/Red curve shows the network/YouTube 1/HTTP 1 throughput. HTTP 1 download completed in 1200 s with YouTube 1 incurring multiple buffering, adversely affecting the QoE as HTTP 1 (red curve) uses the entire available capacity (hits the black curve).

The same situation of the network in Figure 3.8 with the parameters in the table is repeated when the resource allocation method is applied to have the router shape the applications traffic by optimally assigning them rates based on their requirements. To apply the resource allocation method, the phones UE 1 and UE 2 register with the RB entity, where the rate allocation code calculates the rates to be enforced at the router. The average bit rates for the YouTube 1 and HTTP 1 are respectively 731 and 267 kbps, the convergence time for the algorithm is measured at 528 milliseconds (ms), and the overall throughput becomes 0.758 Mbps which is less than the maximum 1 Mbps available capacity due to periods over which

no YouTube traffic is present on the network. Using Wireshark, the rates of the YouTube 1 and HTTP 1 applications when the centralized resource allocation is leveraged in the network under an overall $R = 1$ Mbps constraint are depicted in Figure 3.10.

Here, the black curve indicates the overall available bandwidth and the ordinate (throughput axis) shows that the network bandwidth is confined to 1 Mbps. As we can see, YouTube 1 (green curve) consumes more resources per the rate allocation interval than does the HTTP 1 application, whose download time expectedly takes longer to be completed at 2650 s (Table 3.4). Interestingly, there are intervals where YouTube 1 rate becomes 0, over which HTTP 1 obtains more bandwidth; this zero-grounding behavior is analogously observed for the HTTP 1 download. In this experiment, we observe no YouTube buffering occurrences, thereby it provisions a better video watching experience for the user as opposed to the unshaped traffic scenario depicted in Figure 3.9. Such a dearth of buffering speaks directly to the speculated QoE in that the real-time YouTube 1 application is provided with a consistent rate assignment such that it is able to fill the buffer and does not require any more bandwidth usage.

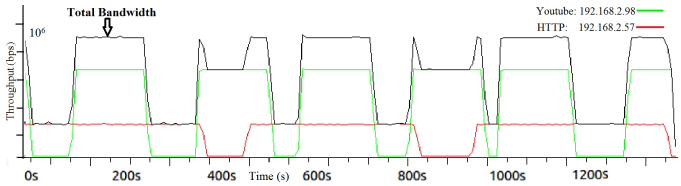


Figure 3.10: Wireshark throughput analysis with applying the resource allocation to the network in Figure 3.8 with the parameters in Table 3.3 under the bandwidth constraint $R = 1$ Mbps and shaping. Black/Green/Red curve shows the network/YouTube 1/HTTP 1 throughput. HTTP 1 download completed in 2650 s with no YouTube 1 buffering occurrences. Occasionally, one application rate zeros and the other application utilizes the maximum bandwidth to achieve a consistent allocation to elevate users' QoE.

As we observed in this real-world implementation, applying the resource allocation architecture elevates the QoE of the users despite the fact that less resources are consumed. This directly helps communications carriers to reduce their operation expenditure (OPEX) and

Table 3.4: Bandwidth Consumption and Download Time.

Performance	Shaped Traffic	Unshaped Traffic
HTTP 1 Download Time (s)	1200	2650
YouTube 1 Buffering	No	Yes
Total Bandwidth (kbps)	758	951

customer churn. As a case in point, the algorithm-induced traffic shaping decreased the bandwidth consumption from 0.963 Mbps to 0.758 Mbps, thereby a 0.205 Mbps reduction of resource consumption is stemmed which has desirable monetary ramifications by utilizing less resources without degrading users' QoE. As such, an QoE remedial measurement is implied in light of employing the developed resource allocation scheme.

Next, section 3.7 summarizes the results obtain in chapter 3.

3.7 Chapter Summary

In this chapter, we introduced a novel QoS-minded utility proportional fairness framework for resource allocation for the cells of a cellular communications system. Users ran both delay-tolerant and real-time applications mathematically modeled correspondingly as logarithmic and sigmoidal application utility functions, where the function values represented the applications QoS percentage. The proposed resource allocation formulation incorporated the service differentiation, application status differentiation modeling the applications usage percentage, and subscriber differentiation within networks into the optimization. We developed a centralized architecture for the resource allocation optimization which assigned the rates to the running applications directly by the allocation entity such as an eNB in a response to the UE utility parameters sent.

Not only did we prove that the proposed centralized resource allocation architecture is a

convex optimization problem and solved it through the Lagrangian of its dual problem, but also we proved the optimality of the rate assignments. Ultimately, we performed simulations in MATLAB to show the application of the proposed centralized architecture to a cellular communications system.

We implemented the centralized architecture to assign bandwidth to real-world real-time and delay-tolerant applications running on mobile devices in a physical network that we created. We did so to evaluate the QoE users perceive when the procedure is implemented in practice. The large-scale network configuration included YouTube and HTTP applications connecting through their WAP and a router, running the resource allocation routine and enforcing the rates, to the Internet. We realized that the absence of the resource allocation algorithm in the network caused multiple buffering instances of the real-time YouTube application, thereby the users QoE was adversely undermined. On the other hand, the presence of the algorithm, through which the rates were assigned to the applications and enforced at the gateway router, eliminated any YouTube buffering at the expense of lengthening the duration of delay-tolerant download applications. Therefore, applying the resource allocation method significantly escalated the users QoE without hurting QoS requirements of applications. Finally, we realized that despite the dearth of YouTube buffering under the presence of the resource allocation algorithm, the applications consumed less resources as opposed to the QoE-hurting algorithm-absent scenario. Consequently, the bandwidth conservation yields in a lower OPEX for the network as less to-be-paid resources are consumed.

Chapter 4

A Distributed Architecture for Resource Allocation

In chapter 3, we introduced a novel convex utility proportional fairness maximization for optimal resource allocation in wireless networks and outfitted the optimization with the subscriber, application status, and service differentiations parameterized respectively as UE subscription weights, application status weights, and application utility functions. Furthermore, we developed a centralized architecture for the proposed resource allocation which assigned application rates by the eNBs in a single stage in response to the application utility parameters sent by the UEs to the eNBs. In this chapter, we provide with a distributed architecture for the same radio resource allocation framework which was introduced in chapter 3. The resource allocation optimization accounts for application types and temporal usages as well as UE priorities. However, the distributed architecture assigns application rates in two stages from the eNBs to the UEs and by the UEs to the running applications. Moreover, the proposed architecture provisions a pricing mechanism for the MNOs to flatten the traffic in peak-hour times.

4.0.1 Contributions

In this chapter:

- We present a utility proportional fairness formulation for resource allocation in cellular communications system.
- We ornament the proposed resource allocation with mechanisms to differentiate traffic based on QoS requirements, applications temporal usages, and UE priorities.
- We develop a distributed architecture for the proposed resource allocation framework.
- We show that the devised distributed architecture is a convex optimization.
- We provide with solution algorithms for the centralized architecture which assigns application rates by the base station (BS) in a single stage.
- we prove that the proposed distributed architecture is not always stable.
- We develop a variation of the distributed architecture which incorporates robustness into the resource allocation procedure.
- We provide with simulations to show the application of the proposed resource allocation in a cellular network simulation scenario.

The remainder of this paper is organized as follows. Section 4.1 presents a distributed architecture for the resource allocation framework that was introduced in chapter 3. Section 4.2 proves that the proposed distributed resource allocation architecture is convex. Furthermore, it shows the potential instability of the distributed architecture and developed a robust variation of the distributed architecture. Section 4.3 provides with solution algorithms for the distributed resource allocation architecture. Section 4.4 proves the mathematical equivalence of the distributed resource allocation to the centralized architecture that was introduced in

chapter 3. Section 4.5 portrays simulations depicting the application of the distributed resource allocation architecture to a cellular system scenario. And, section 4.7 concludes and summarizes the current chapter.

4.1 Distributed Architecture

The centralized architecture for optimal rate allocation in chapter 3 was in accordance with equation (3.10), repeated here for the sake of easy reference as equation (4.1), where for M UEs covered by an eNB in accordance with Figure 4.1, $\mathbf{r} = [r_1, r_2, ..., r_M]$ is the UE allocated rate vector, R is the maximum available resources at the eNB, and β_i is a subscription-dependent weight for the ith UE.

$$\begin{aligned} \max_{\mathbf{r}} \quad & \prod_{i=1}^{M} \left(\prod_{j=1}^{N_i} U_{ij}^{\alpha_{ij}}(r_{ij}) \right)^{\beta_i} \\ \text{subject to} \quad & \sum_{i=1}^{M} \sum_{j=1}^{N_i} r_{ij} \leq R, \\ & r_{ij} \geq 0, \quad i = 1, 2, ..., M, \quad j = 1, 2, ..., N_i \end{aligned} \tag{4.1}$$

To create a distributed architecture, we subdivide the centralized architecture's optimization in equation (4.1) into two simpler optimization problems solved separately. The first optimization concerns with the UE rate allocation by the eNB via collaborations between the eNB and pertinent UEs; hereby, the optimization is referred to as external UE resource allocation (EURA) and is presented in section 4.1.0.1. On the contrary, the second optimization wells up from distributing the application rates by the host UEs, performed internally to the UEs and is named the internal UE rate allocation (IURA) and is portrayed in section 4.1.0.2.

In the system model, the aggregate utility functions for the UEs is the same as chapter 3,

Figure 4.1: System Model: Single cell, within the cellular network, with an eNB covering $M = 6$ UEs each with simultaneously running delay-tolerant and relay-time applications represented by logarithmic and sigmoidal utility functions respectively.

which is repeated below for the sake of reader's convenience.

$$V_i(r_i) = \prod_{j=1}^{N_i} U_{ij}^{\alpha_{ij}}(r_{ij}) \tag{4.2}$$

Next, section 4.1.0.1 presents the EURA formulation for resource allocation from the eNB to the UEs.

4.1.0.1 EURA Optimization Problem

EURA optimization , solved collaboratively amongst UEs and their eNB, can be written as equation (4.3), where $V_i(r_i) = \prod_{j=1}^{N_i} U_{ij}^{\alpha_{ij}}(r_{ij})$ is the ith UE aggregate utility function expressed in equation (4.1), $\mathbf{r} = [r_1, r_2, ..., r_M]$ is the UE rate vector whose i^{th} component represents the rate assigned by the eNB to the ith UE, and M is the number of UEs covered by the eNB. In section 4.2.1, we prove the convexity and tractable optimal solvability of the optimization problem in equation (4.3) and present the algorithm to solve this problem in section 4.3.

$$\begin{aligned} \max_{\mathbf{r}} \quad & \prod_{i=1}^{M} V_i^{\beta_i}(r_i) \\ \text{subject to} \quad & \sum_{i=1}^{M} r_i \leq R, \\ & r_i \geq 0, \quad i = 1, 2, ..., M. \end{aligned} \tag{4.3}$$

Next, section 4.1.0.2 presents the IURA formulation for resource allocation from the UEs to the applications.

4.1.0.2 IURA Optimization Problem

IURA optimization problem, solved internally in each UE, can be written as equation (4.4) for the ith UE with $i \in \{1, 2, ...M\}$, where $\mathbf{r}_i = [r_{i1}, r_{i2}, ..., r_{iN_i}]$ is the application rate allocation vector such that its jth component indicates the bandwidth allotted by the ith UE to its jth application, r_i^{opt} is the ith UE rate allocated by eNB via solving the EURA optimization in equation (4.3), and N_i is the number of applications of the ith UE. Superscript "opt" indicates the optimality of the UE rates which will be proved in section 4.2.1. Besides, section 4.2.2 proves that there exists a tractable global optimal solution to the IURA optimization problem in equation (4.4) and section 4.3 provides the solving algorithm thereof.

$$\begin{aligned} \max_{\mathbf{r}_i} \quad & \prod_{j=1}^{N_i} U_{ij}^{\alpha_{ij}}(r_{ij}) \\ \text{subject to} \quad & \sum_{j=1}^{N_i} r_{ij} \leq r_i^{\text{opt}}, \\ & r_{ij} \geq 0, \quad j = 1, 2, ..., N_i. \end{aligned} \tag{4.4}$$

Next, section 4.2 proves the convexity of the EURA and IURA optimization problems.

4.2 Distributed Architecture's Global Optimal Solution

This section proves the existence of optimal solutions for the distributed resource allocations developed in section 4.1 by substantiating convexity of the EURA and IURA problems in sections 4.2.1 and 4.2.1, respectively.

4.2.1 EURA Global Optimal Solution

Strictly increasing nature of logarithm function yields in an equivalent EURA objective function $\arg\max_{\mathbf{r}} \sum_{i=1}^{M} \beta_i \log(V_i(r_i))$, stemmed from equation (4.3), reformulated and referred to as respectively equation 4.5 and log-EURA problem.

$$\begin{aligned} \max_{\mathbf{r}} \quad & \sum_{i=1}^{M} \beta_i \log(V_i(r_i)) \\ \text{subject to} \quad & \sum_{i=1}^{M} r_i \leq R, \\ & r_i \geq 0, \quad i = 1, 2, ..., M. \end{aligned} \tag{4.5}$$

Before proceeding further, for the simplicity of reference, we repeat the following lemma that we proved in chapter 3 as it will be used in the mathematical analysis to come in this chapter.

Lemma 4.2.1. *The aggregate utility natural logarithm* $\log(V_i(r_i))$ *is strictly concave.*

Then, theorem 4.2.2 proves that the EURA optimization in equation (4.3) is convex and the allocated UE rates are optimal.

Theorem 4.2.2. *The EURA optimization problem in equation (4.3) is convex and has a unique tractable global optimal solution.*

Proof. The aggregate utility concavity based on lemma 4.2.1 concludes that the log-EURA optimization is convex [88], which in turn proves the convexity of the EURA problem in equation (4.3) due to their objective functions equivalence. And, there exists a unique tractable global optimal solution for a convex optimization in general [88], and for EURA in particular. □

Next section proves that the IURA optimization is convex and the application allocated rates are optimal.

4.2.2 IURA Global Optimal Solution

Strictly increasing nature of logarithm function yields in that the IURA objective function in equation (4.4), i.e. $\prod_{j=1}^{N_i} U_{ij}^{\alpha_{ij}}(r_{ij})$, corresponds to $\sum_{j=1}^{N_i} \alpha_{ij} \log(U_{ij}(r_{ij}))$. So equation (4.4) can be reformulated as equation 4.6, referred to as the log-IURA problem for which corollary 4.2.3 is conceivable.

$$\begin{aligned} \max_{\mathbf{r}_i} \quad & \sum_{j=1}^{N_i} \alpha_{ij} \log U_{ij}(r_{ij}) \\ \text{subject to} \quad & \sum_{i=1}^{N_i} r_{ij} \leq r_i^{\text{opt}}, \\ & r_{ij} \geq 0, \qquad j = 1, 2, ..., N_i. \end{aligned} \tag{4.6}$$

Corollary 4.2.3. *The IURA optimization problem in equation (4.4) is convex and has a unique tractable global optimal solution.*

Proof. Substantiating lemma 4.2.1 in chapter 3 was concomitant with proving the concavity of the application utility functions natural logarithm; this ascertains the convexity of the log-IURA problem in equation (4.6) [88]. Since log-IURA and IURA optimizations have equivalent objective functions, IURA optimization in equation (4.4) is also convex. Every

convex optimization has a tractable global optimal solution in general [88], and so does the IURA optimization in particular. □

Theorem 4.2.2 and corollary 4.2.3 indicate that the distributed optimization in section 4.1 is convex and it assigns rates optimally.

Next, section 4.3 solves the distributed optimization problem presented in section 4.1 by a variation of Frank Kelly algorithm presented in chapter 2.

4.3 Distributed Solution Algorithm

The solution to the distributed architecture for resource allocation is provided through the Lagrangian of the dual problems for the EURA and IURA optimizations in sections 4.3.1 and 4.3.2 below. Similarly to ([84, 89]), we deploy the duality for convex optimization problems to solve them efficiently. What proceeds is such an application of the duality to EURA and IURA constituents of the distributed rate allocation problem. We present the EURA algorithm in section 4.3.1 below.

4.3.1 EURA Algorithm

The log-EURA problem (4.5) can be solved by converting it to its dual problem ([84, 89]). We define the Lagrangian as equation (4.7).

$$\begin{aligned} L(\mathbf{r}, p) &= \sum_{i=1}^{M} \log(V_i(r_i)) - p(\sum_{i=1}^{M} r_i + z - R) \\ &= \sum_{i=1}^{M} \left(\log(V_i(r_i)) - pr_i \right) + p(R - z) \\ &= \sum_{i=1}^{M} L_i(r_i, p) + p(R - z) \end{aligned} \tag{4.7}$$

where $z_i \geq 0$ is the slack variable and p is Lagrange multiplier or the shadow price (price per unit bandwidth for all the M channels). Therefore, the ith UE bid for bandwidth can be written as $w_i = pr_i$, where $\sum_{i=1}^{M} w_i = p \sum_{i=1}^{M} r_i$. The first term in equation (4.7) is separable in r_i, so we have $\max_{\mathbf{r}} \sum_{i=1}^{M} (\log(V_i(r_i)) - pr_i) = \sum_{i=1}^{M} \max_{r_i} (\log(V_i(r_i)) - pr_i)$ and the dual problem objective function can be written as equation (4.8).

$$\begin{aligned} D(p) =& \max_{\mathbf{r}} L(\mathbf{r}, p) \\ =& \sum_{i=1}^{M} \max_{r_i} \left(\log(V_i(r_i)) - pr_i \right) + p(R - z) \\ =& \sum_{i=1}^{M} \max_{r_i} (L_i(r_i, p)) + p(R - z) \end{aligned} \tag{4.8}$$

Thus, the dual problem is formulated as equation (4.9).

$$\begin{aligned} \min_{p} \quad & D(p) \\ \text{subject to} \quad & p \geq 0. \end{aligned} \tag{4.9}$$

Leveraging the method of Lagrange multipliers, we have:

$$\frac{\partial D(p)}{\partial p} = R - \sum_{i=1}^{M} r_i - \sum_{i=1}^{M} z_i = 0 \tag{4.10}$$

Substituting by $\sum_{i=1}^{M} w_i = p \sum_{i=1}^{M} r_i$, we have equation (4.11), minimized to $p = \frac{\sum_{i=1}^{M} w_i}{R}$ at $z = 0$ where $w_i = pr_i$ is transmitted by the ith UE to the eNB.

$$p = \frac{\sum_{i=1}^{M} w_i}{R - \sum_{i=1}^{M} z_i} \tag{4.11}$$

As such, we divide the log-EURA problem (4.5) into simpler optimizations at the eNB (eNB EURA problem) and UEs (UE EURA problem). These are respectively shown in

equations (4.13) and (4.12) whose solutions, guaranteeing the utility proportional fairness in equation (4.3), are summarized in Algorithms 5 and 4 in that order.

$$\begin{aligned} \max_{r_i} \quad & \log V_i(r_i) - pr_i \\ \text{subject to} \quad & p \geq 0 \\ & r_i \geq 0, \quad i = 1, 2, ..., M. \end{aligned} \tag{4.12}$$

During the execution of the aforesaid algorithms, starting with $w_i(0) = 0$, the ith UE, transmits an initial bid $w_i(1)$ to the eNB, which in turn subtracts the latterly received bid $w_i(n)$ and the formerly received one $w_i(n-1)$ and ceases the procedure if the difference is less than a threshold δ; Otherwise, it computes and sends a shadow price $p(n) = \frac{\sum_{i=1}^{M} w_i(n)}{R}$ to is covered UEs. The ith UE extracts its rate $r_i(n)$ from the received $p(n)$ such that $\log V_i(r_i) - p(n)r_i$ is maximized. The rate $r_i(n)$ is employed to estimate the new bid $w_i(n) = p(n)r_i(n)$, transmitted to the eNB. This routine repeats until the bid difference $|w_i(n) - w_i(n-1)|$ falls below the threshold δ.

$$\begin{aligned} \min_{p} \quad & D(p) \\ \text{subject to} \quad & p \geq 0. \end{aligned} \tag{4.13}$$

The solution $r_i(n)$ of the ith UE EURA optimization $r_i(n) = \arg\max_{r_i} \left(\log V_i(r_i) - p(n)r_i \right)$ in Algorithm 4 essentially solves the equation $\frac{\partial \log V_i(r_i)}{\partial r_i} = p(n)$, algebraically the Lagrange multiplier solution for equation (4.12) and geometrically the intersection point of the horizontal line $y = p(n)$ with the curve $y = \frac{\partial \log V_i(r_i)}{\partial r_i}$.

A convergence analysis of the EURA algorithms and its resultant snags are discussed in section 4.3.1.1.

Algorithm 4 UE EURA Optimization Algorithm

Send initial bid $w_i(1)$ to eNB.
loop
  Receive shadow price $p(n)$ from eNB.
  if STOP from eNB **then**
    Calculate allocated rate $r_i^{\text{opt}} = \frac{w_i(n)}{p(n)}$.
    STOP
  else
    Solve $r_i(n) = \arg\max_{r_i} \Big(\log V_i(r_i) - p(n) r_i \Big)$.
    Send new bid $w_i(n) = p(n) r_i(n)$ to eNB.
  end if
end loop

Algorithm 5 eNB EURA Optimization Algorithm

loop
  Receive bids $w_i(n)$ from UEs. $\{$Let $w_i(0) = 1 \ \forall i\}$
  if $|w_i(n) - w_i(n-1)| < \delta \ \forall i$ **then**
    Allocate rates, $r_i^{\text{opt}} = \frac{w_i(n)}{p(n)}$ to user i.
    STOP
  else
    Calculate $p(n) = \frac{\sum_{i=1}^{M} w_i(n)}{R}$.
    Send new shadow price $p(n)$ to all UEs.
  end if
end loop

4.3.1.1 EURA Convergence Analysis

To commence analyzing the EURA Algorithms 4 and 5, lemma 3.1.2 from chapter 3 is helpful and is repeated here for the ease of reference. The proof is repeated because some of the equations in the body of the proof will be used in the latter statements in the chapter.

Lemma 4.3.1. *The aggregate utility function $V_i(r_i)$, the slope curvature function $\frac{\partial \log V_i(r_i)}{\partial r_i}$ has inflection points at $r_i = r_{ij}^s \approx r_{ij}^{inf}$ for j^{th} application utility function U_{ij} and is convex for $r_{ij} > \max_j r_{ij}^s$.*

Proof. For the ith UE aggregate utility $V_i(r_i)$, let $S_i(r_i) = \frac{\partial \log V_i(r_i)}{\partial r_i}$ be the aggregate utility slope curvature function , $S_{ij}(r_{ij}) = \frac{\partial \log U_{ij}(r_{ij})}{\partial r_{ij}}$ be the jth application utility slope curvature function, and N_i^S be the number of sigmoidal utility functions. Taking the logarithm and

derivative of both sides of the equation (4.2) yields in equation (4.14).

$$\begin{aligned} S_i(r_i) &= \frac{\partial \log V_i(r_i)}{\partial r_i} = \frac{\partial}{\partial r_i} \sum_{j=1}^{N_i} \alpha_{ij} \log U_{ij}(r_{ij}) \\ &= \sum_{j=1}^{N_i} \alpha_{ij} \frac{\partial \log U_{ij}(r_{ij})}{\partial r_{ij}} = \sum_{j=1}^{N_i} \alpha_{ij} S_{ij}(r_{ij}) \\ &= \sum_{j=1}^{N_i^S} \alpha_{ij} S_{ij}(r_{ij}) + \sum_{j=N_i^S+1}^{N_i} \alpha_{ij} S_{ij}(r_{ij}) \end{aligned} \tag{4.14}$$

Taking the 1th and 2nd derivatives of equation (4.14), we can write:

$$\begin{aligned} \frac{\partial S_i}{\partial r_i} = \sum_{j=1}^{N_i^S} \{ & \frac{-\alpha_{ij} a_{ij}^2 d_{ij} e^{-a_{ij}(r_{ij}-b_{ij})}}{c_{ij}\left(1 - d_{ij}(1 + e^{-a_{ij}(r_{ij}-b_{ij})})\right)^2} \\ & + \frac{\alpha_{ij} a_{ij}^2 e^{-a_{ij}(r_{ij}-b_{ij})}}{\left(1 + e^{-a_{ij}(r_{ij}-b_{ij})}\right)^2} \} \\ & - \sum_{j=N_i^S+1}^{N_i} \{ \frac{\alpha_{ij} k_{ij}^2}{(1 + k_{ij} r^{max}) \log(1 + k_{ij} r_{ij})^2} \} \end{aligned} \tag{4.15}$$

$$\begin{aligned} \frac{\partial^2 S_i}{\partial r_i^2} = \sum_{j=1}^{N_i^S} \{ & \frac{d_{ij} e^{-a_{ij}(r_{ij}-b_{ij})}(1 - d_{ij}(1 - e^{-a_{ij}(r_{ij}-b_{ij})}))}{c_{ij}\left(1 - d_{ij}(1 + e^{-a_{ij}(r_{ij}-b_{ij})})\right)^3} \\ & + \frac{e^{-a_{ij}(r_{ij}-b_{ij})}(1 - e^{-a_{ij}(r_{ij}-b_{ij})})}{\left(1 + e^{-a_{ij}(r_{ij}-b_{ij})}\right)^3} \} \times a_{ij}^3 \alpha_{ij} \\ & - \sum_{j=N_i^S+1}^{N_i} \{ \frac{\alpha_{ij} k_{ij}^2 (\log(1 + k_{ij} r_{ij}) - 1)}{(1 + k_{ij} r_{ij})^2 \log^2(1 + k_{ij} r_{ij})} \} \end{aligned} \tag{4.16}$$

It is easy to show that $\forall\, r_i,\ \frac{\partial S_i}{\partial r_i} < 0$. Denoting the 1th term of equation (4.15) and 2nd and 3rd terms of equation (4.16) as respectively S_i^1, S_i^2, and S_i^3 stems out equation set (4.17), for

which properties in equation set (4.18) are considerable.

$$\begin{cases} S_i^1 = \frac{\alpha_{ij} a_{ij}^3 e^{a_{ij} b_{ij}} (e^{a_{ij} b_{ij}} + e^{-a_{ij}(r_{ij} - b_{ij})})}{(e^{a_{ij} b_{ij}} - e^{-a_{ij}(r_{ij} - b_{ij})})^3} \\ S_i^2 = \frac{a_{ij}^3 \alpha_{ij} e^{-a_{ij}(r_{ij} - b_{ij})} (1 - e^{-a_{ij}(r_{ij} - b_{ij})})}{\left(1 + e^{-a_{ij}(r_{ij} - b_{ij})}\right)^3} \\ S_i^3 = \frac{\alpha_{ij} k_{ij}^2 (\log(1 + k_{ij} r_{ij}) - 1)}{(1 + k_{ij} r_{ij})^2 \log^2(1 + k_{ij} r_{ij})} \end{cases} \tag{4.17}$$

From equation set (4.18), we observe that the slope curvature function S_i has the inflection point $r_i = r_{ij}^s \approx b_{ij} = r_{ij}^{\inf}$ and changes from a convex function close to the origin to a concave function before the inflection point at $r_{ij} = r_{ij}^s$ to a convex function after the inflection point.

$$\begin{cases} \lim_{r_i \to 0} S_i^1 = \infty, \\ \lim_{r_i \to b_{ij}} S_i^1 = 0 \text{ for } b_{ij} \gg \frac{1}{a_{ij}} \ \forall \ j \\ S_i^2(b_{ij}) = 0 \\ S_i^2(r_{ij} > b_{ij}) > 0 \\ S_i^2(r_{ij} < b_{ij}) < 0 \\ S_i^3(r_{ij} > 0) > 0 \end{cases} \tag{4.18}$$

□

Next, corollary 4.3.2 is considerable.

Corollary 4.3.2. *If* $\sum_{i=1}^{M} \max_j \ r_{ij}^s \ll R$, *then Algorithms 4 and 5 converge to the global optimal rates corresponding to the steady state shadow price* $p_{ss} < \frac{a_{i_{\max}} d_{i_{\max}}}{1 - d_{i_{\max}}} + \frac{a_{i_{\max}}}{2}$, *where* $i_{\max} = \arg\max_i \ r_{ij_{\max}}^s$ *and* $r_{ij_{\max}}^s = \max_j \ r_{ij}^s$.

Proof. An essential step to reach the optimal solution in Algorithm (4) is solving $r_i(n) = \arg\max_{r_i} \left(\log V_i(r_i) - p(n) r_i \right)$ using the Lagrange multipliers in equation (4.19).

$$\frac{\partial \log V_i(r_i)}{\partial r_i} - p = S_i(r_i) - p = 0. \tag{4.19}$$

Furthermore, equation (4.18) indicates that the slope curvature function $S_i(r_i)$ is convex for $r_i > \max_j r_{ij}^s \approx \max_j b_{ij}$. Similar to the analyses in [84, 89], the Algorithms 4 and 5 are guaranteed to converge to the global optimal solution when the aggregate slope curvature function $S_i(r_i)$ is in the convex region. Hence, the aggregate utility function's natural logarithm converges to the global optimal solution for $r_i > \max_j r_{ij}^s \approx \max_j b_{ij}$. On the other hand, the sigmoidal utility function inflection point is at $r_i^{\inf} = b_{ij}$. For $\sum_{i=1}^{M} \max_j \; r_{ij}^s \ll R$, Algorithms 4 and 5 allocate rates $r_{ij} > b_{ij}$ for all users and since $S_{ij}(r_{ij})$ is convex for $r_{ij} > r_{ij}^s \approx b_{ij}$, the optimal solution can be achieved by the algorithms. Equation (4.19) and convexity of $S_{ij}(r_{ij})$ for $r_{ij} > r_{ij}^s \approx b_{ij}$ imply that $p_{ss} < S_{ij}(r_{ij} = \max b_{ij})$, where $S_{ij}(r_{ij} = \max b_{ij}) = \frac{a_{i_{\max}} d_{i_{\max}}}{1-d_{i_{\max}}} + \frac{a_{i_{\max}}}{2}$ and $i_{\max} = \arg\max_i b_{ij}$. □

Next, corollary 4.3.3 shows that the EURA optimization can be non-stable.

Corollary 4.3.3. *For $\sum_{i=1}^{M} \max_j r_{ij}^s > R$ and the global optimal shadow price $p_{ss} \approx \frac{a_{ij} d_{ij} e^{\frac{a_{ij} b_{ij}}{2}}}{1-d_{ij}(1+e^{\frac{a_{ij} b_{ij}}{2}})} + \frac{a_{ij} e^{\frac{a_{ij} b_{ij}}{2}}}{(1+e^{\frac{a_{ij} b_{ij}}{2}})}$, the solution by EURA Algorithms 4 and 5 fluctuates about the global optimal solution.*

Proof. It follows from lemma 4.3.1 that for $\sum_{i=1}^{M} r_{ij}^{\inf} > R$, $\exists\ i$ such that the optimal rates $r_{ij}^{\text{opt}} < b_{ij}$. Thus, $p_{ss} \approx \frac{a_{ij} d_{ij} e^{\frac{a_{ij} b_{ij}}{2}}}{1-d_{ij}(1+e^{\frac{a_{ij} b_{ij}}{2}})} + \frac{a_{ij} e^{\frac{a_{ij} b_{ij}}{2}}}{(1+e^{\frac{a_{ij} b_{ij}}{2}})}$ is the optimal shadow price for the optimization problem in equation (4.3). Then, a small change in the shadow price $p(n)$ at the nth iteration can cause the rate $r_{ij}(n)$ (the root of $S_{ij}(r_{ij}) - p(n) = 0$) to fluctuate between the concave and convex curvature of the slope curve $S_{ij}(r_{ij})$ for the ith UE. Therefore, it produces a fluctuation in the bid value $w_i(n)$ sent to the eNB, which in turn induces a vacillation of the shadow price $p(n)$ transmitted by eNB to the UEs. Hence, the iterative solution oscillates about the global optimal rates r_{ij}^{opt}. □

The non-stability of the EURA optimization as it was shown in corollary 4.3.3 leads to the dearth of optimality of the rates assigned by the EURA optimization. This is illustrated in theorem 4.3.4.

Theorem 4.3.4. *EURA Algorithms 4 and 5 do not converge to the optimal solution for all eNB rates R.*

Proof. It directly follows from the corollaries 4.3.2 and 4.3.3 that the EURA algorithm does not converge to the global optimal solution for all values of R. □

The potential EURA seesawing about optimal rates and dearth of convergence thereof motivate us to include some robustness into the procedure. This is done in section 4.3.1.2.

4.3.1.2 EURA Robust Algorithm

Incorporating robustness into the EURA Algorithms 4 and 5 so that they converge for all eNB rates requires the algorithm to refrain from fluctuations in the non-convergent region for $\sum_{i=1}^{M} \max_j r_{ij}^s \ll R$. To do this, a *fluctuation decay function* $\Delta w(n)$ as below reduces the step size between the current and previous bid, i.e. $w_i(n) - w_i(n-1)$, for every user i if a fluctuation occurs. The allocated rates should coincide with those of EURA Algorithms 4 and 5 for $\sum_{i=1}^{M} \max_j r_{ij}^s > R$.

- *Exponential function*: $\Delta w(n) = l_1 e^{-\frac{n}{l_2}}$.
- *Rational function*: $\Delta w(n) = \frac{l_3}{n}$.

wherel_1, l_2, l_3 can be adjusted to change the bids w_i decay rate.

Remark 4.3.5. *The fluctuation decay function can be included in either UE EURA Algorithm or eNB EURA Algorithm.*

In our model, we choose to incorporate the decay function into the UE EURA Algorithm even though, as mentioned before, it can be placed in the eNB EURA as well. The fledgling robust EURA process is illustrated in Algorithms 6 and 7. Here, starting with $w_i(0) = 0$, each UE commences transmitting an initial bid $w_i(1)$ to the eNB, which at each iterate n calculates the difference between the currently and formerly received bids $w_i(n)$ and $w_i(n-1)$, then exits if the difference falls below a threshold δ; otherwise, it computes the shadow price $p_E(n) = \frac{\sum_{i=1}^{M} w_i(n)}{R}$ and send it to its covered UEs, amongst which The ith UE obtains the rate r_i maximizing the statement $\log \beta_i V_i(r_i) - p_E(n) r_i$, estimates its new bid $w_i(n) = p_E(n) r_i(n)$, and sends it to the eNB.

Algorithm 6 UE EURA Robust Algorithm

Send initial bid $w_i(1)$ to eNB.
loop
  Receive shadow price $p(n)$ from eNB.
  if STOP from eNodeB **then**
    Calculate allocated rate $r_{ij}^{\text{opt}} = \frac{w_i(n)}{p(n)}$.
  else
    Solve $r_i(n) = \arg\max_{r_i} \Big(\beta_i \log V_i(r_i) - p_E(n) r_i \Big)$.
    Calculate new bid $w_i(n) = p(n) r_i(n)$.
    if $|w_i(n) - w_i(n-1)| > \Delta w(n)$ **then**
      $w_i(n) = w_i(n-1) + \text{sign}(w_i(n) - w_i(n-1)) \Delta w(n)$ $\{\Delta w = l_1 e^{-\frac{n}{l_2}}$ or $\Delta w = \frac{l_3}{n}\}$
    end if
    Send new bid $w_i(n)$ to eNB.
  end if
end loop

Algorithm 7 eNB EURA Algorithm

loop
  Receive bids $w_i(n)$ from UEs $\{$Let $w_i(0) = 1 \ \forall i\}$
  if $|w_i(n) - w_i(n-1)| < \delta \ \forall i$ **then**
    STOP and allocate rates (i.e r_i^{opt} to user i)
  else
    Calculate $p_E(n) = \frac{\sum_{i=1}^{M} w_i(n)}{R}$
    Send new shadow price $p_E(n)$ to all UEs
  end if
end loop

Remark 4.3.6. *If the subscriber differentiation parameter β_i is available only at the eNB (or other network provider unit), the shadow price p_E is changed to $\frac{p_E}{\beta_i}$.*

Next, section 4.3.2 solves the IURA optimization problem in equation (4.4).

4.3.2 IURA Algorithm

This section presents the second stage of the distributed resource allocation during which the application rates r_{ij} are optimally assigned internally to the UEs in accordance with Algorithm 8, where the ith UE leverages the EURA allocated rate r_i^{opt} to solve $\mathbf{r}_i = \arg\max_{\mathbf{r}_i} \sum_{j=1}^{N_i}(\alpha_{ij} \log U_{ij}(r_{ij}) - p_I r_{ij}) + p_I r_{ij}^{\text{opt}}$.

Algorithm 8 UE IURA Algorithm

loop
  Receive r_i^{opt} from eNB. {by EURA Algorithms}
  Solve
  $\mathbf{r}_i = \arg\max_{\mathbf{r}_i} \sum_{j=1}^{N_i}(\alpha_{ij} \log U_{ij}(r_{ij}) - p_I r_{ij}) + p_I r_i^{\text{opt}}$ $\{\mathbf{r}_i = \{r_{i1}, r_{i2}, ..., r_{iN_i}\}\}$
  Allocate r_{ij} to the jth application.
end loop

Next, section 4.4 proves the mathematical equivalence of the distributed resource allocation architecture introduced in this chapter with the centralized architecture that was developed in chapter 3.

4.4 Mathematical Equivalence

Here, we show the mathematical equivalence of the distributed resource allocation in equations (4.3) and (4.4) with the centralized approach in equation (4.1). The reason to show the mathematical equivalence can be intuitively observed through Figure 4.2. As we can see, the BS in the centralized architecture has to receive the requirements from each of the

applications and then assign the rates to the applications accordingly. On the other hand, the problem can be greatly simplified by decomposing the application rate allocation into two simpler problem. First, the BS assign rates to the UEs, and then, the UE assigns rates to the applications. This can be very helpful in simplifying the BS, such as LTE-in-a-Box.

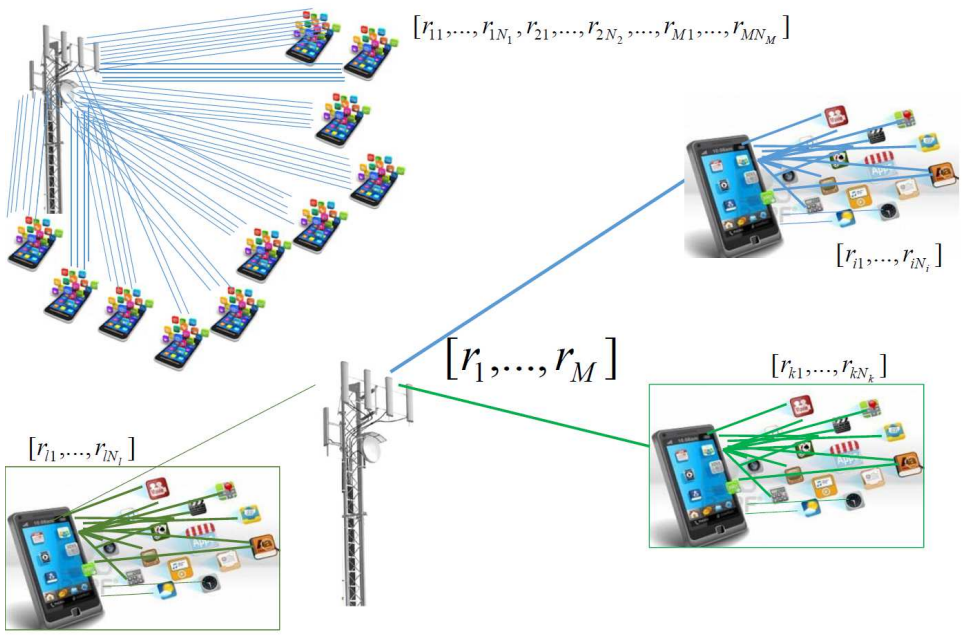


Figure 4.2: The centralized resource allocation assigns application after receiving the request from each application. The problem can simply be decomposed to network and device optimization where the BS only assigns UE rates, and the UE assigns application rates.

Before proceeding further, we repeat the lemma 4.4.1, proved in chapter 3, for the sake of reading convenience.

Lemma 4.4.1. *The aggregate and application utility slope curvature functions* $S_i(r_i) = \frac{\partial \log V_i(r_i)}{r_i}$ *and* $S_{ij}(r_{ij}) = \frac{\partial \log U_{ij}(r_{ij})}{r_{ij}}$ *are invertible and their inverse functions* $r_i = S_i^{-1}(.)$ *and* $r_{ij} = S_{ij}^{-1}(.)$ *are strictly decreasing.*

Proof. The concavity of the logarithmic utility U_ij yields in $U'_{ij}(r_{ij}) = \frac{\partial U_{ij}(r_{ij})}{\partial r_{ij}} > 0$ and $U''_{ij}(r_{ij}) = \frac{\partial^2 U_{ij}(r_{ij})}{\partial r_{ij}^2} < 0$, and lemma 4.2.1 stems out $S_{ij}(r_{ij}) = \frac{\partial \log(U_{ij}(r_{ij}))}{\partial r_{ij}} = \frac{U'_{ij}(r_{ij})}{U_{ij}(r_{ij})} > 0$ and $\frac{\partial S_{ij}(r_{ij})}{\partial r_{ij}} = \frac{U''_{ij}(r_{ij})U_{ij}(r_{ij}) - U'^2_{ij}(r_{ij})}{U^2_{ij}(r_{ij})} < 0$. Also, for the utility function, we have $U_{ij}(r_{ij}) > 0$, $U_{ij}(r_{ij})$

is increasing, and it is twice differentiable with respect to r_{ij} (chapter 2). Therefore, $S_{ij}(r_{ij})$ of the logarithmic utility function is strictly decreasing. From equation (3.2) in chapter 3, for the sigmoidal utility function $U_{ij}(r_{ij})$ where $0 < r_{ij} < R$, we can write inequality set (4.20), giving that $S_{ij}(r_{ij})$ of the sigmoidal utility function is strictly decreasing.

$$S_{ij}(r_{ij}) > 0, \frac{\partial}{\partial r_{ij}} S_{ij}(r_{ij}) < 0 \tag{4.20}$$

Equation (4.14) and inequalities 4.20 yield in inequalities (4.21). Henceforth, $S_{ij}(r_{ij})$ and $S_i(r_i)$ of all the utilities in our problem formulation are strictly decreasing functions; thereby, the slope curvature functions $S_{ij}(r_{ij})$ and $S_i(r_i)$ are invertible and the inverse functions are strictly decreasing.

$$\begin{aligned} S_i(r_i) &= \sum_{j=1}^{N_i^S} \alpha_{ij} S_{ij}(r_{ij}) + \sum_{j=N_i^S+1}^{N_i} \alpha_{ij} S_{ij}(r_{ij}) > 0 \\ \frac{\partial S_i(r_i)}{\partial r_i} &= \sum_{j=1}^{N_i^S} \alpha_{ij} \frac{\partial S_{ij}(r_{ij})}{\partial r_{ij}} + \sum_{j=N_i^S+1}^{N_i} \alpha_{ij} \frac{\partial S_{ij}(r_{ij})}{\partial r_{ij}} < 0 \end{aligned} \tag{4.21}$$

□

Next, corollary 4.4.2 proves that the centralized resource allocation architecture in this chapter leads to the same UE and application rates as the centralized resource allocation architecture that was introduced in chapter 3.

Corollary 4.4.2. *The optimal rates assigned by the distributed resource allocation architecture in equations (4.3) and (4.4) are equal to the one allocated by the centralized resource allocation architecture in equation (4.1).*

Proof. The centralized resource allocation architecture's optimization in equation (4.1) can be written as the log-centralized optimization in equation (3.11) from chapter 3, which

is equation (4.22). The Lagrangian of the log-centralized optimization can be written as equation (4.23) where $z \geq 0$ is the slack variable and p_T is the Lagrange multiplier.

$$\begin{aligned} \max_{\mathbf{r}} \quad & \sum_{i=1}^{M} \beta_i \sum_{j=1}^{N_i} \alpha_{ij} \log U_{ij}(r_{ij}) \\ \text{subject to} \quad & \sum_{i=1}^{M} \sum_{j=1}^{N_i} r_{ij} \leq R, \\ & r_{ij} \geq 0, \quad i = 1, 2, ..., M, \; j = 1, 2, ..., N_i \end{aligned} \tag{4.22}$$

$$L_T(\mathbf{r}) = (\sum_{i=1}^{M} \beta_i \sum_{j=1}^{N_i} \alpha_{ij} \log U_{ij}(r_{ij})) - p_T(\sum_{i=1}^{M} \sum_{j=1}^{N_i} r_{ij} - R + z) \tag{4.23}$$

Then, we have that:

$$\frac{\partial L_T(\mathbf{r})}{\partial r_{ij}} = \beta_i \alpha_{ij} S_{ij}(r_{ij}) - p_T = 0 \Rightarrow p_T = \beta_i \alpha_{ij} S_{ij}(r_{ij}) \tag{4.24}$$

so, the ith UE's j^{th} application rate is:

$$r_{ij} = S_{ij}^{-1}(\frac{p_T}{\beta_i \alpha_{ij}}) \tag{4.25}$$

Using equation (4.14), we can write:

$$N_i p_T = \beta_i S_i(r_i) \tag{4.26}$$

And the ith UE rate can be calculated as equation (4.27).

$$r_i = S_i^{-1}(\frac{N_i p_T}{\beta_i}). \tag{4.27}$$

The Lagrangian of the distributed architecture's log-EURA optimization in equation can be written as equation (4.28) where $z \geq 0$ is the slack variable and p_E is the Lagrange multiplier.

$$L_E(\mathbf{r}) = (\sum_{i=1}^{M} \beta_i \log V_i(r_i)) - p_E(\sum_{i=1}^{M} r_i - R + z) \tag{4.28}$$

Then, we have that:

$$\frac{\partial L_E(\mathbf{r})}{\partial r_i} = \beta_i S_i(r_i) - p_E = 0 \Rightarrow p_E = \beta_i S_i(r_i) \tag{4.29}$$

So, the ith UE rate can be written as:

$$r_i = S_i^{-1}(\frac{p_E}{\beta_i}) \tag{4.30}$$

Replacing S_i from equation (4.14), we can write:

$$p_E = \beta_i \sum_{j=1}^{N_i} \alpha_{ij} S_{ij}(r_{ij}) \tag{4.31}$$

And, we get equation (4.32) below.

$$p_E = \sum_{j=1}^{N_i} p_T = N_i p_T \tag{4.32}$$

Equations (4.27), the UE rate for the centralized architecture, and (4.30), the UE rates for the distributed architecture (EURA component), signify that the centralized and distributed resource allocation architectures lead to identical UE rates. On the other hand, the Lagrangian of the distributed resource allocation architecture's log-IURA optimization

in equation (4.6) can be written as equation (4.33) where $z \geq 0$ is the slack variable and p_I is the Lagrange multiplier corresponding to the internal shadow price, price per bandwidth for all applications in the ith UE.

$$L_I(r_i) = (\sum_{j=1}^{N_i} \alpha_{ij} \log U_{ij}(r_{ij})) - p_I(\sum_{j=1}^{N_i} r_{ij} - r_{ij}^{\text{opt}} + z) \tag{4.33}$$

Then, we have that:

$$\frac{\partial L_I(r_i)}{\partial r_{ij}} = \alpha_{ij} S_{ij}(r_{ij}) - p_I = 0 \Rightarrow p_I = \alpha_{ij} S_{ij}(r_{ij}) \ \ \forall \ j \tag{4.34}$$

And, summing the ith UE applications gives that:

$$\sum_{j=1}^{N_i} p_I = \sum_{j=1}^{N_i} \alpha_{ij} S_{ij}(r_{ij}) \tag{4.35}$$

Using equation (4.14) results in equation (4.36).

$$\beta_i N_i p_I = \beta_i S_i(r_i) = p_E = N_i p_T \Rightarrow p_T = \beta_i p_I \tag{4.36}$$

So, the rate of the ith UE's j^{th} application can be written as equation (4.37).

$$r_{ij} = S_{ij}^{-1}(\frac{p_I}{\alpha_{ij}}) = S_{ij}^{-1}(\frac{p_T}{\beta_i \alpha_{ij}}) \tag{4.37}$$

Considering the constraints of the equation (4.6), the total rate of the ith UE can be written as equation (4.38).

$$r_i^{\text{opt}} = \sum_{j=1}^{N_i} r_{ij} = \sum_{j=1}^{N_i} S_{ij}^{-1}(\frac{p_T}{\beta_i \alpha_{ij}}) \tag{4.38}$$

Comparing equation (4.37), the application rates from the distributed architecture (IURA component), with equation (4.25), the application rate from the centralized architecture, signify that the centralized and distributed resource allocation architectures lead to identical application rates. To summarize, the UE and application rates assigned by the centralized and distributed architectures are the same. □

Next, theorem 4.4.3 proves that the centralized resource allocation architecture in chapter 3 is mathematically equivalent to the distributed resource allocation architecture presented in this chapter.

Theorem 4.4.3. *The distributed resource allocation architecture in in equations (4.3) and (4.4) is equivalent to the centralized resource allocation architecture in equation (4.1).*

Proof. Corollary 4.4.2) proved that the distributed and centralized resource allocation architectures produce identical UE rates as well as equal application rates. Therefore, the centralized resource allocation architecture is equivalent to the distributed resource allocation architecture. □

Next, section 4.5 presents simulations to portray the application of the proposed distributed resource allocation architecture to a cellular communications system.

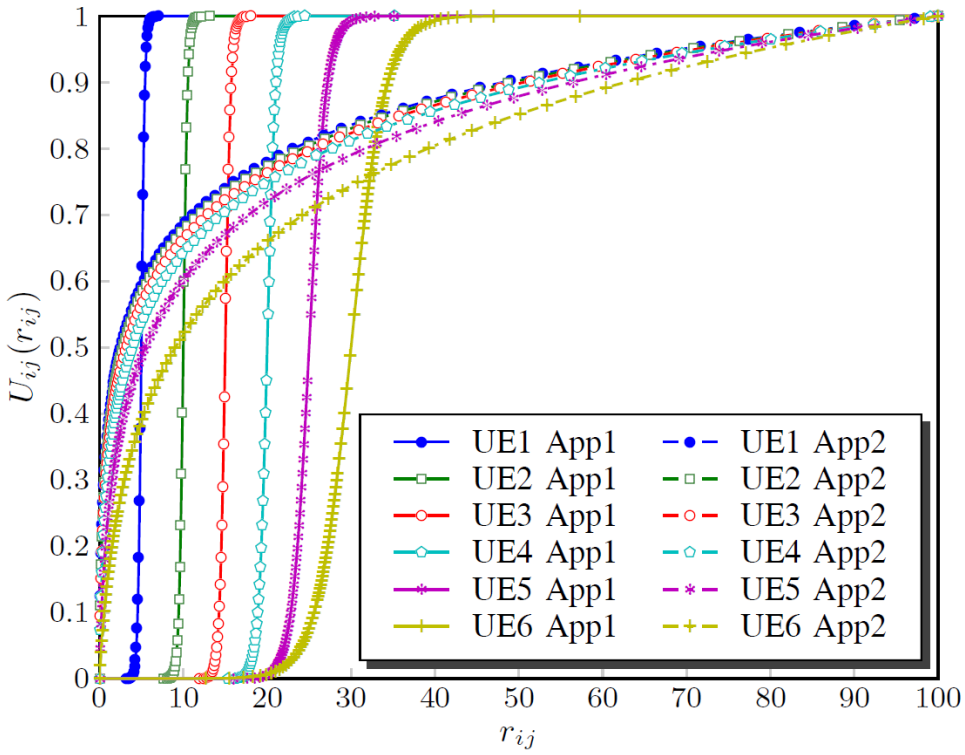
4.5 Distributed Architecture Simulation

Similar to the simulations for the centralized architecture in chapter 3, a cell with $M = 6$ UEs and an eNB, depicted in Figure 4.1, is considered and each UE concurrently runs a delay-tolerant and a real-time application with respectively logarithmic and sigmoidal application utility functions with parameters in Table 4.1. The sigmoidal utility with parameters $a = 5$, $b = 10$ approximates a step function at rate $r = 5$ and is a good model for VoIP , while

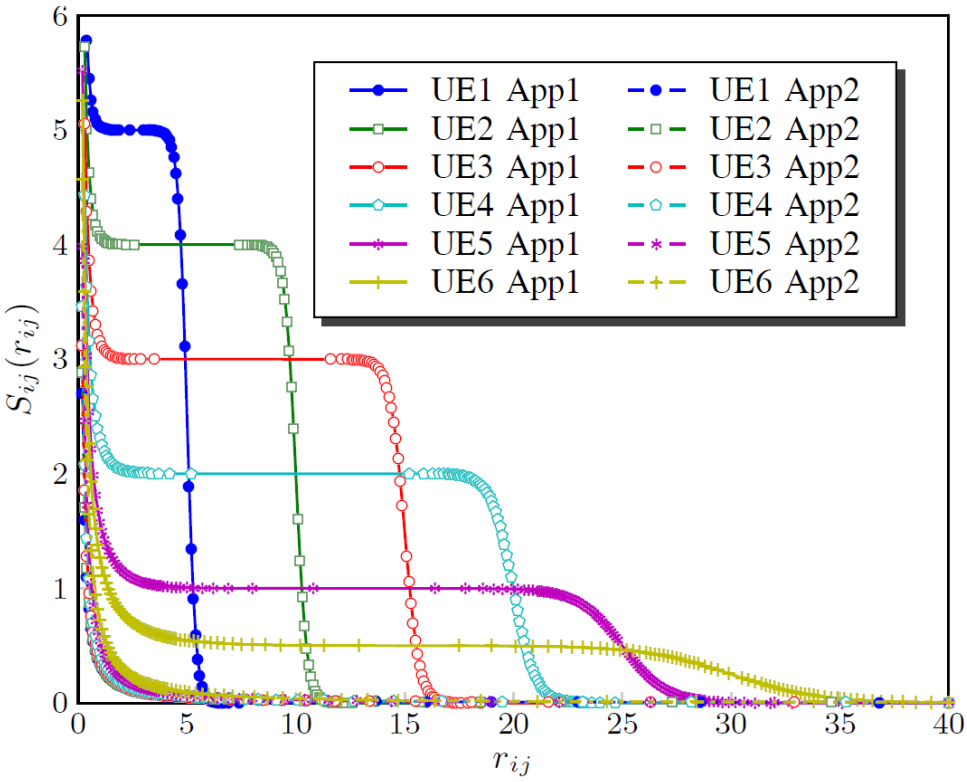
parameters $a = 3$, $b = 15$ is an approximation of a real-time application with an inflection point at rate $r = 15$ and is conducive to modeling standard definition video streaming, whereas parameters $a = 1$, $b = 25$ is an estimation of another real-time application with the inflection point $r = 25$ and is appropriate for the high definition video streaming . Moreover, the logarithmic utilities with $r^{\max} = 100$ and distinct k_i parameters estimate delay-tolerant FTP applications. The plots of the utility functions in Table 4.1 are shown in Figure 4.3(a), from which we can observe that the real-time applications require a minimum rate, i.e. the inflection point, after which the application QoS is fulfilled to a large extent. On the other hand, the logarithmic utility is provided with some QoS even at low rates suitable for the delay-tolerant nature of the applications. Furthermore, as we can observe from Figure 4.3(a), in compliance with the properties mentioned in section 3.1 of chapter 3 the utility functions are strictly increasing continuous functions, zero valued at zero rates. Furthermore, the first derivative of the utility functions natural logarithm, $S_ij(r_ij)$, are shown in Figure 4.3(b), which reflects the positivity and decreasing nature of the first derivative in line with lemma 4.4.1.

Then, the distributed resource allocation architecture's solution Algorithms 6, 7, and 8) were applied to the aforesaid logarithmic and sigmoidal utility functions using MATLAB. To account for the applications usage percentage, we set the application status weight vector in equation (4.2) as $\boldsymbol{\alpha} = \{\alpha_{11}, \alpha_{21}, \alpha_{31}, \alpha_{41}, \alpha_{51}, \alpha_{61}, \alpha_{12}, \alpha_{22}, \alpha_{32}, \alpha_{42}, \alpha_{52}, \alpha_{62}\}$ where α_{ij} represents the status weight of the jth application of the ith UE. It is noteworthy that the addition of application usage percentages per UE is unity, i.e. $\alpha_{i1} + \alpha_{i2} = 1$.

In addition, the aggregate utility functions $V_i(r_i)$ for $i \in \{1, ..., 6\}$ are depicted in Figure 4.4(a) and the first derivative of their natural logarithm, $S_i(r_i)$ for $i \in \{1, ..., 6\}$, are illustrated in Figure 4.4(b). As we can see, in compliance with lemma 4.3.1, the slope curvature functions inflection points occur at the application utility functions' inflection points. Furthermore, in line with lemma 4.4.1, the slope curvature functions are strictly decreasing.

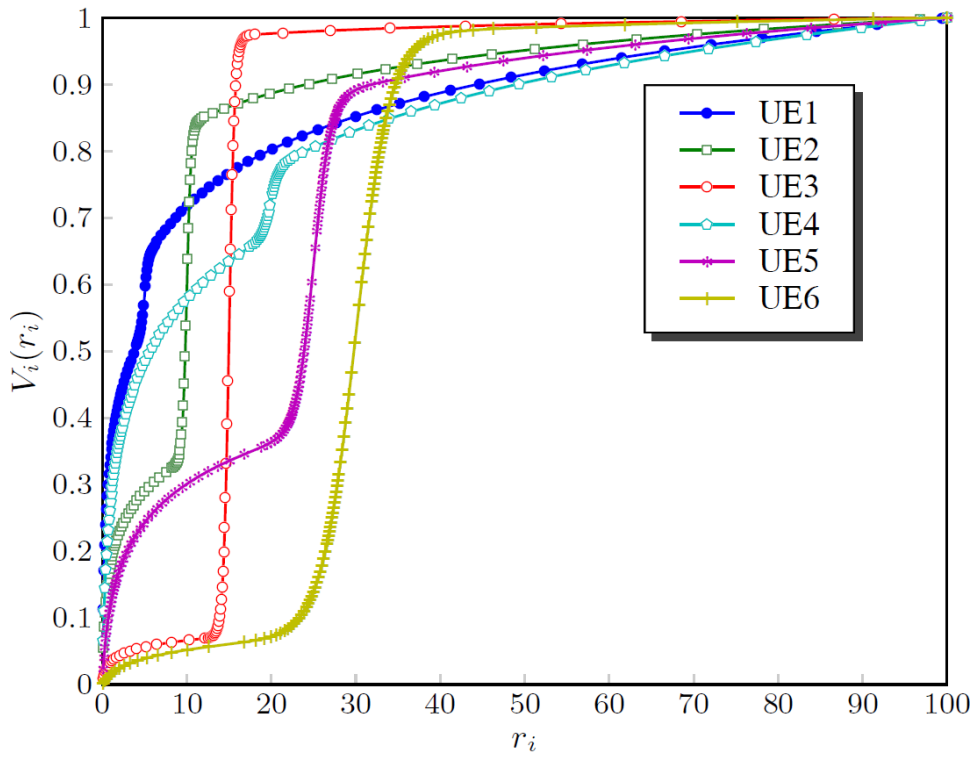


(a) Application Utility Functions - Distributed Architecture

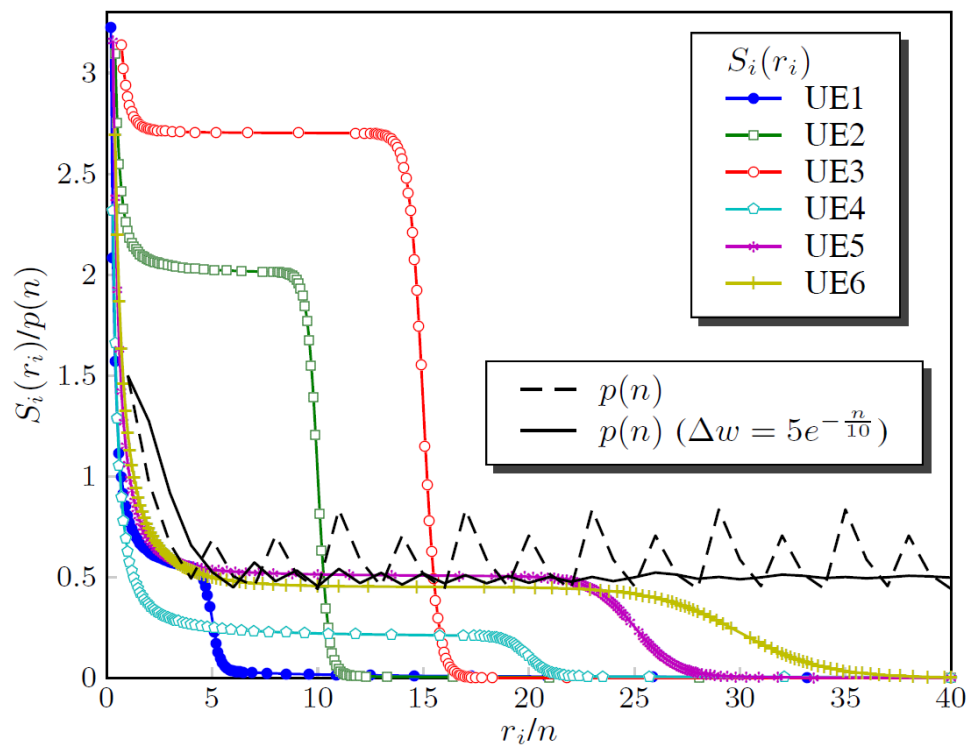


(b) Utility Slope Curvature Functions - Distributed Architecture

Figure 4.3: The system contains 6 UEs, each concurrently running a delay-tolerant and real-time application with respective identically colored logarithmic and sigmoidal utility functions U_{ij} vs. the application-assigned rates r_{ij} plots in Figure 4.3(a). Utility slope curvature functions, the first derivative of the application utility natural logarithms S_{ij} with respect to the application rates r_{ij} are illustrated in Figure 4.3(b) where identical colors relate to the applications on one UE.



(a) Aggregate Utility Functions - Distributed Architecture



(b) Aggregate Slope Curvature Functions - Distributed Architecture

Figure 4.4: Figure 4.4(a) plots the aggregate utilities, multiplications of the usage-percentage-powered application utility functions $V_i(r_i)$ vs. the UE rates r_i, where $i \in \{1, ..., 6\}$. Figure 4.4(b) illustrates the aggregated slope curvature functions , first derivative of the the aggregated utility natural logarithms $S_i(r_i)$. Furthermore, decay function-induced robustness effect is depicted; As we can see, the lack of decay functions yields in the system instability revealed in the shadow price oscillation.

Table 4.1: Applications Utility Parameters - Distributed Architecture

Applications Utilities Parameters	
UE1 App1	Sigmoid $a = 5,\ b = 5$
UE2 App1	Sigmoid $a = 4,\ b = 10$
UE3 App1	Sigmoid $a = 3,\ b = 15$
UE4 App1	Sigmoid $a = 2,\ b = 20$
UE5 App1	Sigmoid $a = 1,\ b = 25$
UE6 App1	Sigmoid $a = 0.5,\ b = 30$
UE1 App2	Logarithmic $k = 15,\ r^{\max} = 100$
UE2 App2	Logarithmic $k = 12,\ r^{\max} = 100$
UE3 App2	Logarithmic $k = 9,\ r^{\max} = 100$
UE4 App2	Logarithmic $k = 6,\ r^{\max} = 100$
UE5 App2	Logarithmic $k = 3,\ r^{\max} = 100$
UE6 App2	Logarithmic $k = 1,\ r^{\max} = 100$

Next section investigates bids and rate allocations for the UEs and applications in our system under varying eNB resource availabilities and the distributed architecture.

4.5.1 Rate Allocation and Bids for $10 \leq R \leq 200$

In the following simulations, we set the termination threshold $\delta = 10^{-4}$ and the eNB rate R to sweep from 10 to 200 with an step size of 5 bandwidth units. Besides, the application status weights is considered to be $\boldsymbol{\alpha} = \{0.1, 0.5, 0.9, 0.1, 0.5, 0.9, 0.9, 0.5, 0.1, 0.9, 0.5, 0.1\}$. It is worth mentioning that the addition of usage percentages per UE is unity, e.g. adding the $1st$ and the $6th$ components of the set (which are indeed the usage percentages for both of applications running on UE1), we get $0.1 + 0.9 = 1$, and so forth. For the distributed resource allocation (Algorithms 6, 7, and 8), UE assigned rates are depicted in Figure 4.5 during the EURA Algorithm with the changes in the eNB available resources R. As we can

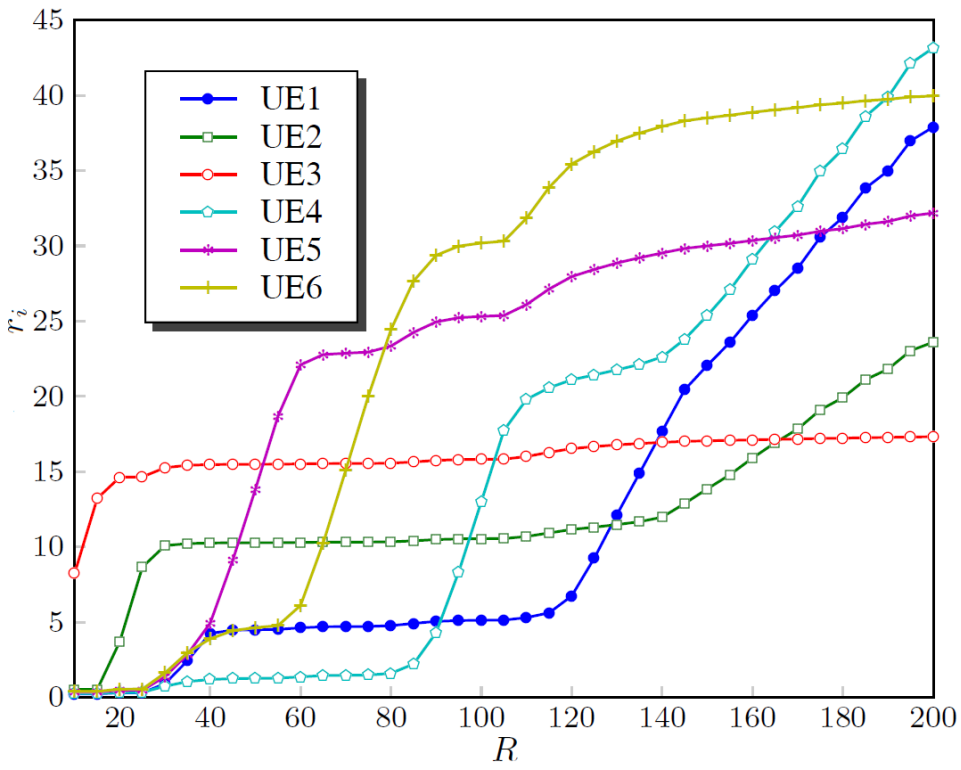


Figure 4.5: UE Optimal Rates: This figure depicts the optimal rates allocated to the UEs by the distributed scheme vs. eNB resources. No user is dropped as no assignment is zero.

observe, initially all the UEs are allocated some rates which is owing to the fact that they all subsume real-time applications in need of immediate rate allocations before any QoS is met. For instance, UE2 has a real-time streaming video application (based on Table 4.1), which requires a bandwidth assignment right away. In Figure 4.6, we show the UEs bids $\{w_i | i \in \{1, ..., 6\}\}$ during the EURA algorithm under changing eNB bandwidth R. First of all, we see that the more resources become available at the eNB, the higher rates are assigned to the UEs. On the other hand, the dearth of the resources (small R) causes those UEs which have applications with higher bit rate requirements to bid higher in order to gain resources. For instance, since UE2 includes a real-time streaming video application, its urgent need for bandwidth allocation causes its initial higher bid for the resources, which is responded by its fast allocation portrayed in the Figure 4.5.

Then, the distributed resource allocation architecture's IURA algorithm has the UEs inter-

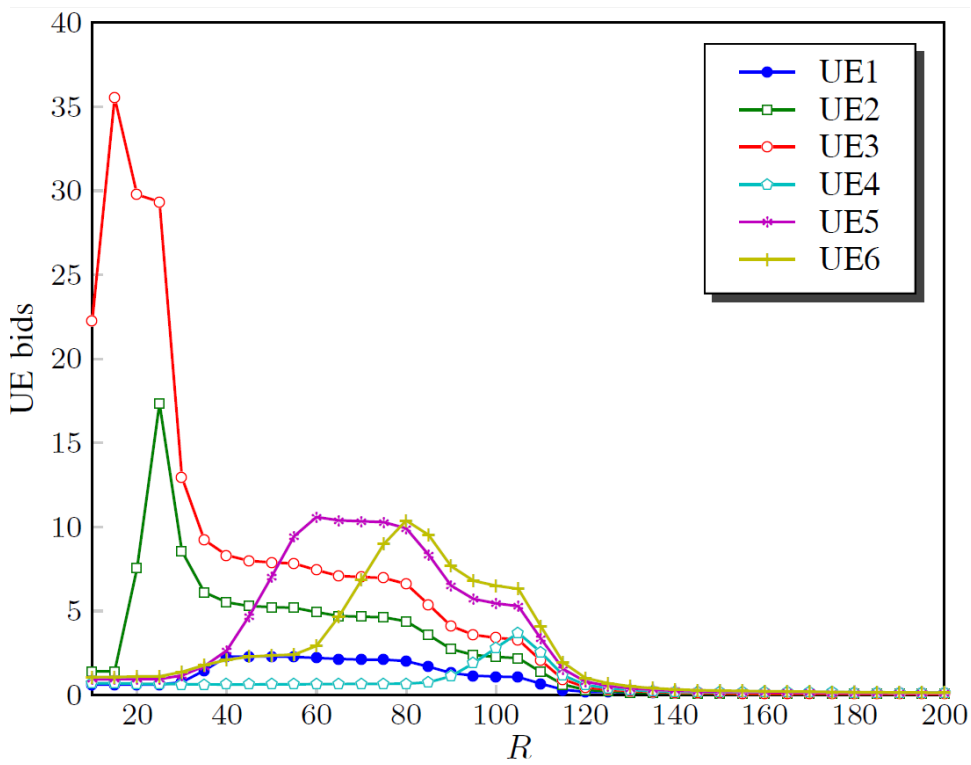


Figure 4.6: UE Bids: This figure illustrates the UE bids for acquiring the resources vs. the eNB rate. The applications requiring more resources bid higher. When bandwidth is scarce, applications needing more resources bid significantly higher than the others. The plots reveal that the higher bid are tantamount to receiving more resources.

nally allocate rates to their applications based on the pledged bids as illustrated in Figures 4.7 and 4.8. In Figure 4.7, we show the allocated applications rates $\{r_{ij}|i \in \{1, ..., 6\} \wedge j \in \{1, 2\}\}$ during the IURA algorithm under changing eNB rate R. As we can observe, initially more resources are allocated to the real-time applications since these have more stringent QoS requirements. In Figure 4.8, we illustrate the applications' internally pledged bids $\{w_{ij}|i \in \{1, ..., 6\} \wedge j \in \{1, 2\}\}$ during the IURA algorithm under changing eNB rate R. Inasmuch as the real-time applications of the UEs need more resources, they bid higher than the delay-tolerant applications specially when the resources are scarce. In fact, we can see that the bid values for the delay-tolerant applications is significantly less than those of the real-time ones such that they are very close to the horizontal axis in Figure 4.8. Furthermore, those applications with higher QoS requirements such as the real-time streaming video in

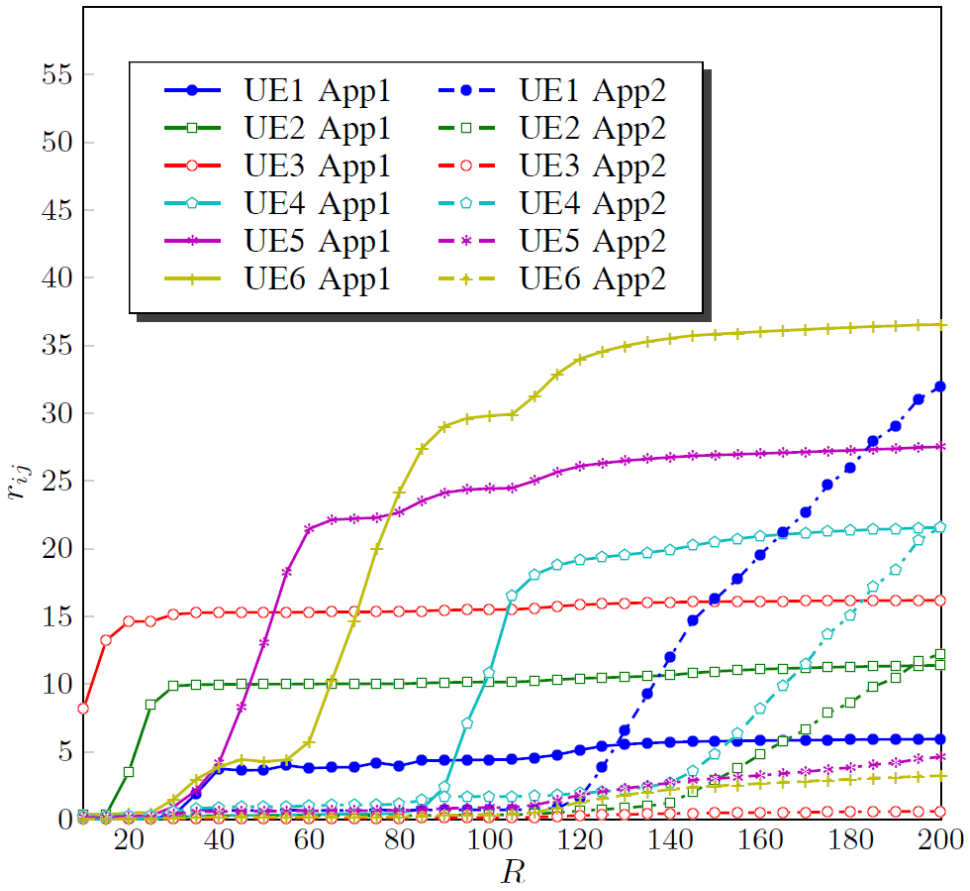


Figure 4.7: Application Optimal Rates: This figure depicts the optimal application rates r_{ij} vs. the eNB rate R. Applications running on an UE are identically colored. As we can see, real-time applications are initially allocated more resources as opposed to the delay-tolerant ones due to their urgent need for resources.

UE1 (red plot) bid higher in order to gain more bandwidth. However, as more resources become available at the eNB, bid values slash down as well.

It is notable that since utility proportional fairness objective functions are leveraged in the formation of the optimizations in equations (4.3) and (4.4), the distributed algorithms do not assign a zero rate to any UEs, thereby no user is dropped and a minimum QoS is warranted. As we mentioned before, an eNB allocates the majority of the resources to the real-time applications until they reach their utility inflection rate $r_{ij} = b_{ij}$. However, when the total eNB rate exceeds the inflection point rates sum $\sum b_{ij}$ of all real-time applications incumbent in the system, eNB can allot more resources to the delay-tolerant applications

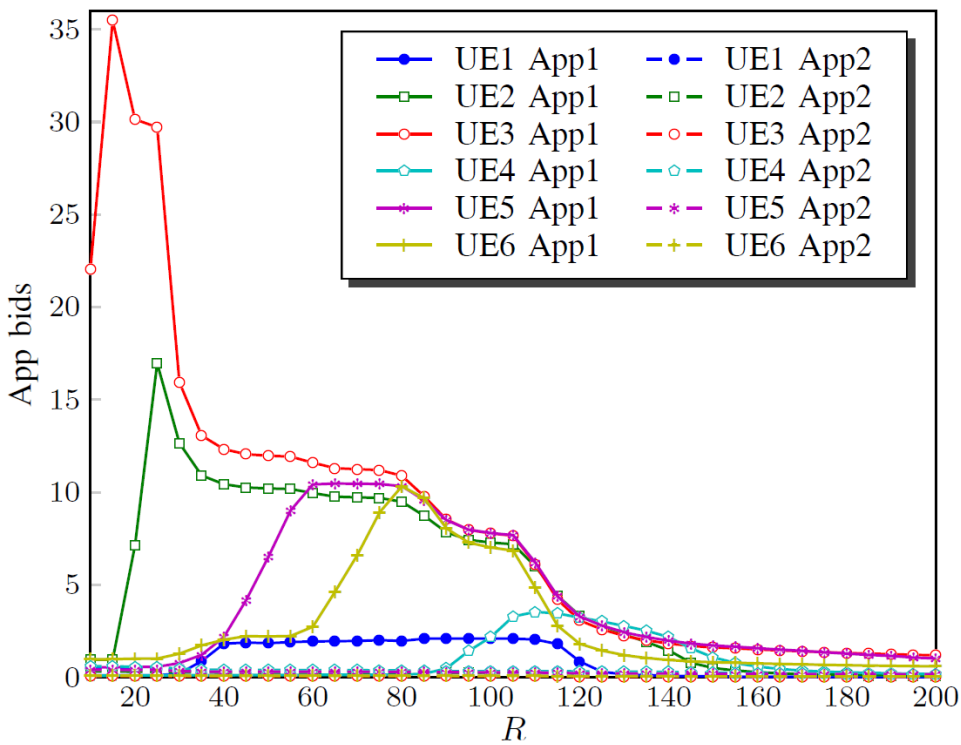


Figure 4.8: Application Bids: This figure illustrates the applications bids in the UEs. The real-time applications bid higher when the resources are scarce, while the opulence of eNB resources escalates the application rates and reduces the UE bids.

with ease of mind. This behavior is observed with the rate increase and bid value plummet that take place after the eNB rate surpasses the inflection points sum, i.e. $R = \sum b_{ij} = 105$, in Figure 4.7.

Furthermore, the improvement in the Algorithms 6 and 7 over the Algorithms 4 and 5 can be observed in the fluctuation reduction of the shadow price depicted in Figure 4.4(b), in which the decay function stabilizes the rate allocation by eliminating oscillations. Such an allocation behavior is similarly seen for Algorithm 6 and 7 over Algorithm 4 and 5 for $R > \sum b_{ij} = 105$, but Algorithm 6 and 7 fails to assign the optimal rates and bids for $R < \sum b_{ij} = 105$. Therefore, Algorithm 6 and 7 is robust under scarce resource availability circumstances.

Next, section 4.5.2 discusses the pricing capability of the proposed distributed resource allocation modus operandi and presents germane simulation results.

4.5.2 Pricing for $10 \leq R \leq 200$

As we explained before, Figures 4.7 and 4.8 show the final rates and bids of different applications with varying eNB bandwidth, and the applications bids are proportional to the allocated rates. For example, the real-time applications (sigmoidal application utility functions) bid higher when the eNB resources are scarce and their bids reduce as R increases. Therefore, the pricing, proportional to the bids, is *traffic-dependent* which outfits service providers with the option to escalate the service price for their subscribers when the traffic load on the system is high. Thereby, service providers can motivate mobile subscribers to utilize the network when the traffic load is low in that they will be paying less for the same services by using the network during off-peak hours.

The shadow price $p(n)$, representing the total price per unit bandwidth for all users and applications, is illustrated in Figure 4.9 when eNB rate changes. As we can observe, the price is high under high-traffic situations, implied by a fixed number of users with less available resources (R is small), and it decreases for low-traffic circumstances when the same number of users have the luxury of more resources (R is large). It is particularly noticeable that large plummets in the shadow price occur after $R = \{15, 25, 85, 105\}$ which are essentially the points at which the rate for one of the real-time application utilities exceeds that of its inflection point. Furthermore, a large decrease is visible at the sum of the inflection points , i.e. $\sum_{i=1}^{k} r_{ij}^{\inf}$. Here, $k = \{1, 2, ..., M\}$ is the users index, M is the number of users, and i is the user with the maximum utility slope $\arg\max_i S_i(r_i)$, in our case user 3 ($b_{3j} = 15$) followed by user 2 ($b_{2j} = 10$) then the three users 1, 5, 6 which have almost the same $S_i(r_i)$ ($b_{1j} = 5, b_{5j} = 25, b_{6j} = 30$), and ultimately user 4 ($b_{4j} = 20$). The larger the difference between slopes $\Delta S_{ij} = |S_i(r_i) - S_j(r_j)|$, the higher the change in the shadow price $p(n)$ plot vs. R.

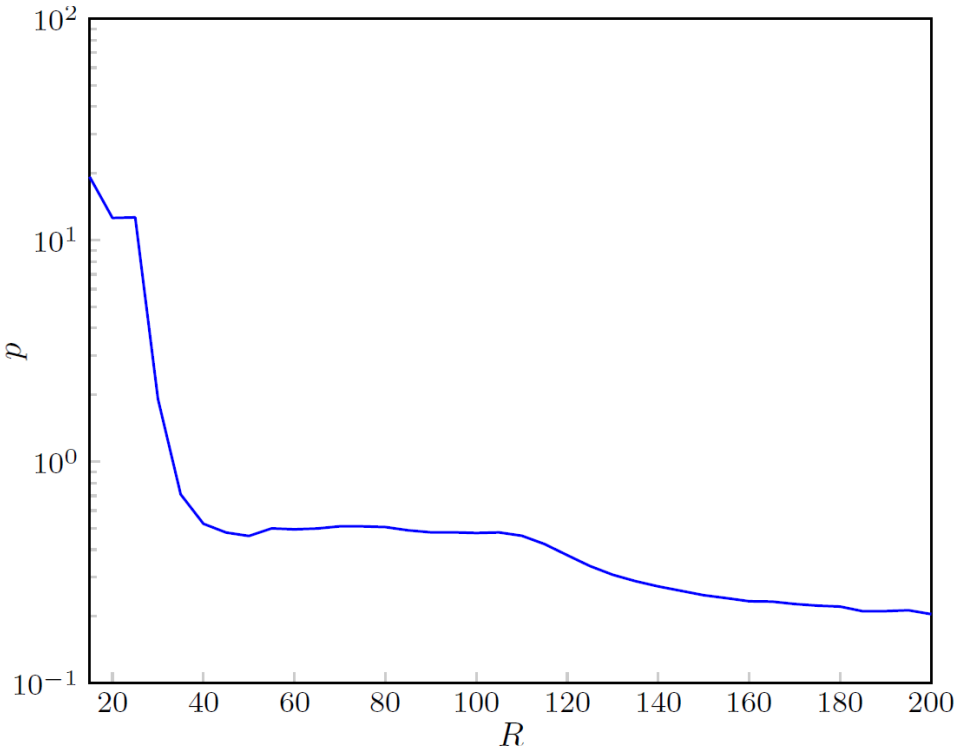


Figure 4.9: Shadow Price p vs. eNB Resources R: Availability of more eNB resources reduces the shadow price.

4.6 Benchmark

In order to evaluate the effectiveness of our proposed resource allocation modus operandi, we look at another proportional fairness method for resource allocation [166]. The authors in this work used the same optimization as we did as in equation 4.39, but they used weighted logarithm functions instead of sigmoidal utility functions. The weighted logarithm function is provided in equation (4.40). The authors proposed using the weight factor w_i to fit the logarithm to sigmoidal traffic. The reason that the authors [166] used weighted logarithm functions is their effort to use convexity in the optimization. The authors proposed that the weights can be adjusted using measurements or curve-fitting.

$$\begin{aligned} \max_{\mathbf{r}} \quad & \prod_{i=1}^{M} \left(\prod_{j=1}^{N_i} U_{ij}^{\alpha_{ij}}(r_{ij}) \right)^{\beta_i} \\ \text{subject to} \quad & \sum_{i=1}^{M} \sum_{j=1}^{N_i} r_{ij} \leq R, \\ & r_{ij} \geq 0, \quad i = 1, 2, ..., M, \quad j = 1, 2, ..., N_i \end{aligned} \tag{4.39}$$

In order to find the wights for the logarithms modeling sigmoidal utilities which model the realtime traffic QS themselves, we used LevenbergMarquardt algorithm (LMA) [167, 168], whose primary application is in the least squares curve fitting problem . Basically, for n datum pairs of independent and dependent variables, (x_i, y_i), LMA optimizes parameters β of the curve $f(x; \beta)$ such that it minimizes $\sum_{i=1}^{n}(y_i - f(x_i; \beta))$.

$$U_i(r_i) = \chi_i log(r_i) \tag{4.40}$$

In order to make the comparison, we used the same logarithmic and sigmoidal utilities as in section 4.5; however, we curve-fitted the sigmoidal utility functions using the LMA to the weighted logarithm function in equation (4.40). The results are shown in Figure 4.10, from which we can see the resources allocated to the realtime applications through our method by the dark blue bar, the resources allocated to the realtime applications by the method in [166] as the light blue bar, the resources allocated to the delay tolerant applications by our method by the yellow bar, and the resources allocated to delay-tolerant traffic by the brown bar. As we can observe from this normalized plot, the method in [166] assigned at least 10% less resources than did our method to the realtime applications. On the other hand, it allocated more resources to the delay-tolerant traffic. This is in spite of the real-time applications need for higher amount of resources as opposed to the delay tolerant-traffic. In fact, the delay-tolerant traffic QoS will not be hurt under lack of enough resources while those of the real-time traffic will be.

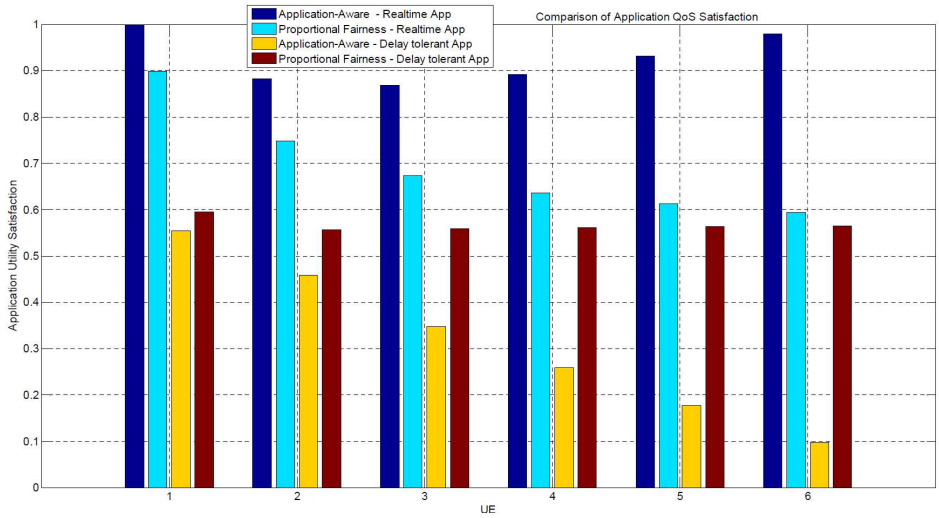


Figure 4.10: Comparison of the resources allocated by our proposed method and sigmoidal-fitted logarithm.

This shows that our proposed method is application-aware in that it looks at the traffic type and it is more cognizant of the QoS requirements of the traffic because it assigns more resources to real-time applications while it allocates less resources to the delay-tolerant applications.

4.7 Chapter Summary

In this chapter, we developed distributed architecture for the QoS-minded utility proportional fairness framework for resource allocation for the cells of a cellular communications system that was introduced in chapter 3. The distributed architecture was composed of a EURA optimization which allocated the UE rates by the eNB and an IURA optimization which assigned application rates by the UEs. Not only did we prove that the proposed distributed resource allocation architecture's EURA and IURA optimization problems are convex and solved them through the Lagrangian of their dual problem, but also we proved the optimality of the rate assignments. Moreover, we proved that the distributed architecture does not lead to optimal rates for all resource availability situations at the eNB. So,

we introduced a variation of the distributed resource allocation which added robustness into the algorithms via decay functions. Furthermore, we proved that the distributed resource allocation architecture that was developed in this chapter is mathematically equivalent to the centralized resource allocation architecture that was introduced in chapter 3. Ultimately, we performed simulations in MATLAB to show the application of the proposed distributed resource allocation architecture to a cellular communications system.

Chapter 5

Traffic and Sensitivity Analysis of Resource Allocation Architectures

Resource allocation methods which aim at fulfilling the QoS requirements of current-day cellular networks have been the focus of many research studies. This attention to QoS-minded resource allocation techniques is partly because present-day cellular communication systems contain smart phones capable of running several applications simultaneously. Since the applications have miscellaneous QoS requirements based on the type of the traffic that they deal with, many modern resource allocation schemes incorporate the application QoS requirements into their operation. Furthermore, a wide variety of subscribers use the current communication systems; as such subscriber types have been included in many of the current resource allocation techniques. Besides, concurrent running of the applications on the smart phones motivates including the application usage temporal changes into novel resource allocation schemes. Much of the state of the art in single carrier resource allocation ([92, 94–96]) considers either the traffic nature or the subscriber type; However, these works do not address whether the allocated rates are optimal and do not account for all the aforementioned QoS-related issues simultaneously. In addition, many research studies

developing a multi-carrier resource allocation ([66,97,98,169]) do not attend to the aforesaid QoS-related concerns, i.e. subscriber type, traffic nature, and traffic dynamics concurrently.

We developed a single-carrier radio resource allocation formulation cast under a utility proportional fairness framework in chapter 3. We proved the optimality of the resource allocation method formulation and incorporated the traffic nature, subscriber type, and application usage percentages in the formulation. Furthermore, we developed the formulation into a centralized and a distributed architecture in chapters 3 and 4, respectively. This chapter complements the centralized and distributed optimal resource allocation that we developed in the aforementioned chapters by presenting a traffic and sensitivity mathematical analysis of the architectures. The traffic analysis includes a transmission overhead analysis of the centralized and distributed resource allocation architectures and derives lower bounds for the transmission overhead. The sensitivity analysis studies whether the UE pledged bids and their allocated rates remain optimal when the number of UEs or their application usage percentages temporally change in the system. Moreover, the sensitivity analysis investigates the variations incurred for the transmission overhead of the centralized and distributed architectures when the UE quantity and/or application usage percentages change in the system. In particular, the aforementioned changes of the transmission overhead is investigated when the UEs rebid or do not rebid in their response to the echosystem changes imposed by the altered UE quantity and/or application usage. Besides, to-be-derived predicates germane to the traffic and sensitivity analyses are confirmed through appropriate simulations.

5.0.1 Contributions

The contributions of this chapter are summarized below.

- We analyze the sensitivity of the centralized optimal resource allocations that we have developed in chapter 3 to temporal variations incurred in the UE quantity or in the

application usage percentages in the system.

- We analyze the sensitivity of the distributed optimal resource allocations that we have developed in 4 to temporal variations incurred in the UE quantity or in the application usage percentages in the system for the cases where all UEs rebid or do not rebid for resources in the face of the imposed system dynamics.
- We analyze the transmission overhead associated with the centralized and distributed optimal resource allocation schemes in chapters 3 and 4 and derive transmission overhead lower bounds.
- We analyze the transmission overhead changes of the centralized and distributed optimal resource allocation schemes in chapters 3 and 4 to temporal variations incurred in the UE quantity or in the application usage percentages in the system for the cases where all UEs rebid or do not rebid for resources in the face of the imposed system dynamics.
- We verify the sensitivity and traffic analyses statements via suitable simulations.

The remainder of this chapter proceeds as follows. Section 5.1 presents a traffic analysis of the centralized and distributed resource allocation architectures that were developed in chapter 3 and 3, respectively. Furthermore, it investigated the sensitivity of the aforementioned architectures to UE quantity dynamics. Then, section 5.2 presents a traffic and sensitivity analysis of the aforementioned centralized and distributed resource allocation architectures to application usage temporal dynamics. Section 5.3 provides with a brief discussion about the computational complexity of the aforementioned centralized and distributed resource allocation architectures. Section 5.4 portrays simulations to depict the traffic/sensitivity analysis results in this chapter. And, section 5.5 concludes the chapter.

Next, section 5.1 develops a traffic analysis of the centralized and distributed resource allocation architectures that were developed in chapters 2 and 3, respectively.

5.1 Traffic/Sensitivity Analysis under UE Quantity Dynamics

We consider a cellular communications systems with M UEs served by a base station (BS) in Figure 5.1. The UEs run real-time and delay-tolerant applications whose QoS is modeled by application utility functions. The application utility function $U(r)$ represents the QoS satisfaction of the application, represented by $U(r)$, in terms of the application allocated rate r.

Figure 5.1: System Model: Single cell of the cellular network with a base station (BS) covering $M = 6$ UEs simultaneously running delay-tolerant and relay-time applications represented by logarithmic and sigmoidal utility functions respectively.

In the centralized architecture, UE application rates are directly assigned by the BS in a single stage. The centralized architecture can be formulated as equation (5.1). Here, M UEs are covered by a BS and $\mathbf{r} = [r_1, r_2, ..., r_M]$ is the rate vector, whose i^{th} component is the rate allocated to the ith UE. Besides, r_{ij}, $U_{ij}(r_{ij})$, and α_{ij} respectively represent the rate allocation, application utility function, and application usage percentage of the jth application running on the ith UE. Hence, we can write $\sum_{j=1}^{N_i} \alpha_{ij} = 1$ which states that the addition of the usage percentages of the applications running on the ith UE is 100%. Furthermore, we can write $r_i = \sum_{j=1}^{N_i} r_{ij}$ where N_i is the number of applications coevally running on the ith UE and the equation implies that the ith UE rate is the addition of the rates for its N_i applications. R is the maximum bandwidth available to the BS and β_i is a subscription-dependent weight for the ith UE. Regarding the centralized architecture, we

proved the following in chapter 3.

- The centralized resource allocation in equation (5.1) is convex and has a tractable solution.
- The resource allocation in equation (5.1) refrains from dropping UEs, prioritizes real-time applications, considers application usage variations, and accounts for UE priorities.
- The centralized resource allocation in equation (5.1) is solved via algorithms 9 and 10 implemented in the UEs and BS.

$$\begin{aligned} \max_{\mathbf{r}} \quad & \prod_{i=1}^{M} \left(\prod_{j=1}^{N_i} U_{ij}^{\alpha_{ij}}(r_{ij}) \right)^{\beta_i} \\ \text{subject to} \quad & \sum_{i=1}^{M} \sum_{j=1}^{N_i} r_{ij} \leq R, \\ & r_{ij} \geq 0, \quad i = 1, 2, ..., M, \quad j = 1, 2, ..., N_i \end{aligned} \tag{5.1}$$

Algorithm 9 UE Centralized Algorithm

loop
  Send application utility parameters $\{a_{ij}, b_{ij}, \alpha_{ij}, k_{ij}, r_{ij}^{\max}\}$ to BS.
  Receive rates $r_i^{\text{opt}} = \{r_{i1}^{\text{opt}}, r_{i2}^{\text{opt}}, ..., r_{iN_i}^{\text{opt}}\}$ from the BS.
  Allocate rate r_{ij}^{opt} internally to j^{th} applications.
end loop

Algorithm 10 BS Centralized Algorithm

loop
  Receive application utility parameters $\{a_{ij}, b_{ij}, \alpha_{ij}, k_{ij}, r_{ij}^{\max}\}$ from UEs.
  Solve $\mathbf{r} = \arg\max_{\mathbf{r}} \sum_{i=1}^{M} \beta_i \sum_{j=1}^{N_i} \alpha_{ij} \log U_{ij}(r_{ij}) - p(\sum_{i=1}^{M} \sum_{j=1}^{N_i} r_{ij} - R)$. {where $\mathbf{r} = \{r_1, r_2, ..., r_M\}$ and $r_i = \{r_{i1}, r_{i2}, ..., r_{iN_i}\}$}
  Send $r_i = \{r_{i1}, r_{i2}, ..., r_{iN_i}\}$ to i^{th} UE.
end loop

In the distributed architecture, there are two optimization problems; the first optimization allocates optimal UE rates by the BS via signaling between the BS and its covered UEs. This optimization, referred to as EURA, is written as equation (5.2). Here, for M UEs, $V_i(r_i) = \prod_{j=1}^{N_i} U_{ij}^{\alpha_{ij}}(r_{ij})$ is the aggregate utility function for the ith UE, and $\mathbf{r} = [r_1, r_2, ..., r_M]$ is the UE rate vector whose i^{th} component is the bandwidth assigned by the BS to the ith UE. On the other hand, the second optimization focuses on distributing the application rates by the UEs hosting the applications and it is performed internally to the UEs. This optimization, denoted as the IURA, is written in equation (5.3). Here, $\mathbf{r}_i = [r_{i1}, r_{i2}, ..., r_{iN_i}]$ is the application rate vector whose j^{th} component indicates the bandwidth allocated by the ith UE to its jth application, r_i^{opt} is rate allocated to the ith UE by the BS via solving the EURA optimization in equation (5.2), and N_i is the number of applications running on the ith UE. Regarding the distributed architecture, we proved the following in chapter 4.

- The distributed resource allocation architecture's EURA optimization in equation (5.2) and IURA optimization in equation (5.3) are convex and have tractable solutions.
- The distributed resource allocation architecture's EURA optimization in equation (5.2) and IURA optimization in equation (5.3) refrain from dropping UEs. IURA prioritize real-time applications over delay-tolerant ones and considers application usage changes. EURA accounts for subscriber differentiations.
- The distributed resource allocation architecture in equations in equations (5.2) and (5.3) are solved via algorithms 11, 12, and 13 implemented in the UEs and BS.

$$\begin{aligned} \max_{\mathbf{r}} \quad & \prod_{i=1}^{M} V_i^{\beta_i}(r_i) \\ \text{subject to} \quad & \sum_{i=1}^{M} r_i \leq R, \\ & r_i \geq 0, \qquad i = 1, 2, ..., M. \end{aligned} \tag{5.2}$$

$$\begin{aligned} \max_{\mathbf{r}_i} \quad & \prod_{j=1}^{N_i} U_{ij}^{\alpha_{ij}}(r_{ij}) \\ \text{subject to} \quad & \sum_{j=1}^{N_i} r_{ij} \leq r_i^{\text{opt}}, \\ & r_{ij} \geq 0, \quad j = 1, 2, ..., N_i. \end{aligned} \tag{5.3}$$

Algorithm 11 UE EURA Algorithm

```
Send initial bid w_i(1) to eNB.
loop
  Receive shadow price p(n) from BS.
  if STOP from eNodeB then
    Calculate allocated rate r_ij^opt = w_i(n)/p(n).
  else
    Solve r_i(n) = arg max_{r_i} ( β_i log V_i(r_i) − p_E(n) r_i ).
    Calculate new bid w_i(n) = p(n) r_i(n).
    if |w_i(n) − w_i(n − 1)| > Δw(n) then
      w_i(n) = w_i(n − 1) + sign(w_i(n) − w_i(n − 1)) Δw(n) {Δw = l_1 e^(−n/l_2) or Δw = l_3/n}
    end if
    Send new bid w_i(n) to eNB.
  end if
end loop
```

Algorithm 12 BS EURA Algorithm

```
loop
  Receive bids w_i(n) from UEs. {Let w_i(0) = 1 ∀i}
  if |w_i(n) − w_i(n − 1)| < δ ∀i then
    STOP and allocate rates (i.e r_i^opt to user i)
  else
    Calculate p_E(n) = (Σ_{i=1}^{M} w_i(n)) / R
    Send new shadow price p_E(n) to all UEs
  end if
end loop
```

For the centralized resource allocation architecture in equation (5.1), the following proposition is conceivable.

Proposition 5.1.1. *The minimum transmission overhead of the centralized resource allocation in equation (5.1) is* $2M$.

Algorithm 13 UE IURA Algorithm

loop
  Receive r_i^{opt} from BS. {by EURA Algorithms}
  Solve
  $\mathbf{r}_i = \arg\max_{\mathbf{r}_i} \sum_{j=1}^{N_i} (\alpha_{ij} \log U_{ij}(r_{ij}) - p_I r_{ij}) + p_I r_i^{\text{opt}}$ $\{\mathbf{r}_i = \{r_{i1}, r_{i2}, ..., r_{iN_i}\}\}$
  Allocate r_{ij} to the jth application.
end loop

Proof. The centralized architecture solution algorithm assigns an application rate in a single iteration. Therefore, M transmissions of the application utility parameters from the M UEs in Figure 5.1 to their BS proceeds with M transmissions of the optimal rates (the rate optimality of the centralized architecture in equation (5.1) is already proved in chapter 3) from the BS to the M UEs. Hence, the the centralized resource allocation architecture's solution algorithm incurs a transmission overhead of $2M$. This is also the minimum possible number of transmissions to permit initiating connections by the M UEs and assigning the rates to them by the BS. □

Now we look at the UE quantity dynamics for both the distributed and centralized resource allocation architectures. Consider a scenario where the number of UEs in the system changes from M_1 at time slot n_1 to M_2 at time slot n_1+1. Therefore, the instantaneous UE quantity $M(n)$ at time slot n can be written as equation (5.4).

$$M(n) = \begin{cases} M_1 \ ; \ n \leq n_1 \\ M_2 \ ; \ n > n_1 \end{cases} \tag{5.4}$$

We assume that steady state rates are reached at the time slots n_1 and n_2 for the M_1 and M_2 users respectively so that the steady state rate vector $\mathbf{r}^{\text{ss}}$ can be expressed as equation (5.5).

$$\mathbf{r}^{\text{ss}} = \begin{cases} \mathbf{r}(n_1) \ ; \ n \leq n_1, \\ \mathbf{r}(n_2) \ ; \ n > n_1. \end{cases} \tag{5.5}$$

Considering the instantaneous number of UEs as equation (5.4) and the steady state rate vector as equation (5.5), we perform a traffic analysis of the distributed resource allocation architecture's solution algorithm when the UEs rebid for resources in the face of the echosystem changes induces by the UE quantity alterations in section 5.1.0.1. Also, we perform a similar analysis when UEs do not rebid for resources in the face of the UE quantity changes within the cellular system in section 5.1.0.2.

5.1.0.1 Distributed Architecture with Rebidding

For the cell with M UEs as in Figure 5.1, we assume that incumbent UEs (users already in the cell and have been assigned optimal rates) rebid for resources when new UEs enter/leave the cell. In this case, proposition 5.1.2 is considerable for the transmission overhead.

Proposition 5.1.2. *When UEs rebid for resources in the face of UE quantity dynamics, the transmission overhead of the distributed resource allocation architecture's solution algorithm is maximum and greater than or equal to* $2M_2 + 2 - M_1$.

Proof. When new UEs enter the cell, $M_2 > M_1$. The new UEs send their initial bids individually requiring $M_2 - M_1$ transmissions to the BS. If β_i is available at the UEs, the BS broadcasts a new shadow price to the UEs which accumulates the number of transmissions to be $M_2 - M_1 + 1$. Next, the M_2 UEs transmit their new bids to the BS which broadcasts another shadow price. This process is repeated k times until the convergence occurs, i.e. optimal rates are assigned to the UEs. This amounts to the transmission overhead $(M_2 - M_1) + 1 + kM_2 + k = (k+1)M_2 + k + 1 - M_1$; this corresponds to the transmission overhead to be minimally $2M_2 + 2 - M_1$ for one iteration ($k = 1$). On the other hand, if β_i is only available at

the BS, it transmits a modified shadow price $\frac{p_E}{\beta_i}$ to the UEs under its coverage. Thus, the transmission overhead will be $(M_2 - M_1) + M_2 + 2kM_2 = (2k+2)M_2 - M_1$, corresponding minimally to $4M_2 - M_1$ for one iteration ($k = 1$). Obviously, $4M_2 - M_1 \leq 2M_2 + 2 - M_1$; hence, the transmission overhead is greater than or equal to $2M_2 + 2 - M_1$ when new UEs enter the system.

When some UEs leave the cell, $M_1 > M_2$. The UEs, leaving the system, send $M_1 - M_2$ transmissions to signal the BS about their service termination request. If β_i is available at the UEs, the BS broadcasts a new shadow price to the UEs which accumulates the number of transmissions to be $M_1 - M_2 + 1$. Next, the M_2 UEs transmit their new bids to the BS which broadcasts another shadow price. This process is repeated k times until the convergence occurs, i.e. optimal rates are assigned to the UEs. This amounts to the transmission overhead $(M_1 - M_2) + 1 + kM_2 + k = (k-1)M_2 + k + 1 + M_1$, corresponding minimally to $2M_2 + 2 - M_1$ for one iteration ($k = 1$). On the other hand, if β_i is only available at the BS, it transmits a modified shadow price $\frac{p_E}{\beta_i}$ to the UEs under its coverage. Thus, the transmission overhead becomes $(M_1 - M_2) + M_2 + 2kM_2 = 2kM_2 + M_1$ transmissions, corresponding minimally to $2M_2 + M_1$ for one iteration ($k = 1$). Obviously, $2M_2 + 2 - M_1 \leq 2M_2 + M_1$ (there is at least one UE in the cell at time n_1, that is $min(M_1) = 1$); hence, the transmission overhead of the distributed approach with rebidding is greater than or equal to $2M_2 + 2 - M_1$ when the UEs leave the system. □

For the distributed scheme with rebidding, the following proposition discusses the dynamics of the allocated rates and pledged bids to the variations incurred in the UE quantity.

Proposition 5.1.3. *When UEs rebid for resources in the face of UE quantity dynamics, the allocated rates and pledged bids of the distributed resource allocation architecture's solution algorithm remain optimal.*

Proof. Without loss of generality, we assume that the number of UEs change from M_1 at

time slot n_1 to M_2 at time slot n_2 and represent this dynamic as $M_1 \rightarrow M_2$. We also presume that the distributed resource allocation architecture's solution algorithm arrives at the steady state rate vector as in equation (5.5). We proved in chapter 4 that the EURA and IURA optimization problems with $M = M_1$ (steady state) are optimal; therefore, the bids and rates for the M_1 UEs before the time slot n_1 can be written as below according to chapter 4.

$$\begin{aligned} r_i(n_1) &= S_i^{-1}(p(n_1)) \\ w_i(n_1) &= p(n_1)S_i^{-1}(p(n_1)) \end{aligned} \tag{5.6}$$

Here, $p(n_1) = \frac{\sum_{i=1}^{M_1} w_i(n_1)}{R}$ and the rate reallocations are done in accordance UE EURA, BS EURA, and UE IURA algorithms with initial bids $w_i(1) = 1$. Since the rates arrive at the steady state at time slot n_2, in accordance with chapter 4, the EURA and IURA optimizations with the M_2 UEs are also optimal and the allocated rates and pledged bids become optimal. Therefore, when the UEs rebid for resources once the UE quantity changes (from M_1) to M_2 UEs before time slot n_2, the optimal rates and bids can be written as equation (5.7) based on chapter 4. Here, $p(n_2) = \frac{\sum_{i=1}^{M_2} w_i(n_2)}{R}$ and the rate reallocations are done in accordance with UE EURA, BS EURA, and UE IURA algorithms with initial bids $w_i(n_1)$.

$$\begin{aligned} r_i(n_2) &= S_i^{-1}(p(n_2)), \\ w_i(n_2) &= p(n_2)S_i^{-1}(p(n_2)), \end{aligned} \tag{5.7}$$

Therefore, the optimal rates and pledged bids remain optimal for the distributed resource allocation architecture's solution algorithm when the UEs rebid for resources in the face of the changes in the the number of UEs. □

Next, section 5.1.0.2 performs a traffic analysis of the distributed resource allocation ar-

chitecture's solution algorithm when the UEs do not rebid for resources in the face of the echosystem changes induces by the UE quantity alterations.

5.1.0.2 Distributed Architecture without Rebidding

For the cell with M UEs as in Figure 5.1, we assume that incumbent UEs (users already in the cell and have been assigned optimal rates) do not rebid for resources when new UEs enter/leave the cell. In this case, proposition 5.1.4 is considerable for the transmission overhead. For every UEs entering/leaving the cell, we assume that prior UEs (i.e. users already in the cell and having been allotted optimal rates) will not rebid for resources so that priori UEs bids do not vary from those in the allocation before time slot $n = n_1$, when UE quantity alters. It is shown in proposition 5.1.4 that the transmission overhead is less than the situation where users rebid for resources.

Proposition 5.1.4. *When UEs do not rebid for resources in the face of UE quantity dynamics, the transmission overhead of the distributed resource allocation architecture's solution algorithm is $M_2 + 1 - M_1$ for $M_2 > M_1$ and $M_1 - M_2$ for $M_1 > M_2$ at minimum.*

Proof. When new UEs enter the cell, $M_2 > M_1$. The new UEs send their initial bids requiring $M_2 - M_1$ transmissions to the BS. If β_i is available at the UEs, the BS broadcasts a new shadow price to the new UEs, which accumulates the number of transmissions to $M_2 - M_1 + 1$. This process is repeated k times until an optimal rate allocation is reached which amounts to the transmission overhead $k(M_2 - M_1) + k$. This corresponds to the minimum $M_2 - M_1 + 1$ for one iteration ($k = 1$). On the other hand, if β_i is only available at the BS, it transmits a modified shadow price $\frac{p_E}{\beta_i}$ to the $M_2 - M_1$ UEs which accumulates to $2(M_2 - M_1)$ transmissions. This process is repeated k times so the transmission overhead becomes $2(M_2 - M_1) + 2k(M_2 - M_1) = (2k + 2)(M_2 - M_1)$ transmissions, corresponding minimally to $4(M_2 - M_1)$ for one iteration ($k = 1$). Obviously, $M_2 - M_1 + 1 \leq 4(M_2 - M_1)$; so the minimum transmission overhead becomes $M_2 + 1 - M_1$ when new UEs enter the cell.

When some UEs leave the cell, $M_1 > M_2$. The UEs, leaving the cell, signal the BS about their service termination requests in $M_1 - M_2$ transmissions. There are no further messages exchanged between the BS and the UEs, so the transmission overhead becomes $M_1 - M_2$. Thus, the minimum transmission overhead becomes $M_1 - M_2$ when some UEs leave the cell. To recap, the minimum transmission overhead is $M_2 - M_1 + 1$ for $M_2 > M_1$ and $M_1 - M_2$ for $M_1 > M_2$. □

It is noticeable that the minimum transmission overhead under a rebidding policy in the face of the UE quantity dynamics is $2M_2 + 2 - M_1$ transmissions - proposition 5.1.2 - which is larger than the minimum transmission overhead when the UEs do not rebid on resources, i.e. $M_1 - M_2$ ($M_2 - M_1 + 1$) (proposition 5.1.4) when the UEs leave (enter) the system. Next, lemma 5.1.5 proves that the rates and bids are not optimal for the distributed scheme with no rebidding when the UE quantity changes.

Lemma 5.1.5. *When UEs do not rebid for resources in the face of UE quantity dynamics, the allocated rates and pledged bids of the distributed resource allocation architecture's solution algorithm are not optimal.*

Proof. We assume that the rates and bids arrive at the steady state before time slot n_1 so that the rates and bids for M_1 UEs at time slot n_1 are optimal and can be written according to the equation (5.5) based on chapter 4.

$$\begin{aligned} r_i(n_1) &= S_i^{-1}(p(n_1)), \\ w_i(n_1) &= p(n_1)S_i^{-1}(p(n_1)), \end{aligned} \tag{5.8}$$

Here, $p(n_1) = \frac{\sum_{i=1}^{M_1} w_i(n_1)}{R}$ and the rate allocations are done in accordance with UE EURA, BS EURA, and UE IURA algorithms with initial bids $w_i(1) = 1$. When new UEs enter the cell, i.e. $M_2 > M_1$, incumbent UE bids ($i \in \{1, 2, ..., M_1\}$) remain fixed for time slots $n > n_1$ and, based on chapter 4, the shadow price can be written as:

$$
\begin{aligned}
p(n > n_1) &= \frac{\sum_{i=1}^{M_1} w_i(n_1) + \sum_{i=M_1+1}^{M_2} w_i(n > n_1)}{R}, \\
p(n > n_1) &= p(n_1) + \frac{\sum_{i=M_1+1}^{M_2} w_i(n > n_1)}{R}.
\end{aligned}
\tag{5.9}
$$

Furthermore, similarly to the equation (according to equation (5.5), the UE pledged bids can be expressed as equation (5.10).

$$
w_i(n > n_1) = \begin{cases} p(n_1) S_i^{-1}(p(n_1)) \ ; \ i = \{1, 2, ..., M_1\} \\ p(n > n_1) S_i^{-1}(p(n > n_1)) \\ \qquad ; \ i = \{M_1 + 1, M_1 + 2, ..., M_2\} \end{cases}
\tag{5.10}
$$

Assuming that rates and bids converge before time slot n_2 when there are M_2 UEs in the system, based on chapter 4, the rates of the M_2 UEs follow equation (5.5) as they reach optimal values. Therefore, the optimal rates and pledged bids at time slot n_2 can be written as equation (5.11).

$$
\begin{aligned}
w_i(n_2) &= \begin{cases} p(n_1) S_i^{-1}(p(n_1)) \ ; \ i = \{1, 2, ..., M_1\} \\ p(n_2) S_i^{-1}(p(n_2)) \ ; \ i = \{M_1 + 1, ..., M_2\} \end{cases} \\
r_i(n_2) &= \begin{cases} \frac{p(n_1)}{p(n_2)} S_i^{-1}(p(n_1)) \ ; \quad i = \{1, 2, ..., M_1\} \\ S_i^{-1}(p(n_2)) \qquad ; \quad i = \{M_1 + 1, ..., M_2\} \end{cases}
\end{aligned}
\tag{5.11}
$$

Here, $p(n_1) = \frac{\sum_{i=1}^{M_1} w_i(n_1)}{R}$ and $p(n_2) = p(n_1) + \frac{\sum_{i=1+M_1}^{M_2} w_i(n_2)}{R}$ and the reallocation is done in accordance with UE EURA, BS EURA, and UE IURA algorithms with the initial bids $w_i(n_1)$. Contrasting the allocated rates and pledged bids at time n_2, i.e. equation (5.11), and the optimal rate allocation and pledged bids, i.e. equation (5.7), we observe the rate reallocation and pledged bids are non-optimal. □

It is noticeable that the rate allocation and pledged bids for the distributed resource allocation architecture do not remain optimal when the incumbent UEs do not rebid for resources in the face of the UE quantity changes. In contrast, the rate allocations and pledged bids remain optimal for the distributed resource allocation architecture when incumbent UEs rebid for resources in response to UE quantity dynamics; however, this is at the expense of the excessive transmission overhead duo to the incumbent UEs rebidding procedures. Next, section 5.1.0.3 discusses the traffic analysis of the centralized resource allocation architecture to the UE quantity changes.

5.1.0.3 Centralized Architecture

A transmission overhead analysis of the centralized resource allocation, assigning optimal rates after one algorithm iteration, is explained in proposition 5.1.6.

Proposition 5.1.6. *The centralized resource allocation architecture's minimum transmission overhead is* $2M_2 - M_1$ *when* $M_2 > M_1$ *and* M_1 *when* $M_2 < M_1$.

Proof. When new UEs enter the cell, $M_2 > M_1$. The new UEs send their utility function parameters in $M_2 - M_1$ transmissions to the BS which returns the application rates for the M_2 UEs. Thus, the transmission overhead becomes $2M_2 - M_1$. On the other hand, when some UEs leave the cell, $M_1 > M_2$. The UEs, leaving the cell, signal their BS with their service termination requests requiring $M_1 - M_2$ transmissions to the BS. The BS returns the rates for the new M_2 UEs. There are no further transmissions and the transmission overhead becomes $M_1 - M_2 + M_2 = M_1$. □

Next, section 5.2 presents a traffic analysis of the distributed and centralized resource allocation architectures to temporal application usage changes.

5.2 Traffic/Sensitivity Analysis under Application Usage Dynamics

We consider a cell of M UEs and a BS as in Figure 5.1. At time slot n_1, the ith UE has the aggregate utility function $V_i(r_i)$. When the UEs change their application usage percentage, the aggregate utility functions for the UEs change. In particular, assume that at time slot n_1+1, M' of the M UEs ($M' < M$) change their application usage percentages. Consequently, those UEs aggregate utility functions change. As an example, the ith UE of the M' UEs gets the aggregate utility function $V_i'(r_i)$. Thus, the time-dependent aggregate utility function for the ith UE of the M' UEs at time slot n, denoted as $V_i(r_i, n)$, can be written as equation (5.12).

$$V_i(r_i, n) = \begin{cases} V_i(r_i) \ ; \ n \leq n_1 \\ V_i'(r_i) \ ; \ n > n_1 \end{cases} \tag{5.12}$$

Furthermore, we assume that a steady state rate allocations are achieved for the ith UE with aggregate utility functions $V_i(r_i)$ and $V_i'(r_i)$ at time slots n_1 and n_2, respectively. Therefore, the rate vector can be expressed as equation (5.13).

$$\mathbf{r}^{\text{ss}} = \begin{cases} \mathbf{r}(n_1) \ ; \ n \leq n_1 \\ \mathbf{r}(n_2) \ ; \ n > n_1 \end{cases} \tag{5.13}$$

Like section 5.1, we consider the centralized resource allocation architecture, the distributed resource allocation architecture when UEs rebid for resources in the face of the application usage dynamics, and the distributed architecture when the UEs do not rebid for the resources once the application usage percentages change.

5.2.0.4 Distributed Architecture with Rebidding

For the cell with M UEs as in Figure 5.1, we assume that fixed-utility UEs, i.e. the UEs with unvaried aggregate utility functions, rebid for resources in the face of the changes to other UEs aggregate utility functions, induced by those UEs' altering their application usage percentages. With regard to the transmission overhead, proposition 5.2.1 is considerable.

Proposition 5.2.1. *When application usage percentage of M' of M UEs changes, the transmission overhead of the distributed resource allocation architecture when all UEs rebid for resources is greater than or equal to $M + M' + 2$.*

Proof. The M' UEs that change their application usages cause changes into their aggregate utility functions. These UEs send their initial bids to the BS in M' transmissions. If β_i is available at the UEs, the BS broadcasts a new shadow price to the UEs. Then, the entire M UEs transmit new bids to the BS. This process is repeated k times until optimal rates are achieved, so the transmission overhead becomes $M' + 1 + kM + k$, corresponding minimally $M + M' + 2$ for one iteration ($k = 1$). On the other hand, if β_i is only available at the BS, it transmits a modified shadow price $\frac{p_E}{\beta_i}$ to the UEs under its coverage requiring M transmissions. The M UEs send new bids to the BS and this process is iterated k times until optimal rates are achieved. Thus, the transmission overhead becomes $M' + M + 2kM$ at minimum; this corresponds to $3M + M'$ for one iteration ($k = 1$). Under normal circumstances, the number of UEs in a cell is much large than 1, so it becomes trivial that $M + M' + 2 < 3M + M$. Hence, the minimal transmission overheads is $M + M' + 2$ when the UEs rebid for resources in the face of application usage dynamics. □

Next, proposition 5.2.2 investigates the sensitivity of the distributed resource allocation architecture to the application usage dynamics when all UEs rebid for resources in response to the application usage changes.

Proposition 5.2.2. *When application usage percentage of M' of M UEs changes, the allocated rates and bids of the distributed resource allocation architecture remain optimal when all of the UEs rebid for resources in the face of the application usage changes.*

Proof. In chapter 4, we proved that the EURA and IURA optimizations are optimal for M users. Hence, without loss of generality and assuming that the ith UE aggregate utility is $V_i(r_i)$ and its slope curvature function is $S_i(r_i)$, we can express the ith UE's optimal rates and bids at time slot n_1 as equation (5.14).

$$\begin{aligned} r_i(n_1) &= S_i^{-1}(p(n_1)) \\ w_i(n_1) &= p(n_1)S_i^{-1}(p(n_1)) \end{aligned} \tag{5.14}$$

Here, $p(n_1) = \frac{\sum_{i=1}^{M} w_i(n_1)}{R}$ and the rate allocations are done in accordance with UE EURA, BS EURA, and UE IURA algorithms with initial bids $w_i(1) = 1$. Once M' UEs ($M' < M$) alter the usage percentage of at least one of their applications, their aggregate utility and slope curvature functions change and, with no loss of generality, we can represent them as $V_i'(r_i)$ and $S_i'(r_i)$, respectively. However, $M - M'$ UEs keep their aggregate utilities $V_i(r_i)$ and slope curvature functions $S_i(r_i)$. Thus, the optimal rates and bids at time slot n_2 can be expressed as equation (5.15).

$$w_i(n_2) = \begin{cases} p(n_1)S_i'^{-1}(p(n_2))\ ; & i = \{1, 2, ..., M'\} \\ p(n_2)S_i^{-1}(p(n_2))\ ; & i = \{M'+1, M'+2..., M\} \end{cases}$$

$$r_i(n_2) = \begin{cases} \frac{p(n_1)}{p(n_2)}S_i^{-1}(p(n_1)) & ; \quad i = \{1, 2, ..., M_1\} \\ S_i'^{-1}(p(n_2)) & ; \quad i = \{1, 2, ..., M'\} \\ S_i^{-1}(p(n_2)) & ; \quad i = \{M'+1, M'+2..., M\} \end{cases} \tag{5.15}$$

Here, $p(n_2) = \frac{\sum_{i=1}^{M} w_i(n_2)}{R}$ and the reallocation procedure is done in accordance with UE EURA, BS EURA, and UE IURA algorithms with initial bids $w_i(n_2)$. Contrasting equations (5.15) and (5.7), we observe that the rates and bids at time n_2 are optimal. □

Next, section 5.2.0.5 investigates the transmission overhead and sensitivity of the distributed resource allocation architecture when UEs do not rebid in the face of application usage dynamics.

5.2.0.5 Distributed Architecture without Rebidding

For the cell with M UEs as in Figure 5.1, we assume that fixed-utility UEs , i.e. the UEs with unvaried aggregate utility functions, do not rebid for resources in the face of the changes to other UEs aggregate utility functions, induced by those UEs' altering their application usage percentages. With regard to the transmission overhead, lemma 5.2.3 is considerable.

Proposition 5.2.3. *When application usage percentage of M' of M UEs changes, the transmission overhead of the distributed resource allocation architecture when fixed-utility UEs do not rebid for resources is equals $M' + 1$ at minimum and is not larger than the one with rebidding.*

Proof. M' UEs that change application usage percentages cause alterations in their aggregate utility functions. These UEs send their initial bids to the BS requiring M' transmissions. If β_i is available at the UEs, the BS broadcasts a new shadow price to the UEs, this routine is repeated k times until optimal rates are achieved, and the transmission overhead becomes $kM'+k$, which at minimum corresponds to $M'+1$ transmission for one iteration ($k = 1$). On the other hand, if β_i is only available at the BS, the BS transmits a modified shadow price $\frac{p_E}{\beta_i}$ to the M' UEs, the algorithm is iterated k times until optimal rates are achieved, and the transmission overhead becomes $2M' + 2kM' = (2k + 2)M'$, which at minimum corresponds to $4M'$ transmissions for one iteration ($k = 1$). Since, it is trivial that $M' + 1 < 4M'$, the

minimum transmission overhead for the distributed resource allocation architecture becomes $M' + 1$, when fixed-utility UEs do not rebid for resources in the face of changed application usage within other UEs.

In contrast to proposition 5.2.1, i.e. the minimum transmission overhead $M + M' + 2$, we see $M' + 1 < M + M' + 2$; therefore, the minimum transmission of the distributed resource allocation architecture when fixed-utility UEs do not rebid on resources is less than the one for the distributed resource allocation architecture when all UEs rebid for resource allocation. □

Next, lemma 5.2.4 investigates the sensitivity of the distributed resource allocation architecture to the application usage dynamics when fixed-utility UEs do not rebid for resources under the application usage changes.

Lemma 5.2.4. *When application usage percentage of M' of M UEs changes, the allocated rates and bids of the distributed resource allocation architecture do not remain optimal when the fixed-utility UEs do not rebid for resources in the face of the application usage changes.*

Proof. Assuming that the distributed resource allocation converges before the time slot n_1, the allocated rates and pledged bids at time slot n_1 are optimal according to chapter 4 and they can then be expressed as equation (5.16), where $p(n_1) = \frac{\sum_{i=1}^{M} w_i(n_1)}{R}$. The resource allocation is done using the UE EURA, BS EURA, and UE IURA algorithms with initial bids $w_i(1) = 1$.

$$\begin{aligned} r_i(n_1) &= S_i^{-1}(p(n_1)) \\ w_i(n_1) &= p(n_1) S_i^{-1}(p(n_1)) \end{aligned} \tag{5.16}$$

The ith UE bids ($\{i|i = \{M' + 1, M' + 2, ..., M\}\}$) does not change for time slots $n > n_1$ as it does not rebid for resources. When the application usage percentage of M' of the M UEs changes, we can write the shadow price as equation (5.17).

$$
\begin{aligned}
p(n > n_1) &= \frac{\sum_{i=1}^{M'} w_i(n > n_1) + \sum_{i=1+M'}^{M} w_i(n_1)}{R} \\
p(n > n_1) &= \frac{\sum_{i=1}^{M'} w_i(n > n_1)}{R} + p(n_1)
\end{aligned}
\tag{5.17}
$$

And the UE bids can be expressed as equation (5.18).

$$
w_i(n > n_1) = \begin{cases} p(n > n_1) S_i'^{-1}(p(n > n_1)); i = \{1, ..., M'\} \\ p(n_1) S_i^{-1}(p(n_1)) \quad ; \quad i = \{M' + 1, ..., M\} \end{cases}
\tag{5.18}
$$

Assuming that the distributed resource allocation converges to optimal values before the time slot n_2, final rates and bids can be expressed as equation (5.19), where $p(n_1) = \frac{\sum_{i=1}^{M} w_i(n_1)}{R}$ and $p(n_2) = \frac{\sum_{i=1}^{M'} w_i(n_2)}{R} + p(n_1)$.

$$
\begin{aligned}
w_i(n_2) &= \begin{cases} p(n_1) S_i'^{-1}(p(n_2)); i = \{1, 2, ..., M'\} \\ p(n_1) S_i^{-1}(p(n_1)) \, ; i = \{M' + 1, M' + 2, ..., M\} \end{cases} \\
r_i(n_2) &= \begin{cases} \frac{p(n_2)}{p(n_1)} S_i'^{-1}(p(n_2)) \, ; i = \{1, 2, ..., M'\} \\ S_i^{-1}(p(n_1)) \qquad ; i = \{M' + 1, M' + 2, ..., M\} \end{cases}
\end{aligned}
\tag{5.19}
$$

The procedure for the resource reallocation is in accordance with the UE EURA, BS EURA, and IURA algorithms with initial bids $w_i(n_1)$. Contrasting the rates and bids in equation (5.19) and the optimal rates and bids in equation (5.7), we observe that the rates and pledged bids for the distributed resource allocation architecture do not remain optimal in the face of application usage changes when fixed-utility UEs do not rebid for resources. □

Next, section 5.2.0.6 investigates the sensitivity of the centralized resource allocation architecture to the system dynamics introduced by the application usage changes.

5.2.0.6 Centralized Architecture

For the cell with M UEs as in Figure 5.1, we assume that the resource allocation is done by means of the centralized resource allocation architecture, i.e. equation (5.1). As such the incumbent UEs have been assigned optimal rates because we proved the optimality of the centralized resource allocation in chapter 3). Next, we assume that M' of the M UEs in the system change their application usages and $M' < M$. For the centralized resource allocation architecture, optimal rates are assigned in one iteration of the algorithms, i.e. one transmission to the BS from the UEs and one reception per UE from the BS. Proposition 5.2.5 analyzes the transmission overhead of the centralized resource allocation architecture to the changes introduced into the system by the application usage changes.

Proposition 5.2.5. *When application usage percentage of M' of M UEs changes, the centralized resource allocation architecture assigns optimal rates with a minimum transmission overhead $M' + M$.*

Proof. The M' UEs whose application usages have changed observe variations in their aggregate utility functions as well. They send their application utility function parameters in M' transmissions to their covering BS, which computes and transmits the new rates to the M UEs that it is covering. Therefore, the transmission overhead is $M' + M$. In chapter 3, we proved that the centralized resource allocation architecture is convex; hence, the new rates assigned to the UEs' applications by the BS are optimal. □

Next, section 5.3 provides some insights into the computational complexity associated with the centralized and distributed resource allocation architectures.

5.3 Computational Complexity Considerations

The centralized resource allocation architecture undergoes less computational complexity on the UE side compared to the distributed one inasmuch as the centralized architecture does not include the UE EURA algorithm computations performed in the the distributed architecture. Since such intra-UE computations are eliminated in the centralized resource allocation scheme, the UE power consumption required for implementing the centralized resource allocation method is less than that of the distributed architecture. Next, section 5.4 presents simulation results for the traffic and sensitivity analyses of the distributed and centralized resource allocation architectures.

5.4 Traffic Analysis Simulation

We consider the cell with 6 UEs and a BS as in Figure 5.1. Each UE concurrently runs a delay-tolerant and a real-time application. The application utility function parameters for the UEs are depicted in Table 5.1. The real-time applications, represented by the sigmoidal utility functions, are VoIP ($b = 5$), standard video ($b = 10, 15, 20$), and high definition video ($b = 25, 30$) - chapter 2. The delay-tolerant applications, represented by logarithmic utility functions, are FTP applications - chapter 2.

Next, section 5.4.1 presents the simulation results depicting the transmission overhead of the centralized and distributed resource allocation architectures presented in chapters 3 and 4, respectively.

Table 5.1: Applications Utility Function Parameters - Traffic Analysis

Applications Utilities Parameters	
UE1 App1	Sigmoid $a = 5,\ b = 5$
UE2 App1	Sigmoid $a = 4,\ b = 10$
UE3 App1	Sigmoid $a = 3,\ b = 15$
UE4 App1	Sigmoid $a = 2,\ b = 20$
UE5 App1	Sigmoid $a = 1,\ b = 25$
UE6 App1	Sigmoid $a = 0.5,\ b = 30$
UE1 App2	Logarithmic $k = 15,\ r^{\max} = 100$
UE2 App2	Logarithmic $k = 12,\ r^{\max} = 100$
UE3 App2	Logarithmic $k = 9,\ r^{\max} = 100$
UE4 App2	Logarithmic $k = 6,\ r^{\max} = 100$
UE5 App2	Logarithmic $k = 3,\ r^{\max} = 100$
UE6 App2	Logarithmic $k = 1,\ r^{\max} = 100$

5.4.1 Transmission Overhead Analysis

For the cell with 6 UEs in Figure 5.1, Figure 5.2 compares the transmission overhead incurred by the centralized resource allocation architecture to the one experienced as a result of deploying the distributed resource allocation approach. As we can see, the centralized resource allocation (green curve) incurs a significantly less transmission overhead as opposed to the distributed resource allocation scheme (blue curve). In fact, the transmission overhead of the centralized resource allocation approach is very close to the abscissa whereas the distributed resource allocation transmission overhead is far away from the abscissa. This behavior is in light of the fact that the centralized resource allocation subsumes many less messages exchanged between the UEs and BS and the optimal rate allocation is fulfilled in a single stage.

Next, section 5.4.2 looks at the simulations done to investigate the traffic and sensitiv-

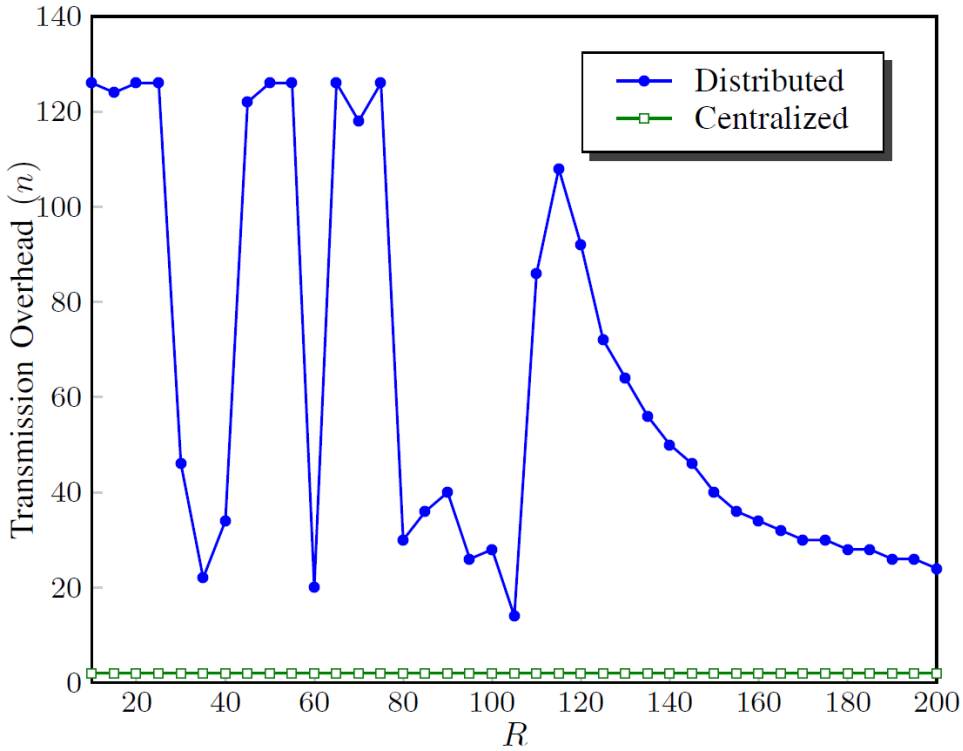


Figure 5.2: Transmission Overhead vs. BS Rate R: The centralized resource allocation scheme incurs a drastically less transmission overhead as opposed to the distributed approach, as it requires less message exchanges before it converges to the optimal rates.

ity analysis predicates, derived in section 5.2, concerning the temporal alterations in the application usage percentages.

5.4.2 Sensitivity to Application Usage Changes

For the cell with 6 UEs in Figure 5.1, we measure the sensitivity of the centralized and distributed resource allocation architectures to the changes incurred in the usage percentage of the applications running on the UEs. Similarly to section 5.4.1, the application utility function parameters are listed in Table 5.1. In the following simulations, we set the termination threshold of the algorithms to $\delta = 10^{-3}$ and the total achievable rate at the BS to $R = 180$.

The application usage percentage is represented as an application-status differentiation

Table 5.2: Application Status Differentiation

Applications Usage-Percentage	
$\boldsymbol{\alpha}_1$	{0.1, 0.5, 0.9, 0.1, 0.5, 0.9, 0.9, 0.5, 0.1, 0.9, 0.5, 0.1}
$\boldsymbol{\alpha}_2$	{0.5, 0.3, 0.2, 0.5, 0.3, 0.2, 0.5, 0.7, 0.8, 0.5, 0.7, 0.8}
$\boldsymbol{\alpha}_3$	{0.5, 0.9, 0.8, 0.5, 0.9, 0.8, 0.5, 0.1, 0.2, 0.5, 0.1, 0.2}
$\boldsymbol{\alpha}_4$	{0.5, 0.3, 0.2, 0.5, 0.3, 0.2, 0.5, 0.7, 0.8, 0.5, 0.7, 0.8}
$\boldsymbol{\alpha}_5$	{0.5, 0.9, 0.8, 0.5, 0.9, 0.8, 0.5, 0.1, 0.2, 0.5, 0.1, 0.2}
$\boldsymbol{\alpha}_a$	{0.1, 0.5, 0.9, 0.1, 0.5, 0.0, 0.9, 0.5, 0.1, 0.9, 0.5, 0.0}
$\boldsymbol{\alpha}_b$	{0.1, 0.5, 0.9, 0.1, 0.5, 0.9, 0.9, 0.5, 0.1, 0.9, 0.5, 0.1}

weight, and the M UEs switch between their applications with the usage percentages according to equation (5.20) whose values are adopted from the first five rows in Table 5.2. As an illustration, during the simulation time $0 < t < 100$, $\boldsymbol{\alpha} = \boldsymbol{\alpha}_1 = \{0.1, 0.5, 0.9, 0.1, 0.5, 0.9, 0.9, 0.5, 0.1, 0.9, 0.5,$ which indicates that the 1st UE utilizes its real-time application 90% of the time duration $\{t|0 < t < 100\}$ and its delay-tolerant application 10% of the entire time interval $\{t|0 < t < 100\}$.

$$\boldsymbol{\alpha}(t) = \begin{cases} \boldsymbol{\alpha}_1 & ; \quad 0 < t \leq 100 \\ \boldsymbol{\alpha}_2 & ; 100 < t \leq 200 \\ \boldsymbol{\alpha}_3 & ; 200 < t \leq 300 \\ \boldsymbol{\alpha}_4 & ; 300 < t \leq 400 \\ \boldsymbol{\alpha}_5 & ; 400 < t \leq 500 \end{cases} \tag{5.20}$$

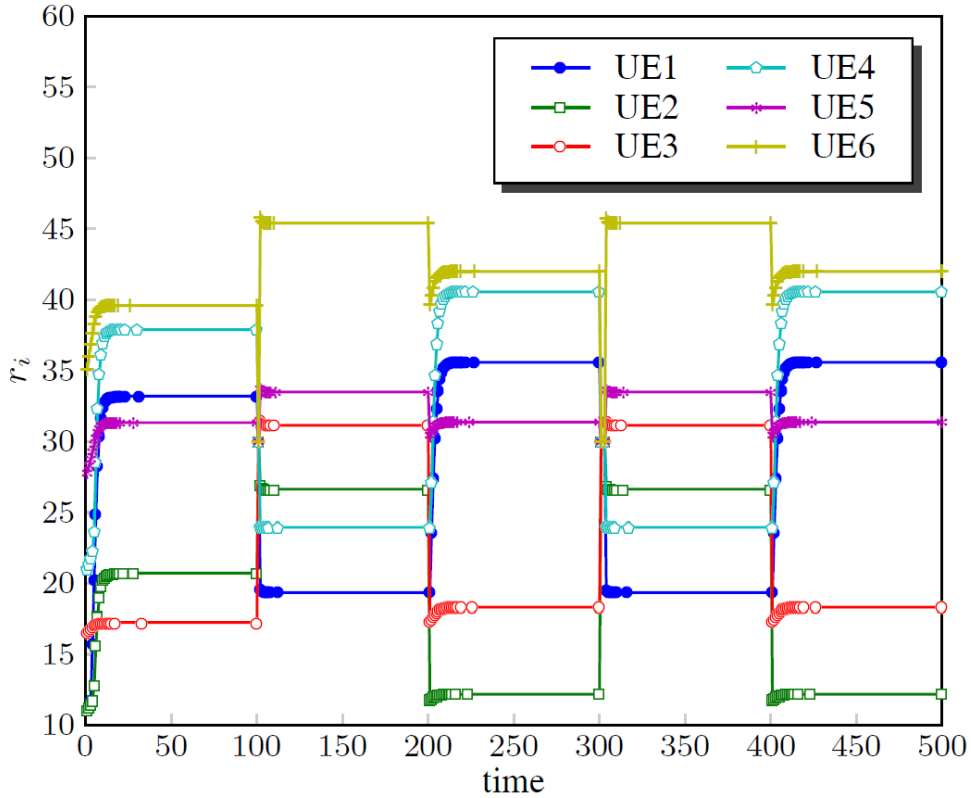
The sensitivity of the centralized and distributed resource allocation architectures is depicted in Figure 5.3. As we can observe from Figure 5.3(a), the UE rates, e.g. r_i for the ith UE, changes in time due to the changing application usage percentages of those UEs in accordance with $\boldsymbol{\alpha}(t)$. During each time interval, the rates converge to optimal ones as the

resource allocation algorithms execute. Furthermore, Figure 5.3(b) portrays the UE pledged bids, e.g. w_i for the ith UE, as the application usage percentages change in accordance with $\boldsymbol{\alpha}(t)$. Like the rates, the bids converge to optimal values when the application usage percentages change. In particular, we observe that the jumps between the bids are larger at the commencement of the resource allocation algorithm, whereas they significantly shrink as time elapses and the rates/bids converge to their optimal values.

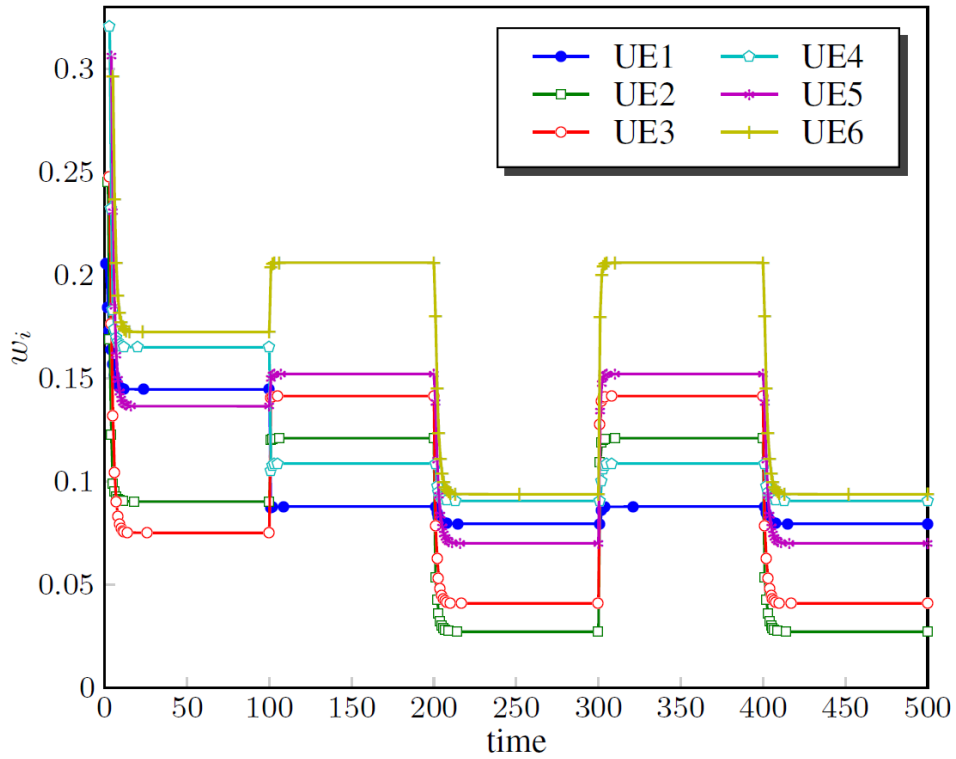
In Figure 5.4, we show the changes in the shadow price p vs. time when the application usage percentages change. Here, we can observe a shrinkage of the shadow price variations as the algorithms execute in time. Furthermore, higher usage corresponds to a higher shadow price. Moreover, a comparison between the centralized and distributed architectures' transmission overhead while the application-usage percentages change is illustrated in Figure 5.5, which shows that the centralized architecture incurs a significantly lower transmission overhead due to the less number of transmissions it needs as opposed to those of the distribution approach.

In fact, the centralized resource allocation's transmission overhead plot is very close to the abscissa. This behavior is because the centralized method has less message exchanges and the transmission overhead is independent of the termination threshold δ. On the flip side, the lower termination thresholds augments the transmission overhead for the distributed resource allocation architecture because the smaller thresholds make the algorithms execute more iterations before the resource allocation scheme converges to optimal rates. More iterations are tantamount to a larger number of messages exchanged between the BS and UEs and escalates the transmission overhead dramatically.

Next, section 5.4.3 discusses the sensitivity of the distributed and centralized resource allocation architectures to the dynamics introduced by the changes in the number of UEs in the system.



(a) UE Rates with Application Usage Dynamics.



(b) UE Bids with Temporally Changing Applications Usages.

Figure 5.3: The UEs allocated rates r_i and pledged bids w_i vary in time as the application usage percentages change. Figures 5.3(a) and 5.3(b) respectively depict the UE rate and bid changes as the applications usages vary in time. The time procession reduces the rate and bid jumps due to the convergence.

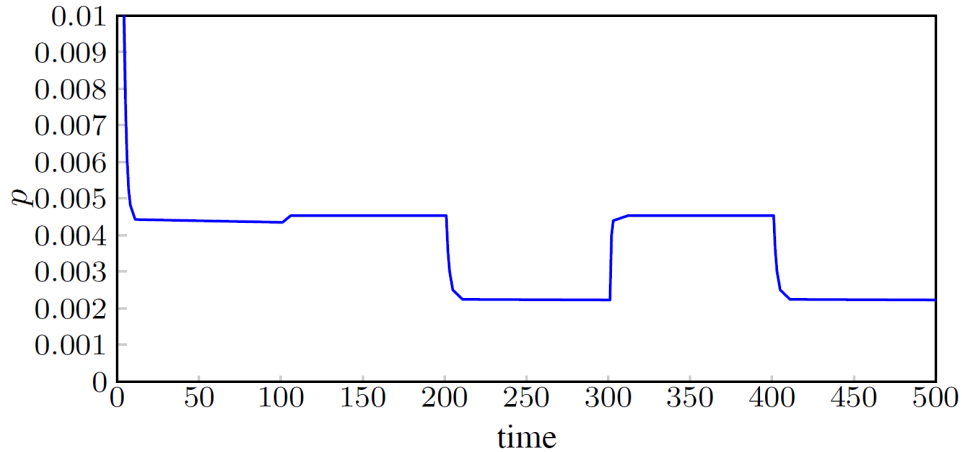


Figure 5.4: The shadow price varies with temporal changes of the application usage percentages. The more usage, the higher the shadow price.

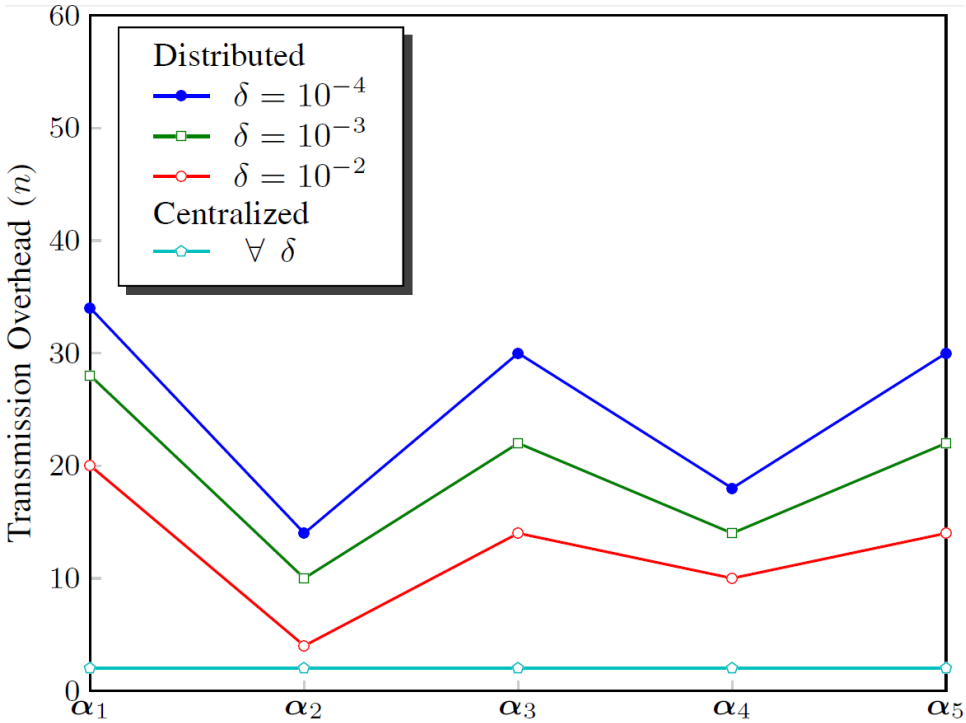


Figure 5.5: Transmission Overhead: The plots compares the centralized and distributed algorithms with respect to their transmission overhead when application usage percentages change. The centralized algorithm is independent of the termination threshold δ, whereas the smaller thresholds causes more iterations of the algorithm leading to escalated transmission overheads.

5.4.3 Sensitivity to UE Quantity Changes

For the cell with 6 UEs in Figure 5.1, we measure the sensitivity of the centralized and distributed resource allocation architectures to the changes incurred in the quantity of UEs

in the cell. Similarly to sections 5.4.1 and 5.4.2, the application utility function parameters are listed in Table 5.1 and we set the termination threshold of the algorithms to $\delta = 10^{-3}$ and the total achievable rate at the BS to $R = 180$. In particular, we measure the sensitivity of the centralized and distributed architectures when new users enter the cell so that the UE quantity at time slot $n_1 = 100$ is $M_1 = 5$ and at time slot $n_1 + 1$ the is $M_2 = 6$. Then, the instantaneous number of users, denoted as $M(n)$, is given by the equation (5.21).

$$M(n) = \begin{cases} M_1 \ ; \ n \leq n_1 \\ M_2 \ ; \ n > n_1 \end{cases} \tag{5.21}$$

Furthermore, we set the temporal application status differentiation weight $\boldsymbol{\alpha}(t)$ as equation (5.22), where $\boldsymbol{\alpha}_a$ and $\boldsymbol{\alpha}_b$ are given in Table 5.2. This means that the temporal usage of the applications in the system is according to the $\boldsymbol{\alpha}_a$ during the time interval $0 < t \leq 100$, while it is according to $\boldsymbol{\alpha}_b$ in the time interval $100 < t \leq 200$.

$$\boldsymbol{\alpha}(t) = \begin{cases} \boldsymbol{\alpha}_a & ; 0 < t \leq 100, \\ \boldsymbol{\alpha}_b & ; 100 < t \leq 200. \end{cases} \tag{5.22}$$

The results for the sensitivity analysis are depicted in Figures 5.6, 5.6, and 5.8. The simulation results show that the distributed resource allocation architecture's assigned rates, pledged bids, and shadow prices deviate from optimal values when the UEs do not rebid for resources in the face of UE quantity changes. On the other hand, the assigned rates, pledged bids, and shadow prices do not deviate from the optimal values for the centralized resource allocation architecture. Figures 5.6, 5.7, and 5.8 respectively show the errors in UE rates $|r_i - r_i^{\text{opt}}|$, in the bids $|w_i - w_i^{\text{opt}}|$, and in the shadow prices $|p - p^{\text{opt}}|$ for the ith UE when a new user (here UE6) enters the cell. Since the centralized architecture retains its optimal UE rates and bids, the errors become 0 for the centralized approach. In contrast, the distributed architecture is concomitant with peak errors before converging to the steady state values. In

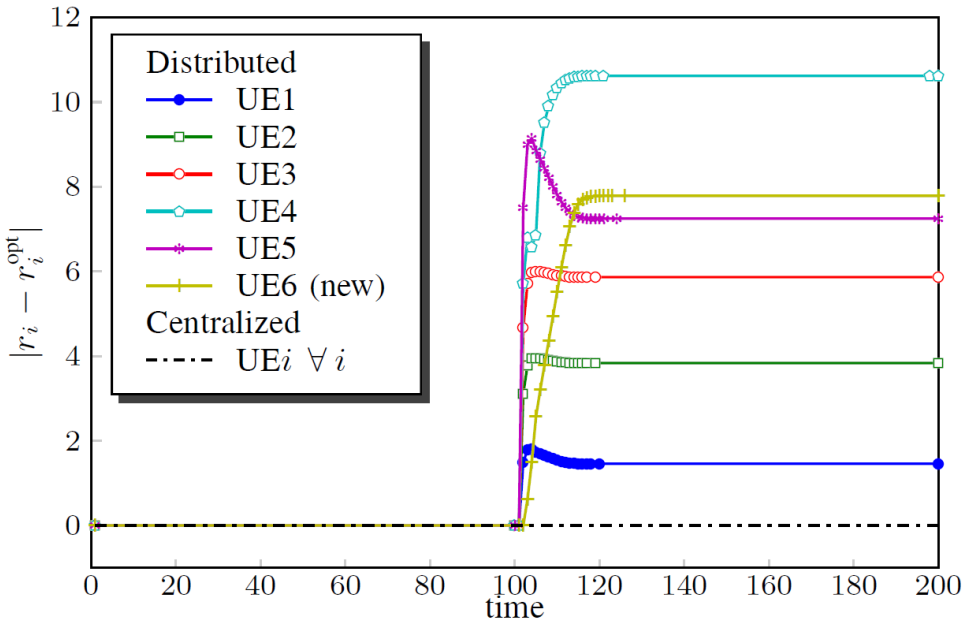


Figure 5.6: Comparisons between the centralized and distributed architectures divergence from optimal solutions when a new users enters the cellular network. This figure depicts the rate error r_i when UE6 enters the network.

addition, looking at the rate and bid errors in Figures 5.6 and 5.7, we observe that the UEs whose real-time applications have higher QoS requirements, tantamount to larger inflection points in their respective sigmoidal application utility functions, undergo larger errors vis-a-vis the other UEs. Last, Figure 5.8 portrays the error in the announced shadow price when UE6 enters the cell. As we can see, the centralized resource allocation architecture boasts a zero pricing error, whereas the distributed one is susceptible to about 30% shadow price error.

Next, section 5.5 summarizes the results that we have developed in this chapter.

5.5 Chapter Summary

In this chapter, we performed a traffic and sensitivity analysis of the centralized and distributed resource allocation architectures that we developed in chapters 3 and 4, respectively.

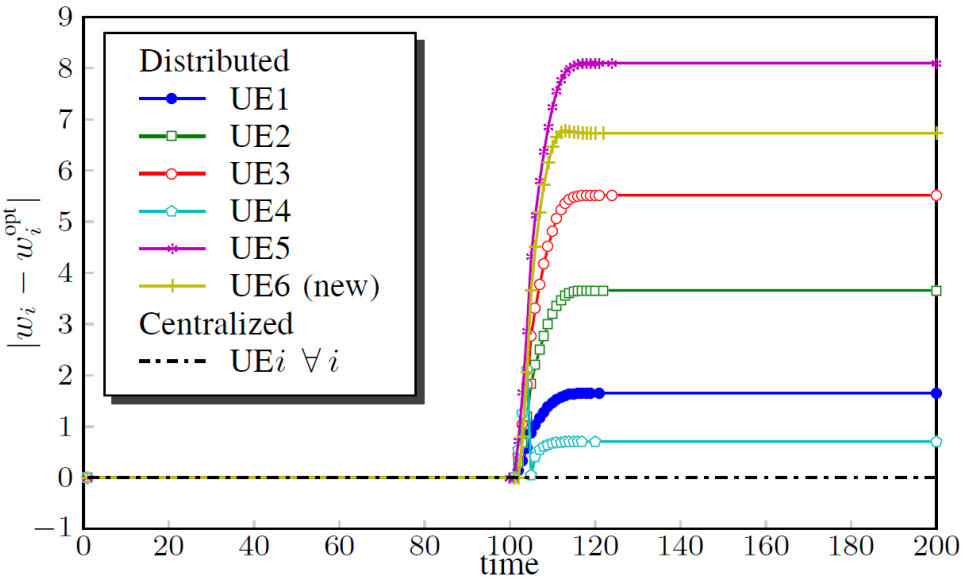


Figure 5.7: UE Bids with Temporally Changing Applications Usages: This figure shows the bid errors when UE6 enters the network and depicts that the rates/bids remain optimal for the centralized architecture.

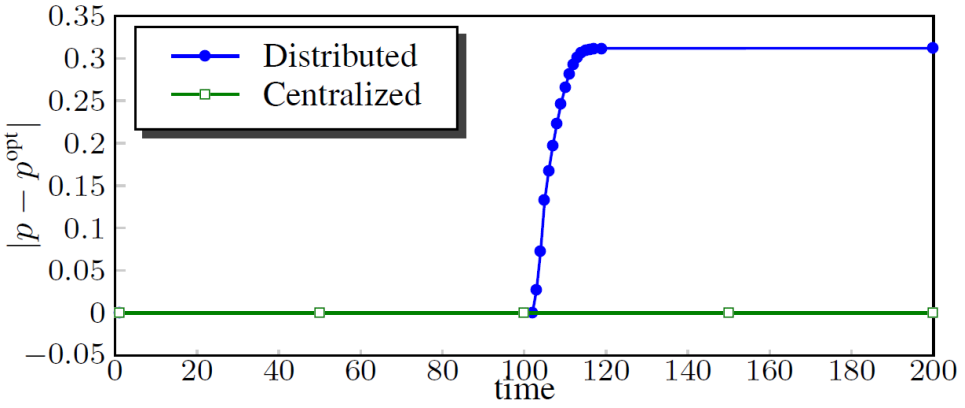


Figure 5.8: UE Bids with Temporally Changing Applications Usages: This figure portrays the shadow price error for the distributed and centralized architectures, where the latter introduces no errors.

First, we the proved that the centralized resource allocation architecture in chapter 3 incurs the minimum possible transmission overhead during the rate assignment procedure. Moreover, we derived the transmission overhead based on changes in the number of UEs in the cell under the existence of the dearth a rebidding policy for the distributed resource allocation architecture developed in chapter 4. We proved that the absence of a rebidding policy entails a transmission overhead less that or equal to that of the one with presence of a rebidding

policy. In addition, we proved that the rate assignments remain optimal for variations in the network's UE quantity for the distributed resource allocation approach under a rebidding policy, whilst the rates are not optimal when the number of UEs change for the distributed approach with a no-rebidding policy and also for the centralized resource allocation scheme. Besides, we proved the rate assignments remain optimal when the application usage percentages changes in case of the centralized resource allocation scheme, but it is not optimal when the aforesaid parameter varies in case of the distributed resource allocation method. We also analyzed the transmission overhead and its changes for both the centralized and distributed method when the application usage percentages vary in the network.

Chapter 6

Radio Resource Block Allocation Architecture

This chapter leverages the proportional fairness resource allocation that we developed in chapter 3 and expands it to include a resource block allocation mechanism for cellular communications networks, where the UEs include delay tolerant and real-time applications generating elastic and inelastic data traffic. The applications QoS is represented as sigmoidal and logarithmic utility functions, and the optimization procedure aims at maximizing the proportional fairness of the utility functions for the cellular environment meanwhile it allocates radio resource blocks (RBs) to the UEs efficiently in terms of the computation time and complexity. The sensitivity of the proposed mechanisms to the variations in total resources is investigated with regard to shadow prices, UE rates and bids, and assigned resource blocks.

6.1 Introduction

During the last few years, the number of mobile subscribers and their traffic has increased enormously at a global level. In accordance with Fig. 6.1, the number of mobile subscribers

in various parts of the world has reached 276 - 1260 million subscriptions (histograms below the name of the region) and these account for an augmentation of 2 - 31 million subscriptions compared with Q1 2013. Furthermore, the total monthly wireless traffic has observed an astronomical growth of 7800% in data and 250% for voice as it is depicted in Fig. 6.1. Simultaneously, the prevalence of smartphones, combining the computational power of personal computers with the mobility of traditional cellphones, enables on-the-go fulfillment of many more user functionalities by leveraging the wealth of information available from the Internet. In fact it is estimated that the number of mobile subscriptions by 2017 will surpass 3700 million users as opposed to only 250 million users in 2008, of which over 3000 million will be smartphone subscribers [3].

In view of such gigantic increments in the quantity of the subscriptions and immense explosion of the network traffic, a significant demand for more mobile broadband service resources is expected. On the other hand, mobile devices run different applications such as mobile TV, multimedia telephony , banking, and many more. The applications generate types of traffic which are very distinctive in their nature [157]. For instance, e-mail applications are delay-tolerant producing elastic traffics whilst real-time services, e.g. VoIP and HDV create inelastic data traffics. This circumstance causes applications to have minimum performance requirements which should be fulfilled in order to provide with a reasonable QoS for the applications and a desirable QoE for their users.

In chapters 3 and 4, we developed radio resource allocation method based on a network utility maximization formulation. The adopted proportional fairness policy allowed for a minimal rate to be allocated to every UE in the system and guaranteed a minimum QoS provision to all services in the network since no user would be dropped. However, the advent and spread of radio resource block (RB)-based communications infrastructures such as LTE [32] mobile broadband services need the dynamic rate allocations be tailored to assigning RBs so as to making a theoretical rate allocation modus operandi pragmatically

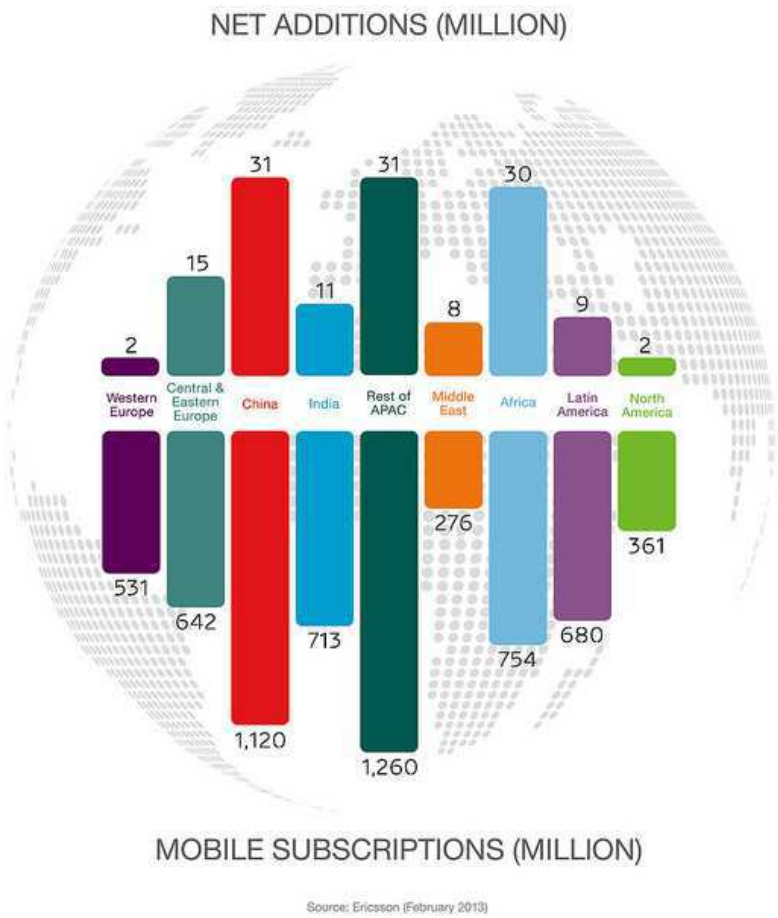


Figure 6.1: Global Mobile Subscriptions between Q1 - Q2 2013 (Adopted from [3]): Lower (Upper) histograms depict the total subscription quantity (subscription augmentation quantity). Such enormous escalations of the number of subscriptions make dynamic resource allocation indispensable.

feasible.

In this chapter, we study the RB allocation by means of the Lagrangian relaxation [170] into a convex continuous rate allocation optimization problem. In particular, we formulate the RB allocation as an integer nonlinear rate allocation optimization, where the positive integer rates represent the RBs which are to be allocated by a cellular network. Employing the Lagrangian relaxation, we obtain and solve the continuous rate allocation equivalent optimization problem by resorting to its dual problem [171]. Then, we utilize the discrete spacial boundary points of the continuous rates in order so as to extract candidates for RBs

to be assigned to the UEs in the network by the base station. Once qualified candidates are selected from the potential RBs based on the constraints on the relaxed continuous rate allocation optimization problem, those candidates which maximize the proportional fairness of the utility function for the cellular network system represent the RBs that can be allocated to the UEs by the eNB. Whilst this approach guarantees a minimum QoS for the UEs, in the light of its proportional fairness policy which never allocates a zero rate to an UE, it is able to efficiently (in time and calculation) allocate the RBs to the UEs in a dynamic fashion.

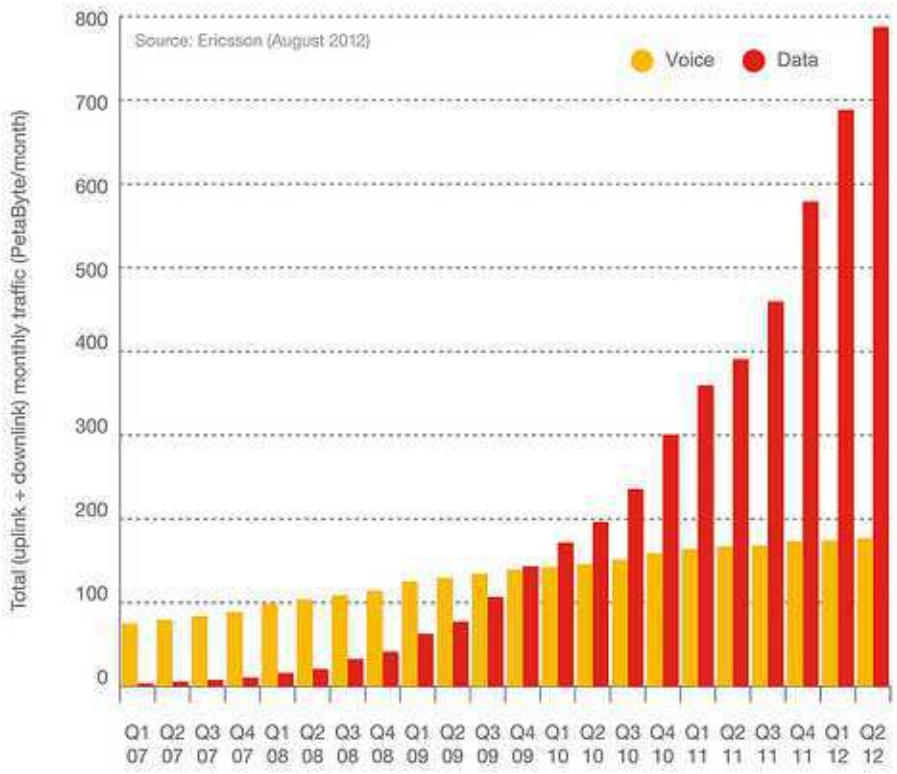


Figure 6.2: Monthly mobile subscribers' traffic has undergone drastic increases between 2007-2012 (Adopted from [3]). Data, as opposed to voice, has the lion's share of the generated traffic increase, so resource allocation strategies with an attention to the type of data traffic is desirable.

In this chapter:

- We introduce an RB allocation scheme for wireless communications systems in which

RBs are assigned to the UEs by maximizing the proportional fairness of utility functions of wireless system, formed from UEs utility function which consist of both sigmoidal and logarithmic utilities to express delay tolerant and real-time applications respectively.

- We demonstrate that the proposed RB allocation scheme is can be transformed into a convex rate allocation problem.
- We explain that our algorithm prioritizes real-time applications above delay-tolerant ones and that it guarantees a minimum QoS by assuring that no user is dropped.
- We show that the problem of RB allocation can be efficiently solved in terms of both computation time and complexity by considering the boundary discrete points as the potential candidates for an assigned RB.

The remainder of this chapter is organized as follows. Section 6.2 presents the mathematical formulation for RB allocation optimization. Chapter 6.3 provides with the solution algorithm for the RB allocation optimization in this chapter. Section 6.4 presents suitable simulations to show the performance of the RB allocation method developed in this chapter. And, section 6.5 concludes this chapter.

Next, section 6.2 presents the system model for the RB allocation study in this chapter.

6.2 Discrete Optimization and Continuous Optimization Relaxation

Without loss of generality, we consider a single cell in a cellular network infrastructure which subsumes a single eNB covering for M UEs. Notationally, the bandwidth allocated by the eNB to the ith UE is given by r_i. Furthermore, each UE has its own utility function $U_i(r_i)$

which indicates the user i^{th} satisfaction percentage for an allocated rate r_i and corresponds to the type of the traffic being generated by the UE at a particular time. Our objective is to determine the RBs (bandwidths units) that the eNB should allocate to the M UEs in our system under a utility proportional fairness regime so that no UE receives a zero bandwidth. Such a policy guarantees a minimum QoS to all UEs based on chapter 3. Like previous chapters, we assume that the application utility functions $U_i(r_i)$ are sigmoidal and logarithmic functions with the equations (6.1) and (6.2), respectively.

$$U_i(r_i) = c_i\left(\frac{1}{1+e^{-a_i(r_i-b_i)}} - d_i\right) \tag{6.1}$$

where $c_i = \frac{1+e^{a_i b_i}}{e^{a_i b_i}}$ and $d_i = \frac{1}{1+e^{a_i b_i}}$. It can be easily verified that $U(0) = 0$ and $\lim_{r_i \to \infty} U(r_i) = 1$, where the latter indicates that a 100% satisfaction is achieved as infinite resources are allocated to the utility function under question. Chapter 2 shows that b_i is the inflection point for the sigmoidal utility function 6.1, after which user is satisfied with the service and before which the QoS for the service is adversely impacted. Since the various parameters a_i, b_i, c_i, and d_i directly impact the shape of the utility function, they are tantamount to distinctive real-time applications with regard to the natures of their generated traffics.

$$U_i(r_i) = \frac{\log(1+k_i r_i)}{\log(1+k_i r_{max})} \tag{6.2}$$

In equation (6.2), r_{max} is the required rate for the user to achieve 100% utility percentage and k_i is the rate at which the utility percentage increases with the augmentation of the allocated rate r_i. Again, it can be easily verified that $U(0) = 0$ and $U(r_{max}) = 1$, where the latter implies that after being provided with a certain rate, the application is receiving a good QoS. Similarly to the sigmoidal utility functions parameters, parameter k_i directly impacts the shape of the logarithmic utility function and represents delay tolerant applications.

In chapter 3, we formulated our utility proportional fairness resource allocation as equation (6.3), where $\mathbf{r} = [r_1, r_2, ..., r_M]$ is a row vector presenting the rates allocated to all the M users and M signifies the number of UEs in the coverage area of the eNB under question. The goal of such a resource allocation objective function is to allocate the eNB resources to each UE such that the allocations maximize the total mobile system objective function, while maintaining the proportional fairness between individual utility functions.

$$\max_{\mathbf{r}} \prod_{i=1}^{M} U_i(r_i) \tag{6.3}$$

In chapter 3, we explained that such a resource allocation formulation ensures non-zero allocations to all UEs as even one zero UE rate will suppress the objective function to zero. Consequently, the aforementioned resource allocation optimization problem guarantees a minimum QoS for all UEs. Furthermore, in chapters 3and 4, we depicted that this approach allots more resources to users who have real-time applications on their devices which in turn leads to residing an inherent remedial to the QoS of wireless system within the structure of the resource allocation methodology. However, since we want to tailor the allocation scheme to wireless technologies such as LTE, which essentially allocate discrete rates as opposed to continuous ones, we consider RB assignments vis-a-vis simple the rate allocations shown in equation (6.3). Besides, we have to account for the fact that an eNB has limited resources R, from which it allocates the RBs to various UEs under its coverage area. Thus, we should modify the formulation in equation (6.3) to consider these two factors as well. The bandwidth available at the eNB can be incorporated as a constraint into the optimization problem in equation (6.3), whereas RB allocations can be indicated by forcing another nonlinear constraint so that all the rates are positive integer numbers. Adding these constraints to the optimization in equation (6.3) renders the resource allocation optimization problem discrete.

The next section presents the formulation for such a discrete RB allocation optimization

which allocates the RBs to different UEs with an eye on the maximal available eNB bandwidth.

6.2.1 Resource Block Optimization Formulation

The basic formulation of the utility proportional fairness RB allocation optimization is given in equation (6.4), where R is the total bandwidth of an eNB which covers the M UEs, and $\mathbf{r} = [r_1, r_2, ..., r_M]$ is the vector of the RBs allocated the UEs such that r_i is the rate which is allocated to the ith UE represented by the application utility function $U_i(r_i)$. Note the addition of a nonlinear constraint to the optimization problem in equation (6.4).

$$\begin{aligned} \max_{\mathbf{r}} \quad & \prod_{i=1}^{M} U_i(r_i) \\ \text{subject to} \quad & \sum_{i=1}^{M} r_i \leq R \\ & r_i \geq 0, \\ & r_i \in \mathbb{N}, \quad i = 1, 2, ..., M. \end{aligned} \tag{6.4}$$

In chapters 3 and 4, we proved that logarithms of the sigmoidal and logarithmic utility functions, shown respectively by equations 6.1 and 6.2, are strictly concave. Furthermore, we deduced that the continuous optimization in equation 6.4 without the condition $r_i \in \mathbb{N}$ is a convex optimization, which in turn establishes the existence and tractability of a global solution for the optimization problem where rates are not positive integers. Should we be able to relax the rate discreteness constraint from our RB allocation problem in equation 6.4, we can solve the problem easily and obtain the continuous rates obtained by the algorithms that we obtained in chapter 3. Such continuous theoretical rates will form the multi-dimensional global solutions at which the proportional utility fairness of the cellular network is maximized.

Next section explains our proposed modus operandi to transform the discrete optimization

problem in equation (6.4) into one with continuous rates, i.e. it eliminates $r_i \in \mathbb{N}$ constraint.

6.2.2 Continuous Rate Allocation Relaxed Optimization Problem

Relaxing the $r_i \in \mathbb{N}$ constraint from the RB allocation optimization problem in equation (6.4) in accordance with Lagrangian relaxation [170], the formulation in equation (6.5) presents the utility proportional fairness continuous rate allocation optimization problem. As we observe, we temporarily get rid of the constraint $r_i \in \mathbb{N}$ of equation 6.4 in order to obtain an optimization which is easy to solve. In fact, in chapter 3, we proved that the optimization problem in equation (6.5) is convex. We further presented algorithms to obtain its global extremum.

$$\begin{aligned} \max_{\mathbf{r}} \quad & \prod_{i=1}^{M} U_i(r_i) \\ \text{subject to} \quad & \sum_{i=1}^{M} r_i \leq R \\ & r_i \geq 0, \qquad i = 1, 2, ..., M. \end{aligned} \tag{6.5}$$

We leverage the Lagrangian multiplier of the dual problem of equation (6.5). To do so, we define the Lagrangian as equation (6.6), where $z \geq 0$ is the slack variable and p is the Lagrange multiplier. In chapter 3, We showed that this is indeed the total price per unit bandwidth for all the M users and denoted it as the shadow price. Therefore, $w_i = pr_i$ gives out the ith UE's bid for bandwidth, and we can write that down as $\sum_{i=1}^{M} w_i = p \sum_{i=1}^{M} r_i$.

$$\begin{aligned} L(\mathbf{r}, p) &= \sum_{i=1}^{M} \log(U_i(r_i)) - p(\sum_{i=1}^{M} r_i + z - R) \\ &= \sum_{i=1}^{M} \left(\log(U_i(r_i)) - pr_i \right) + p(R - z) \end{aligned} \tag{6.6}$$

Then, the dual optimization can be expressed as equation (6.7).

$$\begin{aligned} &\min_{p} \max_{\mathbf{r}} L(\mathbf{r}, p) \\ &\text{subject to} \quad p \geq 0. \end{aligned} \tag{6.7}$$

It is noteworthy that, in view of separability of $\sum_{i=1}^{M}(\log(U_i(r_i)) - pr_i)$ in r_i, we can write $\max_{\mathbf{r}} \sum_{i=1}^{M}(\log(U_i(r_i)) - pr_i) = \sum_{i=1}^{M} \max_{r_i}(\log(U_i(r_i)) - pr_i)$. So the dual problem in equation (6.7) can be written as equation (6.8).

$$\begin{aligned} &\min_{p} \sum_{i=1}^{M} \max_{r_i} \left(\log(U_i(r_i)) - pr_i \right) + p(R - z) \\ &\text{subject to} \quad p \geq 0. \end{aligned} \tag{6.8}$$

We can solving the optimization in equation (6.8) by leveraging the method of Lagrange multipliers. Considering that $\frac{\partial \max_{\mathbf{r}} L}{\partial p}$ and $\sum_{i=1}^{M} w_i = p \sum_{i=1}^{M} r_i$, we can write equation (6.9).

$$p = \frac{\sum_{i=1}^{M} w_i}{R - z} \tag{6.9}$$

It is noticeable that the shadow price p is minimized when $z = 0$ such that we get $\sum_{i=1}^{M} w_i = pR$. In essence, doing so, we have divided the primal problem of equation (6.5) into two simpler optimization problems for the UEs and the eNB similarly to those of chapter 3. Then, the ith UE optimization problem can be written as the following equation .

$$\begin{aligned} \max_{r_i} \quad & \log U_i(r_i) - pr_i \\ \text{subject to} \quad & p \geq 0 \\ & r_i \geq 0, \quad i = 1, 2, ..., M. \end{aligned} \tag{6.10}$$

Furthermore, the eNB optimization problem can be expressed as equation (6.11).

$$\begin{aligned} &\min_{p} \max_{\mathbf{r}} L(\mathbf{r}, p) \\ &\text{subject to} \quad p \geq 0. \end{aligned} \tag{6.11}$$

The solution r_i of the optimization problem $r_i(n) = \arg\max_{r_i} \left(\log U_i(r_i) - p(n) r_i \right)$ is the value of r_i that solves equation $\frac{\partial \log U_i(r_i)}{\partial r_i} = p(n)$. Geometrically, this is the intersection of the horizontal line $y = p(n)$ with the curve $y = \frac{\partial \log U_i(r_i)}{\partial r_i}$ which is calculated in the ith UE.

The utility proportional fairness in the objective function of the optimization problem in equation (6.5) is guaranteed in the solution of the optimization problems for the UE in equation (6.10) and for eNB in equation (6.11), whose solution algorithms are respectively illustrated in (14) and (15). We developed this algorithms in chapter 3 for the continuous resource allocation optimization problem. As we can observe from these algorithms, each UE sends an initial bid $w_i(1)$ to its eNB, which in turn calculates the absolute difference between the currently and previously received bids in order to exit if the difference is lower than a threshold δ. Otherwise, the eNB calculates the shadow price by employing the equation $p(n) = \frac{\sum_{i=1}^{M} w_i(n)}{R}$ and transmits it to all the UEs under its coverage area.

The UEs leverage the shadow price to obtain the continuous rate r_i that maximizes $\log U_i(r_i) - p(n) r_i$, and that rate is deployed to calculate a new bid $w_i(n) = p(n) r_i(n)$ to be sent to the eNB. Similarly to chapter 4, we have used a robust algorithm to ensure that the continuous rate allocation algorithm converges for all eNB bandwidths R. To do so, a convergence measure $\Delta w(n)$, that senses bid fluctuations, decreases the step size between consecutive bids by utilizing a fluctuation decay function defined as either of the following forms, where adjusting l_1, l_2, l_3 alters the rate for the bids decay. The first decay function introduces an exponential decrease in the step sizes between the bids, whereas the second one causes the steps to slash down fractionally as n (algorithm iterates) proceeds.

- $\Delta w(n) = l_1 e^{-\frac{n}{l_2}}$.

- $\Delta w(n) = \frac{l_3}{n}$.

Remark 6.2.1. *The fluctuation decay function can be included in algorithm (14) of the UE or algorithm (15) of the eNB (chapter 4).*

In our model, we add the decay part in algorithm (14) of the UE. In Figure 6.10, we show the new shadow price $p(n)$ of the fluctuation example in Section 6.4 when using algorithms (14) and (15). The shadow price $p(n)$ fluctuation decreases with every iteration n and converges to the optimal shadow price that corresponds to the optimal rates. Detailed simulations are presented in the simulation section (6.4).

Algorithm 14 UE Algorithm (from chapter 3)

```
Send initial bid w_i(1) to eNB.
loop
  Receive shadow price p(n) from eNB.
  if STOP from eNB then
    Calculate allocated rate r_i^opt = w_i(n)/p(n).
  else
    Calculate new bid w_i(n) = p(n) r_i(n).
    if |w_i(n) − w_i(n−1)| > Δw(n) then
      w_i(n) = w_i(n−1) + sign(w_i(n) − w_i(n−1)) Δw(n). {Δw = l_1 e^(−n/l_2) or Δw = l_3/n.}
    end if
    Send new bid w_i(n) to eNB.
  end if
end loop
```

Algorithm 15 eNB Algorithm (from chapter 3)

```
loop
  Receive bids w_i(n) from UEs. {Let w_i(0) = 0 ∀i}
  if |w_i(n) − w_i(n−1)| < δ ∀i then
    STOP and calculate rates r_i^opt = w_i(n)/p(n).
  else
    Calculate p(n) = (Σ_{i=1}^{M} w_i(n))/R.
    Send new shadow price p(n) to all UEs.
  end if
end loop
```

Once the continuous rates from the algorithms (14) and (15) are allocated to the UEs in the equivalent relaxed continuous rate allocation optimization problem in equation (6.5), we

should obtain discrete rates which are indeed the solutions to the RB allocation optimization problem in equation (6.4). Next section presents our proposed methodology to map the continuous rates into possible discrete RBs, which represent the RB units of the wireless infrastructure, and to solve the discrete optimization problem of equation (6.4) based on a series of qualified candidates obtained from the continuous rates.

Next, section 6.3 presents the RB allocation optimization problem that will be obtained from the relaxed continuous optimization problem which we solve by Algorithms 14 and 15 to obtain the continuous rates.

6.3 Resource Block Allocation Optimization Solution

According to chapter 3, the application utility functions in our system are either logarithmic or sigmoidal and their logarithms are concave functions, which means that they have global minimum points. In this chapter, the minimum point will be the optimal rates allocated to UEs by the eNB in the wireless infrastructure. For example, we assume that the proportional fairness utility function for an LTE infrastructure with one eNB and M UEs is denoted by $U(\mathbf{r}) = U(r_1, ..., r_M) = \prod_{i=1}^{M} U_i(r_i)$. Since we proved that such a utility function has a global minimum such as $[r_{c1}, ..., r_{cM}]$, where $r_{c1}, ..., r_{cM}] = arg\max_{\mathbf{r}} \prod_{i=1}^{M} U_i(r_i)$. $[r_c1, ..., r_cM]$ is the vector of continuous optimal rates calculated by the eNB and UE Algorithms 14 and 15.

At this point, we can use all the discrete points within the M dimensional domain of $U(r_1, ..., r_M)$. The set of all possible points is the Cartesian product of the set of possible values for the wireless infrastructure utility. As a case in point, if the values of r_i for the ith UE ranges from 1 (proportional fairness requires a nonzero allocation which in a discrete sense requires at least RB 1) to 100, there will be 100^M possibilities for the discrete rates.

Obviously, an exhaustive search of all the possible discrete rates is computationally very

complex as the size of the system grows. However, we can consider the neighboring points of the continuous optimally allocated rates; that is to say for each component of the allocated rate vector we choose the lower (upper) integer points less (more) than the continuous rate. In spite of its simplicity, such an approach reduces the number of possibilities of the discrete rate candidates to 2^M sequences, which has very much less computation complexity as apposed to the original 100^M points. In essence, this corresponds to calculating the floors and ceilings of the continuous rates, i.e. $[\lfloor r_{c1} \rfloor, ..., \lfloor r_{cM} \rfloor]$and $[\lceil r_{c1} \rceil, ..., \lceil r_{cM} \rceil]$ respectively. Then, the candidate discrete rates will be all the M dimensional vectors whose elements are the floors or ceilings of the corresponding continuous rates, which yield in 2^M possible vectors for the discrete rates (RBs) which can be allocated.

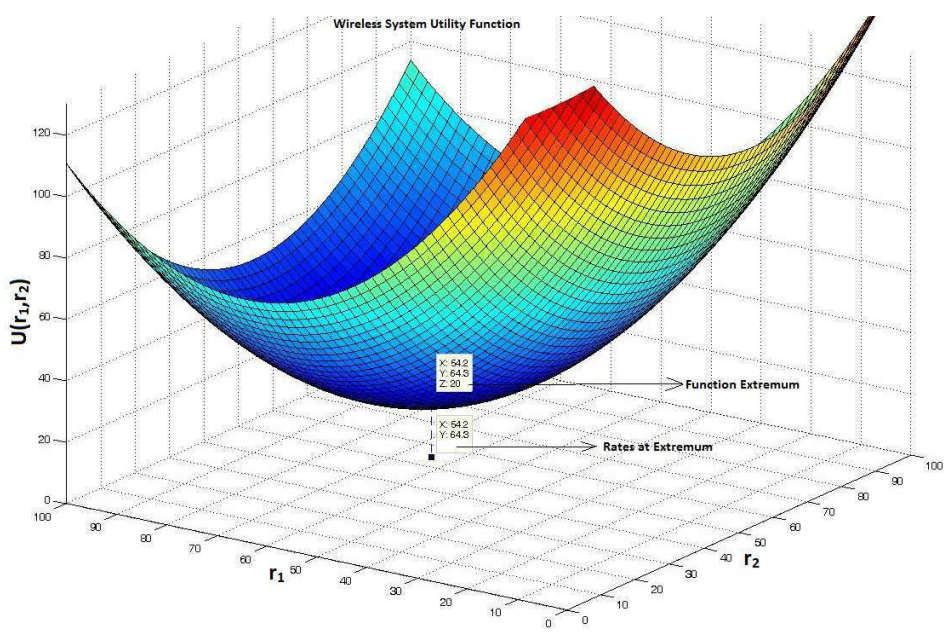


Figure 6.3: Exemplar System Utility Function: Two UE system with rates r_1 and r_2 where $U(r_1, r_2) = U(r_1)U(r_2)$ is the system utility function. As annotated, the minimum in the plot happens at $X = r_2 = 54.2$ and $Y = r_1 = 64.3$, so it is the optimum rate allocation, and the value of the utility function is $U(r_1, r_2) = 20$.

For the sake of clarity, for a system consisting of only two UEs, a system utility as in Figure 6.3 is given. This is in fact the negative of the actual utility function which has a maximum in view of the concavity proved in chapter 3. As we can observe, the utility function is a

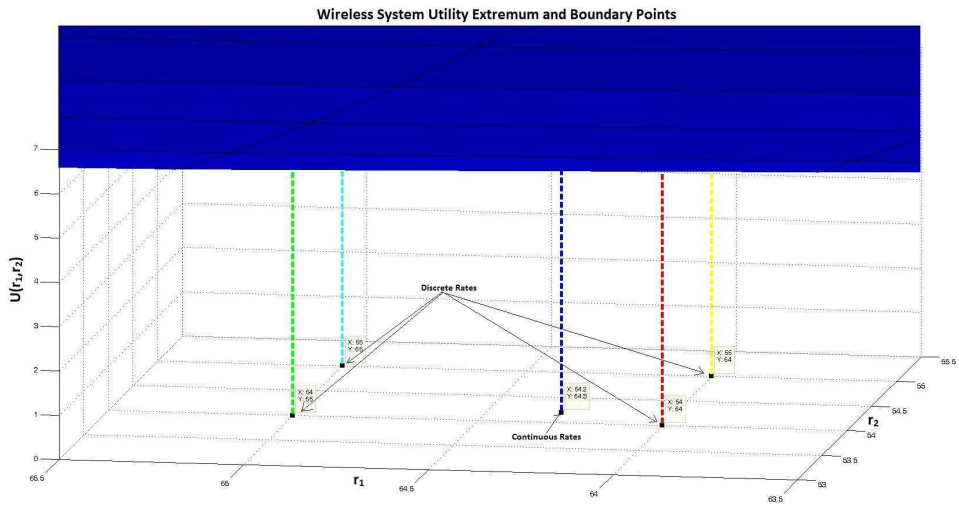


Figure 6.4: Exemplar Boundary Points: The blue point on r_1 - r_2 plane represents the continuous rate assigned to r_1 and r_2 $(r_1, r_2) = (54.2, 64.3)$. The red point $(r_1, r_2) = (54, 64)$, green point $(r_1, r_2) = (54, 65)$, yellow point $(r_1, r_2) = (55, 64)$, and cyan point $(r_1, r_2) = (55, 65)$ represent the discrete rates possible and the closest to the blue point.

surface of two dimensional points whose coordinates r_1 and r_2 which are real numbers which are equivalent to the continuous rates allocated to $U_1(r_1)$ and $U_2(r_2)$. However, for the RB allocation to be meaningful, such rates should be translated to discrete points. While there are many valid discrete points on both r_i axes as part of the $U(r_1, r_2)$ domain, part of the three dimensional space, considering the boundary points for $[r_1, r_2]$ yields in four integer points as in Figure 6.4 which are indeed $[\lfloor r_1, \lfloor r_2]$, $[\lceil r_1, \lfloor r_2]$, $[\lfloor r_1, \rceil r_2]$, and $[\lceil r_1 \rceil r_2]$.

As we observe in Figure 6.4, these boundary points, which are the closest integers to the continuous rates r_1 and r_2, present four possible combinations (four discrete rates in red, green, yellow, and cyan) which are the closest to the continuous point (rates), the blue color. Consequently, our search space constricts to only four points here, or the boundary points in general as opposed to all possible discrete points on the surface of the wireless system's utility function. In order to further restrict our search space, we look at the constraints of the continuous problem which says that $\sum_{i=1}^{M} r_i \leq R$. Even though the boundary points are closest to the optimal continuous solution, they may not fulfill this constraint as we

are essentially continuous rounding rates up or down. In view of this matter, we test the candidate boundary points and only select those who pass this requirement of the relaxed continuous rate allocation optimization problem.

The procedure to extract the distinct RB candidates from the continuous rates obtained from the continuous optimization problem of equation (6.5) is illustrated in Algorithm 16. As we can see, after obtaining the candidates for RB rates, we evaluate the cellular network's utility function for the possible candidates and we will store those who maximize the utility function as feasible RB rates.

Algorithm 16 Resource Block Allocation

```
loop
   Continuous value of r_i is r_i^opt. {Map the floors of continuous rates less than unity to one
   unit.}
   Calculate the ceilings and floors of the r_i.
end loop
List all possible sequences that can be obtained from the floors and ceilings.
if sum of the discrete rates surpasses R then
   Eliminate that sequence of discrete rates.
else
   Store that sequence of discrete rates as a candidate for RB allocation.
end if
loop
   Calculate the system utility for the stored RB candidates.
   Store RBs which maximize the utility.
end loop
```

It is noteworthy that in Algorithm 16 we have mapped the continuous rates which are less than one unit bandwidths to the unity both in the ceiling and roof as RB 1 is the minimum possible discrete rate allocation in the wireless network infrastructure with a proportional fairness policy to eschew the discrete rates to be mapped to zeros. Interestingly, many applications which do not require many resources may be mapped to the RB 1. In essence, such an allocation may result in equal discrete sequences for a continuous rate sequence. On the other hand, this algorithm's capability to generate numerous RB sequences renders eNB with more options and flexibility as to allocating the resources. This particularly can be of

significant consequence in a situation where we have several eNBs in the cellular network environment in that it can reduce the inter-cell and intra-cell interferences by having eNBs assigning different RBs to the UEs under their coverage area from the pool of feasible options for the discrete rates available to them.

In the next section, we present the simulation results for the Algorithms 15, 14, and 16 and analyze the results for experiments done to validate the proposed modus operandi.

6.4 Resource Block Allocation Simulation Results

In his section, we will evaluate the Algorithms 15, 14, and 16. These algorithms were applied to a representative scenario where, without loss of generality, 6 UEs were under the coverage area of an eNB in an LTE infrastructure as in Figure 6.5. The UEs were running 6 applications one each such that 3 applications were delay tolerant FTP services and the other 3 were real-time applications using VoIP, standard streaming video, and HDV. The aforementioned sigmoidal utility functions were expressed by the equation (6.1) with parameters $a = 5$, $b = 10$ to represent a VoIP application (inflection point at rate $r = 10$), $a = 3$, $b = 20$ to model a real-time standard streaming video (the inflection point at rate $r = 20$), and $a = 1$, $b = 30$ to be an HDV application (the inflection point at rate $r = 30$). Furthermore, logarithmic utility functions were formulated in accordance with the equation (6.2) with $k = \{15, 3, 0.5\}$, and they represented delay tolerant applications such as FTP. The simulations were being accomplished in MATLAB.

It is noteworthy that the rates in the LTE infrastructure are in terms of RBs, such that the value of the utility function, obtained from the equations (6.1) and 6.2, remain constant over RBs. The utility functions in terms of RBs are depicted in Figure 6.6. The plotted 6 application utility functions correspond to the applications running on the 6 UEs shown in Figure 6.5 that is we use three normalized sigmoidal functions and three logarithmic utility

functions.

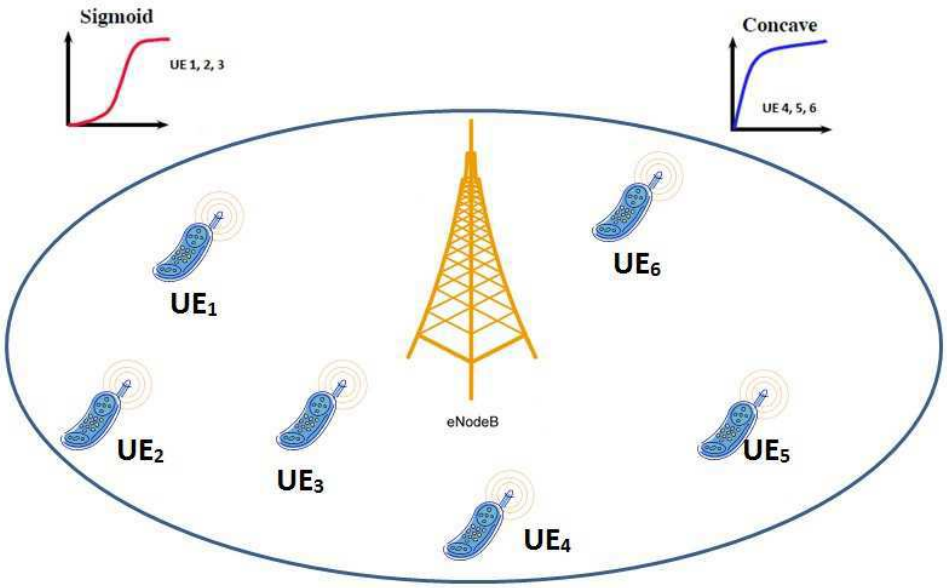


Figure 6.5: Experiment Set-up: 6 UEs are under coverage of an eNB, where UEs 1, 2, and 3 (4, 5, and 6) are running real-time (delay tolerant) applications represented by sigmoidal (logarithmic) utility functions. The goal is for eNB to allocate resource blocks to UEs.

The red sigmoidal in Figure 6.6 represents the HDV application the one with inflection point 30, whereas the green one is for the standard streaming video, and the blue one represents a VoIP application. As we can see from the Fig. 6.6 after exceeding the inflection points, utility percentages are almost 100% which is to be expected. The cyan, purple, and yellow logarithmic utilities FTP applications respectively sorted in a ascending order of delay tolerance. For instance, for the same rate, the utility function for the yellow curve is below the other logarithmic ones which manifests that it has less user satisfaction (utility) and, therefore, it is less delay-tolerant.

As it was explained in section 6.2.2, we relaxed the discrete RB allocation optimization problem into a continuous rate allocation optimization problem which is much easier to solve. The utility functions corresponding to the RB allocation within the relaxed continuous rate allocation problem are depicted in Figure 6.7. Contrasting Figure 6.6 and Figure 6.7, we

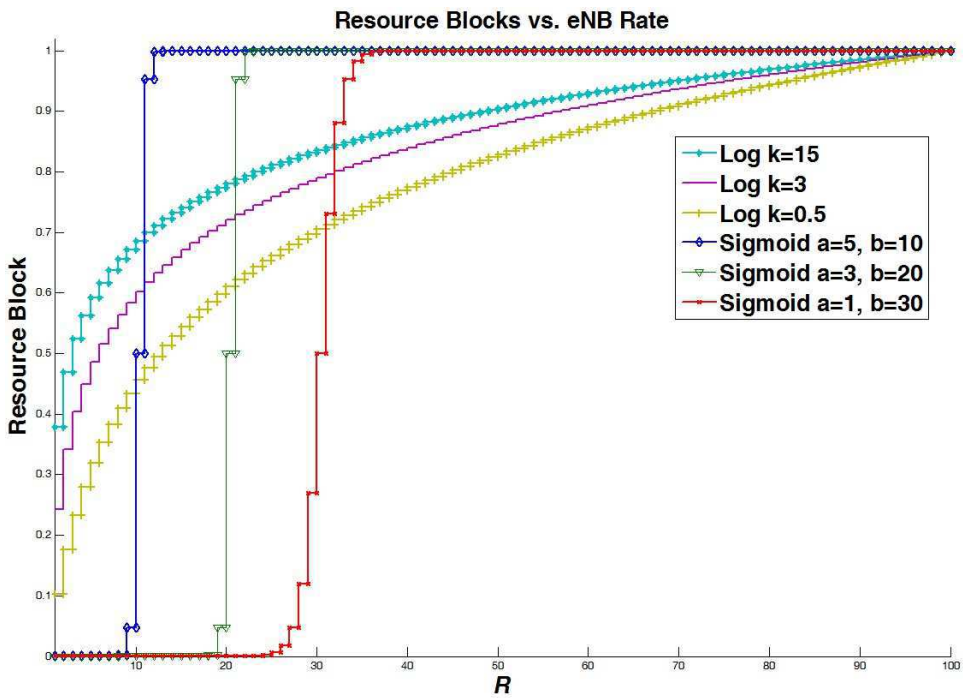


Figure 6.6: Utility Functions for the Discrete Optimization Problem: Each curve represents utility function for one of the UEs in the experiment, where 3 are delay tolerant applications corresponding to logarithmic utility functions and 3 are real-time applications corresponding to sigmoidal utility functions. The horizontal axis represents the resource blocks of size one unit bandwidth.

observe that the continuous relaxation of the RB allocation scheme produces utilities which follow the discrete versions very closely such that the ones for RB allocation look like samples from the continuous version.

Next, we show how the algorithm allocates continuous rates and converges once the RB allocation optimization problem is relaxed.

6.4.1 Algorithm Convergence for R =50

In the simulations in this subsection, we set the eNB's available bandwidth equal to $R = 50$ and the number of iterations of the Algorithms (14), (15), and (16) as $n = 35$. Figure 6.8 shows the continuous rates allocated to the 6 UEs versus the iterations of the algorithms for

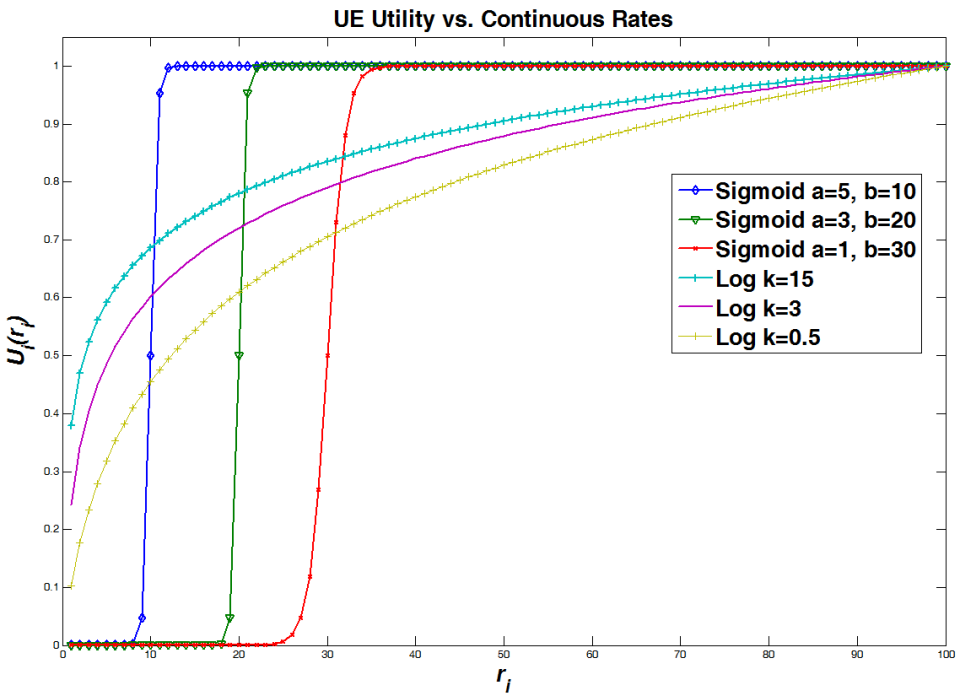


Figure 6.7: Utility Functions for the Continuous Optimization Problem: The horizontal axis represents continuous valued rates, and utilities as continuous functions of the rates. Each curve represents utility function for one of the UEs, where 3 are delay tolerant applications corresponding to logarithmic utility functions and 3 are real-time applications corresponding to sigmoidal utility functions.

the relaxed continuous rate allocation optimization problem. As we can observe from these figures, the sigmoidal utility functions are prioritized by the algorithm over the logarithmic ones. The reason for such an assignment lays in the fact that the sigmoidal utility unctions represent the real-time applications which need significantly higher bit rates compared with the delay tolerant logarithmic utility function so that their QoS is provisioned.

It is further noteworthy that once allocations are stabilized into the stead states, they are above inflection points of the sigmoidal utility functions such that the applications receive their required QoS; however, this is contingent upon the fact that the eNB's resources can satisfy the inflection points of the sigmoidal utility functions. Should this not be the case, the applications QoS is gravely impacted.

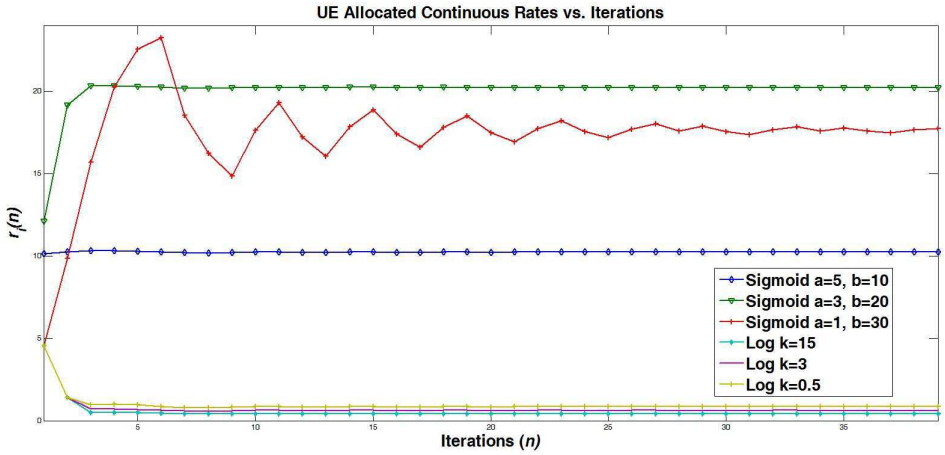


Figure 6.8: Continuous Rate Allocations vs. Iterations for the Relaxed Problem: Red, green, and blue curves are for VoIP, standard streaming video, and HDV applications in UEs 1, 2, and 3, and cyan, purple, and yellow curves are the rates allocated to FTP applications in UEs 4, 5, and 6. Higher rates are allocated to real-time applications (UEs 1, 2, and 3) vis-a-vis delay tolerant ones (UEs 4, 5, and 6).

On the other hand, we look at the UEs' bids for rate allocations during the various iterations of the algorithms. The bid values and their corresponding allocated rates are proportional, meaning that the larger bids yield in higher rates being allocated to the UEs. As it can be observed in Figure 6.9, the UEs with real-time applications (represented by sigmoidal utility functions) initially bid higher so that higher rates are allocated to them by the eNBs. This is in the light of the fact that real-time applications applications need higher bit rates with regard to their QoS requirements. Likewise, the delay tolerant applications also bid higher at the beginning of the algorithms iterations as opposed the later iterates, even though they do bid much smaller than do the real-time applications in the network.

In addition, the shadow price of the system (price per unit bandwidth for all the communications channels between UEs and an eNB) as a function of the number of the iterations of the algorithm is depicted in Figure 6.10. As it can be observed from this illustration, at the beginning of the iterates the real-time applications need more resources, so they are bidding larger which in turn increases the shadow price of the entire system. The eNB-calculated shadow prices will be transmitted to all the UEs under the eNB's coverage area, and the

shadow price will finally stabilize in the steady state.

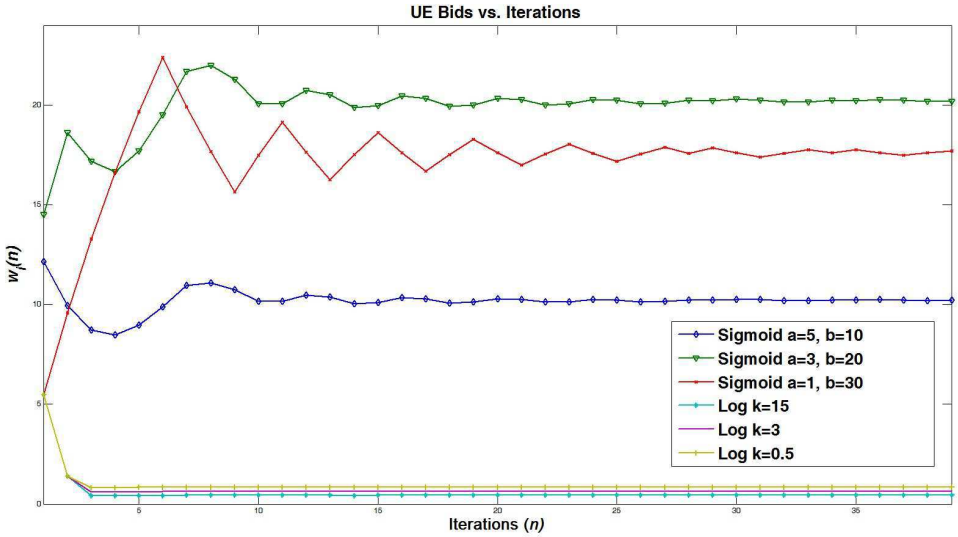


Figure 6.9: Continuous Allocations Bids vs. Iterations for the Relaxed Problem: Red, green, and blue curves are bids for UEs running VoIP, standard streaming video, and HDV applications that is UEs 1, 2, and 3, and cyan, purple, and yellow curves are the bids performed by FTP applications in UEs 4, 5, and 6. Real-time applications (UEs 1, 2, and 3) initially bid higher so that more rates are allocated to them as opposed to the delay tolerant ones (UEs 4, 5, and 6).

Next, we will investigate the sensitivity of our resource and RB allocation methodology to the changes in the available eNB bandwidth.

6.4.2 Algorithm Convergence for $50 \leq R \leq 100$

In order to evaluate the performance of our proposed methodology as to allocating resources and RBs, we performed the following simulations where we set the number of iterations to be 40 and the eNB rate to be R whose values range a gamut from 50 to 100 with unit steps. In Figure 6.11, we can observe that our algorithms refrain from allocating a zero rate to any of the UEs; this itself signifies the fact that no user will be dropped. This is a direct result of the proportional fairness policy that we chose to optimize.

Similarly to Figure 6.8, eNB allocates the bulk of the resources to the UEs which run real-

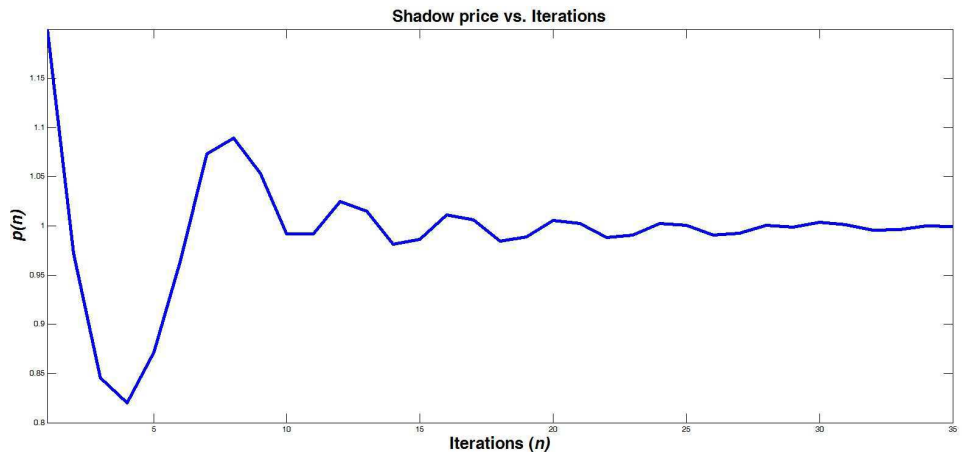


Figure 6.10: Shadow Price vs. Iterations for the Relaxed Problem: As algorithm calculates continuous rates and shadow prices, the value of the shadow price stabilizes in the steady state. These values are sent to all the UEs by the eNB in each iteration of the algorithms.

time applications represented by sigmoidal utility function. This procedure continues until these applications' utility functions reach their corresponding inflection points; however, if R is less than the sum of inflection points of the utility functions of the UEs, then real-time users will get less than their inflection rate which gravely impacts the applications QoS and users' QoE. When the eNB rate R exceeds the sum of the inflection point rates, i.e. $\sum b_i$, of all incumbent real-time applications, the eNB can allocate more resources to the UEs with delay tolerant application which essentially generate elastic traffic. In our case, this situation takes place when the bandwidth of the eNB exceeds 65 units of bandwidth.

We also look a the UEs bidding on the rate allocations as the available resources of the eNB alters. As the available resources of the eNB increase, UEs bids smaller for bandwidth allocations. This situation is depicted in Figure 6.12. As before larger bids are tantamount to higher allocated rates, and as the eNB bandwidth escalates, obviously available resources elevate and, therefore, bandwidth becomes less scarce. Under such circumstance UEs' bids slash down severely.

Moreover, along side the fact that increasing the eNB's bandwidth R decreases the UE bids, reduction of the shadow price also seems straightforward. This is in the light of the

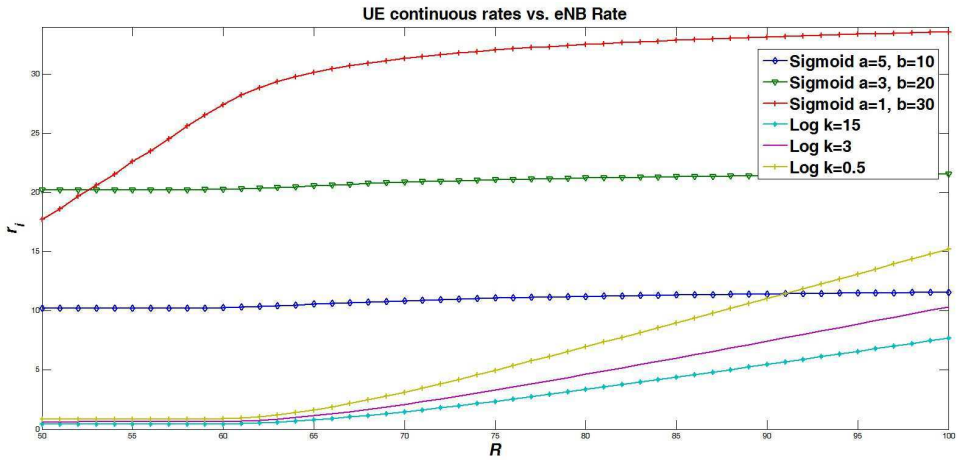


Figure 6.11: Continuous Rate Allocations vs. eNB Bandwidth for the Relaxed Problem: Steady state continuous allocated rates to the UEs increase as the eNB has more available resources. Red, green, and blue curves are bids for UEs running VoIP, standard streaming video, and HDV applications that is UEs 1, 2, and 3, and cyan, purple, and yellow curves are the bids performed by FTP applications in UEs 4, 5, and 6.

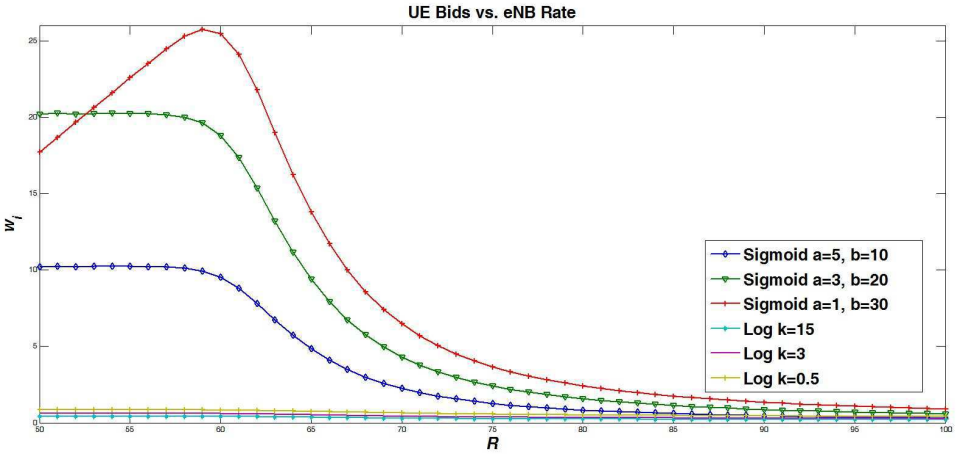


Figure 6.12: UE Bids vs. eNB Bandwidth for the Relaxed Problem: As more resources become available to the eNB, applications bid lower to have bandwidths allocated to them. Red, green, and blue curves are bids for UEs running VoIP, standard streaming video, and HDV applications that is UEs 1, 2, and 3, and cyan, purple, and yellow curves are the bids performed by FTP applications in UEs 4, 5, and 6.

fact that the higher R, the lower UE bids, and the lower the shadow price. The diagram for the shadow price variation vis-a-vis the eNB bandwidth is illustrated in Figure 6.13. As we can observe from this illustration, the shadow price significantly lowers done once the

available resources of the eNB exceed the inflection points of the utility functions.

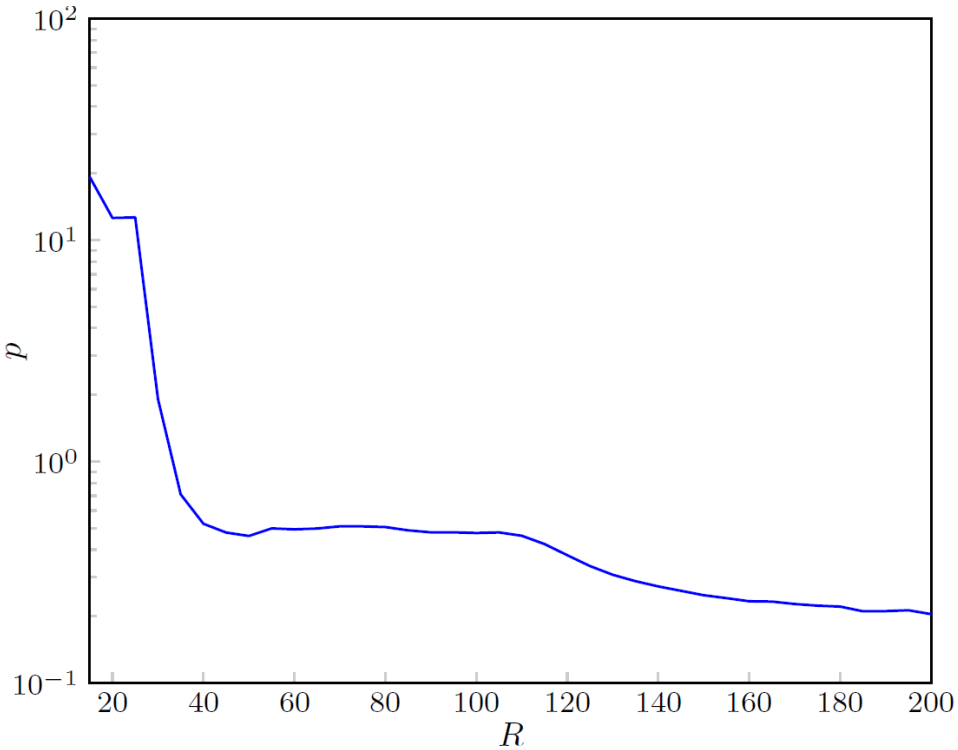


Figure 6.13: Shadow Price vs. eNB Bandwidth for the Relaxed Problem: As more resources become available for the eNB, the shadow price decreases. This is in line with the fact that applications bid lower once resources are opulent.

Once the relaxed problem is solved, as we have seen in the simulation results so far, the continuous rates are obtained . Then, the RB allocation simulation results based on the Algorithm (16) is being considered in the next section.

6.4.3 Resource Block Allocation

As it is illustrated in Figure 6.14, RBs are allocated to different UEs as the eNB rates changes based on the Algorithm (16). Since the structure of many cellular communications networks such as LTE relies on the concept of discrete rates or RBs to allocate radio resources, it is the RB allocations which makes the continuous rate resource allocation methodologies pragmatic. As we can observe, UEs which are running real-time applications get more

RBs allocated to them initially as the resources are scarce. However, as R increases, the availability of resources affords the eNB to allocate further RBs to the UEs with delay-tolerant applications as well.

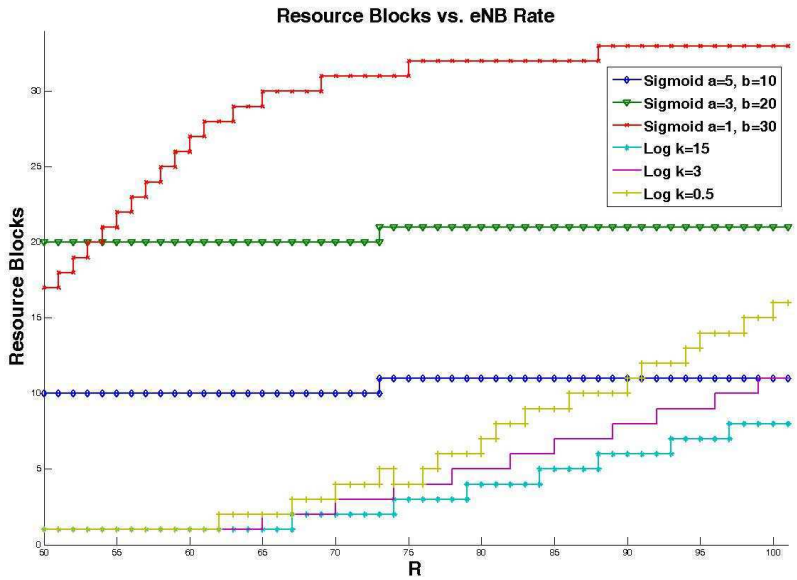


Figure 6.14: Resource Block Allocation versus eNB Bandwidth for the Discrete Optimization Problem: Red, green, and blue curves are bids for UEs running VoIP, standard streaming video, and HDV applications that is UEs 1, 2, and 3, and cyan, purple, and yellow curves are the bids performed by FTP applications in UEs 4, 5, and 6. Other discrete allocations are also feasible.

Now we take a look into an exemplar continuous and RB allocations for certain arbitrarily selected values of the eNB bandwidth R numerically in table 6.1. For instance, in table 6.1, the first column represents an arbitrary eNB bandwidth R, the second column is a 6 dimensional continuous vector whose i^{th} element represents the rate which has been allocated to the ith UE in the relaxed continuous rate allocation optimization problem, and the last column is a discrete 6 dimensional vector whose i^{th} element reflects the RB allocated the the ith UE by means of the Algorithm (16). As we can observe, an optimal continuous can be mapped to one or more RB allocations, which renders the eNB the flexibility to be able to opt for an RB allocation from a pool of feasible discrete allocations. This is very handy

in scenarios where more than one eNB are being considered in a cellular environment in which intercell and intracell interference is detrimental to the QoE of users. Nonetheless, this situation is not considered in this paper.

Table 6.1: An Example of the Continuous Rate Allocations versus the Resource Block Assignments for eNB Rate 100 and 50: The former yields in 8 possible discrete rate allocations whilst the latter generates 1 discrete rate allocation alone.

R	r_i	RB
100	$11.57, 21.57, 33.58, 7.72, 10.36, 15.21$	$11, 21, 33, 8, 11, 16$
100	$11.57, 21.57, 33.58, 7.72, 10.36, 15.21$	$11, 21, 33, 8, 11, 15$
100	$11.57, 21.57, 33.58, 7.72, 10.36, 15.21$	$11, 21, 33, 8, 10, 16$
100	$11.57, 21.57, 33.58, 7.72, 10.36, 15.21$	$11, 21, 33, 8, 10, 15$
100	$11.57, 21.57, 33.58, 7.72, 10.36, 15.21$	$11, 21, 33, 7, 11, 16$
100	$11.57, 21.57, 33.58, 7.72, 10.36, 15.21$	$11, 21, 33, 7, 11, 16$
100	$11.57, 21.57, 33.58, 7.72, 10.36, 15.21$	$11, 21, 33, 7, 10, 16$
100	$11.57, 21.57, 33.58, 7.72, 10.36, 15.21$	$11, 21, 33, 7, 10, 15$
50	$10.46, 20.46, 24.21, 9.13, 9.13, 9.13$	$10, 20, 17, 1, 1, 1$

As we can see in table 6.1, for an NB bandwidth of 100, the relaxed continuous rate allocation optimization problem optimally allocates rates to the 6 UEs as $[11.57, 21.57, 33.58, 7.72, 10.36, 15.21]$ which is going to maximize the utility proportional fairness. On the other hand, for this eNB rate, 8 possible RB allocations can be considered which are $[11, 21, 33, 8, 11, 16]$, $[11, 21, 33, 8, 11, 15]$, $[11, 21, 33, 8, 10, 16]$, $[11, 21, 33, 8, 10, 15]$, $[11, 21, 33, 7, 11, 16]$, $[11, 21, 33, 7, 11, 16]$, $[11, 21, 33, 7, 10, 16]$, and $[11, 21, 33, 7, 10, 15]$. Each of these vectors is a feasible RB allocation. On the other hand, an eNB rate of 50 leads in continuous rate allocation $[10.46, 20.46, 24.21, 9.13, 9.13, 9.13]$ and discrete RB allocation $[10, 20, 17, 1, 1, 1]$ (only one discrete possibility).

Lastly, it is worthwhile to look at the computational complexity of the proposed RB Algorithm (16). Since for every continuous rate r_i which is allocated to the ith UE's utility function, i.e. U_i, there is at least one and there are at most two possible boundary points. Thus, for a wireless network with M UEs, the computational complexity will be at most $O(2^M)$. That is not all 2^M possibilities for the combinations of the lower and upper bounds of the continuous allocated rates will necessarily satisfy the relaxed continuous problem's

constraints such that the sum of the discrete rates be still less than or equal to the available eNB bandwidth. Henceforth, some of the initially feasible discrete RB candidates will not be considered for in view their eNB rate violations.

As a matter of fact, excluding the discrete sequences which violate the rate constraint for the relaxed equivalent continuous rate optimization problem can significantly improve the computational complexity such that, on the assumption that n possibilities for the RB to be allocated to each UE exists, the computational complexity for an M UE system reduces from precisely $O(n^M)$ to mostly $O(2^M)$. For a UE including 100 RB allocation possibilities and 100 UEs under its coverage area, this means that we have slashed the number of computations from 10^{200} to only 2^{100}. This situation's computational complexity is depicted in Figure 6.15 as a semi-log diagram due to the exponential increase of the number of computations vs. the number of UEs.

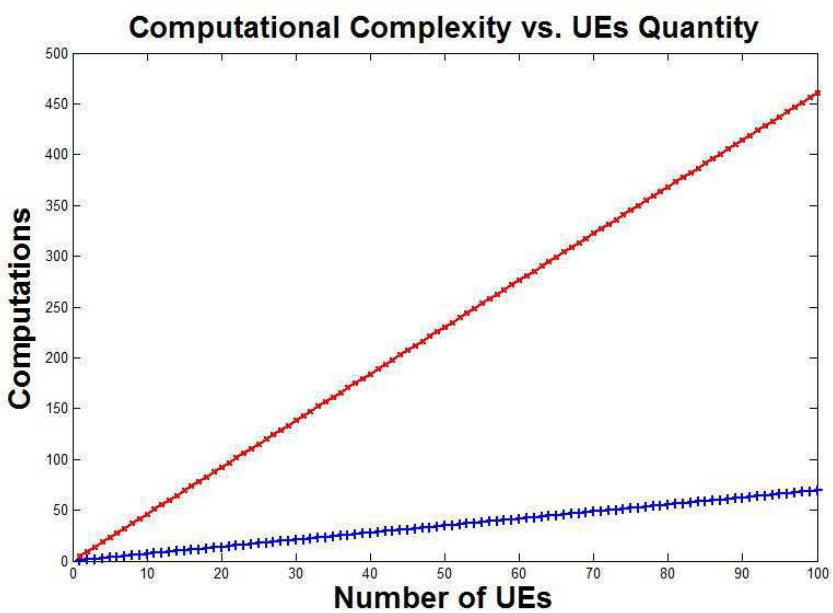


Figure 6.15: Computational Complexity: Red curve shows the semi-log computational complexity as a function of the number of UEs once we do not exclude non-violating sequence possibilities from the proposed methodology. The blue curve is the semi-log computational complexity once we have applied the technique to exclude the sequences which violate the constraint for the continuous relaxed problem.

Furthermore, the runtime of the algorithm is short. In the case where the eNB bandwidth was changing from 50 to 100 with steps of unity, the total runtime was only 16 seconds, which is reasonable as it may take much larger time than 16 seconds for a base station's bandwidth to change 51 times. Besides, the runtime for RB allocation when the eNB has a fixed bandwidth ranged value from 0.24 - 0.30 seconds which is very fast for the RB allocation.

Next, section 6.5 concludes this chapter by summarizing the results briefly.

6.5 Chapter Summary

In this chapter, we studies a resource block allocation optimization problem with a proportional fairness policy to allocate resources (bandwidths) to UEs in a cellular communications network infrastructure. Initially, we formalized the resource block allocation as an integer optimization problem, which was difficult to solve. In order to find a solution to this optimization problem more easily, we used a Lagrangian relaxation to transform the discrete resource block allocation optimization problem into a continuous rate allocation convex optimization problem, which is soluble by means of Lagrange multipliers of the dual problem to the continuous problem. The solution to the relaxed continuous rate allocation was a series of rates assigned by the base station to the UEs under the coverage of the wireless environment's base station.

Once continuous rates were allocated to the UEs, the original discrete resource block allocation optimization problem was approached by leveraging the boundary discrete points precisely below and above the formerly obtained continuous rates. These in turn yield in a series of candidate discrete rates for the UEs without much computation complexity. Furthermore, we sifted those candidates resource blocks which satisfied the constraints of the relaxed continuous rate allocation optimization problem. Such discrete constraint-satisfing

resource blocks are further evaluated in order to locate those among them which maximize the proportional fairness utility functions of the cellular network. These resource blocks constitute a set of feasible discrete rate allocations to the UEs.

We also investigated the sensitivity of our algorithm to changing the available bandwidth in the base station of the cellular communications system. Particularly, we obtained the shadow prices, user bids, continuous rates, and resource blocks calculated and allocated. We realized that as the base station resources become less scarce, both UEs' bidding values and the system's shadow prices decrease, whereas the allocated continuous rate and discrete resource blocks increase.

In addition, we were able to point out to the fact that the proposed resource block allocation optimization algorithm gave a priority to the UEs which were running real-time applications on their devices so that their quality of service is provided for. These applications utility functions were represented by sigmoidal utility functions in this article, whilst delay tolerant applications were expressed as logarithmic utility functions.

Furthermore, we realized that as the BSs' available bandwidth surpasses the addition of the inflection points for the utility functions in the relaxed continuous resource allocation optimization problem, it allocates further resources to UEs running delay-tolerant applications. And last but not the least, we investigated the complexity of the algorithm and realized that the problem is computationally efficient, although it is of $O(2^M)$ as opposed to being polynomial in computation. We further realized that the running time of the algorithm is reasonably short.

Chapter 7

Practical Challenges in Resource Allocation

7.1 Radio Resource Allocation in Congested/Contested Environments

Mobile broadband networks will face a tremendous increase in data traffic volumes over the next 20 years [10]. In order to meet this need, large amounts of spectrum will be a key prerequisite for any RAN evolution. To satisfy this demand, MNOs will need new spectrum allocations. However, the created demand for more bandwidth far exceeds the available commercial spectrum and has spurred the spectrum-governing agencies to consider spectrum sharing, which is an elegant solution to utilizing shareable spectrum bands efficiently. In spite its attractiveness, spectrum sharing is challenging as incumbent systems must be shielded from harmful interference from the entrant systems and vice versa. As a case in point, NTIA conducted a research [2] which showed that fledgling spectrum bands allocated for mobile broadband are occasionally occupied by the incumbents such as satellite and radar systems.

The incumbents would like to use the spectrum whenever they need, and neither incumbents nor entrant systems can tolerate interference from the other side. Henceforth, spectrum as the blood of the wireless communications system is a congested environment constricted by the notion of spectrum sharing, i.e. by the band incumbents from the perspective of the band entrant and vice versa. On the other hand, mobile broadband spectrum becoming a contested environment in the realm of modern-day communications networks inasmuch as an empty band, providing resources to either an incumbent or entrant system, can be contested for when both the incumbent and entrant operate in the vicinity of one another.

Amongst various band incumbents, radar systems are of high consequence as they have conventionally occupied large portions of spectrum as shown in Figure 7.1. So the radar-occupied spectrum can be a great candidate for spectrum sharing in that it can add a huge bandwidth to mobile broadband systems. However, radars can be very mobile and produce strong radio waves which are a threat to the QoS in cellular communications systems. Moreover, electromagnetic interference from the communications devices can jeopardize radar missions. Since radars are often not in the vicinity of wireless networks, the communication systems should be able to utilize the radar spectrum. On the other hand, an occasional radar proximity to the network should be incorporated into any radio resource allocation technique designed for modern cellular communications systems.

In chapters 3 and 4, we developed QoS-minded radio resource allocation architectures for cellular communications systems. The proposed resource allocation methods accounted for UE priorities, applications temporal usages, and traffic type. This chapter expands the aforementioned resource allocation frameworks to allow for sharing the spectrum between sectorized cellular communications systems and radars. The proposed architecture explores allocating optimal resources to the UEs to provision QoS for the running applications meanwhile sharing the communications resources with radar systems. The extended resource allocation architecture is formulated as two convex optimization problems, where

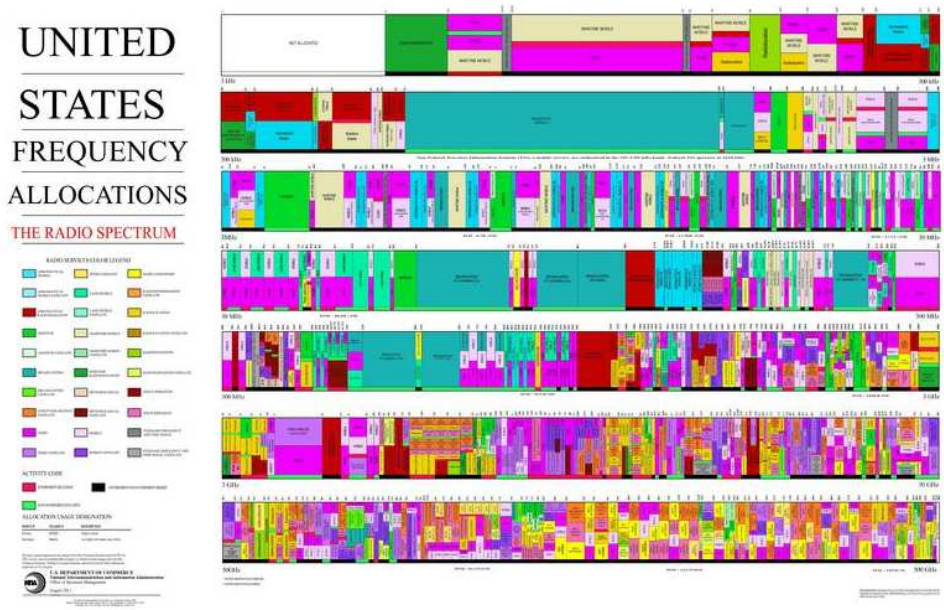


Figure 7.1: FCC spectrum allocation map defines the spectrum assignments in the US.

the rates for the radar-interfering sectors are extracted from the portion of the spectrum non-overlapping with the radar operating frequency. The proposed double-stage resource allocation procedure inherits fairness into the rate allocation scheme by first assigning rates from the portion of the spectrum shared with the radar. In this chapter:

7.1.1 Contributions

- We introduce an optimal rate allocation for radar-coexistent cellular networks with interference-refraining sectors.
- We demonstrate that the interfering sectors rates change as a radar approaches the cellular infrastructure.
- We elucidate that the proposed method fairly allocates resources even when a radar operates close-by the cellular system.

The remainder of this chapter proceeds as follows. Section 7.2 provides with the system

model for the extended resource allocation mechanism in this paper and provides with the formulation of the rate allocation scheme. This section also develops a solution algorithm for the proposed resource allocation architecture. Section 7.5 portrays the application of the resource allocation work in this paper to a simulation scenario where a communications network operates in the vicinity of a spectrally coexistent radar.

7.2 Architecture for Fair Allocation

Envisage a hexagonal K-cell L-sector communications network with the frequency reuse and a per-cell base BS, controlled by the Mobile Management Entity (MME) . A radar operation close by the cellular system causes interference with sectors working at the radar frequency. We assume a relatively stationary radar interferes with deterministic sectors inasmuch as topologically identical sectors deploy the frequency reuse. For the exemplar network in Figure 7.2 with the BSs shown as the gray triangular shapes at the circle centers, the f_3 Hz-operating pink sectors and radar can interfere while the green and pink sectors work safely in different frequency bands (f_1,f_2) belonging solely with the cellular network. It is worth mentioning that such a sector frequency pattern reduces the inter-cell interference by spatially maximizing co-channels, i.e. identical frequency sectors.

Moreover, the cellular system includes M UEs running a delay-tolerant or real-time application mathematically representable by sigmoidal and logarithmic application utility functions as in chapter 2, shown respectively in equations (7.1) and (7.2) for convenience. [104]. Here, $c_i = \frac{1+e^{a_i b_i}}{e^{a_i b_i}}$, $d_i = \frac{1}{1+e^{a_i b_i}}$, r_{max} is the 100% satisfaction-achieving rate ($U(r_{max}) = 1$), and k_i is the utility increase with enlarging the r_i. Based on chapter 3, $r_i = b_i$ is the inflection point of the sigmoidal utility function in equation (7.1) such that only for $r_i > b_i$ the user is satisfied. Parameters $(a_i, b_i, c_i, k_i, d_i)$ direct impact on the utility shape can model various applications such as VoIP ($a_i = 5$ and $b_i = 10$), video streaming ($a_i = 0.5$ and $b_i = 20$),

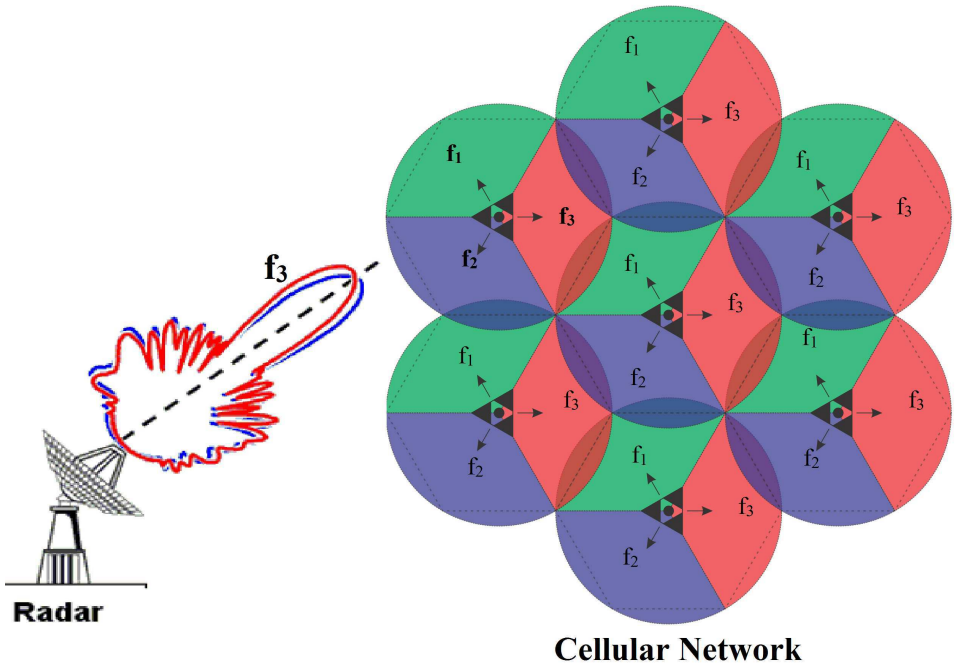


Figure 7.2: Cellular communications system: Cell sector colors indicate frequency bands (f_1, f_2, f_3) and identically colored sectors imply frequency reuse whose topological pattern minimizes the inter-cell interference. The radar and pink sectors operations in the same frequency band f_3 create interferes.

and FTP ($k = 15$) (chapter 3). As an illustration, 6 sample application utility functions are depicted in Figure 7.3.

$$U_i(r_i) = c_i \left(\frac{1}{1 + e^{-a_i(r_i - b_i)}} - d_i \right) \tag{7.1}$$

$$U_i(r_i) = \frac{\log(1 + k_i r_i)}{\log(1 + k_i r_{max})} \tag{7.2}$$

The ith UE rate which is allocated by its BS's l^{th} sector, where $l \in \{1, ..., L\}$, can include constituents from radars spectrum, denoted as $r^l_{i,\text{radar}}$, and communications spectrum, denoted as $r^l_{i,\text{comm}}$, such that the aggregate UE rate can be written as $r^l_{i,ag} = r^l_{i,\text{radar}} + r^l_{i,\text{comm}}$. Therefore, similarly to the formulations in chapter 3, the ith UE aggregate utility function (a function of its aggregate rate assigned by all sectors) can be expressed as in equation (7.3)

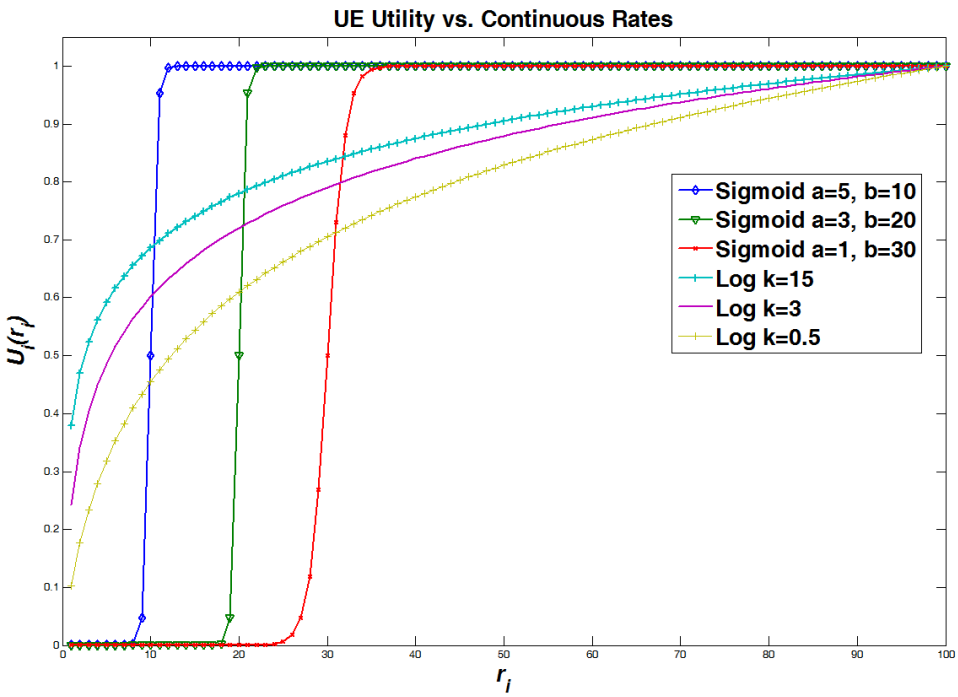


Figure 7.3: Three sigmoidal (purple, green, red) and logarithmic (yellow, pink, cyan) utilitiy functions for real-time and delay tolerant applications.

below.

$$U_i(r_{i,ag}) = U_i(\sum_{l=1}^{L} r_{i,ag}^l) = U_i(\sum_{l=1}^{L}(r_{i,\text{comm}}^l + r_{i,\text{comm}}^l)). \tag{7.3}$$

Next, section 7.2.1 develops the mathematical formulation for the extened rate allocation optimization instrumented with a spectrum sharing mechanism.

7.2.1 Congested Environment Radio Resource Allocation Optimization

To guarantee a minimum service QoS, a minimal nonzero rate per UE is realized through a proportional fairness (no user drop occurs) as in chapter 3. Moreover, spectral coexistence with nearby radars resembles the dual carrier rate allocation optimization in [172], where

UEs gain escalated rates by leveraging available bandwidths simultaneously from primary and secondary carriers, where the former and latter are respectively the radar and cellular systems in our proposed system model. Although a primary carrier is conventionally the communications side in the literature, section 7.5 shows this situation lacks fairness of rate assignments in that radar spectrum-coinciding frequencies can only be allotted to sectors which do not interfere with the radar operation. We propose a two-stage resource allocation, whose stage 1 assigns resources from the radar spectrum as in the equation (7.4) where $r_{i,\text{radar}}$ is the radar spectrum allocated to the ith UE, R^l_{radar} is the stage 1 maximum achievable rate allocated by the MME to the cells l^{th} sectors, M_k is the kth cell UE quantity, $\mathbf{r_{radar}} = \{r_{1,\text{radar}}, \ldots, r_{M,\text{radar}}\}$, and $R^{l_{\text{interference}}}_{\text{radar}} = 0$ implies that no rate is allocated to radar-interfering sectors at this stage which excerpts resources from the radar spectrum.

$$\begin{aligned}
\max_{\mathbf{r_{radar}}} \quad & \prod_{i=1}^{M} U_i\Big(\sum_{l=1}^{L} r^l_{i,\text{radar}}\Big) \\
\text{subject to} \quad & \sum_{l=1}^{L} r^l_{i,\text{radar}} = r_{i,\text{radar}}, \sum_{l=1}^{M_k} r^l_{i,\text{radar}} \leq R^l_{\text{radar}} \\
& \sum_{l=1}^{L} R^l_{\text{radar}} = R_{\text{radar}}, R^{l_{\text{interference}}}_{\text{radar}} = 0, r^l_{i,\text{radar}} \geq 0 \\
& i = 1, ..., M, l = 1, ..., L.
\end{aligned} \tag{7.4}$$

Next, at stage 2 of the resource allocation, the carrier (communications) spectrum is allocated to the UEs based on the equation (7.5) where $r_{i,\text{comm}}$ is the ith UE allocated communications spectrum, R^l_{comm} is the lth sector maximum achievable rate once resources are excerpted from the carrier spectrum (not radar), $\mathbf{r_{comm}} = \{r_{1,\text{comm}}, \ldots, r_{M,\text{comm}}\}$, and R^l_{comm} is the communications spectrum allocated to the lth sector of all the cells by the MME. It is notable that shifting non-interfering cell rates by the stage 1-assigned amounts is included in the equation (7.5)to insure fairness (we will see in section 7.5).

$$\begin{aligned}
\max_{\mathbf{r_{comm}}} \quad & \prod_{i=1}^{M} U_i(r_{i,\text{comm}} + r_{i,\text{radar}}^{\text{opt}}) \\
\text{subject to} \quad & \sum_{l=1}^{L} r_{i,\text{comm}}^{l} = r_{i,\text{comm}}, \sum_{l=1}^{M_k} r_{i,\text{comm}}^{l} \leq R_{\text{comm}}^{l} \\
& \sum_{l=1}^{L} R_{\text{comm}}^{l} = R_{\text{comm}}, r_{i,\text{radar}}^{\text{opt}} = \sum_{l=1}^{L} r_{i,\text{radar}}^{l,\text{opt}}, \\
& r_{i,\text{comm}}^{l} \geq 0, i = 1, ..., M, l = 1, ..., L.
\end{aligned} \tag{7.5}$$

We proved in chapter 3 that a resource allocation problem in the form of the equations (7.4) and (7.5) is a convex optimization which means that it has a global maximum and has a tractable solution. We write these solutions respectively as $\{r_{i,\text{radar}}^{l,\text{opt}} | i \in \{1, \ldots, M\} \wedge l \in \{1, \ldots, L\}\}$ and $\{r_{i,\text{comm}}^{l,\text{opt}} | i \in \{1, \ldots, M\} \wedge l \in \{1, \ldots, L\}\}$, which are the set of optimal rates allocated to the ith UE by l^{th} sector from respectively the radar and communications spectrum. The UE total optimal rate in the system is sum of the aforementioned solutions. To summarize the resource allocation procedure, for the ith UE, we first obtain $r_{i,\text{radar}}^{l,\text{opt}}$ (which is 0 bandwidth units for the UEs in the radar-interfering sectors of the communications system). Then, we obtain the UE rate obtained from the communications-only portion of the spectrum as $r_{i,\text{comm}}^{l,\text{opt}}$; And, the total optimal rate at the UE will be $r_{i,ag}^{l,\text{opt}} = r_{i,\text{comm}}^{l,\text{opt}} + r_{i,\text{radar}}^{l,\text{opt}}$.

Next, section 7.2.2 presents solution algorithms to solve the extended resource allocation optimizations in equations (7.4) and (7.5).

7.2.2 Solution Algorithm

Equation (7.4) solution relies on the Algorithms (17), (19), and (20). The solution for equation (7.5) relies on Algorithms (18), (19), and (20). Among these algorithms, the MME Algorithm is a variation of the algorithm in [4] and the other two are modifications of the dual carrier resource allocation algorithms presented in [172]. The subscript "radar"/"comm" in the

Algorithm (17)/(18) indicates resource allocations from radars/communications spectrum, which replaces the general subscript j in the Algorithms (19) and (20) at stage 1/stage 2 of the resource allocation to imply that resources are assigned from the radar/communications spectrum only. To recap, at stage 1 we allocate the radio resources from the radar spectrum to non-interfering sectors of the cells and no resources are passed to radar-interfering sectors' UEs. However, to ensure the fairness, less new resources should be given to the non-interfering sectors' UEs at stage 2 as radar-interfering UEs did not yet receive any resources at stage 1. This is realized by the shift in the Algorithm (18) with the subscript "radar" incorporating rate shifts due to the stage 1 assignments from the radar spectrum.

Algorithm 17 The lth Sector i^{th} UE Algorithm - **Stage 1**

Send an initial bid $w_i^l(1)$ to the BS l^{th} sector.
loop
  Receive a shadow price $P_l(n)$ from the BS l^{th} sector.
  if STOP from the BS l^{th} sector, **then**
    Calculate the allocated rate $r_{i,\text{radar}}^{l,\text{opt}} = \frac{w_{i,\text{radar}}^l(n)}{P_{\text{radar}}^l(n)}$.
  else
    Calculate $r_{i,\text{radar}}^l(n) = \arg\max_{r_{i,\text{radar}}^l} \Big(\log U_i(r_{i,\text{radar}}^l) - P_{\text{radar}}^l(n) r_{i,\text{radar}}^l \Big)$.
    Calculate a new bid $w_{i,\text{radar}}^l(n) = P_{\text{radar}}^l(n) r_{i,\text{radar}}^l(n)$.
    Send the new bid $w_{i,\text{radar}}^l(n)$ to the BS l^{th} sector.
  end if
end loop

Algorithm 18 The lth Sector i^{th} UE Algorithm - **Stage 2**

Send an initial bid $w_{i,\text{comm}}^l(1)$ to the BS l^{th} sector.
loop
  Receive a shadow price $P_{\text{comm}}^l(n)$ from the BS l^{th} sector.
  if STOP from the BS l^{th} sector. **then**
    Calculate the allocated rate $r_{i,\text{radar}}^{l,\text{opt}} = \frac{w_{i,\text{radar}}^l(n)}{P_{\text{radar}}^l(n)}$.
  else
    Calculate $r_{i,\text{comm}}^l(n) = \arg\max_{r_{i,\text{comm}}^l} \Big(\log U_i(r_{i,\text{comm}}^l + r_{i,\text{radar}}^{l,\text{opt}}) - P_{\text{comm}}^l(n) r_{i,\text{comm}}^l \Big)$.
    Calculate a new bid $w_{i,\text{comm}}^l(n) = P_{\text{comm}}^l(n) r_{i,\text{comm}}^l(n)$.
    Send the new bid $w_{i,\text{comm}}^l(n)$ to the BS l^{th} sector.
  end if
end loop

Algorithm 19 The BS l^{th} Sector Algorithm - **Stage 1,2**

loop
  Receive UE bids $w^l_{i,j}(n)$.
  Calculate aggregate bids $W^l_{k,j}(n)$ and send them to the MME. {Let $w^l_{i,j}(0) = 0 \ \forall i$}
  Receive the sector rate $R^l(n)$ from the MME.
  if STOP received from the MME, **then**
    STOP and send STOP to all UEs.
  else
    Calculate $P^l_j(n) = \frac{\sum_{i=1}^{M} w^l_{i,j}(n)}{R^l_j}$.
    Send the new shadow price to all UEs.
  end if
end loop

Algorithm 20 The MME Algorithm - - **Stage 1,2**, [4]

Send the sector rate $R^l(0)$ to the lth sector. {Let $R^l(0) = \frac{R}{L}$.}
loop
  Receive aggregated bids from $W^l_{k,j}(n)$ from the lth sector.
  Calculate total aggregated bids $W^l_j(n)$. {Let $W^l_j(0) = 0 \ \forall l$}
  Receive a sector rate $R^l(n)$ from the MME.
  if $|W^l_j(n) - W^l_j(n-1)| < \delta \ \forall l$, **then**
    STOP and send STOP to all sectors.
  else
    Calculate $R^l_j(n)$ and send to the lth sector.
  end if
end loop

The resource allocation process is as follows. The ith UE sends its initial bid $w^l_{i,j}$ to the kth BS lth sector, which calculates the aggregate sector bid $W^l_{k,\text{radar}}(n)$ at time n and transmits the aggregate sector bids $\{W^l_{k,\text{radar}}(n)|l = \{1, \ldots, L\}\}$ to the MME. This entity computes total aggregate sector bids W^l_{radar} and the difference from its former value $|W^l_{\text{radar}} - W^{l-1}_{\text{radar}}|$ for all the sectors. Should the difference be less than a pre-set threshold δ for all the sectors, an *exit* criterion is met; Otherwise, the MME evaluates sector rates R^l_{radar} and transmit them to corresponding BSs. Furthermore, the kth BS lth sector calculates the shadow price (price per unit bandwidth for all communications channels) $P^l_{radar}(n) = \frac{\sum_{i=1}^{M} w^l_{i,\text{radar}}(n)}{R^l_{\text{radar}} s}$ and transmits it to its covered UEs which solve $r^l_{i,\text{radar}}(n) = \arg\max_{r^l_{i,\text{radar}}} \left(\log U_i(r^l_{i,\text{radar}} - P^l_{\text{radar}}(n) r^l_{i,\text{radar}}) \right)$. Then, the calculated bid $w^l_{i,radar}(n) = P^l_{\text{radar}}(n) r^l_{i,\text{radar}}(n)$ is sent to the lth sector BS and the

procedure is repeated until the termination of the stage 2, $|W^l_{\text{radar}} - W^{l-1}_{\text{radar}}| < \delta$. The final optimum rates are calculated as $r^{l,\text{opt}}_{i,\text{radar}} = \frac{w^l_{i,\text{radar}}(n)}{P^l_{\text{radar}}(n)}$.

Then, at stage 2, each UE transmits an initial bid $w^l_{i,\text{comm}}(1)$ to the kth BS lth sector, which calculates the aggregate sector bids $W^l_{k,\text{comm}}(n)$ at time n and transmits it to the MME, which, in case $|W^l_{\text{comm}} - W^{l-1}_{\text{comm}}| > \delta$, evaluates sector rates R^l_j and sends them to corresponding BSs, which transmit the UEs their calculated shadow price $P^l_{\text{comm}}(n) = \frac{\sum_{i=1}^{M} w^l_{i,\text{comm}}(n)}{R^l_{\text{comm}}}$ leveraged to evaluate UE rates $r^l_{i,\text{comm}}(n) = \arg\max_{r^l_{i,\text{comm}}} \Big(\log U_i(r^l_{i,\text{comm}} + r^{l,\text{opt}}_{i,\text{radar}}) - P^l_{\text{comm}}(n) r^l_{i,\text{comm}} \Big)$. It is noteworthy that $r^{l,\text{opt}}_{i,p}$ in the Algorithm 18 is stage 1-obtained rate (solution of the Algorithm 17) and is part of the radar operating frequency. Then, the calculated bid $w^l_{i,comm}(n) = P^l_{\text{comm}}(n) r^l_{i,\text{ comm}}(n)$ is sent to the lth sector BS and the procedure is repeated until the termination of the stage 2, $|W^l_{\text{comm}} - W^{l-1}_{\text{comm}}| < \delta$. The final optimum rates are calculated as $r^{l,\text{opt}}_{i,\text{comm}} = \frac{w^l_{i,\text{comm}}(n)}{P^l_{\text{comm}}(n)}$. This entire process is summarized in Figure 7.4.

7.3 Practical Examples for Congested Environments and Resource Assignment

Spectrum sharing is an elegant solution to the dramatic increase in the data traffic volume of mobile broadband networks over the next 20 years [173]. It helps meet demands of Mobile Network Operators (MNOs) by assigning them new pieces of spectrum. A pioneering effort toward realizing the spectrum sharing was made by the Council of Advisers on Science and Technology (PCAST) [46] to leverage the full potential of the government-held spectrum. PCAST spurred the Federal Communications Commission (FCC) to issue a Notice of Proposed Rulemaking (NPRM) [47] to designate the 3550 - 3650 MHz range, abbreviated as the 3.5 GHz band, for mobile broadband. In sequel, the National Telecommunications and Information Administration (NTIA) recognized radar as major band incumbents and

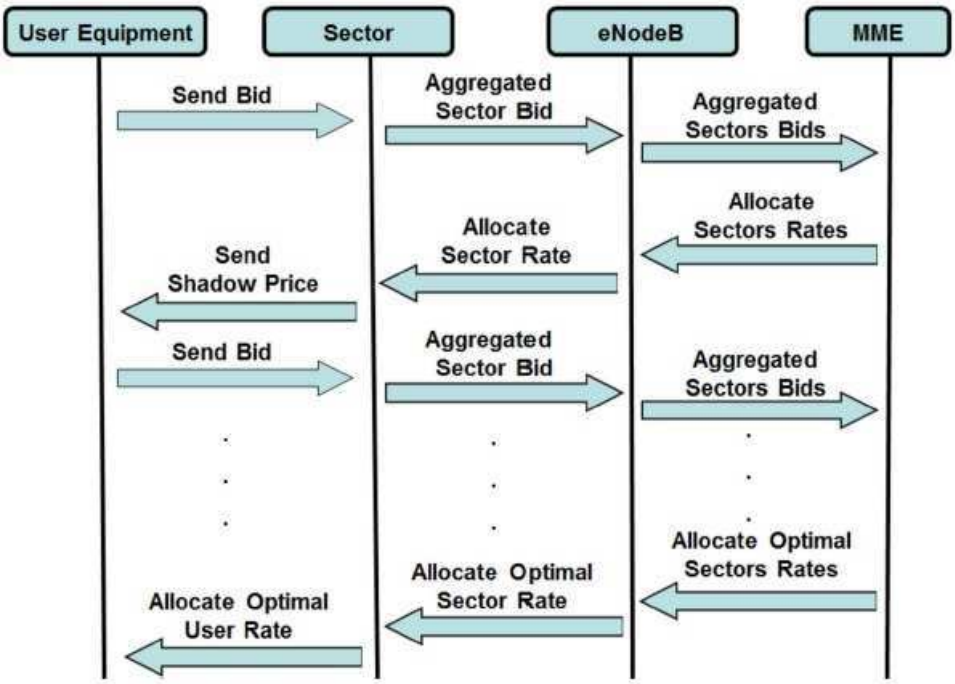


Figure 7.4: Resource Allocation Stage Diagram: This illustration only depicts one stage of the allocation whereas the assignment is performed once over the radar spectrum (stage 1) then over the carrier spectrum where radar is not operating (stage 2). The procedure in both stages is similar with the difference that the stage 2 includes a rate shift to include the stage 1-allotted bandwidth from radars spectrum (adopted from [4]).

conducted a measurement campaign [174], which revealed a low average temporal utilization of the band by the incumbents and the potential for spectrum sharing. However, an effective spectrum sharing refrains from destructive interference between incumbents and entrants. Concerning the 3.5 GHz band, NTIA investigated the interference between radar and WiMAX systems [31] and suggested exclusion zones reaching 557 km inland (Figure 7.5) [2] where no 3.5 GHz communications systems can be used. This hinders deploying the 3.5 GHz communications in coastal regions of the United States (US) where over 55% of the American reside [175]. Hence, judiciously reducing the exclusion zones affords MNOs to utilize this band to enhance mobile broadband coverage. The geographic separations in [2] came from link budget analyses of radar-WiMAX ecosystems; since LTE [18] is the expected cellular technology in the 3.5 GHz band, the nuances of the LTE link-level protocol (turbo coding, advanced scheduling, Hybrid Automatic Repeat Request - HARQ, etc.) can change

the exclusion zones. So, analyzing radar-to-LTE interference (not WiMAX) gives out relevant exclusion zones in the 3.5 GHz band . Besides, small cell implementation of LTE is becoming more popular as it extends the network capacity and the service coverage. As such, existence of band incumbents and particularly radar makes the spectrum a congested environment. On the other hand, incumbents may want to utilize the spectrum as needed; this make the resource allocation in cellular networks to be within a contested environments.

This chapter looks into the interference from S-band rotating radars into a Time Division Duplex (TDD) LTE in the uplink (UL) direction . The investigation relies on macro and outdoor small cell LTE system level simulations compliant to the 3rd Generation Partnership Project (3GPP) [5] . Moreover, radars, free space path loss (FSPL) [176] , and irregular terrain model (ITM) diffraction and troposcatter losses [177] are simulated using parameters from the NTIA [2] which led to the exclusion zones . The simulations show that LTE macro and small cells can operate within the aforementioned exclusion zones. Furthermore, the out-of-band interference into LTE macro cells is experimented. To our best knowledge, no prior work on radar-to-LTE interference renders a full consideration to radar operating characteristics along with LTE protocol details. This chapter motivates that resource allocation in congested environments is feasible.

To show some practical examples, we:

- We simulated an S-band radar system with every operational parameter.
- We simulated the FSPL and ITM propagation loss models.
- We leveraged a 3GPP-compliant detailed LTE simulator and simulated the interference from the radar into LTE macro, outdoor small cell, and indoor small cells scenarios.

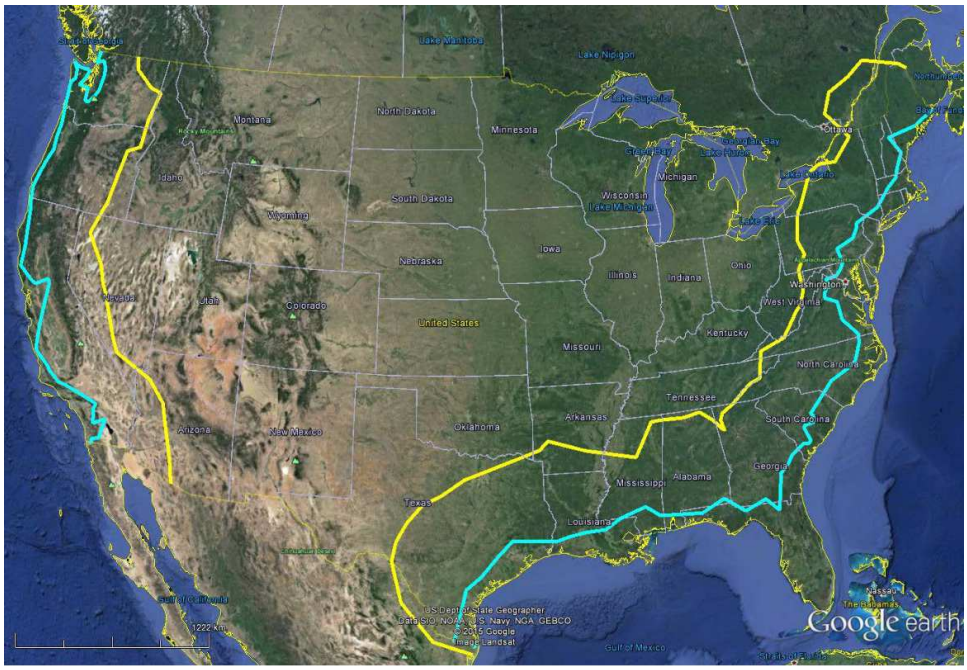


Figure 7.5: Radar-WiMAX exclusion zones (yellow curves) reach 577 km, where no 3.5 GHz communications system can currently be deployed [2]. This omits major US cities.

7.4 Radar Systems and Interference into LTE

7.4.1 Radar Simulation

The radar parameters are listed in Table 7.1 based on NTIA [2]. The items marked with asterisk were not disclosed in [2] due to the tactical sensitivity and were set to typical operating values for medium-to-large shipborne S-band radars [8]. Radar is set at 50, 100, 150, and 200 km away from the LTE and sends pulses into the LTE network as in Figure 2 (a). Scanning 360 deg azimuthally with 30 rotations per minutes (rpm), the radar scan time becomes 2 s during which 4000 pulses each 83 dBm are emitted as the pulse repetition interval (PRI) is 0.5 ms ($\frac{1}{0.5}$ ms = 2000 Hz). The horizontal beamwidth 0.81 deg necessitates 445 beam positions to cover the 360 deg search fence [8]; so the antenna dwell time becomes 4.5 ms in which 9 pulses - Figure 2(b) - are transmitted by the radar. Here, the abscissa and ordinate axes represent time in seconds and amplitude in Volt respectively. The radar

antenna has cosine pattern with equation (1) [2] plotted in Figure 3 as the normalized gain in terms of the off-boresight angle θ. Here, the first, second, and third expressions are the theoretical directivity pattern, a mask equation for pattern deviation from the theoretical value at the side-lobe (-14.4 dB main beam) , and back-lobe , respectively. As we can see, the back-lobe is fixed at 50 dBm below the main lobe. It is noteworthy that we are simulating the rotations as well, and the rotations during the dwell time are accounted for. The interference is calculated every transmission time interval (TTI) which means that we are simulating rotations every TTI even during the dwell time .

$$G(\theta) = \begin{cases} \frac{\pi}{2}\left(\frac{cos(\frac{68.8\pi sin(\theta)}{\theta_{3dB}})}{(\frac{\pi}{2})^2-(\frac{68.8\pi sin(\theta)}{\theta_{3dB}})^2}\right) \\ -17.5ln(\frac{2.33|\theta|}{\theta_{3dB}}) \\ -50dB \end{cases} \tag{7.6}$$

Table 7.1: Radar parameters from [2], except those marked with * where we used typical parameters for medium-to-large shipborne S-band radars [8].

Parameters	Value
Operating Frequency	$3.5GHz$
Pulse Peak Power	83 dBm
Antenna Gain	45 dBi
Antenna Pattern	Cosine
Antenna Height	50 m
Insertion Loss	2 dB
Pulse Repetition Interval (PRI)	0.5 ms
Pulse-Width	78 microseconds
Azimuth Beamwidth	0.81 deg
Elevation Beamwidth	0.81 deg
Search Fence	360 deg
EIRP	128 dBm
Scan Time	2 s
Pulse/Rotation	4000
Distance (r) to LTE	50, 100, 150, 200 km
Radiation Width (d) for r	1.5, 3.0, 4.5, 6.0 km

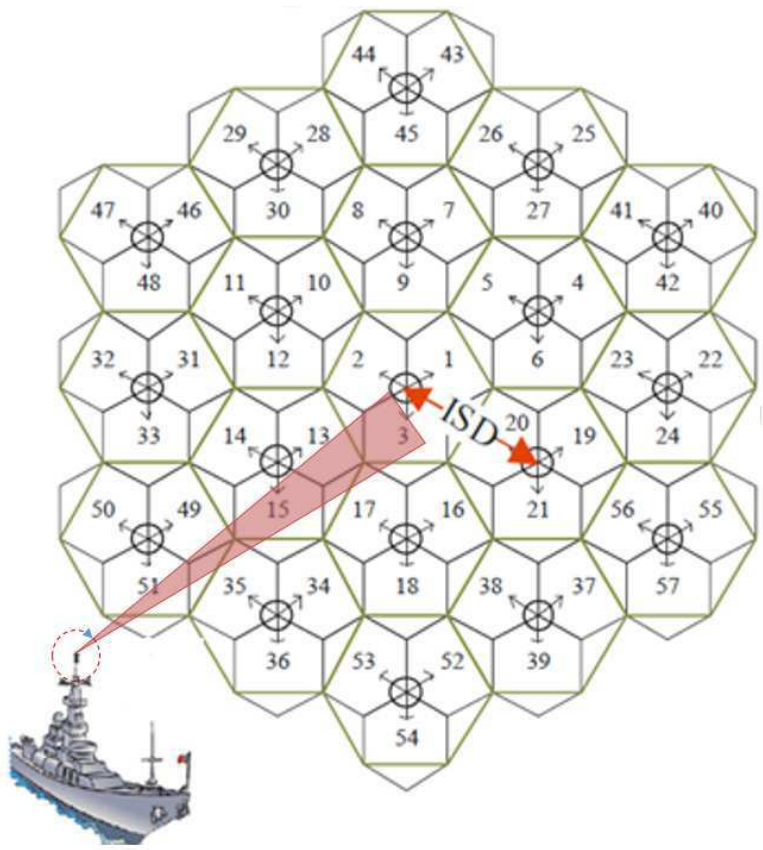


Figure 7.6: Radar simulation scenario subsumes a radar next to the 3.5 GHz LTE.

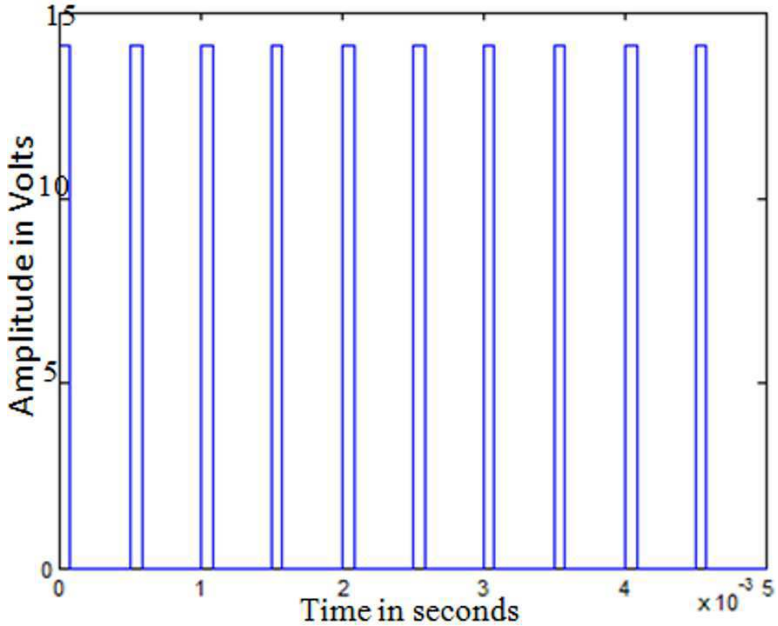


Figure 7.7: Radar pulses radiated on the LTE during the antenna dwell time.

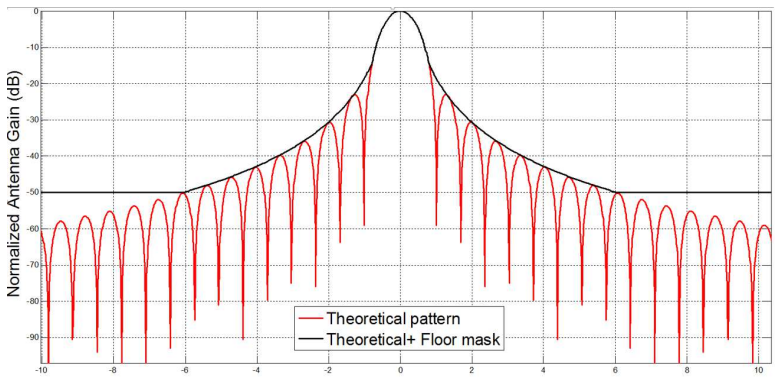


Figure 7.8: Radar Antenna Pattern [2]: The black curve represents the pattern after the first side lobe is 14.4 dB below the main lobe. The back lobe is constant at -50 dB.

7.4.2 Macro cell, Micro cell, and Indoor LTE Simulation

We leverage an LTE system level simulation which is 3GPP compliant [5]. It contains 3GPP-defined outdoor and indoor small cells as well as macro cell infrastructures. Furthermore, it utilizes a proportionally fair scheduler in both time and frequency domains, and includes a detailed UL air interface modeling. Besides, it subsumes UL multi input multi output (MIMO) and receiver diversity. Our LTE simulation has non-ideal link adaptation with HARQ and leverages an exponential effective SNR mapping (EESM) link-to-system mapping [178]. In addition, we model RF receiver saturation as a threshold -30 dB and turbo decoder saturation 50 dB. The UL uses single carrier orthogonal frequency division multiple access (SC-OFDMA) [18], and deploys a full-buffer traffic model. First, we deployed the LTE simulator with a macro cell model , referred to as urban macro (UMa) [5], with an intersite distance 500 m . We created a 7-site system whose cells have 120 deg sectors at 90, 210, and 330 deg, so there exist 21 cells in the system (7×3). The BS antenna pattern per sector is given in equations (7.4.2) and (7.4.2) where G_A and θ_A (G_E and θ_E) are respectively the antenna azimuth (elevation) pattern and azimuth (elevation) angle off the boresight [5] . The normalized composite antenna pattern is plotted in Figure 7.9. The BS antenna gain is 17 dBi and there are 10 UEs per cell which are at least 25 m away from the BSs. UE have omnidirectional 0 dBi antennae. In the macro cell scenario, 80% of the

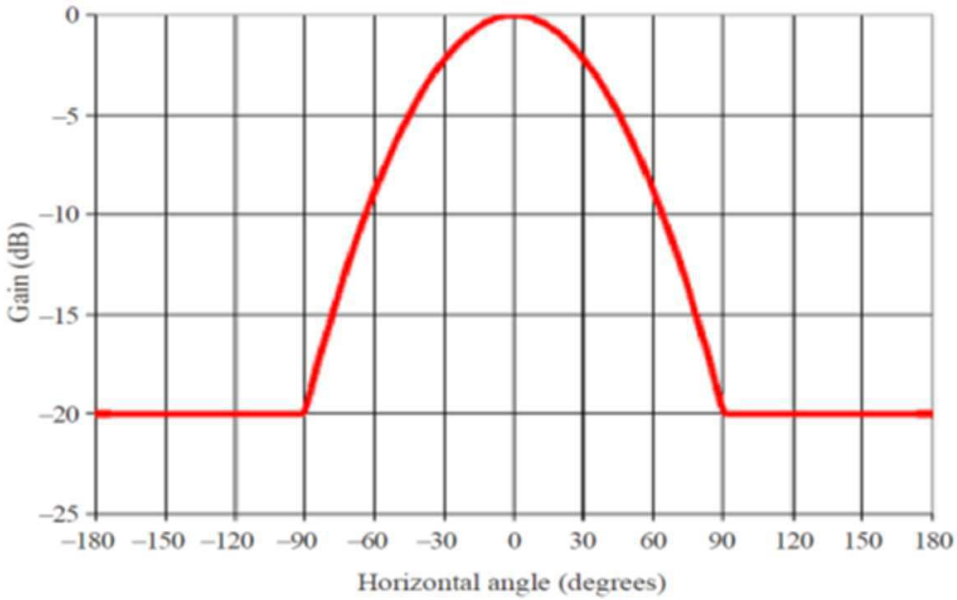


Figure 7.9: Composite antenna pattern for the macro cell LTE BSs for 3-sector cells from equation (3) [5].

UEs are indoor, LTE bandwidth is 10 MHz , and it operates at 3.5 GHz in TDD mode with a DL:UL split 2:3 . The BS and UE transmit powers are 46 dBm and 23 dBm respectively, and the EIRP transmit power is maximally 63 dBm and 23 dBm in that order. Moreover, the BS and UE antenna heights are 25 m and 1.5 m, and their noise figures are 5 dB and 9 dB respectively.

$$G_i(\theta_i) = -min12(\frac{\theta_i - \theta_{i,t}}{\theta_{3dB}})^2, A_m \tag{7.7}$$

$$G = -min-(G_A(\theta_A) + G_E(\theta_E)), A_m \tag{7.8}$$

For UMa, there are 7 sites with red circles representing BSs and UEs uniformly distributed around them. The UEs are depicted with the black addition symbols in the layout which

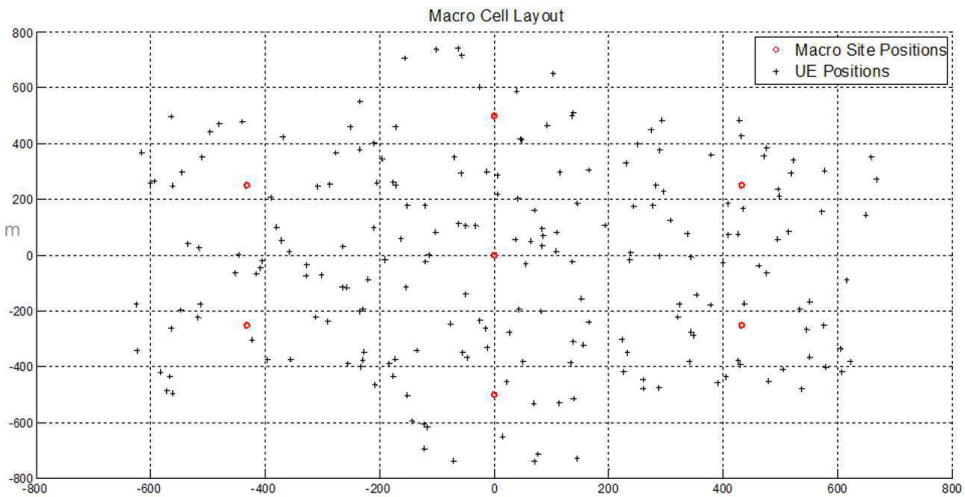


Figure 7.10: : LTE macro cell layout simulation.

spans about 1600 m × 1600 m. Next, we did an LTE simulation with outdoor small cells, referred to as urban micro (UMi) [5], whose main distinction from the macro cells is that its cells are not sectorized, use 10 m omnidirectional BSs as opposed to the 25 m in the UMa. Here, the blue squares represent the macro cell BSs operating at 2 GHz, so they do not interfere with the small cells and the radar. The red circles show outdoor small cell BSs, and their UEs are shown as the black addition symbols. The layout spans an area as large as 1600 m × 1600 m. The operation is at 3.5 GHz in TDD mode with 2:3 UL:DL split and 20% of the UEs are indoor. There exist 84 small cells as there exist 4 small cells per macro cell and 21 macro cells. The BS gain is 5 dBi and the minimum distance between the UEs and BSs is 5 m. There are 30 UEs per small cell, which creates $30 \times 84 = 2520$ UEs moving with uniform direction at 3 km/h. The UEs are clustered around the BSs. The BSs noise figure is 5 dB, LTE bandwidth is 10 and 20 MHz, and thermal noise is -174 dBm/Hz. The modeled traffic is the same as the macro cell scenario, full buffer best effort. The LTE parameters for both UMa and UMi simulations are listed in Table 7.2.

Table 7.2: Macro Cell LTE Parameters for Our Simulations (adopted from 3GPP [5]).

Parameters	Value
Operating Frequency	$3.5GHz$
UMa (UMi) [InH] Layout	Hexagonal (Clustered) [Indoor Hall]
Bandwidth	20 MHz
Operation Mode	TDD
UE Transmit (TX) Power	23 dBm
Cell Sites	7
UMa (UMi) Cell Quantity	21 (84)
UMa (UMi) [InH] Indoor UE	80% (20%) [100%]
UMa (UMi) [InH] BS Antenna Height	25 (10) [6] m
UMa (UMi) [InH] BS Antenna Pattern	Cosine (Omni) [Omni]
UMa (UMi) [InH] BS Antenna Gain	17 (5) [5] dBi
UE Antenna Gain	0 dBi
UE Antenna Height	1.5 m
Inter-site Distance (ISD)	500 m
UMa (UMi) BS-UE Minimum Distance	25 (5) m
UMa (UMi) BS Antenna Downtilt	12 (0) deg
UE Antenna	Omnidirectional
UE Distribution	Uniform
UE Mobility Speed, Speed	3 km/h, Uniform
UE Noise Figure (NF)	9
BS NF	5
Thermal Noise	-174 dBm/Hz
Service Profile	Full Buffer Best Effort
UMa (UMi) [InH] UE per Cell	10 (30) [20]
Channel Model	UMa (UMi)[InH]

7.4.3 Propagation Model Simulation

The radar signals that are emitted into the outer space undergo various attenuations before reaching the LTE sites [8]. The most paramount attenuation is the propagation pathloss for which various models relying on the distance and terrain between the radar and LTE can be leveraged to obtain how strong the radar pulses are once they get to the LTE BSs. Similarly to the Fast Track Evaluation by NTIA [2], we use the FSPL and ITM for the line-of-sight (LoS) and non-LoS (NLoS) regions , respectively. These models are predominantly used by the FCC and NTIA and were used to extract the exclusion zones depicted in Figure 7.5.

FSPL can be expressed as equation (7.9) [176] where f is the radar operating frequency in MHz, r is the distance in km at which the loss is of interest, LdB,FSPL in dB is the FSPL, and rLoS is the LoS region border in km as equation (7.10) [2]. Here, where h_{radar} and h_{LTE} is the radar and LTE antenna heights where the former is 50 m in our simulations (shipborne radar) , and the latter is 25 m and 10 m for UMa and UMi in that order. Hence, the LoS region border grows 49.59 km and 41.96 km for macro cell and small cells respectively, meaning that in all distances away from the radar until the 49.59 km for macro cells and 41.96 km for small cells, the propagation loss comes from the FSPL model.

Similarly to [2], we leverage the ITM for the loss in the NLoS region in its area prediction mode (APM) [177] with the terrain roughness 10 and 20 m, LTE macro and small cell antenna heights 25 and 10m, radar antenna height 50 m, ground dielectric constant 15, ground conductivity 0.005 S/m, refractivity 301 N-units , continental temperate climate, and single message mode as in Table 3. The aforesaid parameters are extracted from [2] which led to Figure 7.5 exclusion zones. The ITM model gives the diffraction and troposcatter losses introduced into the radar signals in the NLoS region after 49.59 km and 41.96 km for correspondingly the macro and small cells. The plots for the FSPL and ITM losses are depicted in Figure 7.11, where the green curve indicates FSPL, red curve is the NLoS loss for the macro cell case, and blue curve is the NLoS loss in dB for the small cell case as a function of the travelled distance by radar pulses in km. As we see, initially pathloss increases using the FSPL model, then it increases dramatically and almost linearly in the NLoS region which is the diffraction loss [177], and ultimately the loss grows less rapidly which is the troposcatter loss. It is notable that the FSPL and ITM models predict very close propagation losses in the LoS region whilst the the introduced loss sharply elevates in the NLoS region, where ITM model is utilized.

$$L_{dB,FSPL} = 20log(f) + 20log(r) + 32.45, r \leq r_{LoS} \tag{7.9}$$

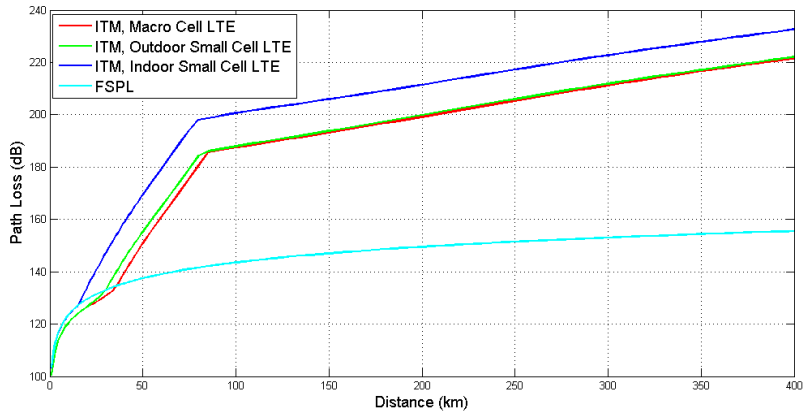


Figure 7.11: LoS FSPL in green and NLoS ITM loss in blue and red for macro and small cells respectively represent the radar signal degradation in dB vs. distance in km.

$$r_{LoS} = 4.1(sqrt(h_{radar}) + sqrt(h_{LTE})) \tag{7.10}$$

We set the simulation time to 5 seconds, during which the impact of the radiations from the radar into the BSs of an LTE cellular system with parameters is investigated. We test both cases where the radar is cochannel or out-of-band with respect to the LTE system and rotates 360 deg in azimuth covering 445 beam positions, where the radar sojourns for the dwell time 4.5 ms and sends 9 pulses 78 s wide and 83 dBm through its 45 dBi antenna to the BSs. The BSs covered by the radiation will suffer from the radar pulses which are amplified by the TX and antenna gain and undermined by the propagation loss.

7.4.4 Radar to Macro Cell LTE UL Interference

With regard to the interference, the main performance metric for LTE systems is the mean throughput [2]. We assume that the radar is cochannel with the LTE system. The normalized mean throughput is depicted in Figure 7.12. As we see, slight throughput losses incur when the radar is 50, 100, 150, and 200 km away vis--vis the baseline (brown bar). In fact, the LTE system as close as 100 km away from the radar undergoes less than 10%

throughput loss with respect to the baseline, and the loss is less than 30% when the radar is only 50 km away from the LTE. As we can see, UMa can operate within the exclusion zones identified by NTIA [2]. In addition, we plot the SINR of an LTE macro BS versus LTE symbol and subcarrier indices in Figure 7.13 where SINR drops due to radar pulses affecting LTE symbols in UL. Interestingly, even when the radar is present, the SINR recovers back to its normal baseline situation until the next pulse hits the same region. As we see, the radar hits symbol 1; since the radar pulses (78 s) are larger than the LTE symbols (71.4 s), symbol 2 is also affected. Then, the next pulse hits symbol 8, which spans over to symbol 9 too for the same reason. So, most radar energy is concentrated on symbol 1 and symbol 8, with the remaining pulse energy present on symbols 2, 9 and 14. This is promising as only certain symbols within an LTE sub-frame are affected by the radar. Also, there are 7 symbols between symbol 1 (start of the first radar hit) to symbol 8 (start of the second radar hit), corresponding to $7 \times 71.4 = 0.5$ ms, the radar PRI . Because the radar signal is assumed to be centered in the LTE band, most pulse energy is around subcarrier 300 (in the middle of the LTE band). We also look at the out-of-band interference from the radar into the LTE system in Figure 7.14, where we set the radar cochannel, 3 , 5, and 10 MHz offset from the LTE operating frequency. The throughput loss for 3 MHz and cochannel is similar; however, the throughput with respect to the baseline loss drastically alleviates as the offset becomes 5 and 10 MHz.

7.4.5 Radar to Outdoor Small Cell LTE UL Interference

Next, we show the simulation results for the outdoor small cell LTE. Similarly to the macro cells, we look at the mean UE throughputs . The radar is 50, 100, 150, and 200 km away from the small cell LTE. The normalized mean UE throughput is illustrated in Figure 7.15, which represents relative values. As we observe, the throughput loss for the LTE small cells is not dramatic until 100 km away from the radar. In fact, for an LTE outdoor small cell as

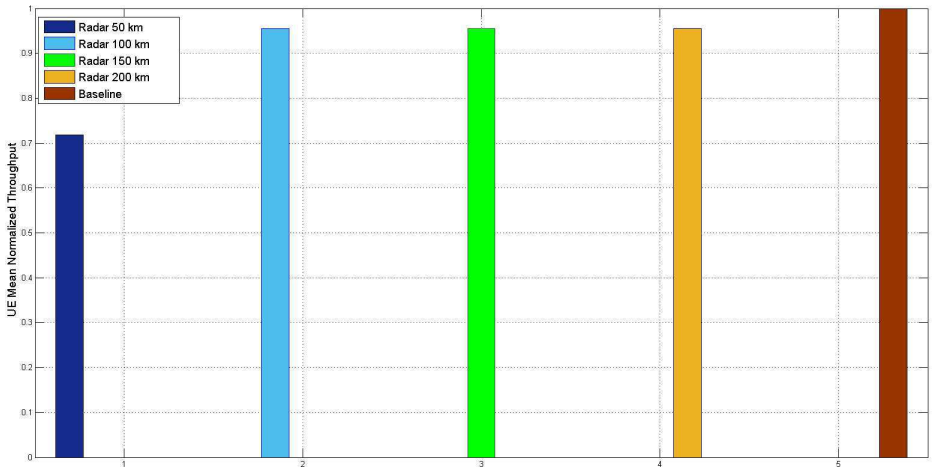


Figure 7.12: Normalized mean UE throughput for LTE macro cell and cochannel radar. Macro cell LTE operates within NTIA exclusion zones, and at 50 km the throughput loss is less than 30%.

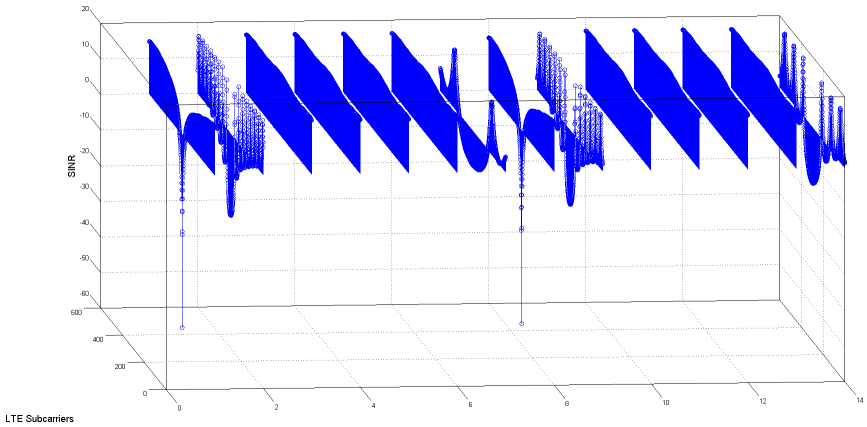


Figure 7.13: SINR for the LTE system in the UL direction when a cochannel radar at the center of the LTE band affects the BSs. SINR recover between the radar pulses and drops during the time the BS is hit by a radar pulse. This is promising as only certain symbols are affected in an LTE subframe.

close as 150 km away from a cochannel radar the throughput loss is less than 10%, whereas at 100 km the throughput loss becomes slightly less than 20%. At 50 km way the throughput loss is approximately 55% in contrast to the baseline (brown bar). We believe that the more decaying trend that we observe for the outdoor small cell LTE vs. the macro cells is because small cell LTE BSs are omnidirectional and are not downtilted ; so they are more exposed

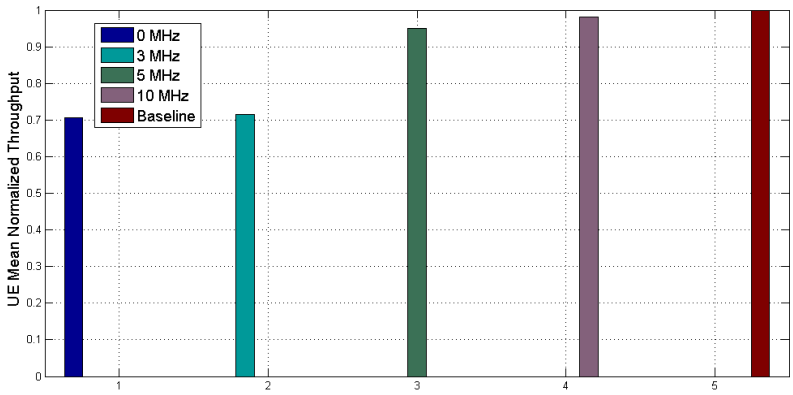


Figure 7.14: UE mean throughput for out-of-band interference from radar 50 km away from the macro cell LTE system. The throughput loss is drastically alleviated.

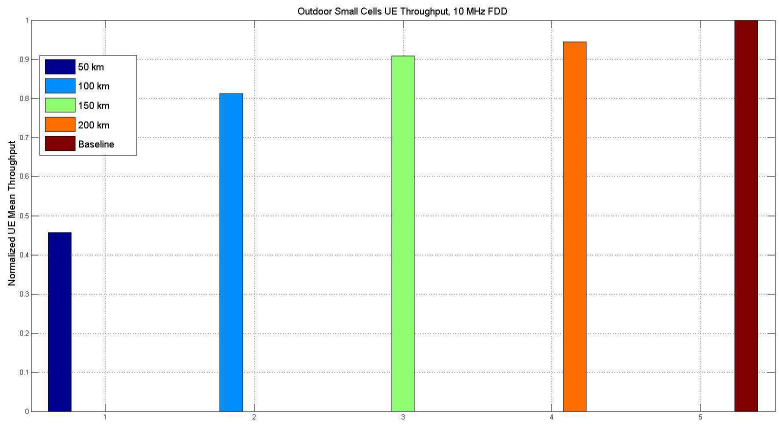


Figure 7.15: Normalized UE mean throughput for outdoor small cells and a cochannel radar. At 50 km, there is 55% throughput loss for the small cells.

to the radar radiations and causes a more dramatic loss as the radar gets closer to the LTE system.

7.4.6 Radar to Indoor Small Cell LTE UL Interference

Next, we show the simulation results for the outdoor small cell LTE. Similarly to the macro cells, we look at the mean UE throughputs. The radar is 50, 100, 150, and 200 km away from the small cell LTE. The normalized mean UE throughput is illustrated in Figure 7.16.

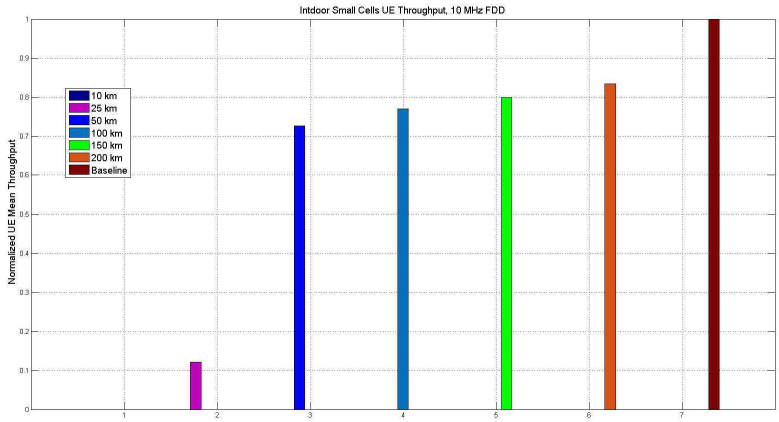


Figure 7.16: Normalized UE mean throughput for indoor small cells and a cochannel radar.

As we observe, the throughput loss for the LTE small cells is not dramatic until 100 km away from the radar. In fact, for an LTE outdoor small cell as close as 150 km away from a cochannel radar the throughput loss is less than 10%, whereas at 100 km the throughput loss becomes slightly less than 20%. At 50 km way the throughput loss is approximately 55% in contrast to the baseline (brown bar). We believe that the more decaying trend that we observe for the outdoor small cell LTE vs. the macro cells is because small cell LTE BSs are omnidirectional and are not downtilted; so they are more exposed to the radar radiations and causes a more dramatic loss as the radar gets closer to the LTE system.

Table 7.3: ITM Parameters

Parameters	Value
Operation Mode	Area Prediction Mode
Dielectric Constant	15
Conductivity	0.005 S/m
Refractivity	301 N-units
Climate	Continental Temperate
Variability Mode	Single Message
Surface Refractivity	15
Sitting Criteria	Random

7.5 Resource Allocation Simulation under Radar Interference

This section starts with simulating delay tolerant and real-time applications based on the sigmoidal and logarithmic utility functions in our system. Then, it implements the algorithms presented in the section 7.2.2 for a cellular network in the vicinity of a spectrally coexistent radar. All simulations are performed in MATLAB.

We consider a cellular network as Figure 7.17 with 3 sectors, operating in different frequency bands, whose topologically identical sectors form co-channels. For instance, the blue, red, and green cell phones indicate that their host sectors operate in distinct frequencies repeated with the same pattern over other cells of the infrastructure. Each cell is equipped with a BS, in charge of the UEs under its coverage area, collaborating with an MME unit which monitors the operation of all the BSs in the cellular network. The UEs run different applications whose utility function parameters are shown in the table, which represent sectors of the cells A, B, and C of the network in Figure 7.17 and 6 applications inside each sector. The 3 sigmoidal/logarithmic application utility functions per sector are characterized with the abbreviation "Sig"/"Log". For example, Table 7.4 says that the UE C5 is in cell C sector 1 and runs a logarithmic (delay-tolerant) application with parameter $k = 1.8$. Furthermore, we assume an approaching radar and red UE sectors use the same spectrum and 200/400 bandwidth units are available to the radar/communications system at maximum.

The sector rates are plotted in Figure 7.18, which shows the more resources available to BSs, the higher rate the sector rates. Furthermore, the radar-interfering red sector rates are 0 at stage 1 (Algorithms (17), (19), and (20)), whilst the non-interfering blue and green sectors are assigned resources from the radar spectrum. At stage 2 (Algorithms (18, (19), and (20)) non-radar carrier spectrum is allocated to all sectors. Particularly, the sharp increase of the red sectors rate vs. the other ones is due to no resource assignments to the red

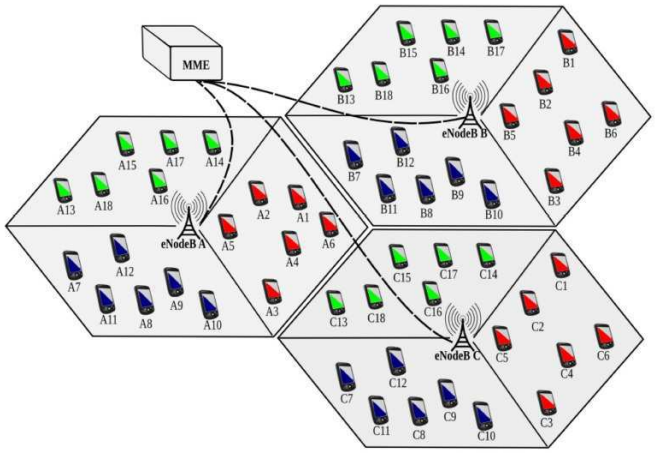


Figure 7.17: Cellular communications system: Three sector cells are covered by BSs, controlled by an MME unit. The cell phone callers indicate frequency bands such that sectors with the same colored cell phones are reusing identical bands. Such a topological reuse pattern is aimed at minimizing the inter-cell interference.

sectors at the preceding stage 1. At this time, no R^1_{comm} and R^2_{comm} allocations occur until R^3_{comm} equals their rates after which these non-interfering sectors also obtain more resources; Hence, giving the radar spectrum to non-interfering sectors initially then assigning much more radar-spectrum resources to interfering sectors hold an intrinsic fairness into the rate assignment procedure. Once some resources are passed to the interfering red sectors, the rate plots grow very close to each other due to the traffic analogy in the sectors (3 sigmoidal and 3 logarithmic utility functions).

To compare the deployment of the devised resource allocation and lack of it, we do the same experiment under the assumption of no shared spectrum, for which rate allocations are plotted in Figure 7.19. As we see, the red sector rates are allocated from the commencement of the allocation since no interfering sector exists. Besides, Fig. 7.18 shows that, at the rate 250, R^3_{comm} utilizes 50 bandwidth units from the communications spectrum whereas, before the rate 200, its allocated rate is 0. On the flip side, 7.19 depicts that at the rate 250, R^3_{comm} uses much more units of bandwidth.

Next, section 7.6 summarizes the results obtained in this chapter.

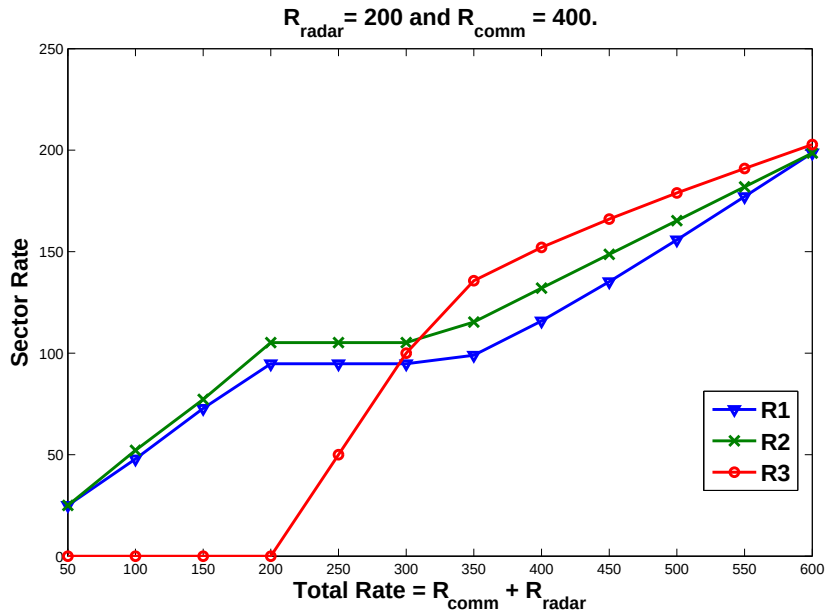


Figure 7.18: Rate allocation to the UEs when a radar interfering with sectors 3 of cells is in the vicinity of the cellular communications system. The allocated rates are initially allocated to non-interfering sectors from the radar operating frequency bands and no bandwidth is allocated to the interfering sectors. In the second stage of the allocation, the rates start allocating to the interfering sectors only in order to insure fairness. Once certain bandwidth is allocated to the sector, the remainder of the bandwidths can be allocated to all the sectors.

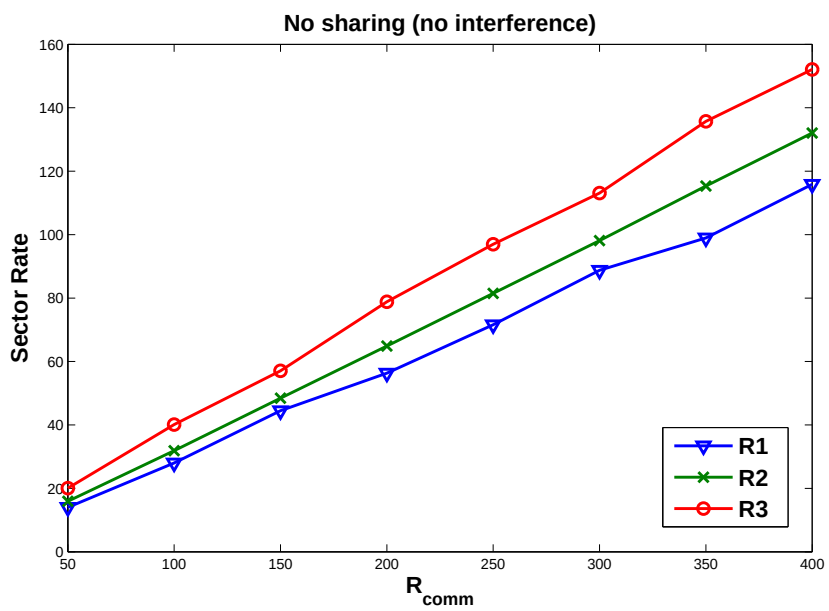


Figure 7.19: Rate allocation to the UEs when no radar is in the vicinity of the cellular communications system. The allocated rates are similar since the pattern of the applications in the cells are alike.

7.6 Chapter Summary

In this chapter, we studied the impact of shipborne S-band radars that are co-channel with and in the vicinity of a 3.5 GHz macro cell and outdoor small cell LTE systems. We also looked at the out-of-band interference for the macro cell LTE in the 3.5 GHz band. We leveraged the 3GPP to simulate the LTE at a system level in the UL direction. Furthermore, we simulated a rotating radar with parameters from the NTIA report [2] which had led to large exclusion zones. Moreover, we simulated FSPL and ITM to model the LoS and NLoS diffraction and troposcatter losses that attenuate radar signals to obtain relevant signal levels at the LTE sites. In the simulations, we assessed the radar impact by observing the SINR for symbol and subcarrier indices per TTI. We realized the radar presence causes BSs SINR drops during pulses, but the LTE SINR recovers in the time between the radar pulses. Furthermore, we looked at the UE mean throughputs when radar interference occurs.

Contrasting the baseline with interference cases at distinct distances between the radar and LTE systems which included operation within NTIAs exclusion zones, we found that both macro cell and outdoor small cell LTE can operate inside the exclusion zones, and these zones are overly conservative. So, it would be premature to lock the exclusion zones at such large distances. Furthermore, we looked at the out-of-band radar interference effects and showed that operating out-of-band significantly improves the LTE throughput loss operation in close distances between the radar and LTE systems. In view of the results of the current article, the authors foresee a significant reduction of the NTIA exclusion zones and propose that reducing the zones is constructive to sharing the 3.5 GHz band with government and motivates the spectrum sharing in the other bands as well. Furthermore, resource allocation in congested/contested environments is feasible.

In this chapter, we presented a two-stage novel resource allocation optimization method which assigned bandwidth to the UEs of cellular communications networks operating in

the vicinity of a radar such that the radar spectrum was shared with certain sectors of the communications infrastructure. First, we formulated the rate allocation process as two convex optimization problems which initially assigned resources from the radar spectrum to the UEs in the non-interfering sectors. Then, we allocated the communications spectrum , not coinciding with that of the radar, to all the sectors. The final solution was a set of optimal rates for the UEs in the cellular environment based on their running applications. Moreover, we discussed that, intrinsic to the proportional fairness formulation, the rate allocations dropped no users so that it could warrant a minimum QoS for all running applications.

We demonstrated that the devised two-stage resource allocation scheme not only refrained from any interference between the radar system and sectors of the cellular infrastructure, but it also rendered the rate allocation mechanism fair by incorporating the amount of radar spectrum assigned to non-interfering sectors during the first stage of the allocation into the resource allocation optimization problem at the second stage. Thereby, the interfering sectors were given more resources at the commencement of the second stage to compensate for the lack of assignment in the first stage. Our simulation results validated that the proposed modus operandi afforded the cellular communications network to allocate resources to UEs both to provision applications QoS and to eschew from any interference with neighboring spectrally coexistent radar systems, simultaneously.

Table 7.4: Application Utility Functions of the Cellular Network.

Sector 1 - Cell A						
A1	Sig	a = 3	b = 10.0	A4	Log	k = 1.1
A2	Sig	a = 3	b = 10.3	A5	Log	k = 1.2
A3	Sig	a = 1	b = 10.6	A6	Log	k = 1.3
Sector 2 - Cell A						
A7	Sig	a = 3	b = 10.0	A10	Log	k = 1
A8	Sig	a = 3	b = 15.3	A1	Log	k = 2
A9	Sig	a = 3	b = 12.0	A12	Log	k = 3
Sector 3 - Cell A						
A13	Sig	a = 3	b = 15.1	A16	Log	k = 10
A14	Sig	a = 3	b = 15.3	A17	Log	k = 11
A15	Sig	a = 3	b = 15.5	A18	Log	k = 12
Sector 1 - Cell B						
B1	Sig	a = 3	b = 15.9	B4	Log	k = 1.4
B2	Sig	a = 3	b = 11.2	B5	Log	k = 1.5
B3	Sig	a = 1	b = 11.5	B6	Log	k = 1.6
Sector 2 - Cell B						
B7	Sig	a = 3	b = 13	B10	Log	k = 4
B8	Sig	a = 3	b = 14	B11	Log	k = 5
B9	Sig	a = 1	b = 15	B12	Log	k = 6
Sector 3 - Cell B						
B13	Sig	a = 3	b = 15.7	B16	Log	k = 13
B14	Sig	a = 3	b = 15.9	B17	Log	k = 14
B15	Sig	a = 3	b = 17.3	B18	Log	k = 15
Sector 1 - Cell C						
C1	Sig	a = 3	b = 11.8	C4	Log	k = 1.7
C2	Sig	a = 3	b = 12.1	C5	Log	k = 1.8
C3	Sig	a = 1	b = 12.4	C6	Log	k = 1.9
Sector 2 - Cell C						
C7	Sig	a = 3	b = 16	C10	Log	k = 7
C8	Sig	a = 3	b = 17	C11	Log	k = 8
C9	Sig	a = 1	b = 18	C12	Log	k = 9
Sector 3 - Cell C						
C13	Sig	a = 3	b = 17.5	C16	Log	k = 16
C14	Sig	a = 3	b = 17.7	C17	Log	k = 17
C15	Sig	a = 3	b = 17.9	C18	Log	k = 18

Chapter 8

Resource Allocation, LTE, and Channel Effect

In chapters 3 and 4, we introduced a novel convex utility proportional fairness maximization for optimal resource allocation in wireless networks and outfitted the optimization with the subscriber, application status, and service differentiations parameterized respectively as UE subscription weights, application status weights, and application utility functions. Over there, we developed a centralized architecture for the proposed resource allocation which assigned application rates by the eNBs in a single stage in response to the application utility parameters sent by the UEs to the eNBs. Moreover, we provided with a distributed architecture for the same radio resource allocation framework which was introduced in chapter 3, which accounted for application types and temporal usages as well as UE priorities, and assigned application rates in two stages from the eNBs to the UEs and by the UEs to the running applications. While we saw the efficacy of the proposed methodology in a real-world Wifi network in chapter 3, in a realistic large scale wireless communications system, there are other factors to be accounted for. Radio waves undergo various propagation effects [176] including path loss, absorption by Oxygen and water vapor [8], and diffraction loss, collectively

referred to as the channel.

When said that a UE has a bad channel, it means that based on its current position with respect to its serving BS, the signals transmitted between the BS and the UE and vice verse suffer from severe path loss, and possibly diffraction loss so that the throughput for the UE is much lower. Transmitting under bad condition leads to transmission loss and bit errors, therefore in order to transmit under bad channel conditions, lower order modulations/coding schemes (MCS) are used. Lower coding MCS means that lower number of bits can be transmitted with one symbol.

8.0.1 Contributions

In this chapter:

- We show the inefficacy of the resource allocation methods introduced in chapters 3 and 4 under various channel conditions.
- We develop a distributed channel-aware architecture for the proposed resource allocation framework.
- We prove that the new channel-aware proposed mechanism is convex.
- We provide with solution algorithms for the channel-aware resource allocation.
- we provide with simulation results under isolated scenarios to show the effectiveness of the proposed method in resource allocation with channel considerations.
- We show the results of the resource allocations under the abundance as well as scarcity of resources with channel effects considered.
- We provide with a large scale simulation over a large geographic area and show the application of the proposed method.

8.0.2 Organization

The remainder of this chapter is organized as follows. Section 8.1provides the motivation for including the channel in the resource allocation and its necessity. Section 8.2 discusses radio resource management in LTE systems. . Section 8.3 discusses the inefficiencies that will come to happen by not considering the channel effect. Section 8.4 presents a channel-aware distributed architecture for the resource allocation framework that was introduced in chapter 4. Section 8.5 provides with the solution algorithms for the resource allocation. Section 8.6 provides with the simulation results; And, section 8.7 concludes and summarizes the current chapter.

8.1 Channel Quality, Modulation, and Coding

Channel coding, as a pillar in digital communication systems, can be considered as one of the differences between analog and digital systems and it makes error detection/correction feasible. Error correction can be in the form of an Automatic Repeat Request (ARQ) , where RX requests a retransmission of data in case an error is detected. Another error correction mechanism is Forward Error Correction (FEC) , where redundant bits are added to the data bits under a block or convolutional coding [179]. Another mechanism is Turbo coding [176], whose performance is within a few tenth of a dB from the Shannons limit [6].

Link adaptation is another feature of modern wireless systems and is a mechanism to match transmission parameters to the channel automatically. LTE link adaptation leverages the Adaptive Modulation and Coding (AMC) , in which if the SINR is sufficiently high, higher-order modulation schemes with higher spectral efficiency are used. This leads to a higher bit rates, whereas a lower order modulation scheme, which is essentially more robust to transmission errors, yields in lower spectral efficiency. On the other hand, for a

given modulation scheme, an appropriate code rate can be chosen depending on the channel quality [6]. A better channel quality allows for a higher code rate which in turn results in a higher data rate. LTE does this by means of a Rate Matching module (RMM) after the Turbo encoder to permit choosing proper code rates through puncturing and repetition [176]. Figure 8.1 shows the signal generation chain of an LTE PHY with Turbo coding and modulation modules.

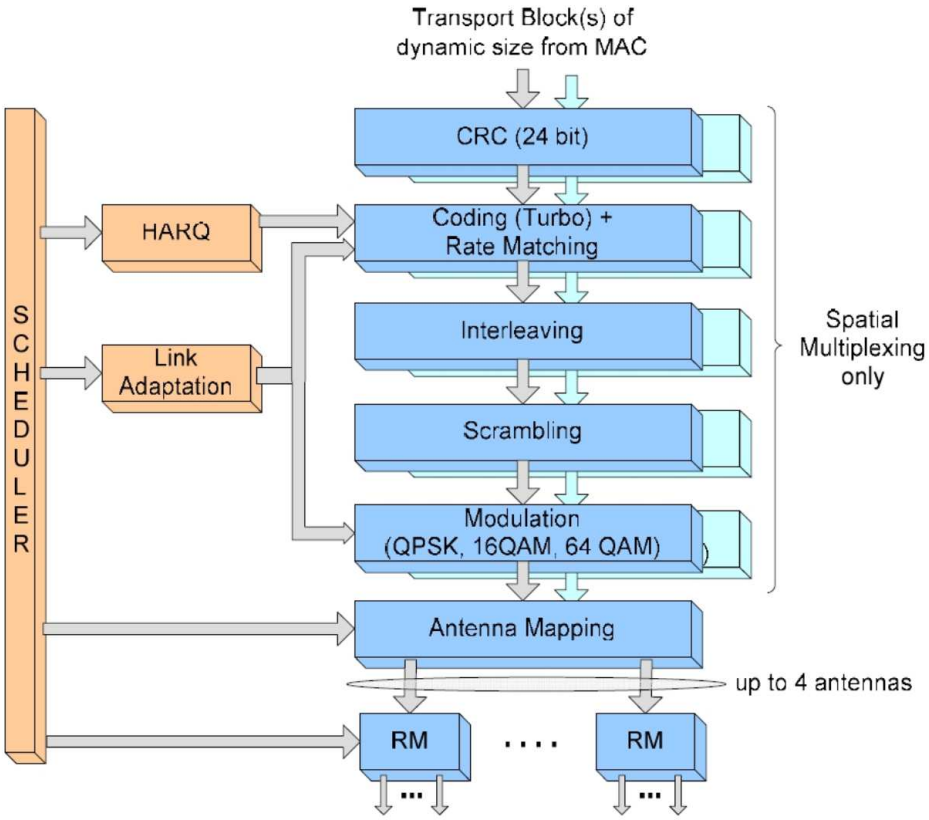


Figure 8.1: Link Adaptation System: LTEs physical layer with Turbo coding and modulation modules [6].

As we mentioned, the channel quality is an important aspect of wireless systems. In LTE, the quality of DL channel is measured in the UE for the reference symbols [9] and transmitted to the eNB as Channel Quality Indicator (CQI) . CQI depends on on the channel, noise, interference, and RX quality (e.g. analog front end noise figure (NF) and digital signal processing (DSP) algorithms performance. That means a receiver with better front end or more powerful signal processing algorithms delivers a higher CQI. The SNR recommends a

Table 8.1: LTE DL Link Adaptation

MCS, Code Rate, Coding Efficiency			
0	No Transmission	-	-
1	$QPSK$	0.76	0.1523
2	$QPSK$	0.12	0.2344
3	$QPSK$	0.19	0.377
4	$QPSK$	0.3	0.6016
5	$QPSK$	0.44	0.877
6	$QPSK$	0.59	1.1758
7	$16QAM$	0.37	1.4766
8	$16QAM$	0.48	1.9141
9	$16QAM$	0.6	2.4063
10	$64QAM$	0.45	2.7305
11	$64QAM$	0.55	3.3223
12	$64QAM$	0.65	3.9023
13	$64QAM$	0.75	4.5234
14	$64QAM$	0.85	5.1152
15	$64QAM$	0.95	5.554

MCS to ensure that Block Error Rate (BLER) [179] is less than a threshold such as 0.1 [6].

In Table 8.1, CQIs corresponding to 16 possible MCSs are shown. CQI 1 is selected for the worst channel quality as it is the most robust transmission parameters, QPSK modulation and the lowest code rate 0.076. The highest MCSs are 64 QAM and 0.93 (CQI 15). This table assumes a slow fading channel so that it does not change between two consecutive CQI measurements, in other words, the channel coherent time does not exceed the CQI measurement period.

The link adaptation in UL is very similar, but eNB estimates channel quality by using the Sounding Reference Signals (SRSs) [9].

8.2 Radio Resource Management

In LTE at the PHY layer, resources are managed using Resource Management (RM) modules which assign data blocks to RRBs. One RRB includes 12 subcarriers and one time slot. The resource management in LTE can be seen in Figure 8.4, where CQI values are leveraged in choosing the RRBs. CQIs can be periodic where the Physical Uplink Control Channel (PUCCH) reports them, or Physical Uplink Shared Channel (PUSCH) reports aperiodic CQIs. The latter is used if BS wants the channel quality at a particular time. The bandwidth is divided into subbands consisting of RRBs, where the number of sub-bands is $N = \lceil \frac{N_{RB}^{DL}}{k} \rceil$, where k is the number of RRBs (Figure fig:Ch13RRB).

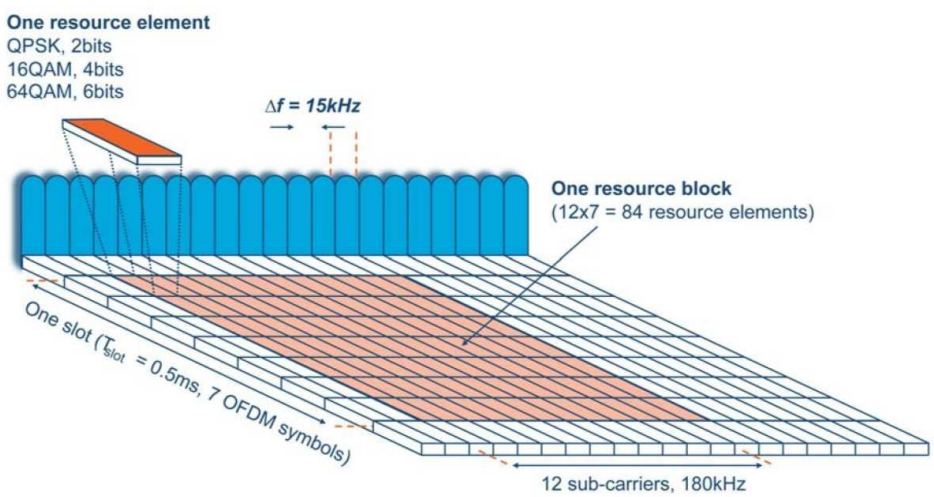


Figure 8.2: LTE Radio Resource Block [6].

The LTE has two modes of operation, FDD and TDD [176]. For the FDD mode, we have the following. Time duration for one frame is 10 ms, which means that there are 100 radio frames per second. There are 307200 samples in the 10 ms LTE frame, i.e. 30.72 million samples in 1 s. There are 10 subframes in a frame and 2 slots in 1 subframe. Thus, a frame consists of 20 slots which are 0.5 ms each. It is noteworthy that slot is not the smallest time unit in LTE. Normally ,each slot contains 7 small time blocks called symbols,

a certain time span of the signal that carry one spot in the I/Q constellation diagram [179]. Furthermore, at the beginning if each symbol, there is a small time span called cyclic prefix and the remaining is the symbol data. The cyclic prefix can be normal or extended, where the former leads to 7 data symbols and the latter yields in 6 data symbols. The explanations above are summarized in Figure 8.3.

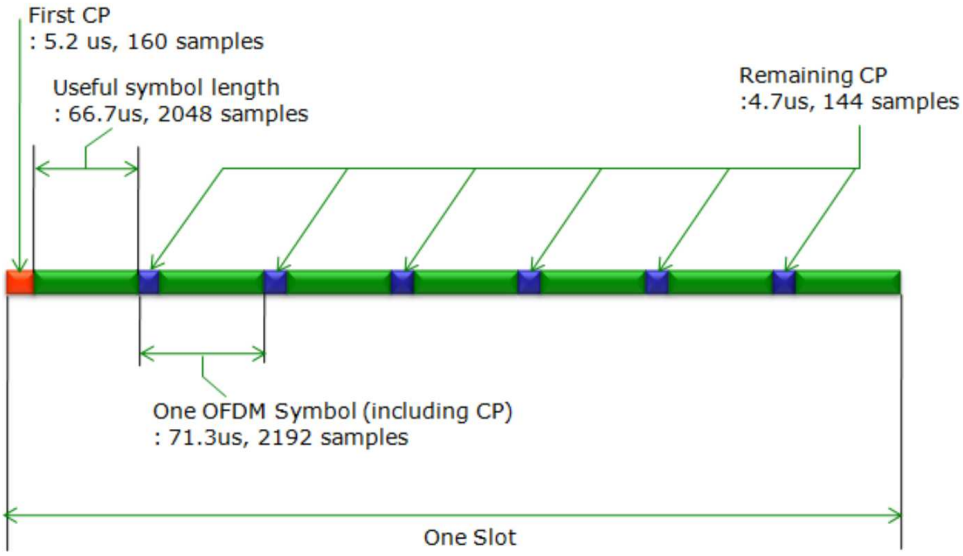


Figure 8.3: LTE Frame Structure [7].

For the DL, symbols are OFDM and for the UL, they are SC-OFDMA [176]. The first symbol is a little bit longer than the others. Since there are 30.072 million samples per second, there would be 2048 bins/IFFT. Since the spacing between the subcarriers is identical (15 kHz), changing the system bandwidth generates different number of subcarriers as we can see in table 8.2. As such, for 5, 10, and 20 MHz LTE, there are respectively 300, 600, and 1200 subcarriers. The symbols are 66.7 mirco ssecond, and each RRB contains 7 symbols, because it is 1 time slot, and 12 subcarriers which amounts to 84 resource elements in a RRB. Thus, 5, 10, and 20 MHz LTE contains 25, 50, and 100 RBs in that order. The details about TDD LTE can be found in [18]. While we present the formulations in this chapter general, our simulations are based on FDD due to simpler frame structure.

Table 8.2: LTE DL Link Adaptation

Number of RRBs, Number of IFFT		
1.4	6	128
3.0	15	256
5.0	25	512
10.0	50	1024
15.0	75	2048
20.0	100	2048

8.3 Resource Allocation Efficacy and Channel Conditions

The distributed architecture for optimal rate allocation in chapter 4 was in accordance with equation (8.1), repeated here for the ease of reference, where for M UEs covered by an eNB in accordance with Figure 4.1, $\mathbf{r} = [r_1, r_2, ..., r_M]$ is the UE allocated rate vector, R is the maximum available resources at the eNB, β_i is a subscription-dependent weight for the ith UE, and $V_i(r_i) = \prod_{j=1}^{N_i} U_{ij}^{\alpha_{ij}}(r_{ij})$ is the ith UE aggregate utility function.

$$\begin{aligned} \max_{\mathbf{r}} \quad & \prod_{i=1}^{M} V_i^{\beta_i}(r_i) \\ \text{subject to} \quad & \sum_{i=1}^{M} r_i \leq R, \\ & r_i \geq 0, \quad i = 1, 2, ..., M. \end{aligned} \tag{8.1}$$

And, the application rates was allocated using equation (8.2), where $\mathbf{r}_i = [r_{i1}, r_{i2}, ..., r_{iN_i}]$ is the application rate allocation vector such that its jth component indicates the bandwidth allotted by the ith UE to its jth application, r_i^{opt} is the ith UE rate allocated by eNB.

$$
\begin{aligned}
\max_{\mathbf{r}_i} \quad & \prod_{j=1}^{N_i} U_{ij}^{\alpha_{ij}}(r_{ij}) \\
\text{subject to} \quad & \sum_{j=1}^{N_i} r_{ij} \leq r_i^{\text{opt}}, \\
& r_{ij} \geq 0, \qquad j = 1, 2, ..., N_i.
\end{aligned}
\tag{8.2}
$$

Figure 8.4: System Model for Resource Allocation.

However, the model specified by equations (8.2) and (8.2) do not account for the channel conditions of the UEs as discussed in section 8.1. This is shown in Figure 8.5, from which we see that eNB A which is serving 2 of its UEs depicted is having a bad channel to the UE in the blue area of the graph, which represents low SNRs from the heat map index on the right side, whereas the other UE is observing an excellent channel by having a high SNR. Applying the algorithm from equations (8.2) and (8.2), does not differentiate between these two users at all since it only looks at UE subscriptions, instantaneous app usages, and application types. Therefore, two identical applications in terms of QoS requirements for two identical UEs in terms of subscription type and application focus will have identical resources allocated to them solely based on the bit rate requirements of the application and the UE and application weights. However, in practice, the throughput (in terms of bit per second) for the UE with bad channel condition might not be met since this UE has a bad channel and it has to use a much lower order of MCS in order to have a low BLER. In fact, for this situation, the UE with bad channel needs far more RRBs in order to have its QoS requirements met.

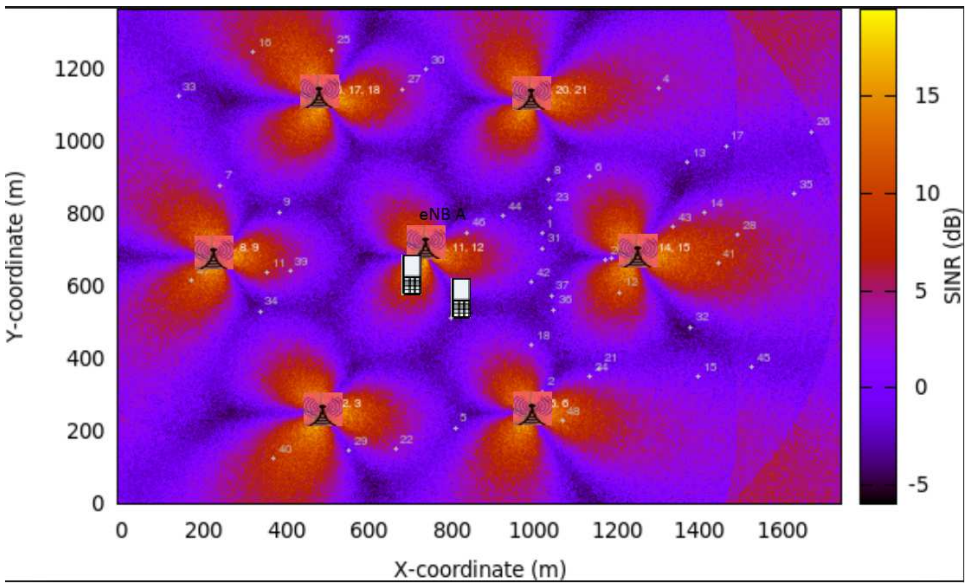


Figure 8.5: System Model and channel for Resource Allocation.

8.4 Channel-Aware Distributed Resource Allocation Formulation

Here, we define the number of bits that we can transmit with k_i resource elements at the signal-to-noise ratio SNR_i as function f shown in equation (8.3). The number of bits transmittable with k_i resource elements is k_i times the number of bits that can be transmitted with only 1 resource element under the same SNR. On the other hand, f_1 is upper bounded according to the Shannon law such that we can write equation (8.6), where β is the resource block bandwidth.

$$r_i = f(k_i, SNR_i) = k_i \underset{f_1}{f}(1, SNR_i) \tag{8.3}$$

Then, we can multiply by a proportionality factor to make it into an equation. In order to do this, we know that there are control and pilot per resource block. So, if we assume there are m non data control and pilot bits of total n resource elements available,then, $\frac{n-m}{n}$

comes into the equation. Furthermore, if the frame time is T_f and the resource block time is T_{RB}, then $\frac{T_{\text{RB}}}{T_f}$ comes to the equation as well.

$$r_i \propto k_i \beta log_2(1 + SNR_i) \tag{8.4}$$

Therefore, we can write the following equation.

$$r_i = \beta k_i \frac{T_{\text{RB}}}{T_f} \frac{n-m}{n} \epsilon_i log_2(1 + SNR_i) \tag{8.5}$$

As a case in point, if there were 4 pilots in a resource block and 84 (7 symbols and 12 subcarriers) resource elements in a resource block, and subcarriers were $15kHz$ so that the resource block was 180 kHz, considering 0.5 ms time slot (resource block time which consists of 7 symbols each 71.4 μ s.) leads to the following equation which is in kbps.

$$r_i = 8.5713 k_i \epsilon_i log_2(1 + SNR_i) \tag{8.6}$$

Then, the resource allocation in equation (8.1) can be written as equation (8.8), where $\mathbf{k}_i = [k_1, r_2, ..., r_M]$ is the resource block vector whose i^{th} element is the number of resource blocks allocated to the ith UE which is under SNR_i channel condition. The constraint in the optimization says that the total number of resource blocks should be less than or equal to the total number of data resource blocks which are available at the eNB and that the number of allocated resource blocks must be positive. We can use $\geq$ instead of $>$ in the latter condition, since nonzero allocation is implied by the multiplication of utilities (chapter 3).

$$\begin{aligned} \max_{\mathbf{k}} \quad & \prod_{i=1}^{M} V_i^{\beta_i}(\beta k_i \frac{T_{\mathrm{RB}}}{T_f} \frac{n-m}{n} \epsilon_i log_2(1+SNR_i)) \\ \text{subject to} \quad & \sum_{i=1}^{M} k_i \leq K, \\ & k_i \geq 0, \quad i = 1, 2, ..., M. \end{aligned} \tag{8.7}$$

It is noteworthy that to apply the proposed algorithms to a realistic scenario, one must obtain the values for the parameters K and epsilon. The latter was added to change the equation (8.6) to 8.5). It implies that the rate should be bounded by the Shannon capacity, and therefore it should be a function of code efficiency and SNR. Furthermore, the parameter K should be obtained by reducing the total number of resource blocks available minus the number of control, pilot, and synchronization channel.

8.4.1 Determining ϵ

The minimum SNRs required for different CQIs reported are given in the 3GPP documentation as in table 8.3. The first four columns are excerpted from table 8.1, and the columns 5 to 9 show the minimum SNRs required for various MIMO/HARQ modes at various CQIs [9]. Them we select the minimum SNR required, and the linear scale of this minimum SNR is depicted at the last column. Then, the parameter ϵ will be: as in equation (8.8).

$$\epsilon_i = \frac{\gamma_i}{log_2(1+SNR_{i,\mathrm{lin}})} \tag{8.8}$$

In order to have a continuo graph, we curve fit the values of this parameter ϵ vs. SNR based on linear regression, which yields in the equation (8.9), depicted in Figure 8.6.

$$\epsilon_i = 0.0087 SNR_{\mathrm{dB}} + 0.6821 \tag{8.9}$$

Table 8.3: Characterizing epsilon for different channels. (M: Modulation, CR: Coding Rate,SE: Spectral Efficiency, T: Transmit Mode [9])

CQI	M	CR	SE	T1	T2	T3	T4	T5	$SNR_{min,lin}$
0	NT	-	-	-	-	-	-	-	-
1	$QPSK$	0.76	0.1523	1.95	2.00	-7.00	-3.10	-4.80	0.199526231
2	$QPSK$	0.12	0.2344	4.00	4.05	-5.00	-1.15	-2.60	0.316227766
3	$QPSK$	0.19	0.377	6.00	5.10	-3.15	1.50	0.00	0.484172368
4	$QPSK$	0.3	0.6016	8.00	8.00	-1.00	4.00	2.60	0.794328235
5	$QPSK$	0.44	0.877	10.00	10.00	1.00	6.00	4.95	1.258925412
6	$QPSK$	0.59	1.1758	11.95	11.80	3.00	8.90	7.60	1.995262315
7	$16QAM$	0.37	1.4766	14.05	13.90	5.00	12.70	10.60	3.16227766
8	$16QAM$	0.48	1.9141	16.00	16.10	6.90	14.90	12.95	4.897788194
9	$16QAM$	0.6	2.4063	17.90	17.45	8.90	17.50	15.40	7.762471166
10	$64QAM$	0.45	2.7305	19.9	19.50	10.85	20.50	18.10	12.16186001
11	$64QAM$	0.55	3.3223	21.5	21.50	12.60	22.45	20.05	18.19700859
12	$64QAM$	0.65	3.9023	23.45	23.10	14.35	23.20	22.00	27.22701308
13	$64QAM$	0.75	4.5234	25.00	24.90	16.15	24.90	24.55	41.20975191
14	$64QAM$	0.85	5.1152	27.30	27.00	18.15	27.00	26.80	65.31305526
15	$64QAM$	0.95	5.554	22.00	29.10	20.00	29.10	29.60	100.00

Therefore, the equation (8.8) can be written as equation (8.10).

$$\begin{aligned} \max_{\mathbf{k}} \quad & \prod_{i=1}^{M} V_i^{\beta_i}(\beta k_i \frac{T_{\text{RB}}}{T_f}\frac{n-m}{n}(0.0087SNR_i + 0.6821)log_2(1+SNR_i)) \\ \text{subject to} \quad & \sum_{i=1}^{M} k_i \leq K, \\ & k_i \geq 0, \quad i = 1, 2, ..., M. \end{aligned} \tag{8.10}$$

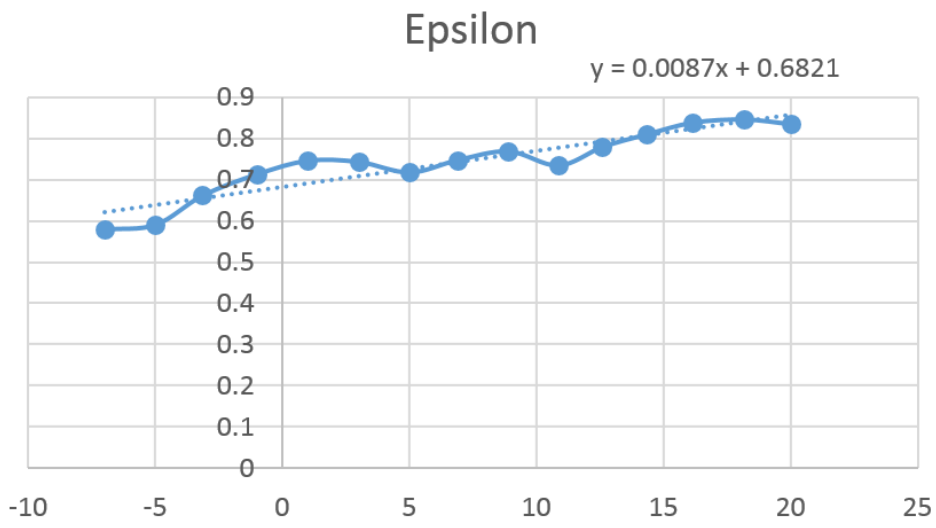


Figure 8.6: Parameter ϵ vs SNR.

It is noteworthy that we assume that the SNR is ergodic and that the resource blocks are assigned within the coherence time/bandwidth so that we only have have flat and slow fading.

8.4.2 Determining K

The number of available resource elements depends on the bandwidth for the LTE system and the mode of operation, i.e. FDD vs. TDD. The total number of resource elements in M resource blocks is $M \times 12 \times 7 \times 2 \times 10 \times 1000$, where 12 is the 12 subcarriers in 1 resource block, 7 is the number of symbols in 1 resource block, 2 is due to having 2 slots in a subframe, and 10 is due to having 10 subframes in 1 frame and 1000 is due to having 1000 frames in 1 second. However, there are data pilot and synchronization symbols which should be reduced from the total available to give out the number of data resource elements. This is depicted in table 8.4.

Table 8.4: Characterizing K for different channels. (BW: Bandwidth, TB: Transmit Bandwidth, TREPF: Total Resource Element per Frame, TREPS: Total Resource Element per Second, OREPF: Overhead Resource Element per Second, OREPS: Overhead Resource Element per Second, DREPS: Data Resource Element per Second).

BW	TB	PUCCH	TREPF	TREPS	OREPF	OREPS	DREPS
10	9	2	84000	84000000	16857	67143	67143000
20	18	2	168000	168000000	27454	140546	140546000

8.5 Global Solution Existence

In order to follow the same procedure as in chapters 3 and 4, we have to prove the the logarithm of the UE utility function is convex. It is noteworthy that factoring the channel effect into the optimization changes the shape of the objective function under equation (8.10) from the original sigmoidal or logarithmic form. Doing so, renders the optimization in equation (8.10) convex, and therefore, proves the existence of a globally optimal solution. In order to do this, first we present the following theorem.

Lemma 8.5.1. *The objective function of the optimization in equation (8.10) is convex.*

Proof. The sigmoidal utility function $U(r)$ from chapter 3 is now replaced with $U(f(k, SNR))$ from equation (8.3). Therefore, for the sigmoidal utility we have that:

$$\frac{\partial f}{\partial k} = \underbrace{\frac{\partial f}{\partial U}\frac{\partial U}{\partial r}}_{\frac{\partial log(U)}{\partial r}>0}\frac{\partial r}{\partial k} = \frac{\partial log(U)}{\partial r}(\beta k_i \frac{T_{\text{RB}}}{T_f}\frac{n-m}{n}\epsilon_i log_2(1+SNR_i)) > 0. \tag{8.11}$$

Now, we only should prove that the second derivative is negative. We have that:

$$\frac{\partial^2 log(U)}{\partial k^2} = \frac{\partial}{\partial k}\underbrace{(\frac{\partial log(U)}{\partial r})}_{g}\frac{\partial r}{\partial k}) = \\ \frac{\partial g}{\partial k}\underbrace{\beta k_i \frac{T_{\text{RB}}}{T_f}\frac{n-m}{n}\epsilon_i log_2(1+SNR_i)}_{.>0} + g\underbrace{\frac{\partial^2 r}{\partial k^2}}_{=0} \tag{8.12}$$

So the positiveness or negativeness depends on $\frac{\partial g}{\partial k}$.

$$\frac{\partial g}{\partial k} = \frac{\partial}{\partial k}(\frac{\partial log(U)}{\partial r}) = \\ \frac{\partial}{\partial k}(\frac{ade^{-a(r-b)}}{1-d(1+e^{-a(r-b)})} + \frac{ae^{-a(r-b)}}{1+e^{-a(r-b)}}) \tag{8.13}$$

Taking $\beta k_i \frac{T_{\text{RB}}}{T_f}\frac{n-m}{n}\epsilon_i log_2(1+SNR_i) - b)$ as Ψ, the derivative of the first term becomes:

$$\frac{\partial}{\partial k}\frac{ade^{-a(r-b)}}{1-d(1+e^{-a(r-b)})} = ade^{-a\Psi}1 - d(1+e^{-a\Psi}) = \underbrace{-a^2 d\Psi}_{A} e^{-a\Psi}\left[1-d(1+e^{-a\Psi})\right] \\ - Ae^{-a\Psi}(ade^{-a\Psi})/\left[1-d(1+e^{-a\Psi})\right]^2 \\ \frac{-aAe^{(}Bk-ba)\left[1-d(1+e^{-(Bk-ba)})\right] + de^{-(Bk-ba)}}{\left[1-d(1+e^{-(Bk-ba)})\right]^2} = \frac{-Aae^{-(Bk-ba)(1-d)}}{\left[1-d(1+e^{-(Bk-ba)})\right]^2} \tag{8.14}$$

And the derivative of the second term becomes:

$$\frac{\partial}{\partial k}\frac{ae^{-a(r-b)}}{1+e^{-a(r-b)}} = \frac{\partial}{\partial k}\frac{ae^{-aB}}{1+e^{-aB}} = \\ \frac{(-a^2\beta k_i \frac{T_{\text{RB}}}{T_f}\frac{n-m}{n}\epsilon_i log_2(1+SNR_i)e^{-aBk-ba})(1+e^{-a(Bk-b)})}{(1+e^{-a(Bk-ba)})^2} = \\ \frac{-aAe^{-(Bk-ba)}(1+e^{-(Bk-ba)}) + Aae^{-(Bk-ba)}e^{-(Bk-ba)}}{\left[1-d(1+e^{-(Bk-ba)})\right]^2} = \frac{-aAe^{-(Bk-ba)}}{\left[1-d(1+e^{-(Bk-ba)})\right]^2} \tag{8.15}$$

Therefore, the derivative comes as the following equation.

$$\frac{\partial g}{\partial k} = \frac{-Aae^{-(Bk-ba)(1-d)}}{[1-d(1+e^{-(Bk-ba)})]^2} + \frac{-aAe^{-(Bk-ba)}}{[1-d(1+e^{-(Bk-ba)})]^2} =$$
$$-Aae^{-(Bk-ba)} + \frac{d(1+e^{-(Bk-ba)})^2(1-d) + \left[1-d(1+e^{-(Bk-ba)})\right]^2}{(1-d(1+e^{-(Bk-ba)}))^2 d(1+e^{-(Bk-ba)})^2} = \tag{8.16}$$
$$\frac{-aAd^2(1+e^{-2(Bk-ba)})e^{-(Bk-ba)} + (2d+1)Aae^{-(Bk-ba)}}{(1-d(1+e^{-(Bk-ba)}))^2 d(1+e^{-(Bk-ba)})^2}$$

Thus, the second derivative is negative and hence it is a concave function. The proof for the logarithmic utility is similar □

Theorem 8.5.2. *The optimization problem in equation (8.8) is convex and has a unique tractable global optimal solution.*

Proof. We can form the equivalent log-EURA optimization (8.17) by taking the logarithm of the (8.8)

$$\begin{aligned} \max_{\mathbf{k}} \quad & \sum_{i=1}^{M} \beta_i \log(V_i(\beta k_i \frac{T_{\text{RB}}}{T_f} \frac{n-m}{n}(0.0087SNR_i + 0.6821)log_2(1+SNR_i))) \\ \text{subject to} \quad & \sum_{i=1}^{M} r_i \leq K, \\ & k_i \geq 0, \quad i = 1, 2, ..., M. \end{aligned} \tag{8.17}$$

The aggregate utility concavity based on lemma 8.5.1 concludes that the channel-aware log-EURA optimization is convex [88], which in turn proves the convexity of the EURA problem in equation (8.8) due to their objective functions equivalence. And, there exists a unique tractable global optimal solution for a convex optimization in general [88], and for EURA in particular. □

8.5.1 Solution for Channel-Aware EURA Optimization

The solution to the distributed architecture channel-aware for resource allocation is provided through the Lagrangian of the dual problems for the EURA optimization. Similarly to ([84, 89]), we deploy the duality for convex optimization problems to solve them efficiently. What proceeds is such an application of the duality to EURA and IURA constituents of the distributed rate allocation problem.

$$\begin{aligned} L(\mathbf{k}, p) &= \sum_{i=1}^{M} \log(V_i(\beta k_i \frac{T_{\text{RB}}}{T_f} \frac{n-m}{n}(0.0087SNR_i + 0.6821)log_2(1+SNR_i))) - p(\sum_{i=1}^{M} k_i + z - K) \\ &= \sum_{i=1}^{M} \left(\log(V_i(\beta k_i \frac{T_{\text{RB}}}{T_f} \frac{n-m}{n}(0.0087SNR_i + 0.6821)log_2(1+SNR_i))) - pk_i \right) + p(K-z) \\ &= \sum_{i=1}^{M} L_i(k_i, p) + p(K-z) \end{aligned} \tag{8.18}$$

Here, $z_i \geq 0$ is the slack variable and p is Lagrange multiplier or the shadow price (price per unit bandwidth for all the M channels). Therefore, the ith UE bid for bandwidth can be written as $w_i = pk_i$, where $\sum_{i=1}^{M} w_i = p\sum_{i=1}^{M} k_i$. The first term in equation (4.7) is separable in k_i, so we have $\max_{\mathbf{k}} \sum_{i=1}^{M}(\log(V_i(\beta k_i \frac{T_{\text{RB}}}{T_f} \frac{n-m}{n}(0.0087SNR_i + 0.6821)log_2(1+SNR_i))) - pk_i) = \sum_{i=1}^{M} \max_{k_i}(\log(V_i(\beta k_i \frac{T_{\text{RB}}}{T_f} \frac{n-m}{n}(0.0087SNR_i + 0.6821)log_2(1+SNR_i))) - pk_i)$ and the dual problem objective function can be written as equation (8.19).

$$
\begin{aligned}
D(p) =& \max_{\mathbf{k}} L(\mathbf{k}, p) \\
=& \sum_{i=1}^{M} \max_{k_i} \left(\log(V_i(\beta k_i \frac{T_{\text{RB}}}{T_f} \frac{n-m}{n}(0.0087SNR_i + 0.6821)log_2(1+SNR_i))) - pk_i \right) \\
&+ p(K-z) = \sum_{i=1}^{M} \max_{k_i}(L_i(k_i, p)) + p(K-z)
\end{aligned}
\tag{8.19}
$$

Thus, the dual problem is formulated as equation (8.20).

$$
\begin{aligned}
\min_{p} \quad & D(p) \\
\text{subject to} \quad & p \geq 0.
\end{aligned}
\tag{8.20}
$$

Leveraging the method of Lagrange multipliers, we have:

$$
\frac{\partial D(p)}{\partial p} = K - \sum_{i=1}^{M} k_i - z = 0 \tag{8.21}
$$

Substituting by $\sum_{i=1}^{M} w_i = p \sum_{i=1}^{M} k_i$, we have equation (8.22), minimized to $p = \frac{\sum_{i=1}^{M} w_i}{K}$ at $z = 0$ where $w_i = pk_i$ is transmitted by the ith UE to the eNB.

$$
p = \frac{\sum_{i=1}^{M} w_i}{K - z} \tag{8.22}
$$

As such, we divide the channel-aware log-EURA problem (8.17) into simpler optimizations at the eNB (eNB EURA problem) and UEs (UE EURA problem). These are respectively shown in equations (8.24) and (8.23) whose solutions, guaranteeing the utility proportional fairness in equation (4.3), are summarized in Algorithms 22 and 21 in that order.

$$\begin{aligned} \max_{r_i} \quad & \log V_i(r_i) - pr_i \\ \text{subject to} \quad & p \geq 0 \\ & r_i \geq 0, \quad i = 1, 2, ..., M. \end{aligned} \tag{8.23}$$

During the execution of the aforesaid algorithms, starting with $w_i(0) = 0$, the ith UE, transmits an initial bid $w_i(1)$ to the eNB, which in turn subtracts the latterly received bid $w_i(n)$ and the formerly received one $w_i(n-1)$ and ceases the procedure if the difference is less than a threshold δ; Otherwise, it computes and sends a shadow price $p(n) = \frac{\sum_{i=1}^{M} w_i(n)}{R}$ to is covered UEs. The ith UE extracts its rate $k_i(n)$ from the received $p(n)$ such that $\log V_i(\beta k_i \frac{T_{\text{RB}}}{T_f} \frac{n-m}{n}(0.0087SNR_i + 0.6821)log_2(1 + SNR_i)) - p(n)k_i$ is maximized. The rate $k_i(n)$ is employed to estimate the new bid $w_i(n) = p(n)k_i(n)$, transmitted to the eNB. This routine repeats until the bid difference $|w_i(n) - w_i(n-1)|$ falls below the threshold δ.

$$\begin{aligned} \min_{p} \quad & D(p) \\ \text{subject to} \quad & p \geq 0. \end{aligned} \tag{8.24}$$

The solution $k_i(n)$ of the ith UE EURA optimization can be written as equation (8.25), and Algorithm 21 essentially solves the equation (8.26) algebraically the Lagrange multiplier solution for equation (8.23) and geometrically the intersection point of the horizontal line $y = p(n)$ with the curve given by equation (8.27).

$$k_i(n) = \arg\max_{k_i} \left(\log V_i(\beta k_i \frac{T_{\text{RB}}}{T_f} \frac{n-m}{n}(0.0087SNR_i + 0.6821)log_2(1 + SNR_i)) - p(n)k_i \right) \tag{8.25}$$

$$\frac{\partial \log V_i(\beta k_i \frac{T_{\text{RB}}}{T_f} \frac{n-m}{n}(0.0087SNR_i + 0.6821)log_2(1 + SNR_i))}{\partial k_i} = p(n) \tag{8.26}$$

$$y = \frac{\partial \log V_i(\beta k_i \frac{T_{\text{RB}}}{T_f} \frac{n-m}{n}(0.0087SNR_i + 0.6821)log_2(1 + SNR_i))}{\partial k_i} \tag{8.27}$$

Algorithm 21 UE Channel-Aware EURA Optimization Algorithm

Send initial bid $w_i(1)$ to eNB.
loop
  Receive shadow price $p(n)$ from eNB.
  if STOP from eNB **then**
    Calculate allocated rate $k_i^{\text{opt}} = \frac{w_i(n)}{p(n)}$.
    STOP
  else
    Solve $k_i(n) = \arg\max_{k_i}\Big(\log V_i(\beta k_i \frac{T_{\text{RB}}}{T_f} \frac{n-m}{n}(0.0087SNR_i + 0.6821)log_2(1 + SNR_i)) - p(n)k_i\Big)$.
    Send new bid $w_i(n) = p(n)k_i(n)$ to eNB.
  end if
end loop

Algorithm 22 eNB EURA Optimization Algorithm

loop
  Receive bids $w_i(n)$ from UEs. {Let $w_i(0) = 1\ \forall i$}
  if $|w_i(n) - w_i(n-1)| < \delta\ \forall i$ **then**
    Allocate rates, $k_i^{\text{opt}} = \frac{w_i(n)}{p(n)}$ to user i.
    STOP
  else
    Calculate $p(n) = \frac{\sum_{i=1}^{M} w_i(n)}{K}$.
    Send new shadow price $p(n)$ to all UEs.
  end if
end loop

8.5.2 IURA Global Optimal Solution

Strictly increasing nature of logarithm function yields in that the IURA objective function in equation (8.2), i.e. $\prod_{j=1}^{N_i} U_{ij}^{\alpha_{ij}}(r_{ij})$, corresponds to $\sum_{j=1}^{N_i} \alpha_{ij} \log(U_{ij}(r_{ij}))$. So equation (4.4) can be reformulated as equation 8.28, referred to as the log-IURA problem for which corollary 4.2.3 is conceivable. It is noteworthy that for equation (8.28), the value of r_i is given as equation (8.29).

$$\begin{aligned} \max_{\mathbf{r}_i} \quad & \sum_{j=1}^{N_i} \alpha_{ij} \log U_{ij}(r_{ij}) \\ \text{subject to} \quad & \sum_{i=1}^{N_i} r_{ij} \leq \beta k_i^{opt} \frac{T_{\text{RB}}}{T_f} \frac{n-m}{n} (0.0087 SNR_i + 0.6821) log_2(1 + SNR_i), \\ & r_{ij} \geq 0, \quad j = 1, 2, ..., N_i. \end{aligned} \tag{8.28}$$

$$r_i^{opt} = (\beta k_i^{opt} \frac{T_{\text{RB}}}{T_f} \frac{n-m}{n} (0.0087 SNR_i + 0.6821) log_2(1 + SNR_i)) \tag{8.29}$$

Corollary 8.5.3. *The IURA optimization problem in equation (8.2) is convex and has a unique tractable global optimal solution.*

Proof. Substantiating lemma 8.5.1 is concomitant with proving the concavity of the application utility functions natural logarithm; this ascertains the convexity of the log-IURA problem in equation (8.28) [88]. Since log-IURA and IURA optimizations have equivalent objective functions, IURA optimization in equation (8.2) is also convex. Every convex optimization has a tractable global optimal solution in general [88], and so does the IURA optimization in particular. □

Theorem 8.5.2 and corollary 8.5.3 indicate that the distributed optimization in section 8.3 is convex and it assigns rates optimally. The application rates r_{ij} are optimally assigned internally to the UEs in accordance with Algorithm 23, where the ith UE leverages the EURA allocated rate r_i^{opt} to solve $\mathbf{r}_i = \arg\max_{\mathbf{r}_i} \sum_{j=1}^{N_i} (\alpha_{ij} \log U_{ij}(r_{ij}) - p_I r_{ij}) + p_I r_{ij}^{\text{opt}}$.

Algorithm 23 UE IURA Algorithm

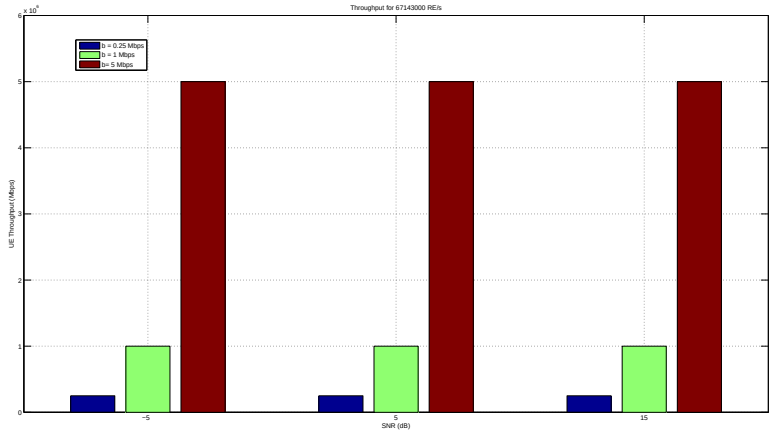
loop
  Receive r_i^{opt} from eNB. {by EURA Algorithms}
  Solve
  $\mathbf{r}_i = \arg\max_{\mathbf{r}_i} \sum_{j=1}^{N_i} (\alpha_{ij} \log U_{ij}(r_{ij}) - p_I r_{ij}) + p_I r_i^{\text{opt}}$ $\{\mathbf{r}_i = \{r_{i1}, r_{i2}, ..., r_{iN_i}\}\}$
  Allocate r_{ij} to the jth application.
end loop

8.6 Simulation Results

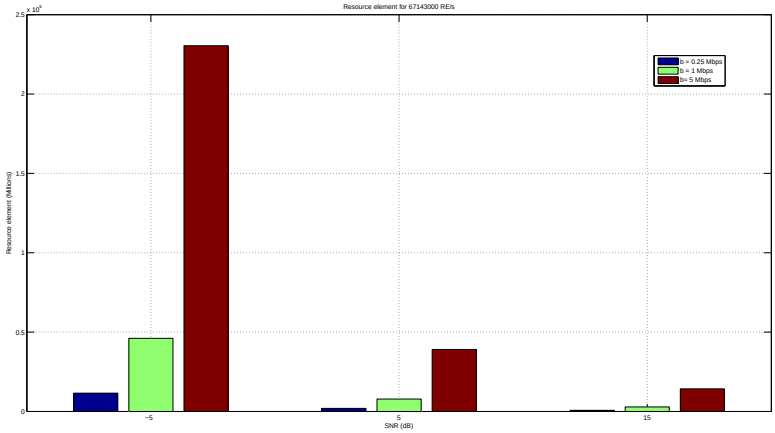
We consider 9 UEs and a BS serving the UEs. We assume a 10 MHz LTE system so that 67143000 resource elements are available. Then, we apply the algorithm to observe the rates allocated to the UEs and to the applications. We assume esodic SNRs. Then, the throughput and number of resources allocated are illustrated in Figures 8.7(a) and 8.7(b). Here, the bit rate requirements of the UEs was 0.25, 1, and 5 Mbps and the SNRs was -5, 5, and 15 dB as we can see on the x axis. As we can observe, for identical QoS requirements, more resources are allocated by the algorithm to the UEs at lower SNRs, i.e. worse channel quality. On the other hand, a smaller number of resources meets the same bit rate requirements as the lower SNR ones as a higher MCS can be used. Furthermore, the abundance of the resources allows the UEs to have their throughput met as we can see in Figure 8.7(a).

Next, we consider the same 9 UEs but we put two applications in each UE and reduce resources to 5000000 resource elements. As we can see, more resources are allocated to the UEs with lower SNRs in order to meet their bit rate requirements (Figure 8.7(a)). The throughput of the UEs are met as it is shown in Figure 8.7(b), and IURA algorithm distributed the resources to its applications. Since the applications are identical, resources are allocated between them equally which is shown by the stack bar chart 8.8(c)

Next, we reduce the resources to only 1000000 resource elements for the 9 UEs. As we can see from thr throughput plot in Figure 8.9(a), the throughput of the UEs under band channel conditions are not met. This is in spite of the fact that more resources were allocated



(a) Throughput of the UEs

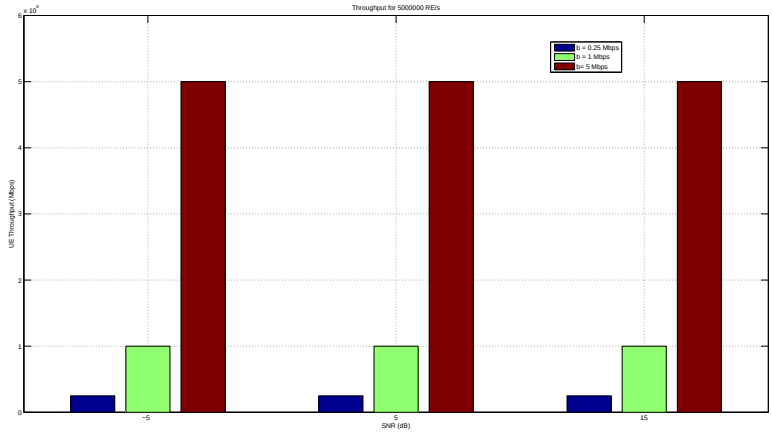


(b) Resources Allocated to UEs

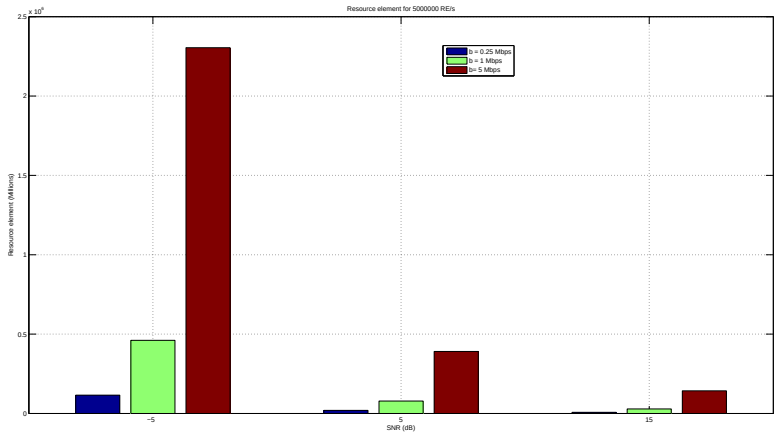
Figure 8.7: The system contains 9 UEs, each concurrently running a real-time application with respective identically colored sigmoidal utility functions. The lower SNR UEs are allocated more resources so that their throughput meet their QoS requirements.

to these UEs according to Figure 8.9(b); however, there were simply not enough resources available to meet the QoS requirements of the applications running on these UEs.

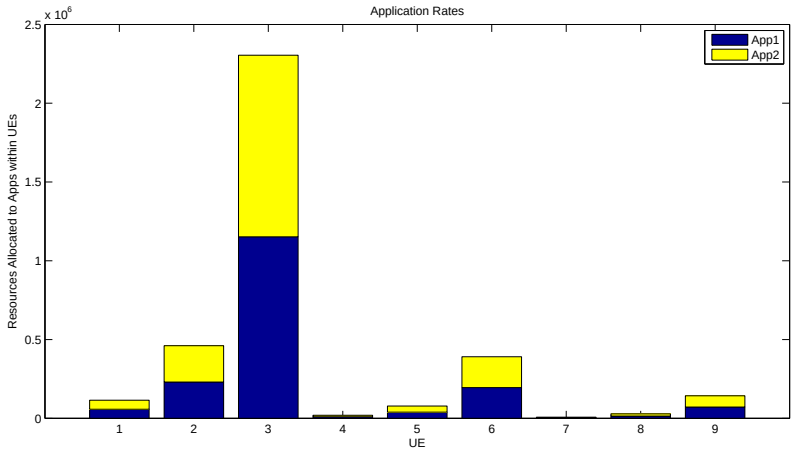
Next, we reduce the number of resources to only 100000. Under this circumstance, network is severely suffering and the throughput of no application is met as there are no enough resources available.



(a) Throughput of the UEs

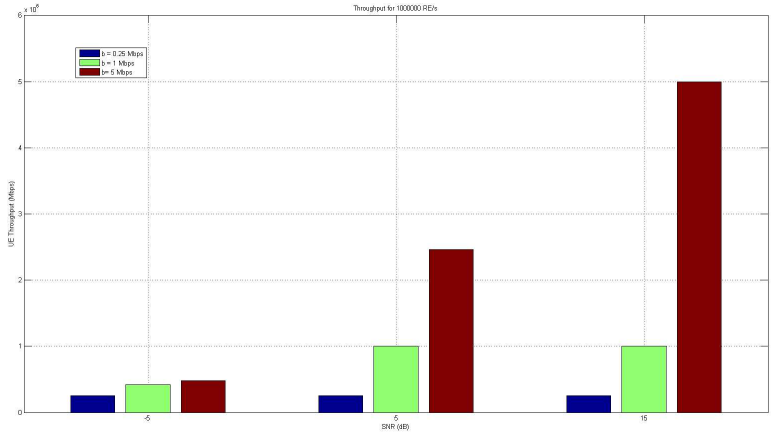


(b) Resources Allocated to UEs

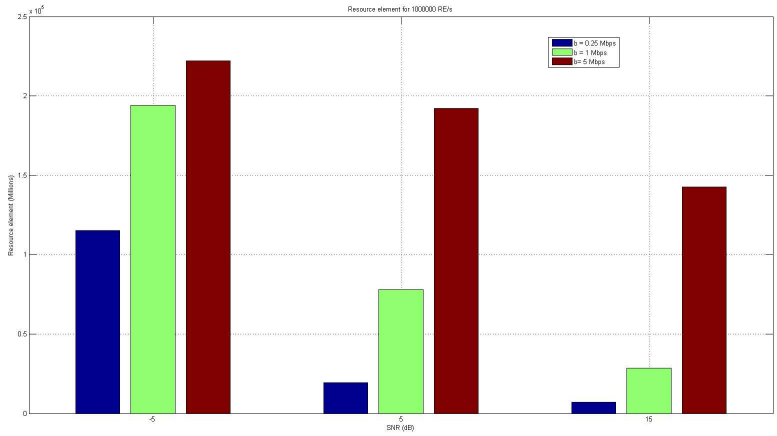


(c) Resources Allocated to UEs

Figure 8.8: The system contains 9 UEs, each concurrently running two real-time application with respective identically colored sigmoidal utility functions. The lower SNR UEs are allocated more resources so that their throughput meet their QoS requirements.



(a) Throughput of the UEs

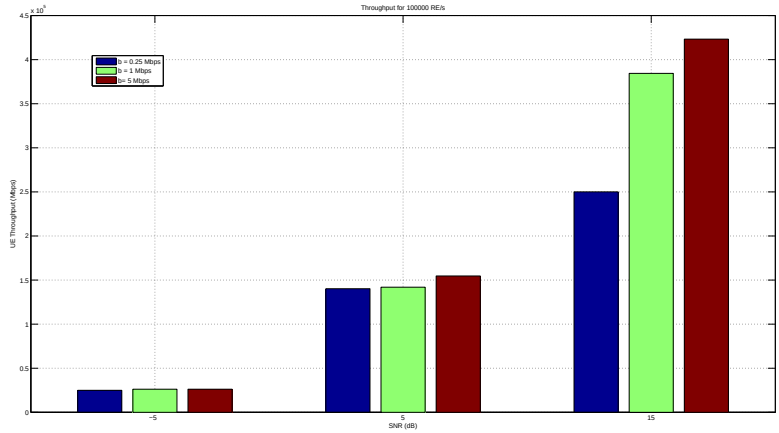


(b) Resources Allocated to UEs

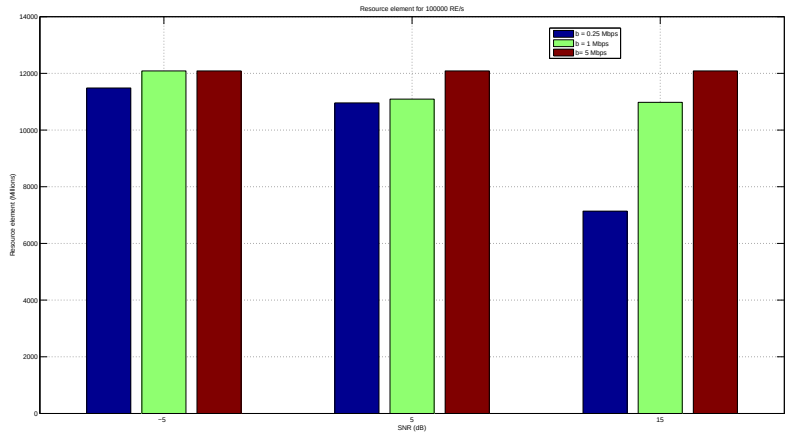
Figure 8.9: The system contains 9 UEs, each concurrently running a real-time application with respective identically colored sigmoidal utility functions. The lower SNR UEs are allocated more resources so that their throughput meet their QoS requirements.

8.6.1 Large Scale Network Planning

Here, we perform a large scale network planning to observe the effect of resource allocation. We consider a 10 km $\times$10 km area in Falls Church, VA. We consider a cellular system to be planned to cover this area as it is depicted in Figure 8.11. This geographic area covers Falls Church and Annandale cities in Northern Virginia. The red dots in the picture illustrate the corners of the grid. In order to do network planning, we need to obtain the radio environment



(a) Throughput of the UEs



(b) Resources Allocated to UEs

Figure 8.10: The system contains 9 UEs, each concurrently running a real-time application with respective identically colored sigmoidal utility functions. The lower SNR UEs are allocated more resources so that their throughput meet their QoS requirements.

map (REM) all over this geographic area. The REM would include propagation pathloss, diffraction loss, as well as troposcatter loss; however, due to the size of the area, only propagation and diffraction loss would be present [134]. In order to obtain the path loss, we use the ITM model in its point-to-point (P2P) [134] mode. The reason to resort to ITM is the fact that ITM is the only publicly available propagation model that considers morphology and elevation. It is also suitable for a wide range of frequencies from VHF to 20 GHz which makes it accommodate a wide range of applications.

The ITM has two modes of operation, which are the APM and P2P, where the former only accounts for the average changes in the elevation in the geographic area while the latter accounts for the precise elevations on the path. In order to obtain elevation, we download the 1/3 arc second digital elevation model (DEM), from the United Stated Geological Survey (USGS) . The data are available for free and provide with elevations with resolution of 10 m. The elevation data are in the form of XML files and they are processed and converted to text files before usage.

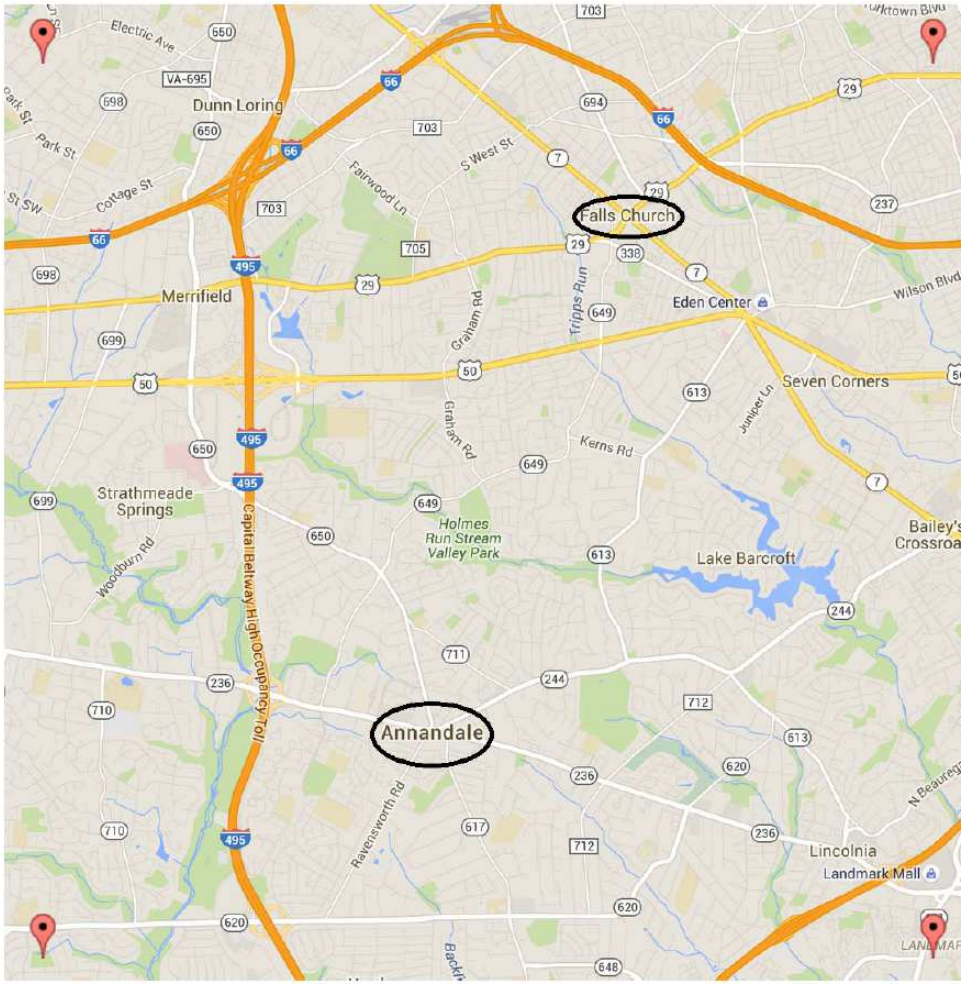


Figure 8.11: 10 km × 10 km area covering Falls Church and Annandale in Northern Virginia, where our cell planning and resource allocation occurs.

Once the REM is obtained, the link budgets in the UL and DL directions are performed in order to obtain SNRs which will be used in the resource allocation algorithms. There will

20 eNodeBs distributed over the geographic area. And we will be using genetic algorithms to distributed the BSs in the region in a fashion to maximize the coverage of the area. It is noteworthy that the footprint of the BSs is based on the DL SNR 1 dB. Due to the difference of the path loss in various directions, the foot prints will be non-circles. The plot of SNR of the UEs for the eNodeBs in the system in depicted in Figure 8.12. As we can see the y-axis (eNodeB) shows the 20 eNodeBs in our system and the x axis (UE) shows the UEs that are served by each corresponding eNodeB in the y-axis. Since, UEs chose their serving eNodeB based on the received SNR, the number of UEs in the different eNodeBs are different as well. Furthermore, we can see some UEs with very low SNRs (such as those in the eNodeB 1). This is because we distributed the UEs purely randomly, and therefore, some Ues are in bad channel conditions.

Once the eNodeBs are distributed all over the area, then we randomly distribute the UEs all over the grid and use the DL SNR from the BSs to help UEs pick their serving BSs. As such, we do not drop the UEs under BSs footprint, instead, we distribute them at random and the SNRs define the serving BS. Then, the UL SNR is leveraged to obtain the coefficients which are used in equation (8.8). We are assuming omnidirectional antennae and no interference is accounted for. The UEs choose their serving BS based on the strongest signal that they receive from each BS in the downlink direction. Therefore, the UEs will be distributed in various cells of the BSs as in Figure 8.13.

The network planning relies on calculation of the link budget parameters in the UP and DL directions. The UL link budget has a fixed part and a varying part. The first part can be calculated as equation (8.30). Here, P_i is the ith UE effective radiated isotropically radiated power (EIRP), which is the sum of the UE transmit power in dB and the gain in the direction of the eNode in dBi. The L_{Cable} is the cable loss in dB, the I is the interference margin in dB, G_{TX} is the eNodeB antenna gain, and G_{MH} is eNodeB master head gain [9]. In the simulation, we take the UE EIRP as 24 dBm, interference margin as 2 dB, cable loss

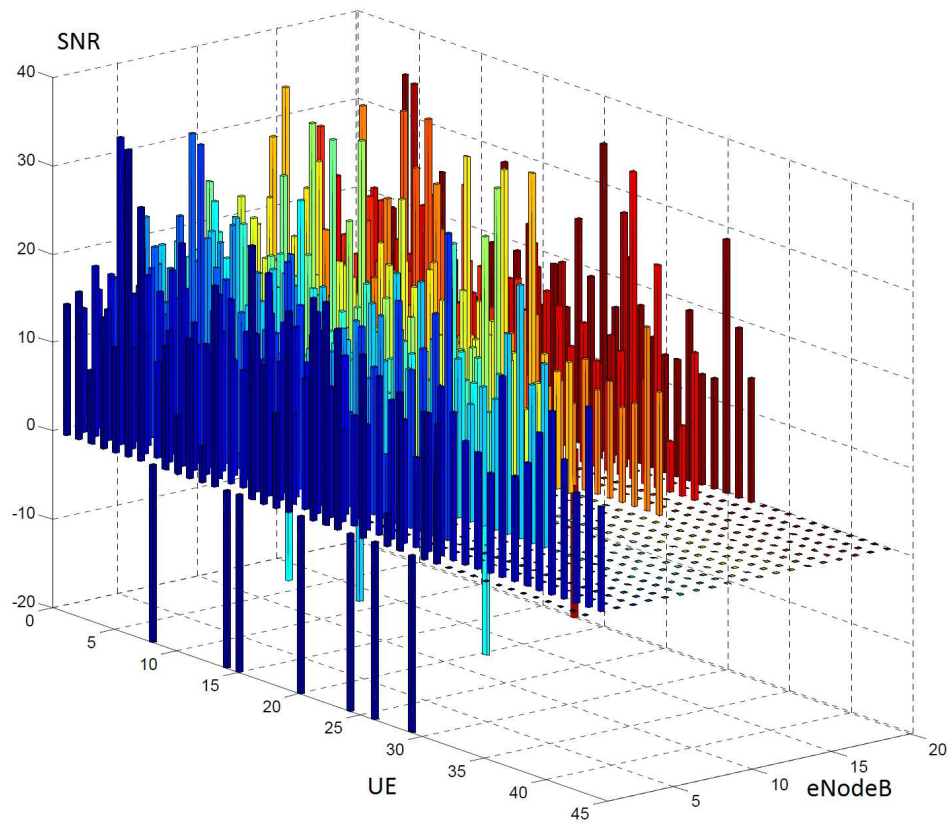


Figure 8.12: Ues are randomly distributed in the area in Figure 8.11, and since they choose the serving base station according to the strongest downlink signal that they receive from all base stations, they are distributed distinctively amongst the base stations as in this figure. The numbers on each pie piece shows the percentage of the UEs of 500 total UEs that belong to a particular cell.

as 1 dB, eNodeB antenna gain as 6 dBi, and master-head gain as 2 dB.

$$P_i^{\mathrm{UL}} = P_{\mathrm{UE}} - I + L_{\mathrm{Cable}} + G_{\mathrm{TX}} + G_{\mathrm{MH}} - L_{\text{Implementation Margin}} - \\ L_{\text{Fast Fading Margin}} - L_{\text{Body Loss}} + G_{\mathrm{RX}} - L_{\mathrm{Loss}} \tag{8.30}$$

Furthermore, in equation (8.30), we take the implementation margin 5 dB, body loss 3 dB, receiver gain 6 dB, fast fading margin 5 dB, and clutter loss 12 dB. In is noteworthy that the 12 dB clutter loss is conservative, and a precise clutter loss can be obtained using USGS Land Use Land Cover (LCLU) data [180]. However, this is out of the scope of this simulation. Besides, the noise figure (NF) for the receiver, i.e. eNodeB in the UL direction,

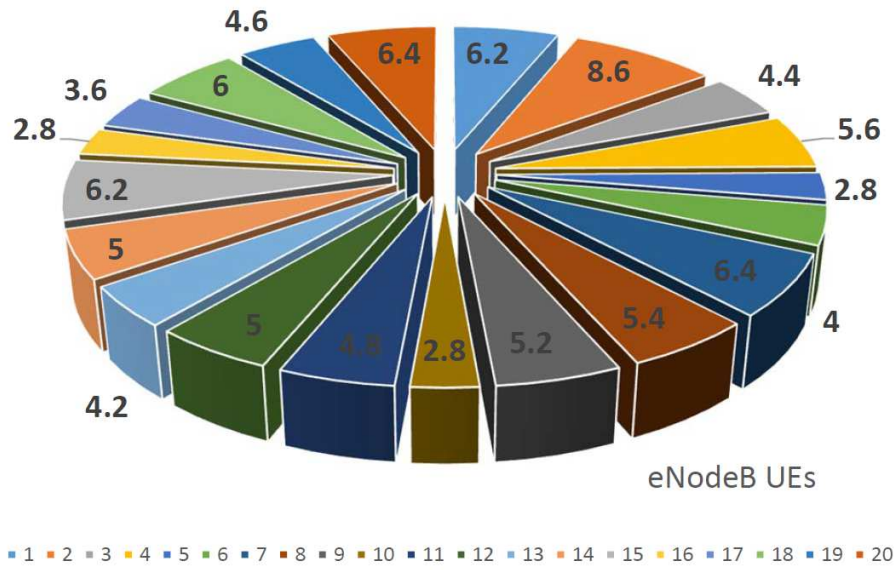


Figure 8.13: Ues are randomly distributed in the area in Figure 8.11, and since they choose the serving base station according to the strongest downlink signal that they receive from all base stations, they are distributed distinctively amongst the base stations as in this figure. The numbers on each pie piece shows the percentage of the UEs of 500 total UEs that belong to a particular cell.

is 2 dB, which accumulates to receiver noise floor density as equation (8.31). The second term and third term in the addition comes from the thermal noise, evaluated at 10 MHz LTE, and at the temperature 290 degrees of Kelvin.

$$ND_{\text{eNodeB}} = NF_{\text{eNodeB}} + 30 + 10\log(1.38e^{-23} \times 290) \tag{8.31}$$

Then, the signal strength at the jth eNodeB position can be calculated as equation (8.32), where $P_{i,j}$ is the signal strength from the ith UE to the jth eNodeB, whose noise floor density is ND_j, and $L_{\text{REM}}(i,j)$ is the pathloss between the ith UE to the jth eNodeB.

$$P_{i,j}^{\text{UL}} = P_i^{\text{UL}} - L_{\text{REM}}(i,j) - ND_j - 10\log(180e^3) \tag{8.32}$$

Furthermore, in the DL direction, the link budget can be calculated as equation (8.33).

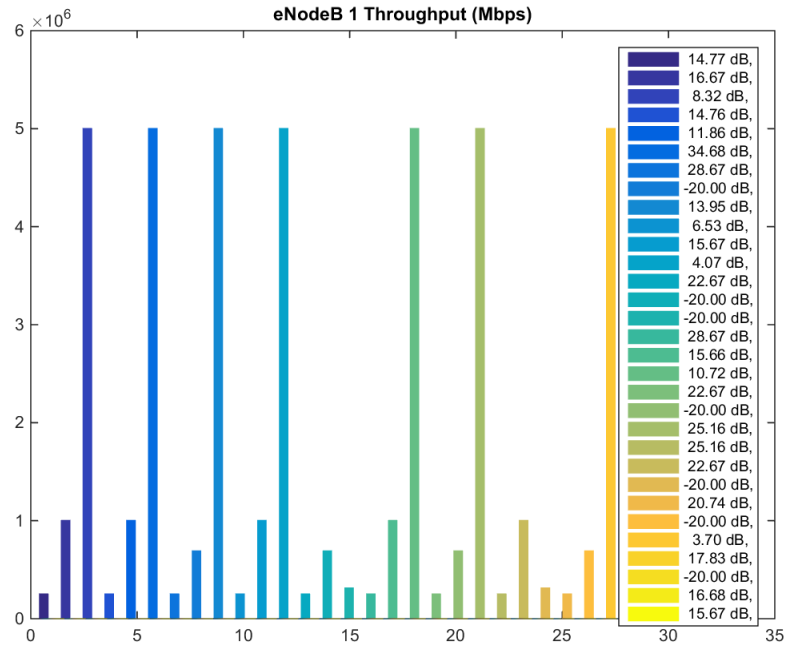
$$P_i^{\mathrm{DL}} = P_{\mathrm{eNodeB}} - I + L_{\mathrm{Cable}} + G_{\mathrm{MH}} - L_{\mathrm{Implementation\ Margin}} - \\ L_{\mathrm{Fast\ Fading\ Margin}} - L_{\mathrm{Body\ Loss}} + G_{\mathrm{RX}} - L_{\mathrm{Loss}} \tag{8.33}$$

Then, the signal strength at the jth UE position can be calculated as equation (8.34), where $P_{i,j}$ is the signal strength from the ith eNodeB to the jth UE, whose noise floor density is ND_j, and $L_{\mathrm{REM}}(i,j)$ is the pathloss between the ith eNodeB to the jth UE.

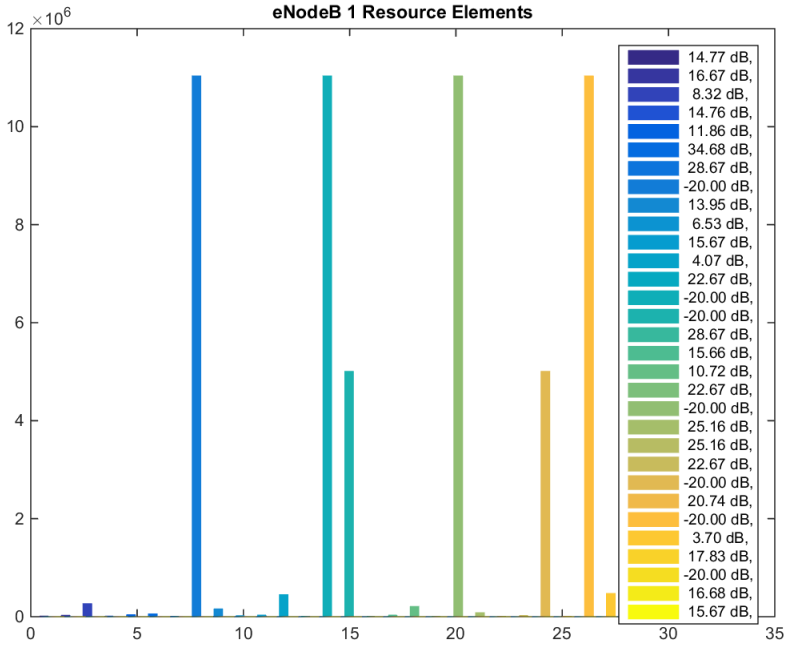
$$P_{i,j}^{\mathrm{DL}} = P_i^{\mathrm{DL}} - L_{\mathrm{REM}}(i,j) - ND_j - 10\log(180e^3) \tag{8.34}$$

For eNodeB 1, the plots of resource elements allocated to the UEs as well as the throughput in Mbps are given in Figures (8.14(a)) and (8.14(a)), respectively. As we can observe from the figures, the horizontal axis is the UE indices which indicate that there are 31 UEs being served by the eNodeB 1. The vertical axis is the number of resource elements allocated by the optimization in Algorithms (21) and (22). Each UE has only 1 application running. This assumption simplifies the simulation because the goal of this chapter is observing the effect of the channel for which channel-aware EURA is performed. On the other hand, inside the UE there is no channel and the IURA is not needed to show any channel effect but mere allocation to the applications, which was presented extensively in chapters 3 and 4. Furthermore, the legend shows the SNR of the UEs in dB. The bit rate requirements of the applications is according to $\{0.25, 1, 5, 0.25, 1, 5, ...\}$ Mbps.

As we can observe from Figure (8.14(b)), the UEs with low SNRs are receiving more resource elements in order to meet their bit rate requirements. On the other hand, UEs with high SNR are receiving less resources. On the other hand, (8.14(a)) shows that the throughputs of UEs 8, 14, 15, 24 are not met. This is due to the -20 dB SNR which is placed on the network by UEs 8, 14, 15, 20, 24, 26, and 29 each with 1, 1, 5, 1, 5, 1, and 1 Mbps which are high rates at low SNRs and indicate using many REs. We can see these in



(a) Throughput of the UEs for eNB 1



(b) Resources Allocated to UEs by eNB 1

Figure 8.14: The system contains 31 UEs, each concurrently running a real-time application.

Figure (8.14(b)) where there is a spike at UEs 8, 14, 15, 20, 24, 26, and 29 which shows the algorithm is allocating more REs to these UEs.

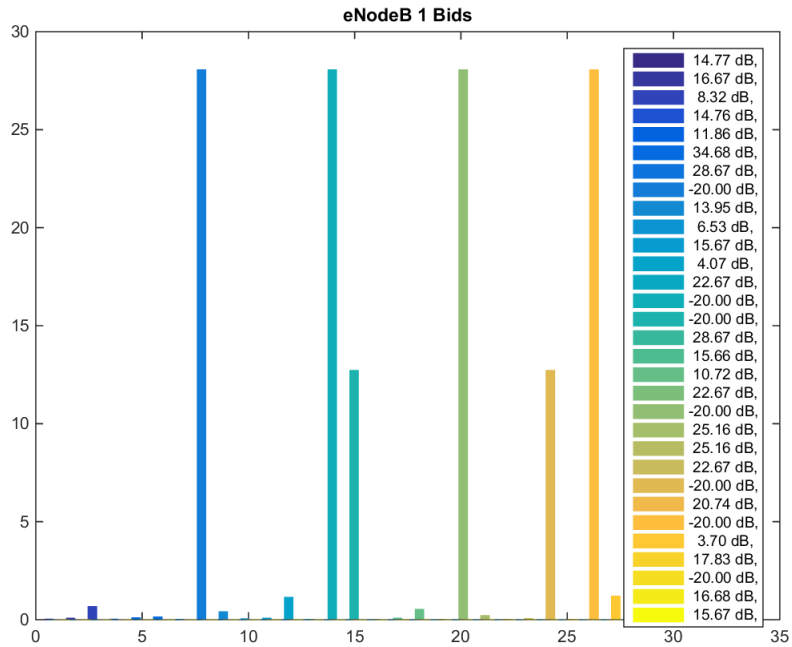


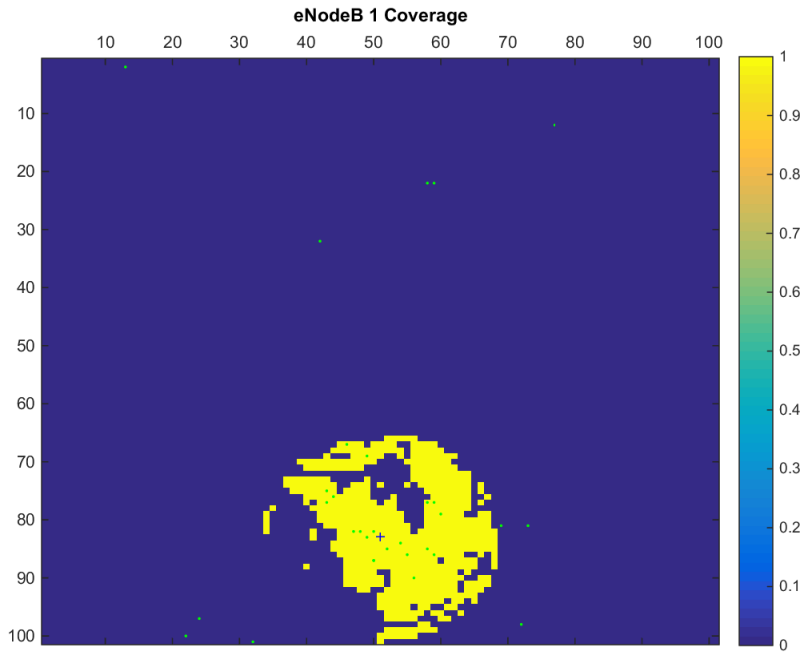
Figure 8.15: UE Bids pledged to eNB 1

Moreover, as we can see from Figure (8.15), UEs 8, 14, 15, 20, 24, 26, and 29 are bidding the highest due to the fact the they require more resources in view of their bad channel conditions. This plot shows the last iteration of the algorithm where the shadow prices are converged. Furthermore, we can observe the coverage area of eNodeB 1 in Figure (8.16(a)), and as we see the coverage is not circular which is a result of REM and elevations in various directions leading to different SNRs. The black addition symbol in the middle of the coverage, shown in yellow, is the eNodeB 1 coordinates in the 101×101 grid. The side bar, shows 0 - dark blue - and 1 - yellow - for the areas which are respectively not covered and are under coverage. The green dots in the region are the UEs scattered all over the area. As we can see, some of the UEs are not in the foot print of the eNodeBs, and this is due to the fact that the UE choose their serving eNodeBs only based on the received SNR and as

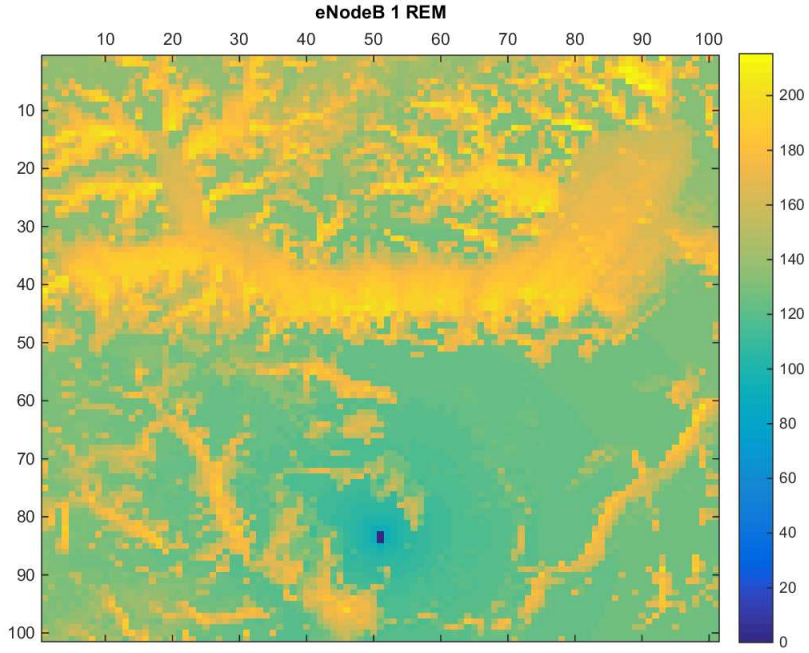
a result, many of the were randomly distributed in the area and were under bad channel conditions (specially those at -20 dB as we discussed). These simply chose eNodeB1 as it was the strongest signal they could ever receive. Furthermore, we can see the REM which shows that a blue dot at coordinate (50, 83), which is for the eNodeB 1 and the blue color from the side bar represents a 0 dB loss which is expected as we are at the eNodeB position. However, as we go further away from the eNodeB, the path loss increases, which is shown by the color spectrum ligt blue, green, and yellow in the order of increase. The highest pathloss in the area is about 215 dB.

For eNodeB 2, the plots of resource elements allocated to the UEs as well as the throughput in Mbps are given in Figures (8.17(b)) and (8.17(a)), respectively. As we can observe from the figures, the horizontal axis is the UE indices which indicate that there are 23 UEs being served by the eNodeB 2. The vertical axis is the number of resource elements allocated by the optimization in Algorithms (21) and (22). Each UE has only 1 application running. This assumption simplifies the simulation because the goal of this chapter is observing the effect of the channel for which channel-aware EURA is performed. Furthermore, the legend show the SNR of the UEs in dB. The bit rate requirements of the applications is according to $\{0.25, 1, 5, 0.25, 1, 5, ...\}$ Mbps.

As we can observe from Figure (8.17(b)), the UEs with low SNRs are receiving more resource elements in order to meet their bit rate requirements. On the other hand, UEs with high SNR are receiving less resources. On the other hand, (8.14(a)) shows that UE throughputs are met. This is due all UEs are at good channel conditions such that the minimum SNR was 10.1885 dB. We can see these in Figure (8.17(b)) that UEs with higher bit rate needs were allocated more resources, i.e. for UE k, $R_{1+3K} < R_{2+3K} < R_{3+3K}$. This statements says that the rate allocated to UE 1, 4, 7, ..., is less than the rate allocated to UEs 2, 5, 8, ..., and is less than the rates allocated to UEs 3, 6, 9, ... compared one to one (i.e. corresponding indices compared to each other).

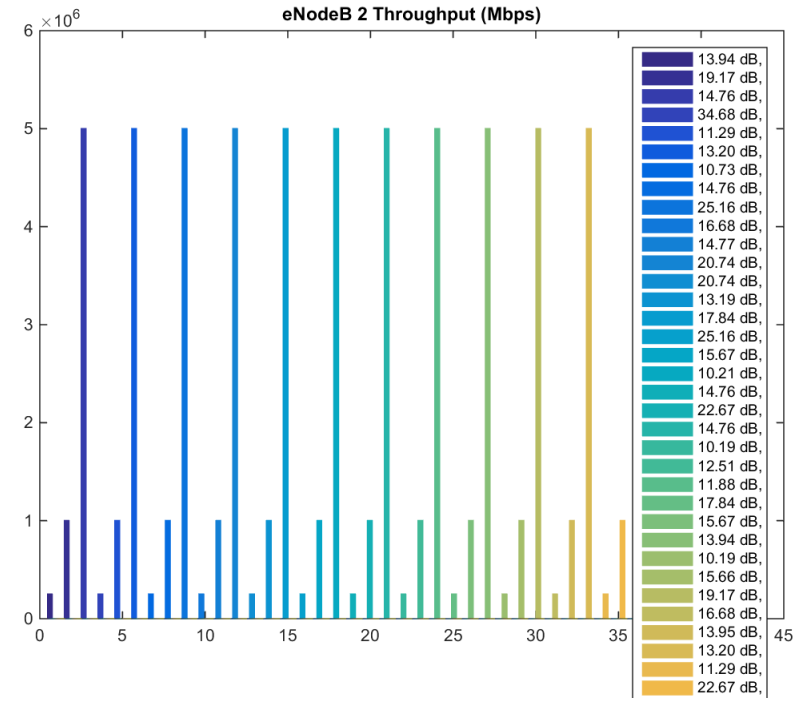


(a) eNB 1 Coverage

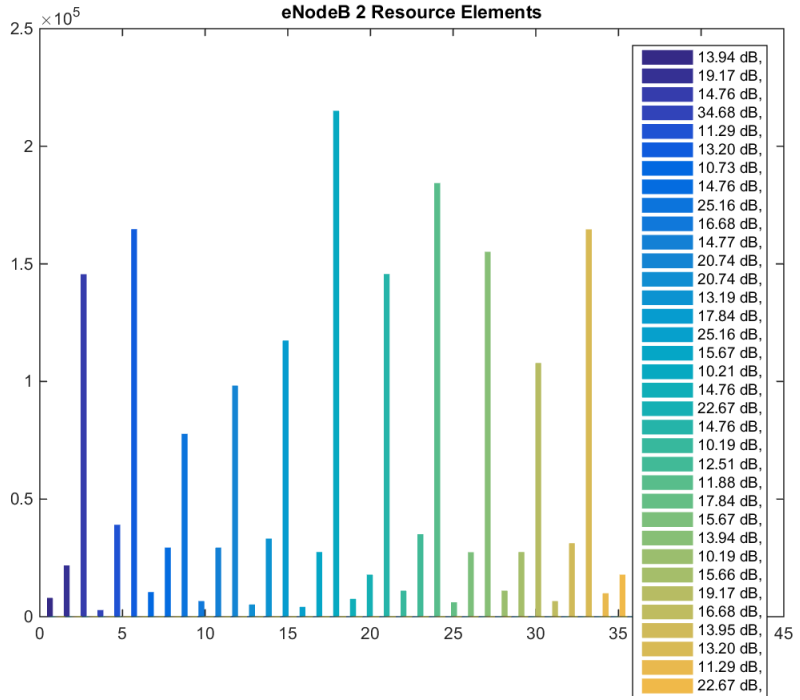


(b) eNB 1 REM

Figure 8.16: The system contains 31 UEs, each concurrently running a real-time application.



(a) Throughput of the UEs for eNB 2



(b) Resources Allocated to UEs by eNB 2

Figure 8.17: The system contains 23 UEs, each concurrently running a real-time application.

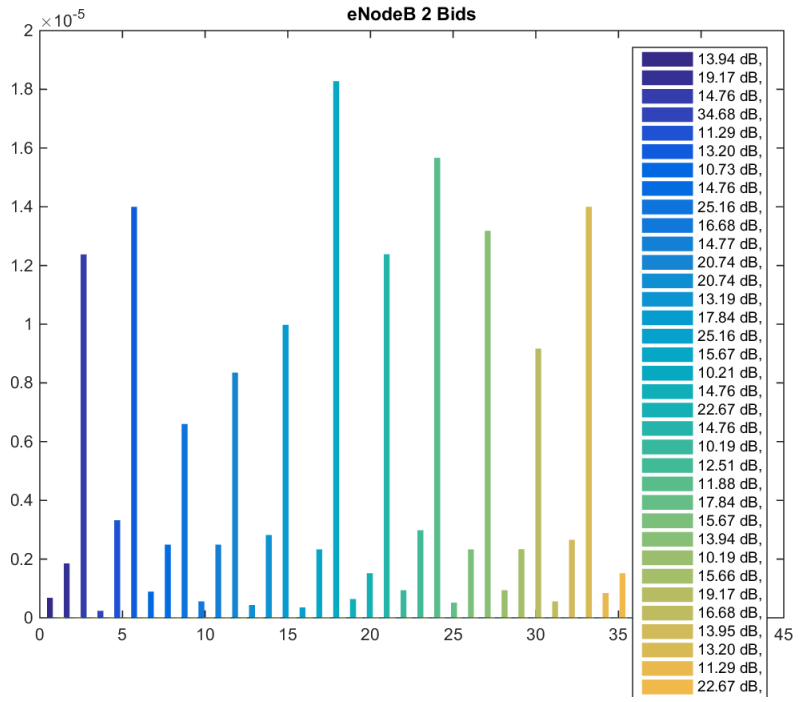


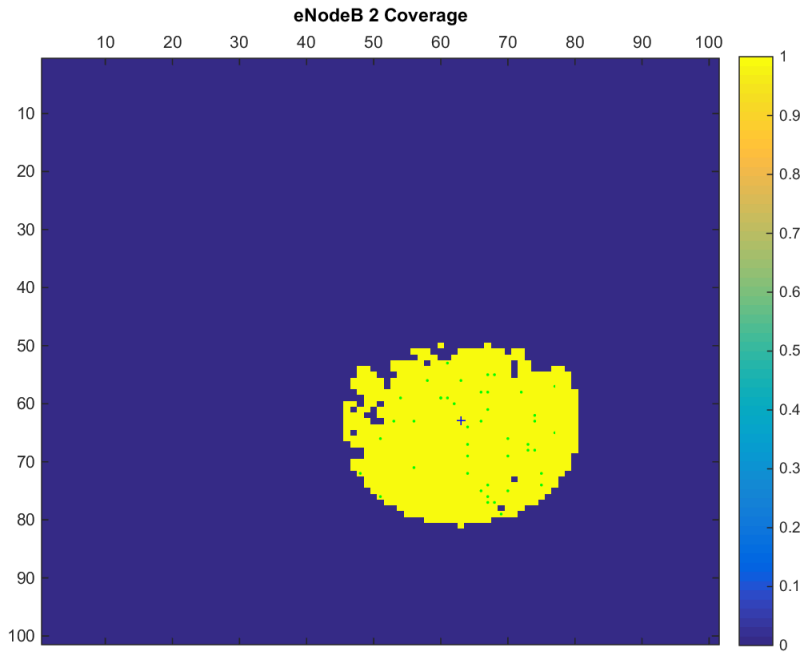
Figure 8.18: UE Bids pledged to eNB 2

Moreover, as we can see from Figure (8.18), UEs $3+3K$ are higher than UEs $2+3K$ which bid higher than UEs $1+3K$ due to the fact the they require more resources in view of their applications (5 Mbps vs. 1 Mbps vs 0.25 Mbps.) This plot shows the last iteration of the algorithm where the shadow prices are converged. Furthermore, we can observe the coverage area of eNodeB 1 in Figure (8.19(a)), and as we see the coverage is not circular which is a result of REM and elevations in various directions leading to different SNRs. The black addition symbol in the middle of the coverage, shown in yellow, is the eNodeB 2 coordinates in the 101×101 grid. The side bar, shows 0 - dark blue - and 1 - yellow - for the areas which are respectively not covered and are under coverage. The green dots in the region are the UEs scattered all over the area. As we can see, UEs are in the foot print of the eNodeBs, and had good channel conditions. These simply chose eNodeB 2 as it was the strongest signal they could ever receive. Furthermore, we can see the REM which shows that a blue dot at the eNodeB location, which is for the eNodeB 2 and the blue color from the side bar

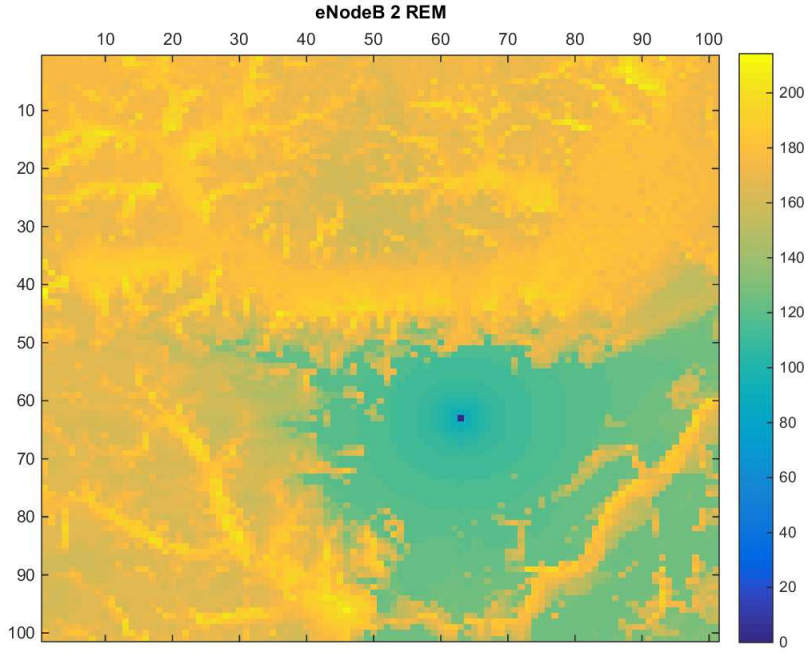
represents a 0 dB loss which is expected as we are at the eNodeB position. However, as we go further away from the eNodeB, the path loss increases, which is shown by the color spectrum ligt blue, green, and yellow in the order of increase. The highest pathloss in the area is about 210 dB.

For eNodeB 3, the plots of resource elements allocated to the UEs as well as the throughput in Mbps are given in Figures (8.20(b)) and (8.20(a)), respectively. As we can observe from the figures, the horizontal axis is the UE indices which indicate that there are 22 UEs being served by the eNodeB 3. The vertical axis is the number of resource elements allocated by the optimization in Algorithms (21) and (22). Each UE has only 1 application running. This assumption simplifies the simulation because the goal of this chapter is observing the effect of the channel for which channel-aware EURA is performed. Furthermore, the legend show the SNR of the UEs in dB. The bit rate requirements of the applications is according to $\{0.25, 1, 5, 0.25, 1, 5, ...\}$ Mbps.

As we can observe from Figure (8.20(b)), the UEs with low SNRs are receiving more resource elements in order to meet their bit rate requirements. On the other hand, UEs with high SNR are receiving less resources. On the other hand, (8.20(a)) shows that UE throughputs are met. This is due all UEs are at good channel conditions such that the minimum SNR was 10.1885 dB. We can see these in Figure (8.17(b)) that UEs with higher bit rate needs were allocated more resources, i.e. for UE k, $R_{1+3K} < R_{2+3K} < R_{3+3K}$. This statements says that the rate allocated to UE 1, 4, 7, ..., is less than the rate allocated to UEs 2, 5, 8, ..., and is less than the rates allocated to UEs 3, 6, 9, ... compared one to one (i.e. corresponding indices compared to each other). Also, we see that for same bit rate requirements, UEs with higher SNR are allocated less resources. For instance, UEs 6, 15 and 18 are allocated less REs than UEs 3, 9, and 12 even though the bit rate requirements for both are 5 Mbps. This is due to the fact the these UEs have higher SNRs. Moreover, amongst UEs 6, 15, and 18, the last one has the smallest SNR 15.67 dB as opposed to the

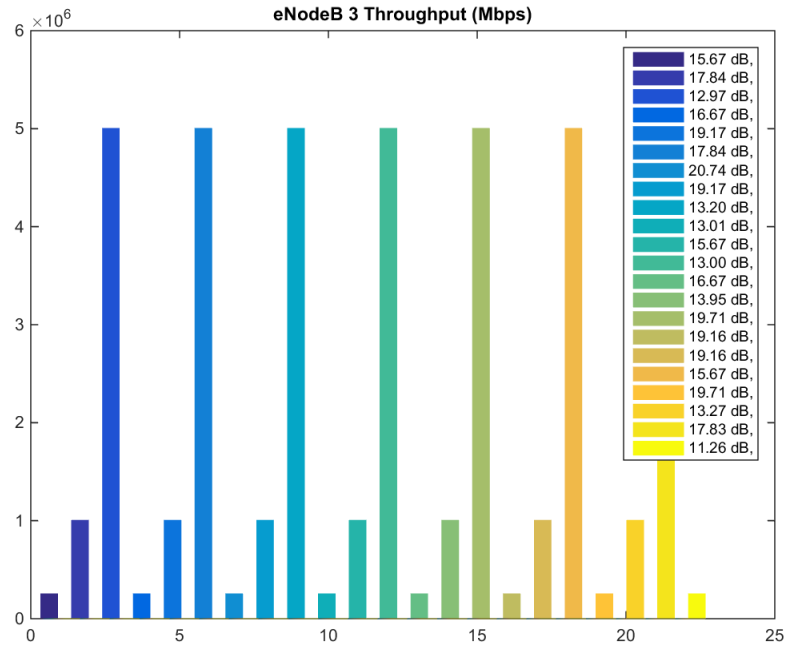


(a) eNB 2 Coverage

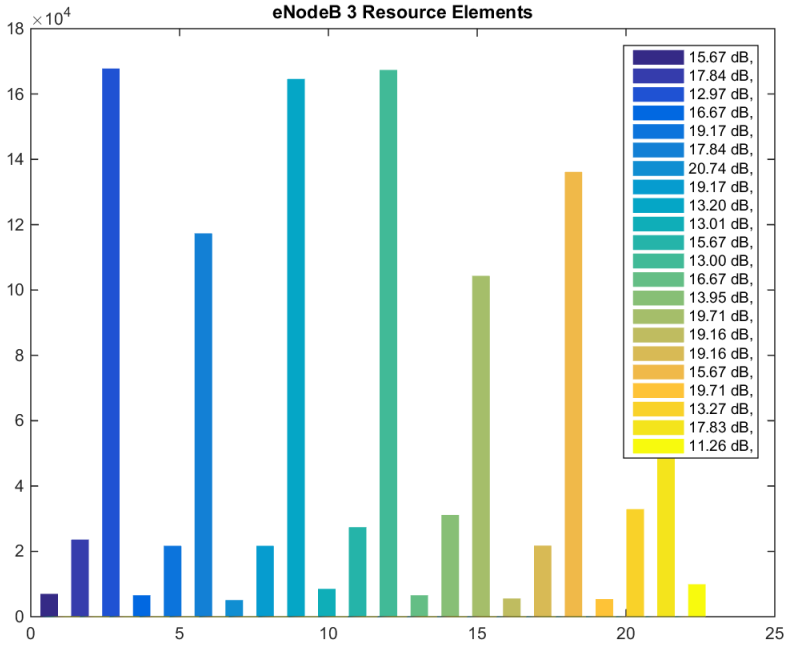


(b) eNB 2 REM

Figure 8.19: The system contains 23 UEs, each concurrently running a real-time application.



(a) Throughput of the UEs for eNB 3



(b) Resources Allocated to UEs by eNB 3

Figure 8.20: The system contains 22 UEs, each concurrently running a real-time application.

19.71 SNR for UE 15 which causes UE 18 to get more resources as opposed to UE 15. On the other hand, UE 12 has the lowest SNR in the system, 12.97 dB while it is a high bit requirement application (5 Mbps), so it needs more REs as it is assigned by the algorithm.

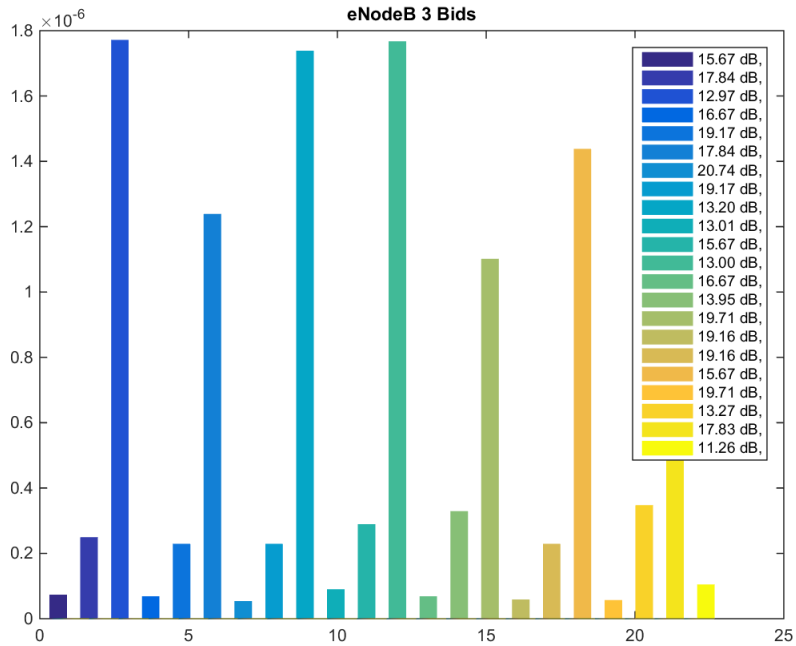


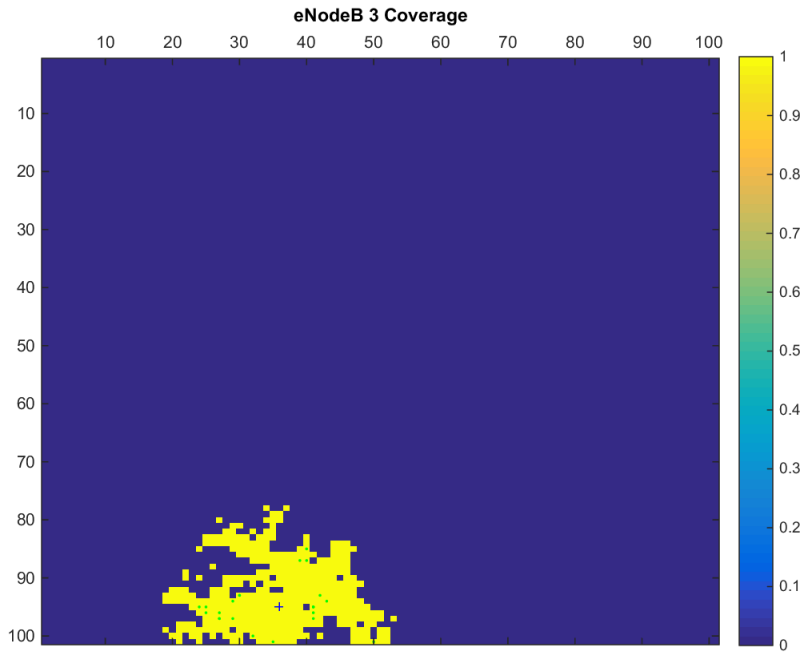
Figure 8.21: UE Bids pledged to eNB 3

Moreover, as we can see from Figure (8.21), UEs $3+3K$ are higher than UEs $2+3K$ which bid higher than UEs $1+3K$ due to the fact the they require more resources in view of their applications (5 Mbps vs. 1 Mbps vs 0.25 Mbps.) This plot shows the last iteration of the algorithm where the shadow prices are converged. Furthermore, we can observe the coverage area of eNodeB 1 in Figure (8.22(a)), and as we see the coverage is not circular which is a result of REM and elevations in various directions leading to different SNRs. The black addition symbol in the middle of the coverage, shown in yellow, is the eNodeB 3 coordinates in the 101×101 grid. The side bar, shows 0 - dark blue - and 1 - yellow - for the areas which are respectively not covered and are under coverage. The green dots in the region are the UEs scattered all over the area. As we can see, UEs are in the foot print of the eNodeBs,

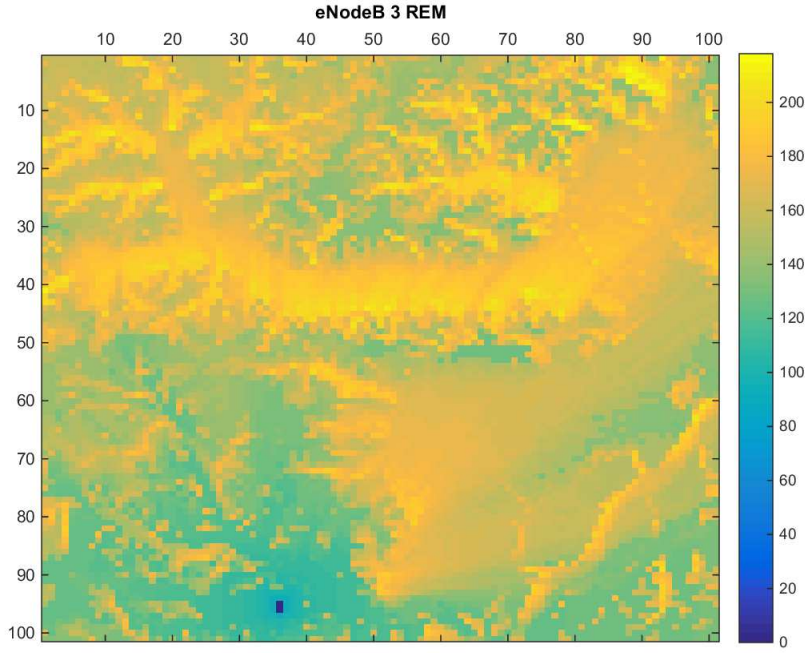
and had good channel conditions. These simply chose eNodeB 3 as it was the strongest signal they could ever receive. Furthermore, we can see the REM which shows that a blue dot at the eNodeB location, which is for the eNodeB 3 and the blue color from the side bar represents a 0 dB loss which is expected as we are at the eNodeB position. However, as we go further away from the eNodeB, the path loss increases, which is shown by the color spectrum ligt blue, green, and yellow in the order of increase. The highest pathloss in the area is about 218 dB.

For eNodeB 4, the plots of resource elements allocated to the UEs as well as the throughput in Mbps are given in Figures (8.23(b)) and (8.23(a)), respectively. As we can observe from the figures, the horizontal axis is the UE indices which indicate that there are 28 UEs being served by the eNodeB 4. The vertical axis is the number of resource elements allocated by the optimization in Algorithms (21) and (22). Each UE has only 1 application running. This assumption simplifies the simulation because the goal of this chapter is observing the effect of the channel for which channel-aware EURA is performed. Furthermore, the legend show the SNR of the UEs in dB. The bit rate requirements of the applications is according to $\{0.25, 1, 5, 0.25, 1, 5, ...\}$ Mbps.

As we can observe from Figure (8.23(b)), the UEs with low SNRs are receiving more resource elements in order to meet their bit rate requirements. On the other hand, UEs with high SNR are receiving less resources. On the other hand, (8.23(a)) shows that UE throughputs are met. This is due all UEs are at good channel conditions such that the minimum SNR was 10.1885 dB. We can see these in Figure (8.23(b)) that UEs with higher bit rate needs were allocated more resources, i.e. for UE k, $R_{1+3K} < R_{2+3K} < R_{3+3K}$. This statements says that the rate allocated to UE 1, 4, 7, ..., is less than the rate allocated to UEs 2, 5, 8, ..., and is less than the rates allocated to UEs 3, 6, 9, ... compared one to one (i.e. corresponding indices compared to each other). Also, we see that for same bit rate requirements, UEs with higher SNR are allocated less resources. For instance, UEs 11 and

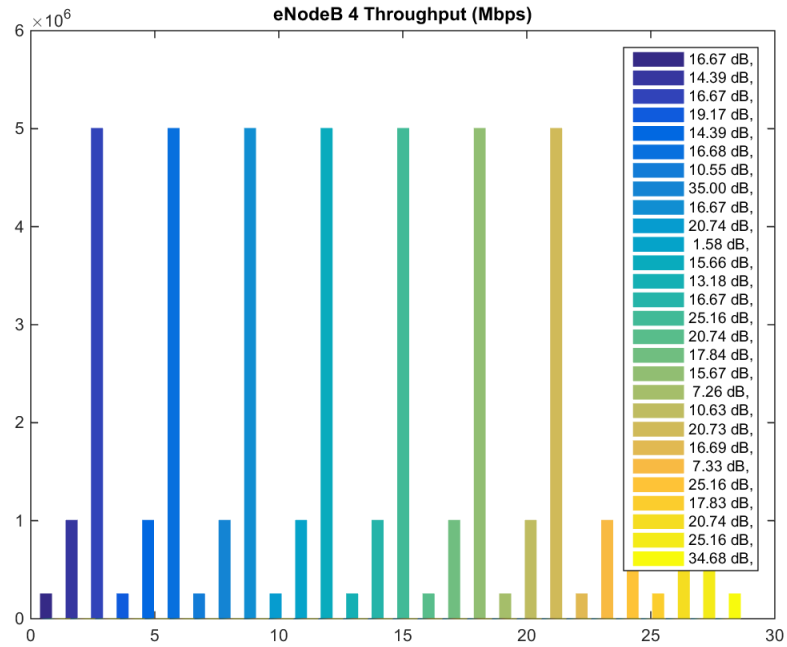


(a) eNB 3 Coverage

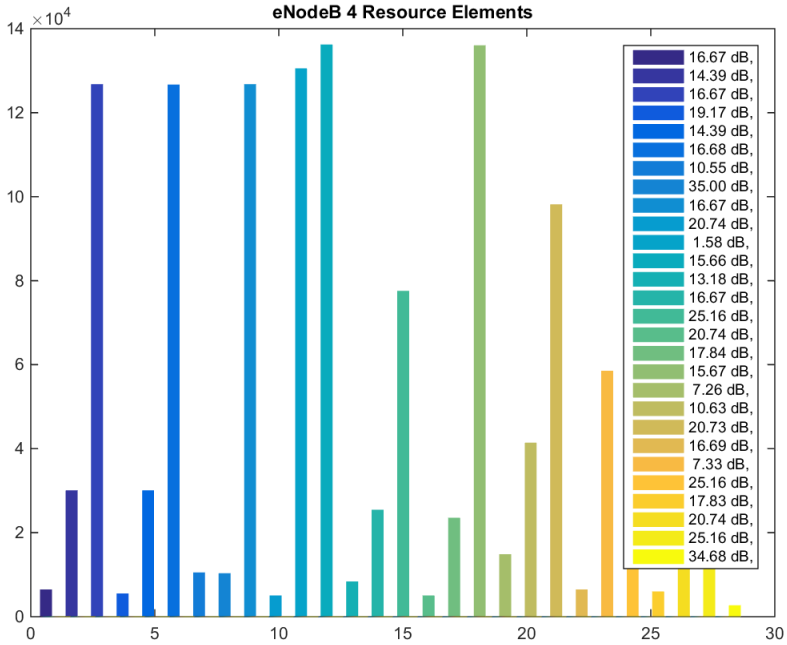


(b) eNB 3 REM

Figure 8.22: The system contains 22 UEs, each concurrently running a real-time application.



(a) Throughput of the UEs for eNB 4



(b) Resources Allocated to UEs by eNB 4

Figure 8.23: The system contains 28 UEs, each concurrently running a real-time application.

12 and 18 are allocated more REs since they are at lower SNRs. In particular, the spike for UE 11 is interesting even though it has less bit needs vis-a-vis UE 12 who needs 5 Mbps. The spike is because this UE is at the lowest SNR situation in the system 1.58 dB. We see a similar behavior for UE 20 which is getting more REs vs other UEs with 1 Mbps requirement since its SNR is lower. On the other hand, UE 15 who has a high bit requirement of 5 Mbps is getting less REs than its counterparts since it is at a good channel condition 25.16 dB.

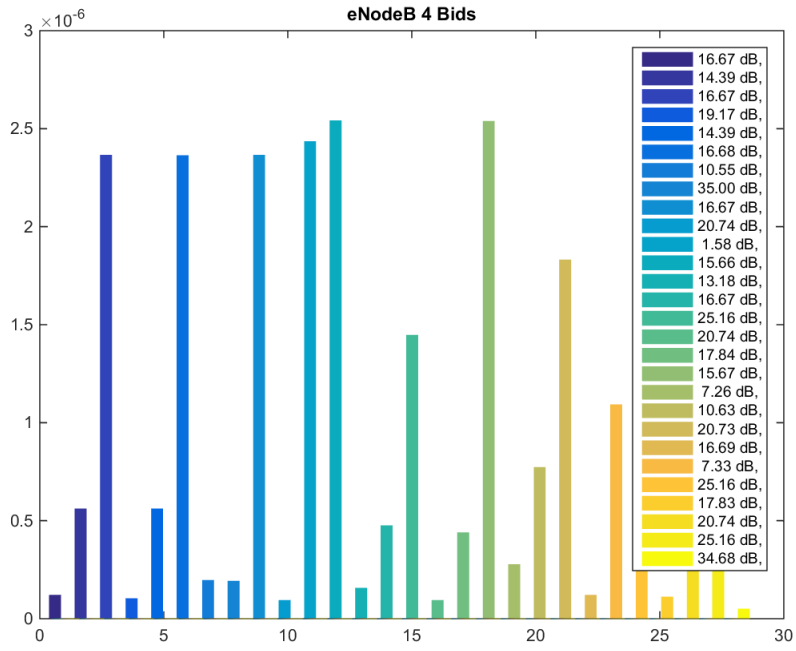


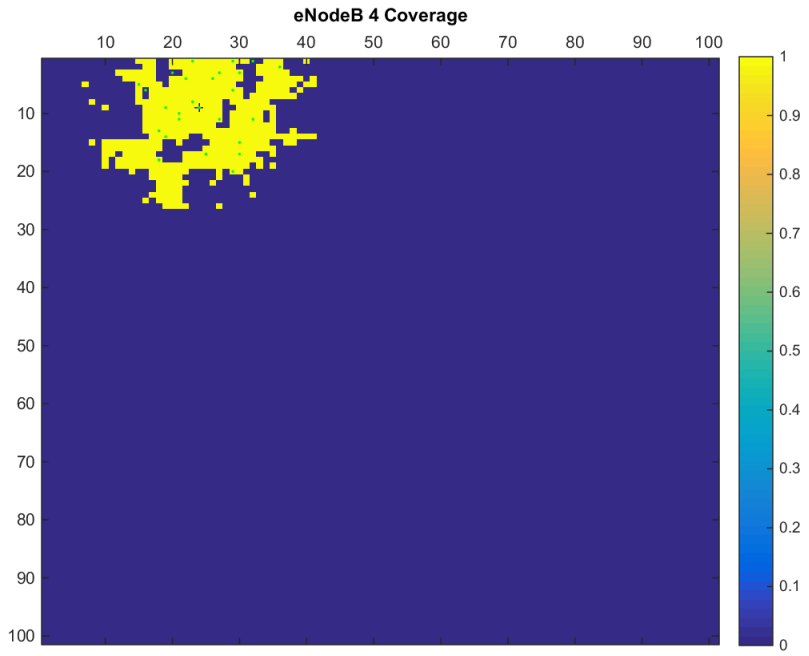
Figure 8.24: UE Bids pledged to eNB 4

Moreover, as we can see from Figure (8.24), UEs $3 + 3K$ are higher than UEs $2 + 3K$ which bid higher than UEs $1 + 3K$ due to the fact the they require more resources in view of their applications (5 Mbps vs. 1 Mbps vs 0.25 Mbps.) This plot shows the last iteration of the algorithm where the shadow prices are converged. Interestingly, UE 15 which needed high resources at good channel is bidding less than its counterparts. Furthermore, we can observe the coverage area of eNodeB 4 in Figure (8.25(a)), and as we see the coverage is not circular which is a result of REM and elevations in various directions leading to different

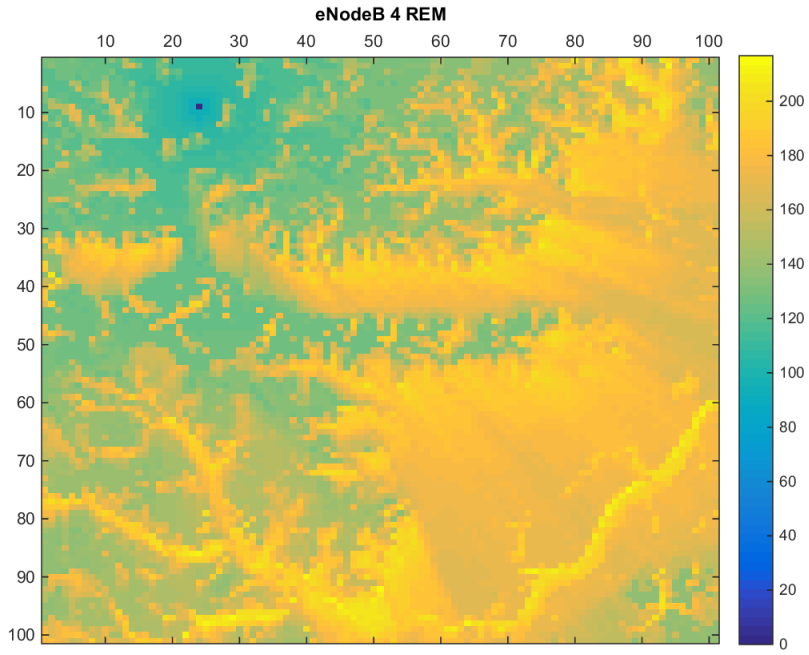
SNRs. The black addition symbol in the middle of the coverage, shown in yellow, is the eNodeB 4 coordinates in the 101×101 grid. The side bar, shows 0 - dark blue - and 1 - yellow - for the areas which are respectively not covered and are under coverage. The green dots in the region are the UEs scattered all over the area. As we can see, UEs are in the foot print of the eNodeBs, and had good channel conditions. These simply chose eNodeB 4 as it was the strongest signal they could ever receive. Furthermore, we can see the REM which shows that a blue dot at the eNodeB location, which is for the eNodeB 3 and the blue color from the side bar represents a 0 dB loss which is expected as we are at the eNodeB position. However, as we go further away from the eNodeB, the path loss increases, which is shown by the color spectrum light blue, green, and yellow in the order of increase. The highest pathloss in the area is about 211 dB.

For eNodeB 5, the plots of REs allocated to the UEs as well as the throughput in Mbps are given in Figures (8.26(b)) and (8.26(a)), respectively. As we can observe from the figures, the horizontal axis is the UE indices which indicate that there are 14 UEs being served by the eNodeB 5. The vertical axis is the number of resource elements allocated by the optimization in Algorithms (21) and (22). Each UE has only 1 application running. This assumption simplifies the simulation because the goal of this chapter is observing the effect of the channel for which channel-aware EURA is performed. Furthermore, the legend show the SNR of the UEs in dB. The bit rate requirements of the applications is according to $\{0.25, 1, 5, 0.25, 1, 5, ...\}$ Mbps. The lowest SNR is 9.6584 dB.

As we can observe from Figure (8.26(b)), the UEs with low SNRs are receiving more resource elements in order to meet their bit rate requirements. On the other hand, UEs with high SNR are receiving less resources. On the other hand, (8.26(a)) shows that UE throughputs are met. This is due all UEs are at good channel conditions such that the minimum SNR was 9.6584 dB. We can see these in Figure (8.26(b)) that UEs with higher bit rate needs were allocated more resources, i.e. for UE k, $R_{1+3K} < R_{2+3K} < R_{3+3K}$. This

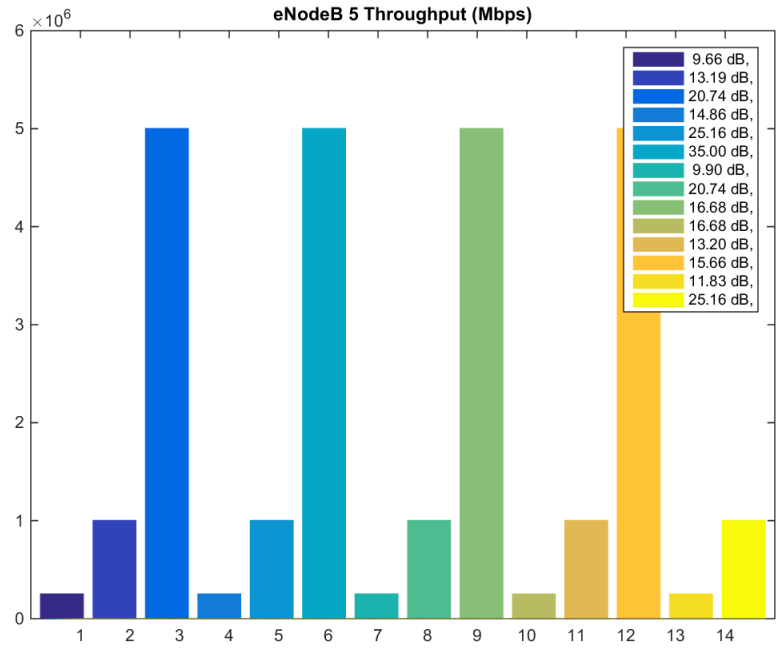


(a) eNB 4 Coverage

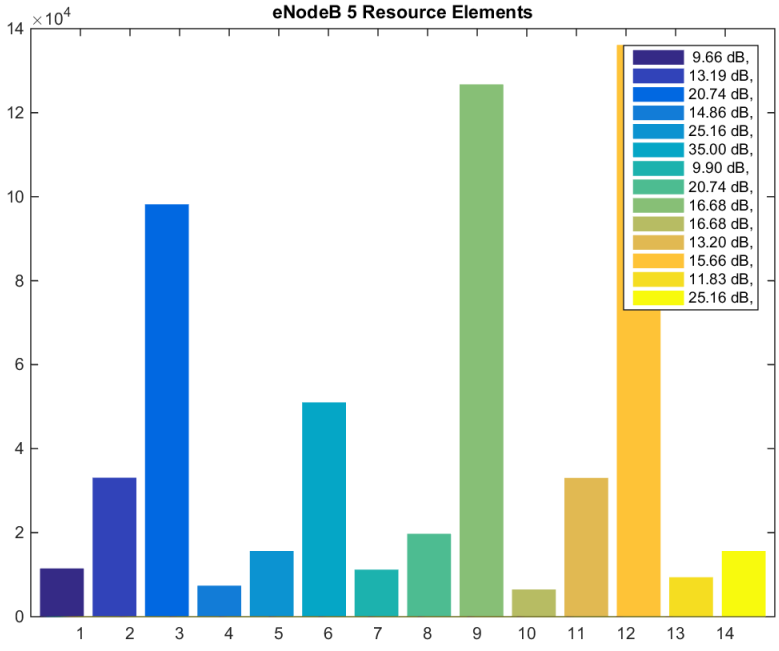


(b) eNB 4 REM

Figure 8.25: The system contains 28 UEs, each concurrently running a real-time application.



(a) Throughput of the UEs for eNB 5



(b) Resources Allocated to UEs by eNB 5

Figure 8.26: The system contains 14 UEs, each concurrently running a real-time application.

statements says that the rate allocated to UE 1, 4, 7, ..., is less than the rate allocated to UEs 2, 5, 8, ..., and is less than the rates allocated to UEs 3, 6, 9, ... compared one to one (i.e. corresponding indices compared to each other). Also, we see that for same bit rate requirements, UEs with higher SNR are allocated less resources. For instance, UEs 5 is allocated more REs than UEs 2 , 8, 11, and 14 since it has a better channel condition.

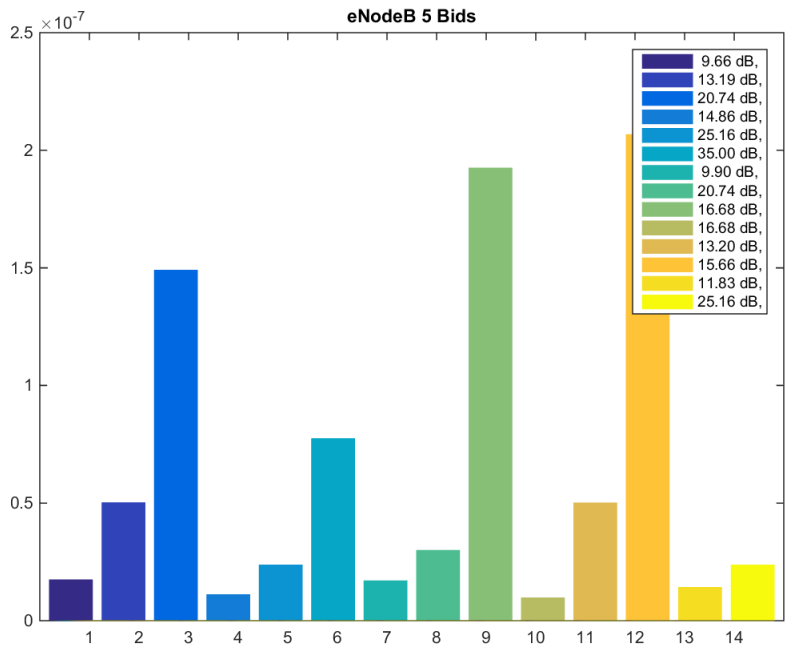


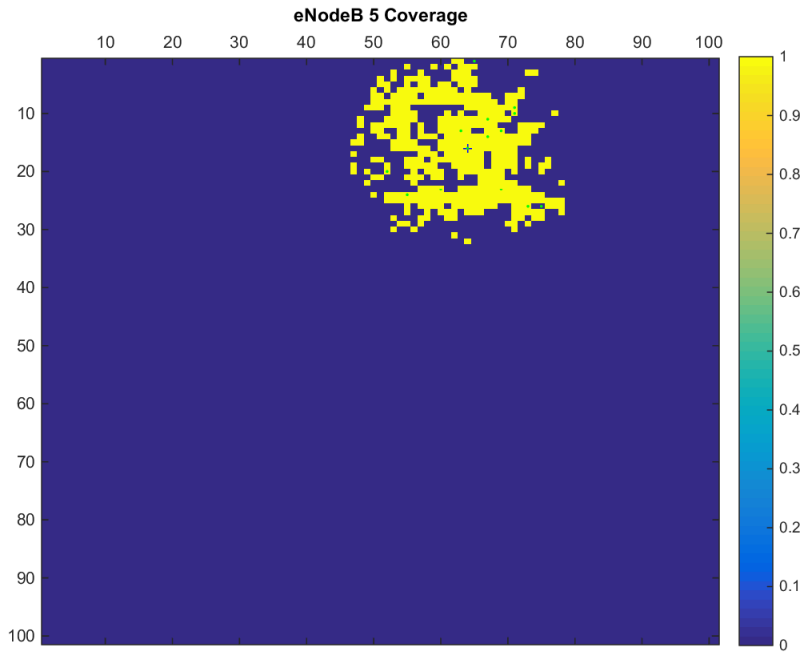
Figure 8.27: UE Bids pledged to eNB 5

Moreover, as we can see from Figure (8.27), UEs $3+3K$ are higher than UEs $2+3K$ which bid higher than UEs $1+3K$ due to the fact the they require more resources in view of their applications (5 Mbps vs. 1 Mbps vs 0.25 Mbps.) This plot shows the last iteration of the algorithm where the shadow prices are converged. Furthermore, we can observe the coverage area of eNodeB 4 in Figure (8.28(a)), and as we see the coverage is not circular which is a result of REM and elevations in various directions leading to different SNRs. The black addition symbol in the middle of the coverage, shown in yellow, is the eNodeB 5 coordinates in the 101×101 grid. The side bar, shows 0 - dark blue - and 1 - yellow - for the areas which

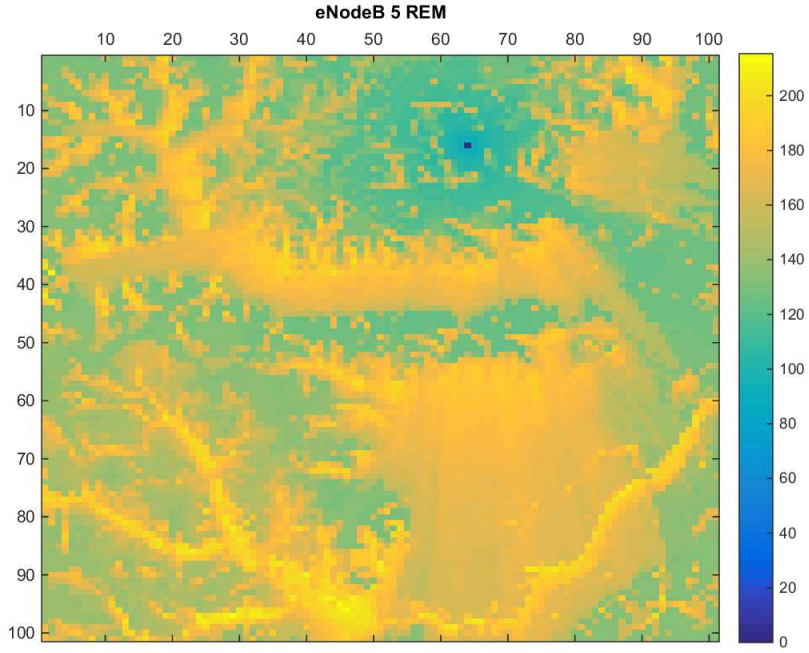
are respectively not covered and are under coverage. The green dots in the region are the UEs scattered all over the area. As we can see, UEs are in the foot print of the eNodeBs, and had good channel conditions. These simply chose eNodeB 5 as it was the strongest signal they could ever receive. Furthermore, we can see the REM which shows that a blue dot at the eNodeB location, which is for the eNodeB 5 and the blue color from the side bar represents a 0 dB loss which is expected as we are at the eNodeB position. However, as we go further away from the eNodeB, the path loss increases, which is shown by the color spectrum light blue, green, and yellow in the order of increase. The highest pathloss in the area is about 210 dB.

For eNodeB 6, the plots of resource elements allocated to the UEs as well as the throughput in Mbps are given in Figures (8.29(b)) and (8.29(a)), respectively. As we can observe from the figures, the horizontal axis is the UE indices which indicate that there are 28 UEs being served by the eNodeB 4. The vertical axis is the number of resource elements allocated by the optimization in Algorithms (21) and (22). Each UE has only 1 application running. This assumption simplifies the simulation because the goal of this chapter is observing the effect of the channel for which channel-aware EURA is performed. Furthermore, the legend show the SNR of the UEs in dB. The bit rate requirements of the applications is according to $\{0.25, 1, 5, 0.25, 1, 5, ...\}$ Mbps.

As we can observe from Figure (8.29(b)), the UEs with low SNRs are receiving more resource elements in order to meet their bit rate requirements. On the other hand, UEs with high SNR are receiving less resources. On the other hand, (8.29(a)) shows that UE throughputs are met. This is due all UEs are at good channel conditions such that the minimum SNR was 9.65 dB. We can see these in Figure (8.29(b)) that UEs with higher bit rate needs were allocated more resources, i.e. for UE k, $R_{1+3K} < R_{2+3K} < R_{3+3K}$. This statements says that the rate allocated to UE 1, 4, 7, ..., is less than the rate allocated to UEs 2, 5, 8, ..., and is less than the rates allocated to UEs 3, 6, 9, ... compared one to one

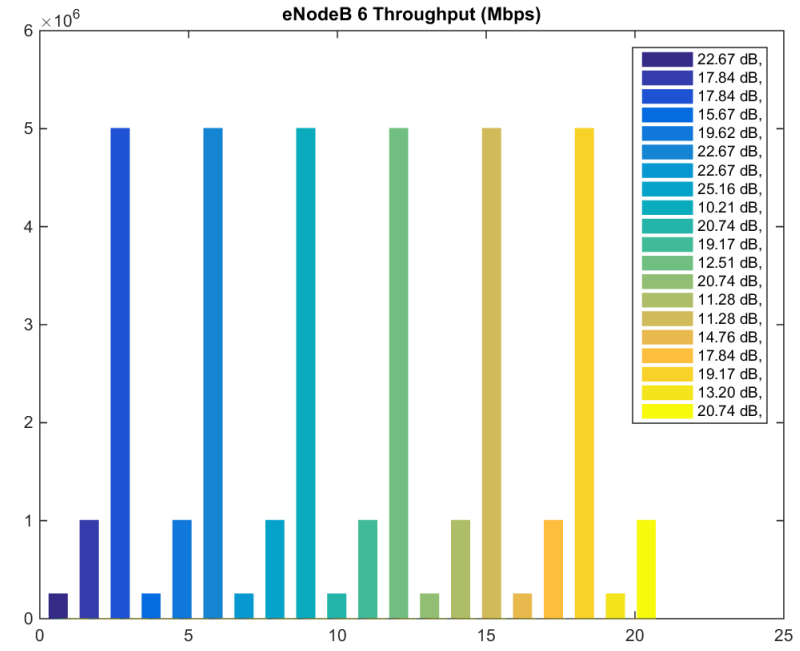


(a) eNB 5 Coverage

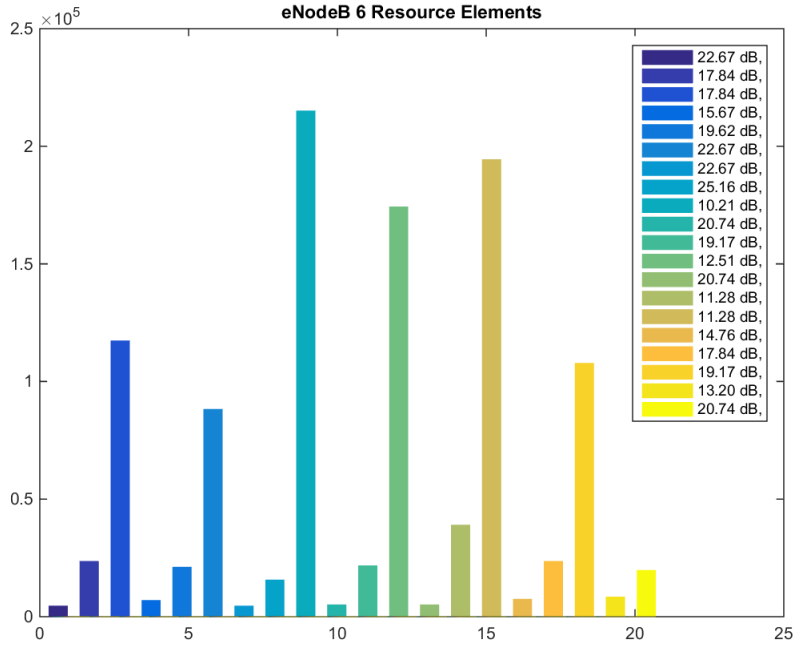


(b) eNB 5 REM

Figure 8.28: The system contains 14 UEs, each concurrently running a real-time application.



(a) Throughput of the UEs for eNB 6



(b) Resources Allocated to UEs by eNB 6

Figure 8.29: The system contains 20 UEs, each concurrently running a real-time application.

(i.e. corresponding indices compared to each other). Also, we see that for same bit rate requirements, UEs with higher SNR are allocated less resources. For instance, UEs 9 and 15 are allocated more REs than UEs 3 and 5 since they are at lower SNRs.

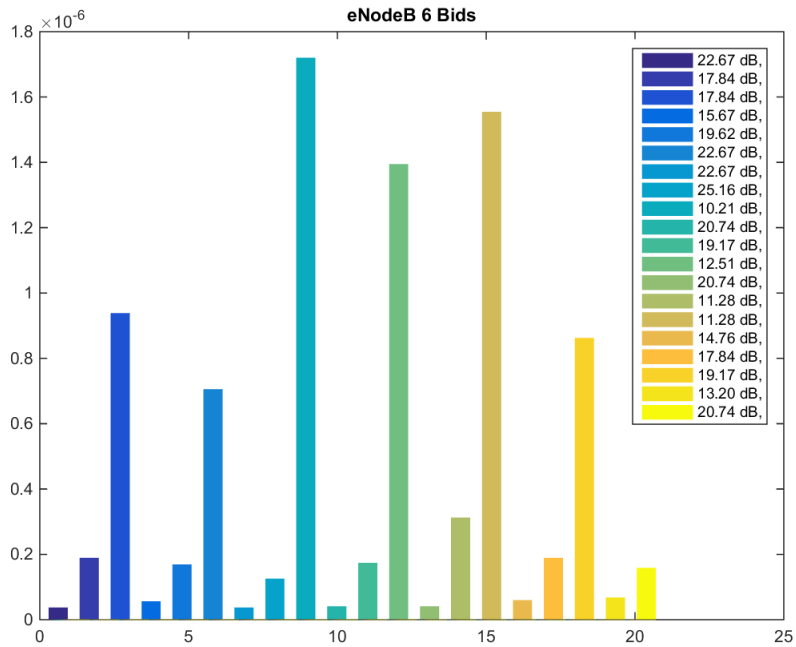


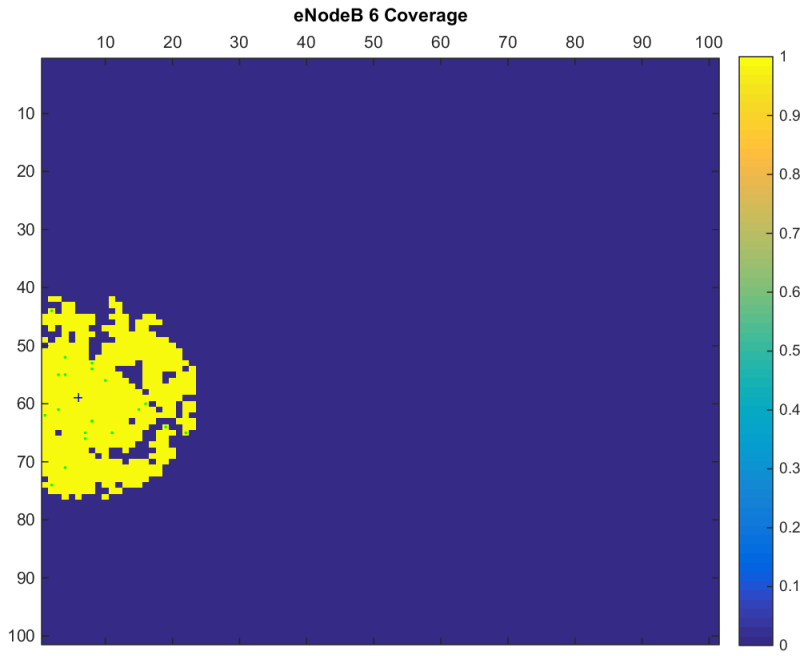
Figure 8.30: UE Bids pledged to eNB 6

Moreover, as we can see from Figure (8.30), UEs $3 + 3K$ are higher than UEs $2 + 3K$ which bid higher than UEs $1 + 3K$ due to the fact the they require more resources in view of their applications (5 Mbps vs. 1 Mbps vs 0.25 Mbps.) This plot shows the last iteration of the algorithm where the shadow prices are converged. Interestingly, UE 15 which needed high resources at good channel is bidding less than its counterparts. Furthermore, we can observe the coverage area of eNodeB 6 in Figure (8.31(a)), and as we see the coverage is not circular which is a result of REM and elevations in various directions leading to different SNRs. The black addition symbol in the middle of the coverage, shown in yellow, is the eNodeB 6 coordinates in the 101×101 grid. The side bar, shows 0 - dark blue - and 1 - yellow - for the areas which are respectively not covered and are under coverage. The green

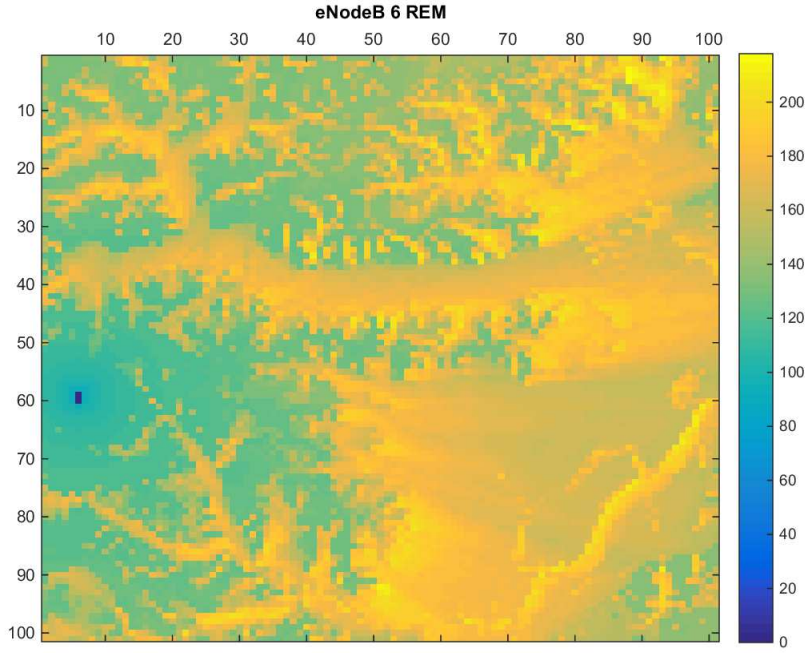
dots in the region are the UEs scattered all over the area. As we can see, UEs are in the foot print of the eNodeBs, and had good channel conditions. These simply chose eNodeB 6 as it was the strongest signal they could ever receive. Furthermore, we can see the REM which shows that a blue dot at the eNodeB location, which is for the eNodeB 6 and the blue color from the side bar represents a 0 dB loss which is expected as we are at the eNodeB position. However, as we go further away from the eNodeB, the path loss increases, which is shown by the color spectrum light blue, green, and yellow in the order of increase. The highest pathloss in the area is about 216 dB.

For eNodeB 7, the plots of resource elements allocated to the UEs as well as the throughput in Mbps are given in Figures (8.32(b)) and (8.32(a)), respectively. As we can observe from the figures, the horizontal axis is the UE indices which indicate that there are 32 UEs being served by the eNodeB 7. The vertical axis is the number of resource elements allocated by the optimization in Algorithms (21) and (22). Each UE has only 1 application running. This assumption simplifies the simulation because the goal of this chapter is observing the effect of the channel for which channel-aware EURA is performed. Furthermore, the legend show the SNR of the UEs in dB. The bit rate requirements of the applications is according to $\{0.25, 1, 5, 0.25, 1, 5, ...\}$ Mbps.

As we can observe from Figure (8.32(b)), the UEs with low SNRs are receiving more resource elements in order to meet their bit rate requirements. On the other hand, UEs with high SNR are receiving less resources. On the other hand, (8.32(a)) shows that UE throughputs are met. This is due all UEs are at good channel conditions such that all SNRs were two digit positive and the only bad SNR was -13.31 dB for UE 17. We can see these in Figure (8.32(b)) that UEs with higher bit rate needs were allocated more resources, i.e. for UE k, $R_{1+3K} < R_{2+3K} < R_{3+3K}$. This statements says that the rate allocated to UE 1, 4, 7, ..., is less than the rate allocated to UEs 2, 5, 8, ..., and is less than the rates allocated to UEs 3, 6, 9, ... compared one to one (i.e. corresponding indices compared to each other). Also,

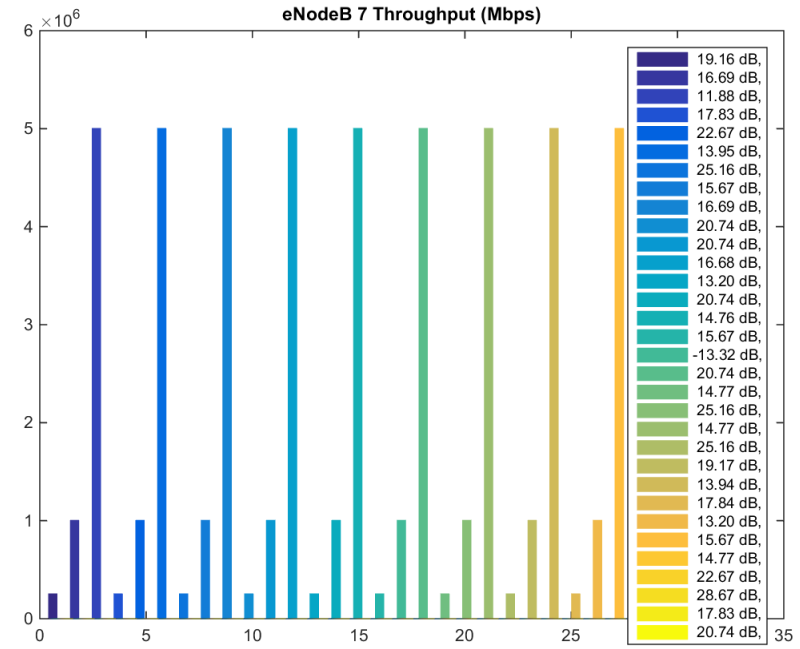


(a) eNB 6 Coverage

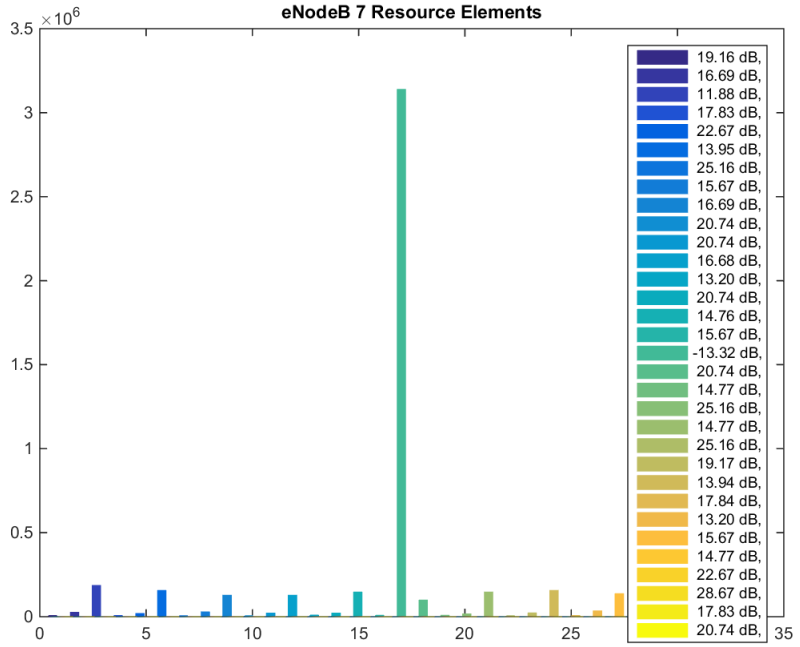


(b) eNB 6 REM

Figure 8.31: The system contains 20 UEs, each concurrently running a real-time application.



(a) Throughput of the UEs for eNB 7



(b) Resources Allocated to UEs by eNB 7

Figure 8.32: The system contains 32 UEs, each concurrently running a real-time application.

we see that for same bit rate requirements, UEs with higher SNR are allocated less resources. In particular, the spike for UE 17 is interesting which is due to its low SNT at -13.32 dB, so the algorithm has to assign more REs to this UE to meet its bit rate requirements.

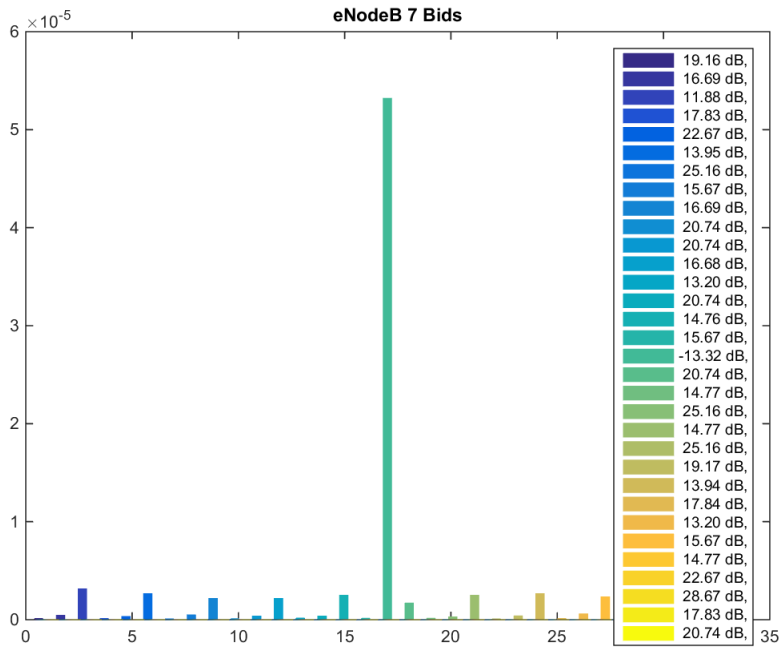


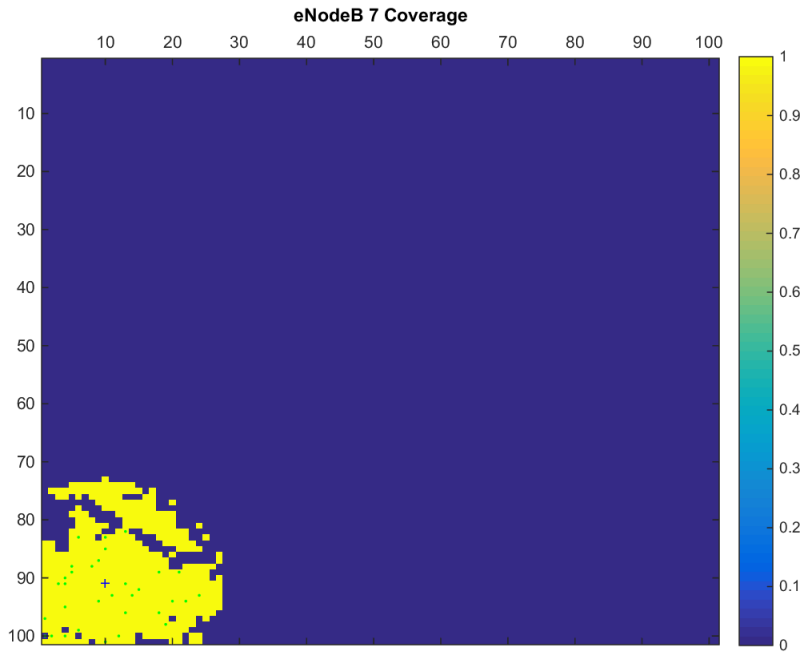
Figure 8.33: UE Bids pledged to eNB 7

Moreover, as we can see from Figure (8.33), UEs $3 + 3K$ are higher than UEs $2 + 3K$ which bid higher than UEs $1 + 3K$ due to the fact the they require more resources in view of their applications (5 Mbps vs. 1 Mbps vs 0.25 Mbps.) This plot shows the last iteration of the algorithm where the shadow prices are converged. Interestingly, UE 15 which needed high resources at good channel is bidding less than its counterparts. Furthermore, we can observe the coverage area of eNodeB 4 in Figure (8.34(a)), and as we see the coverage is not circular which is a result of REM and elevations in various directions leading to different SNRs. The black addition symbol in the middle of the coverage, shown in yellow, is the eNodeB 7 coordinates in the 101×101 grid. The side bar, shows 0 - dark blue - and 1 - yellow - for the areas which are respectively not covered and are under coverage. The green

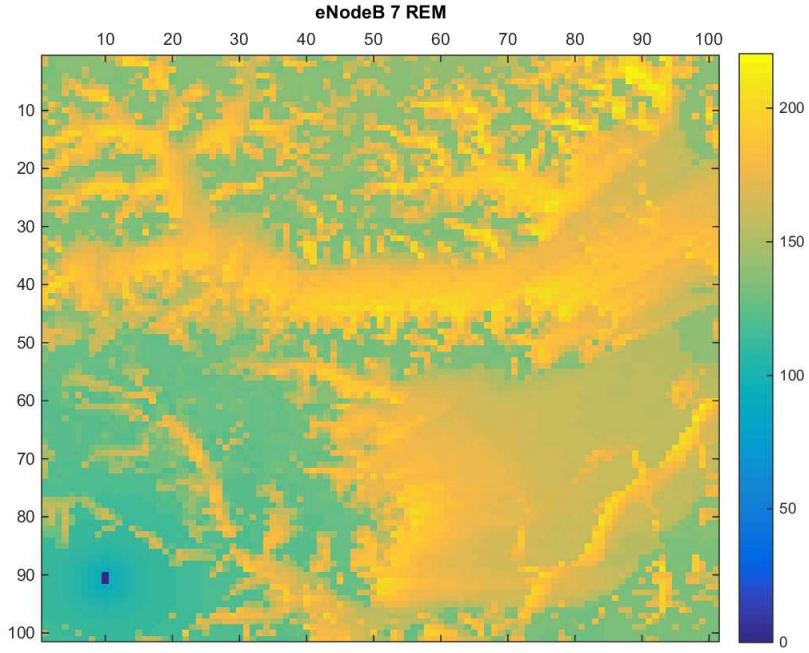
dots in the region are the UEs scattered all over the area. As we can see, UEs are in the foot print of the eNodeBs, and had good channel conditions. These simply chose eNodeB 7 as it was the strongest signal they could ever receive. Furthermore, we can see the REM which shows that a blue dot at the eNodeB location, which is for the eNodeB 7 and the blue color from the side bar represents a 0 dB loss which is expected as we are at the eNodeB position. However, as we go further away from the eNodeB, the path loss increases, which is shown by the color spectrum light blue, green, and yellow in the order of increase. The highest pathloss in the area is about 214 dB.

For eNodeB 8, the plots of REs allocated to the UEs as well as the throughput in Mbps are given in Figures (8.35(b)) and (8.35(a)), respectively. As we can observe from the figures, the horizontal axis is the UE indices which indicate that there are 27 UEs being served by the eNodeB 8. The vertical axis is the number of resource elements allocated by the optimization in Algorithms (21) and (22). Each UE has only 1 application running. This assumption simplifies the simulation because the goal of this chapter is observing the effect of the channel for which channel-aware EURA is performed. Furthermore, the legend show the SNR of the UEs in dB. The bit rate requirements of the applications is according to $\{0.25, 1, 5, 0.25, 1, 5, ...\}$ Mbps.

As we can observe from Figure (8.35(b)), the UEs with low SNRs are receiving more resource elements in order to meet their bit rate requirements. On the other hand, UEs with high SNR are receiving less resources. On the other hand, (8.35(a)) shows that UE throughputs are met. This is due all UEs are at good channel conditions such that the minimum SNR was -15.4759 dB. We can see these in Figure (8.35(b)) that UEs with higher bit rate needs were allocated more resources, i.e. for UE k, $R_{1+3K} < R_{2+3K} < R_{3+3K}$. This statements says that the rate allocated to UE 1, 4, 7, ..., is less than the rate allocated to UEs 2, 5, 8, ..., and is less than the rates allocated to UEs 3, 6, 9, ... compared one to one (i.e. corresponding indices compared to each other). Also, we see that for same bit rate

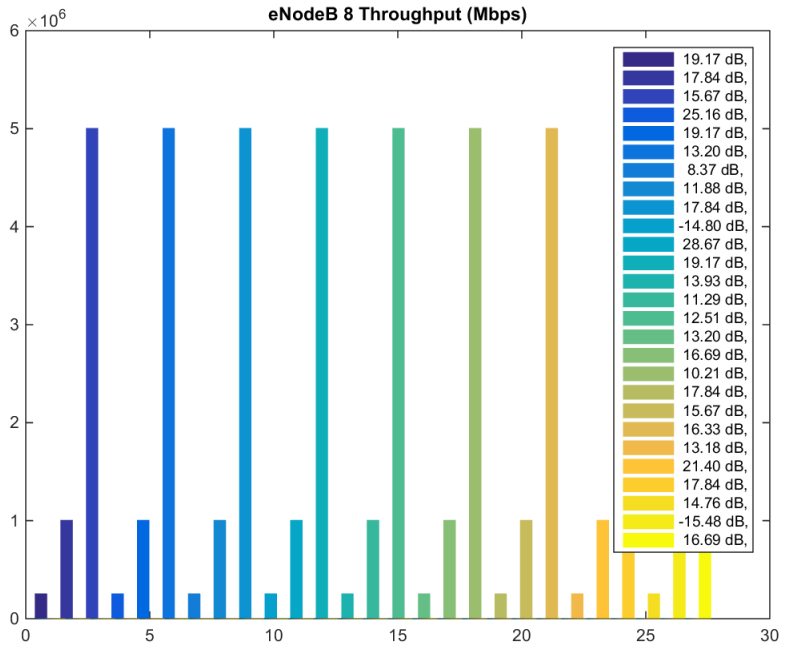


(a) eNB 7 Coverage

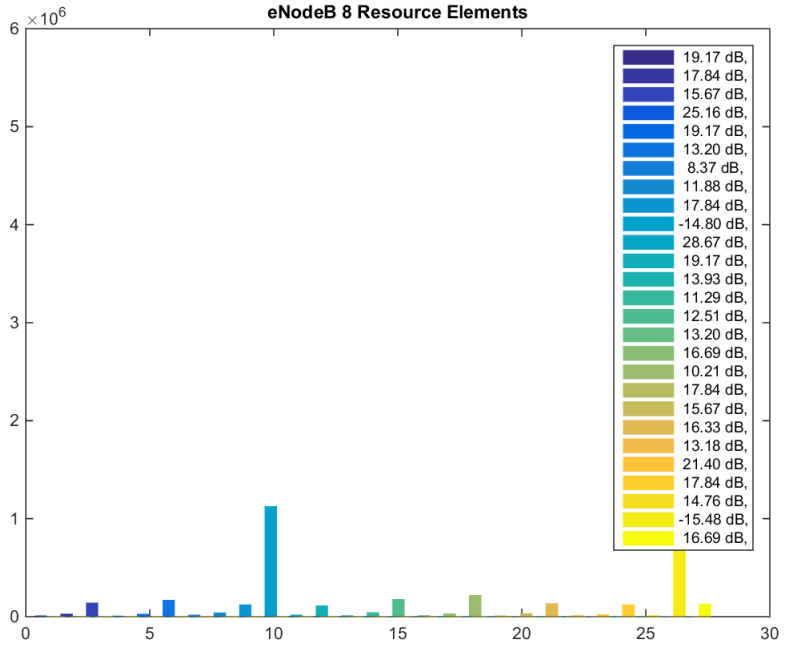


(b) eNB 7 REM

Figure 8.34: The system contains 32 UEs, each concurrently running a real-time application.



(a) Throughput of the UEs for eNB 8



(b) Resources Allocated to UEs by eNB 8

Figure 8.35: The system contains 27 UEs, each concurrently running a real-time application.

requirements, UEs with higher SNR are allocated less resources. For instance, UEs 10 and 26 are allocated more REs since they are at lower SNRs -14.80 and -15.48 dB respectively.

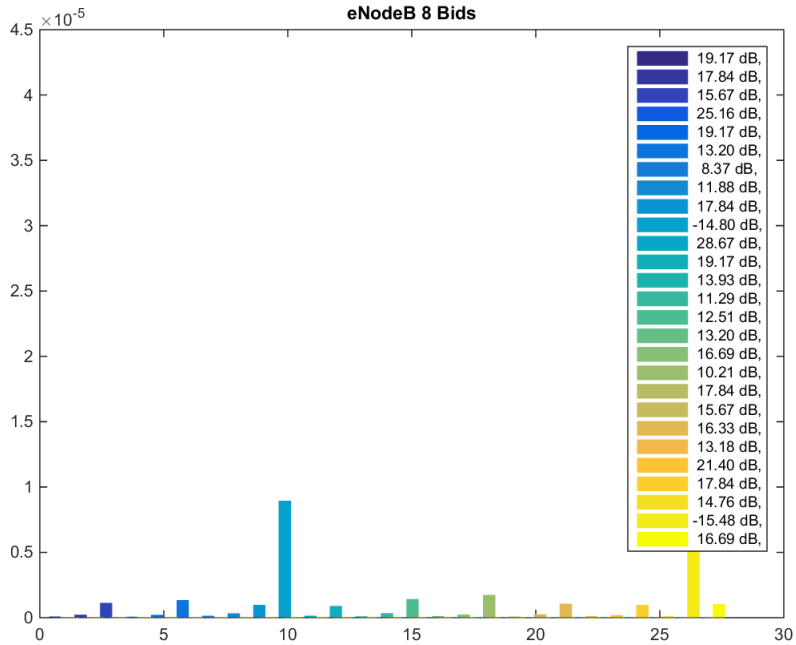


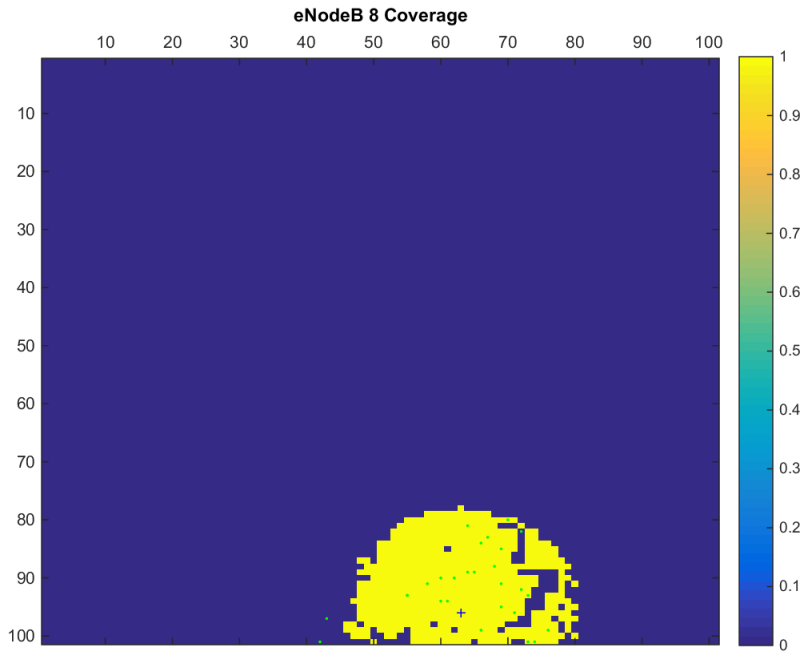
Figure 8.36: UE Bids pledged to eNB 8

Also, as we can see from Figure (8.36), UEs $3 + 3K$ are higher than UEs $2 + 3K$ which bid higher than UEs $1 + 3K$ due to the fact the they require more resources in view of their applications (5 Mbps vs. 1 Mbps vs 0.25 Mbps.) This plot shows the last iteration of the algorithm where the shadow prices are converged. Interestingly, UE 15 which needed high resources at good channel is bidding less than its counterparts. Furthermore, we can observe the coverage area of eNodeB 8 in Figure (8.37(a)), and as we see the coverage is not circular which is a result of REM and elevations in various directions leading to different SNRs. The black addition symbol in the middle of the coverage, shown in yellow, is the eNodeB 8 coordinates in the 101×101 grid. The side bar, shows 0 - dark blue - and 1 - yellow - for the areas which are respectively not covered and are under coverage. The green dots in the region are the UEs scattered all over the area. As we can see, UEs are in the

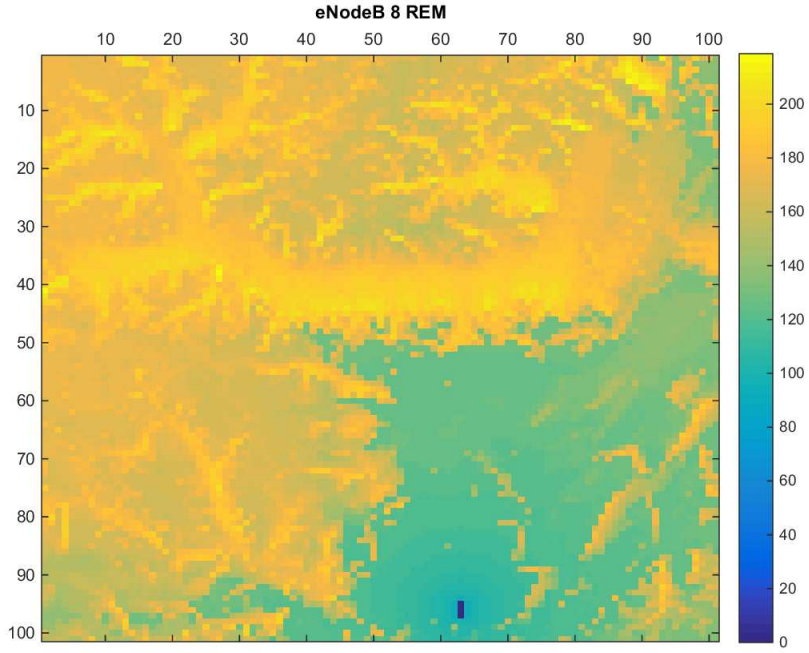
foot print of the eNodeBs, and had good channel conditions. These simply chose eNodeB 8 as it was the strongest signal they could ever receive. Furthermore, we can see the REM which shows that a blue dot at the eNodeB location, which is for the eNodeB 8 and the blue color from the side bar represents a 0 dB loss which is expected as we are at the eNodeB position. However, as we go further away from the eNodeB, the path loss increases, which is shown by the color spectrum light blue, green, and yellow in the order of increase. The highest pathloss in the area is about 221 dB.

For eNodeB 9, the plots of resource elements allocated to the UEs as well as the throughput in Mbps are given in Figures (8.38(b)) and (8.38(a)), respectively. As we can observe from the figures, the horizontal axis is the UE indices which indicate that there are 28 UEs being served by the eNodeB 9. The vertical axis is the number of resource elements allocated by the optimization in Algorithms (21) and (22). Each UE has only 1 application running. This assumption simplifies the simulation because the goal of this chapter is observing the effect of the channel for which channel-aware EURA is performed. Furthermore, the legend show the SNR of the UEs in dB. The bit rate requirements of the applications is according to $\{0.25, 1, 5, 0.25, 1, 5, ...\}$ Mbps.

As we can observe from Figure (8.38(b)), the UEs with low SNRs are receiving more resource elements in order to meet their bit rate requirements. On the other hand, UEs with high SNR are receiving less resources. On the other hand, (8.38(a)) shows that UE throughputs are met. This is due all UEs are at good channel conditions such that the minimum SNR was -0.6 dB. We can see these in Figure (8.38(b)) that UEs with higher bit rate needs were allocated more resources, i.e. for UE k, $R_{1+3K} < R_{2+3K} < R_{3+3K}$. This statements says that the rate allocated to UE 1, 4, 7, ..., is less than the rate allocated to UEs 2, 5, 8, ..., and is less than the rates allocated to UEs 3, 6, 9, ... compared one to one (i.e. corresponding indices compared to each other). Also, we see that for same bit rate requirements, UEs with higher SNR are allocated less resources. For instance, UEs 3 and 15

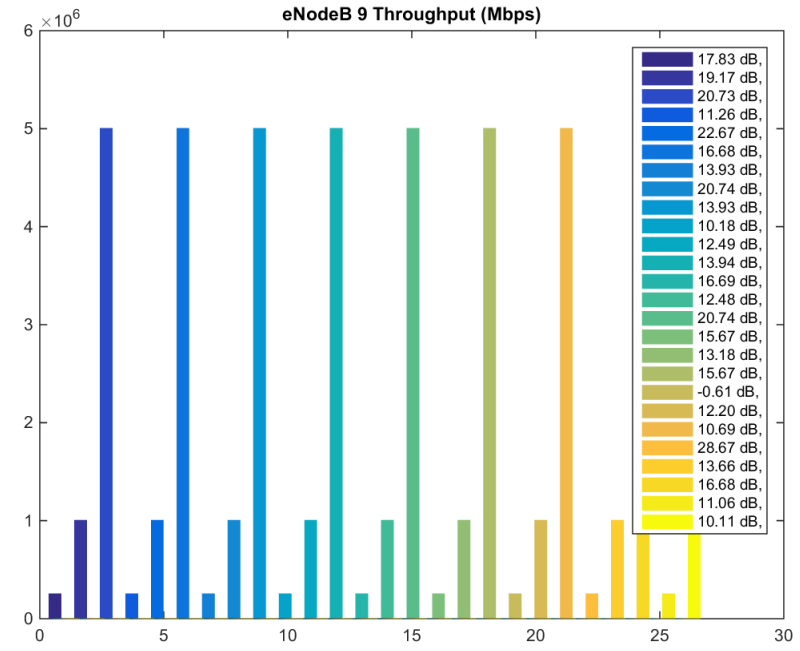


(a) eNB 8 Coverage

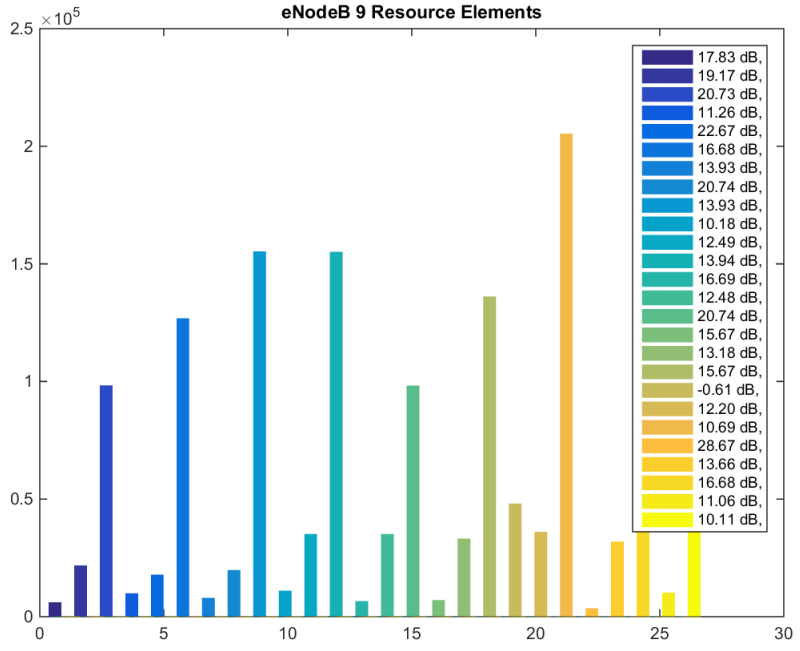


(b) eNB 8 REM

Figure 8.37: The system contains 27 UEs, each concurrently running a real-time application.



(a) Throughput of the UEs for eNB 9



(b) Resources Allocated to UEs by eNB 9

Figure 8.38: The system contains 26 UEs, each concurrently running a real-time application.

are allocated less REs compared to UES 6, 9, 12, 18, and 21 since they are at lower SNRs. In particular, the spike for UE 20 due to its low SNR at -0.6 dB.

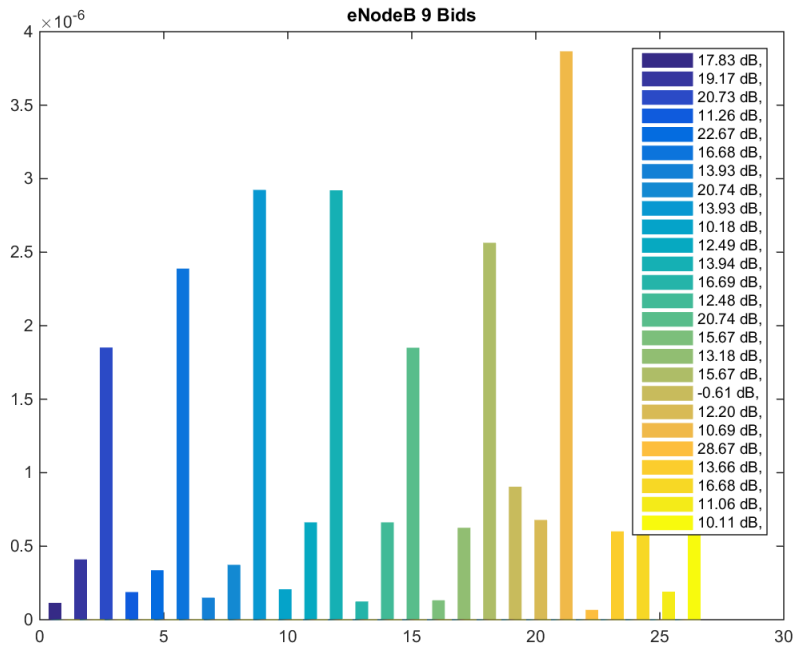


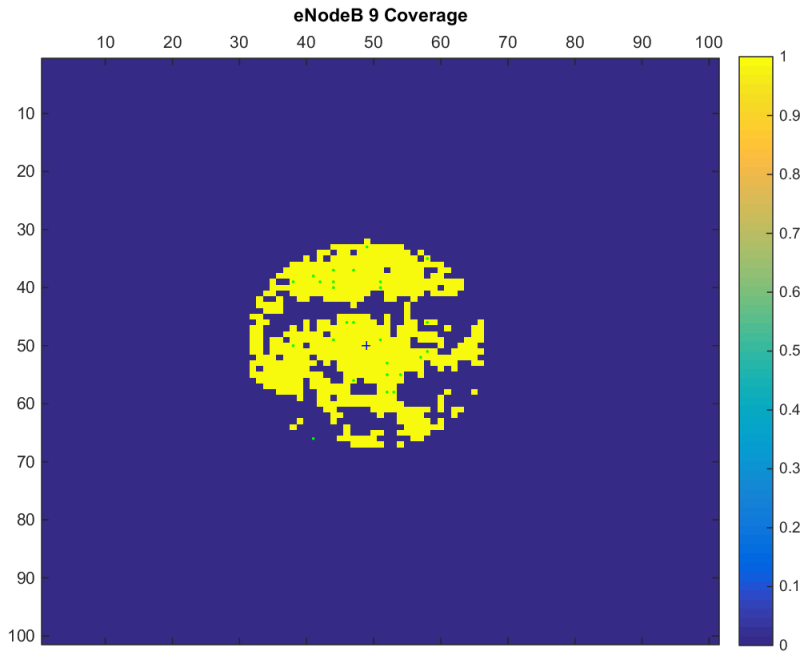
Figure 8.39: UE Bids pledged to eNB 9

Moreover, as we can see from Figure (8.39), UEs $3+3K$ are higher than UEs $2+3K$ which bid higher than UEs $1+3K$ due to the fact the they require more resources in view of their applications (5 Mbps vs. 1 Mbps vs 0.25 Mbps.) This plot shows the last iteration of the algorithm where the shadow prices are converged. Interestingly, UE 15 which needed high resources at bad channel conditions is bidding higher than its counterparts. Furthermore, we can observe the coverage area of eNodeB 9 in Figure (8.40(a)), and as we see the coverage is not circular which is a result of REM and elevations in various directions leading to different SNRs. The black addition symbol in the middle of the coverage, shown in yellow, is the eNodeB 9 coordinates in the 101×101 grid. The side bar, shows 0 - dark blue - and 1 - yellow - for the areas which are respectively not covered and are under coverage. The green dots in the region are the UEs scattered all over the area. As we can see, UEs are in the

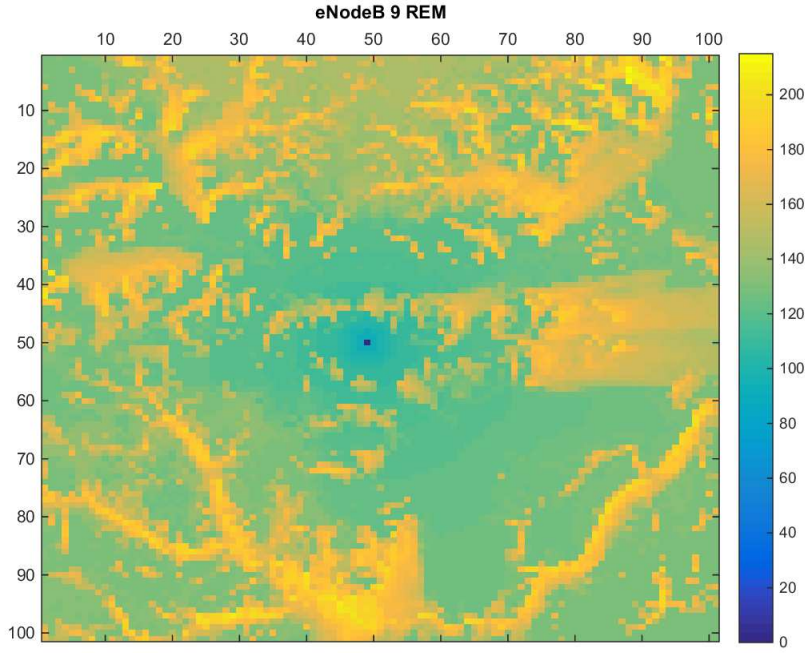
foot print of the eNodeBs, and had good channel conditions. These simply chose eNodeB 9 as it was the strongest signal they could ever receive. Also, UE 15 is outside the footprint of the eNodeB which is refelected in its low SNR at -0.6 dB. Furthermore, we can see the REM which shows that a blue dot at the eNodeB location, which is for the eNodeB 9 and the blue color from the side bar represents a 0 dB loss which is expected as we are at the eNodeB position. However, as we go further away from the eNodeB, the path loss increases, which is shown by the color spectrum light blue, green, and yellow in the order of increase. The highest pathloss in the area is about 208 dB.

For eNodeB 10, the plots of resource elements allocated to the UEs as well as the throughput in Mbps are given in Figures (8.41(b)) and (8.41(a)), respectively. As we can observe from the figures, the horizontal axis is the UE indices which indicate that there are 14 UEs being served by the eNodeB 10. The vertical axis is the number of resource elements allocated by the optimization in Algorithms (21) and (22). Each UE has only 1 application running. This assumption simplifies the simulation because the goal of this chapter is observing the effect of the channel for which channel-aware EURA is performed. Furthermore, the legend show the SNR of the UEs in dB. The bit rate requirements of the applications is according to $\{0.25, 1, 5, 0.25, 1, 5, ...\}$ Mbps.

As we can observe from Figure (8.41(b)), the UEs with low SNRs are receiving more resource elements in order to meet their bit rate requirements. On the other hand, UEs with high SNR are receiving less resources. On the other hand, (8.41(a)) shows that UE throughputs are met. This is due all UEs are at good channel conditions such that the minimum SNR was 5.62 dB. We can see these in Figure (8.41(b)) that UEs with higher bit rate needs were allocated more resources, i.e. for UE k, $R_{1+3K} < R_{2+3K} < R_{3+3K}$. This statements says that the rate allocated to UE 1, 4, 7, ..., is less than the rate allocated to UEs 2, 5, 8, ..., and is less than the rates allocated to UEs 3, 6, 9, ... compared one to one (i.e. corresponding indices compared to each other). Also, we see that for same bit rate

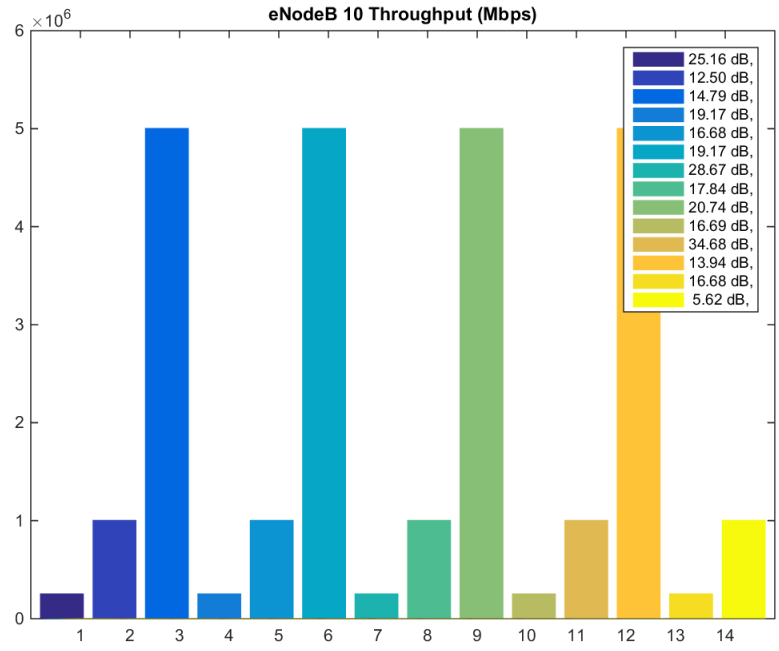


(a) eNB 9 Coverage

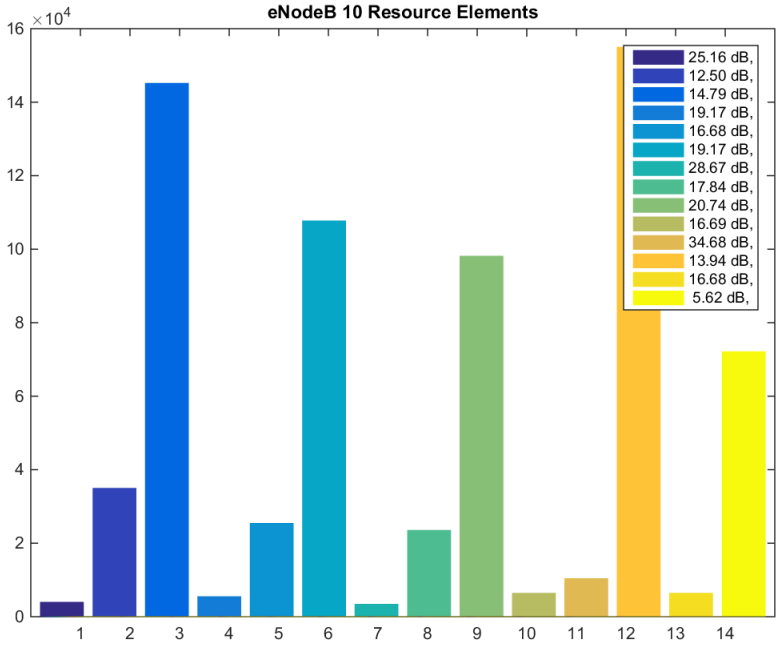


(b) eNB 9 REM

Figure 8.40: The system contains 26 UEs, each concurrently running a real-time application.



(a) Throughput of the UEs for eNB 10



(b) Resources Allocated to UEs by eNB 10

Figure 8.41: The system contains 14 UEs, each concurrently running a real-time application.

requirements, UEs with higher SNR are allocated less resources. For instance, UEs 3 and 12 are allocated more REs than UEs 6 and 9 since they are at lower SNRs. In particular, the spike for UE 14 is interesting as opposed to other rates of UEs R_{2+3K}. The spike is because this UE is at the lowest SNR situation in the system 5.62 dB.

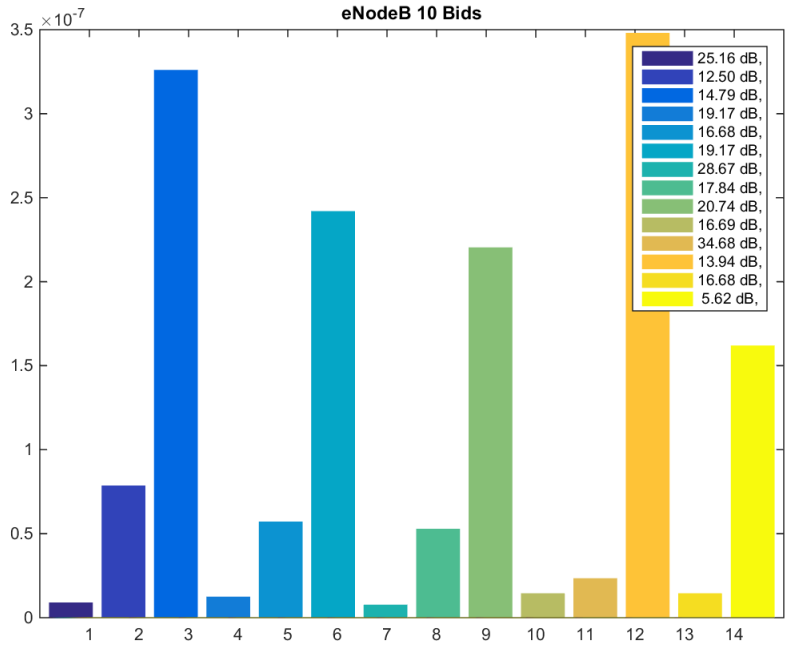


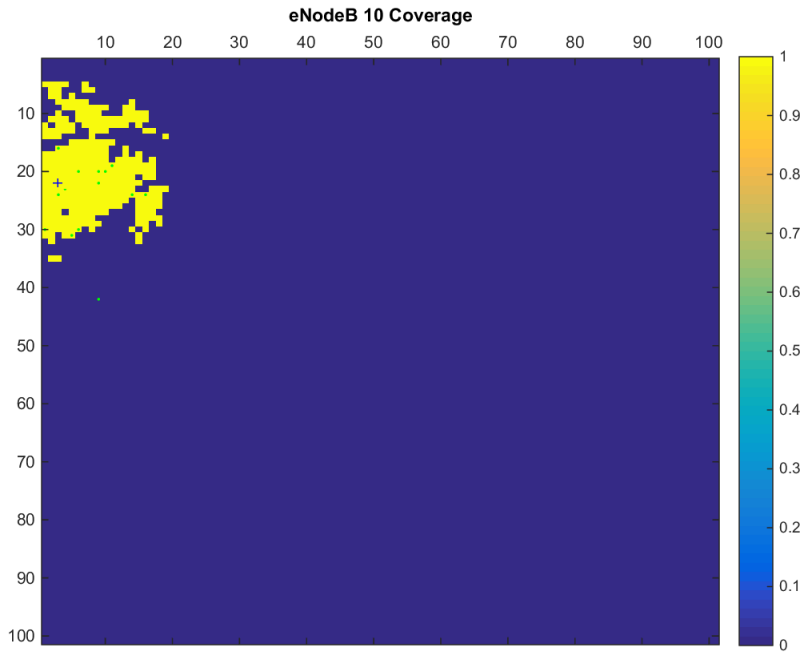
Figure 8.42: UE Bids pledged to eNB 10

Moreover, as we can see from Figure (8.42), UEs $3+3K$ are higher than UEs $2+3K$ which bid higher than UEs $1+3K$ due to the fact the they require more resources in view of their applications (5 Mbps vs. 1 Mbps vs 0.25 Mbps.) This plot shows the last iteration of the algorithm where the shadow prices are converged. Interestingly, UE 14 which needed high resources at good channel is bidding less than its counterparts. Furthermore, we can observe the coverage area of eNodeB 10 in Figure (8.43(a)), and as we see the coverage is not circular which is a result of REM and elevations in various directions leading to different SNRs. The black addition symbol in the middle of the coverage, shown in yellow, is the eNodeB 10 coordinates in the 101×101 grid. The side bar, shows 0 - dark blue - and 1 - yellow - for

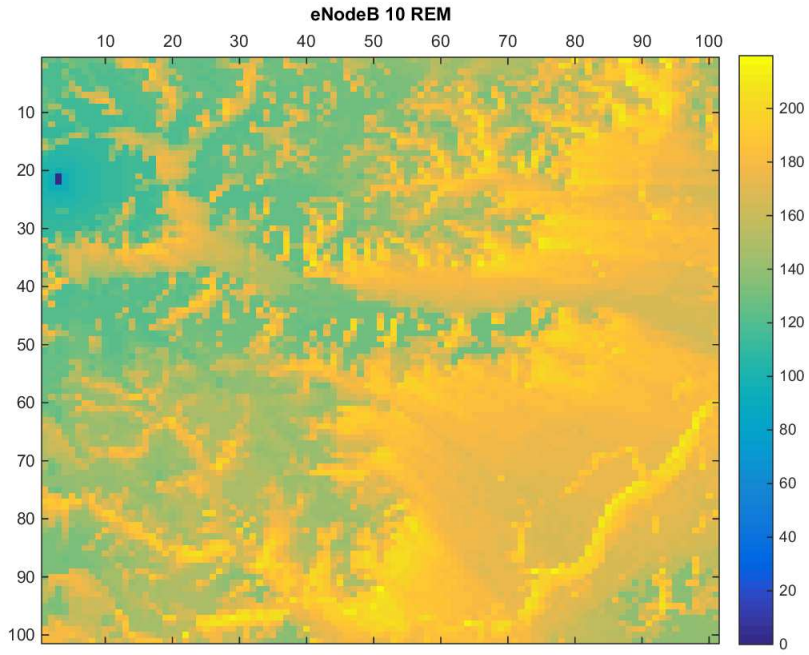
the areas which are respectively not covered and are under coverage. The green dots in the region are the UEs scattered all over the area. As we can see, UEs are in the foot print of the eNodeBs, and had good channel conditions. These simply chose eNodeB 10 as it was the strongest signal they could ever receive. Furthermore, we can see the REM which shows that a blue dot at the eNodeB location, which is for the eNodeB 10 and the blue color from the side bar represents a 0 dB loss which is expected as we are at the eNodeB position. We see that UE 14 which has the lowest SNR is outside the footprint of the eNodeB. However, as we go further away from the eNodeB, the path loss increases, which is shown by the color spectrum light blue, green, and yellow in the order of increase. The highest pathloss in the area is about 210 dB.

For eNodeB 11, the plots of resource elements allocated to the UEs as well as the throughput in Mbps are given in Figures (8.44(b)) and (8.44(a)), respectively. As we can observe from the figures, the horizontal axis is the UE indices which indicate that there are 24 UEs being served by the eNodeB 11. The vertical axis is the number of resource elements allocated by the optimization in Algorithms (21) and (22). Each UE has only 1 application running. This assumption simplifies the simulation because the goal of this chapter is observing the effect of the channel for which channel-aware EURA is performed. Furthermore, the legend show the SNR of the UEs in dB. The bit rate requirements of the applications is according to $\{0.25, 1, 5, 0.25, 1, 5, ...\}$ Mbps.

As we can observe from Figure (8.44(b)), the UEs with low SNRs are receiving more resource elements in order to meet their bit rate requirements. On the other hand, UEs with high SNR are receiving less resources. On the other hand, (8.44(a)) shows that UE throughputs are met. This is due all UEs are at good channel conditions such that the minimum SNR was 10.7 dB. We can see these in Figure (8.44(b)) that UEs with higher bit rate needs were allocated more resources, i.e. for UE k, $R_{1+3K} < R_{2+3K} < R_{3+3K}$. This statements says that the rate allocated to UE 1, 4, 7, ..., is less than the rate allocated to

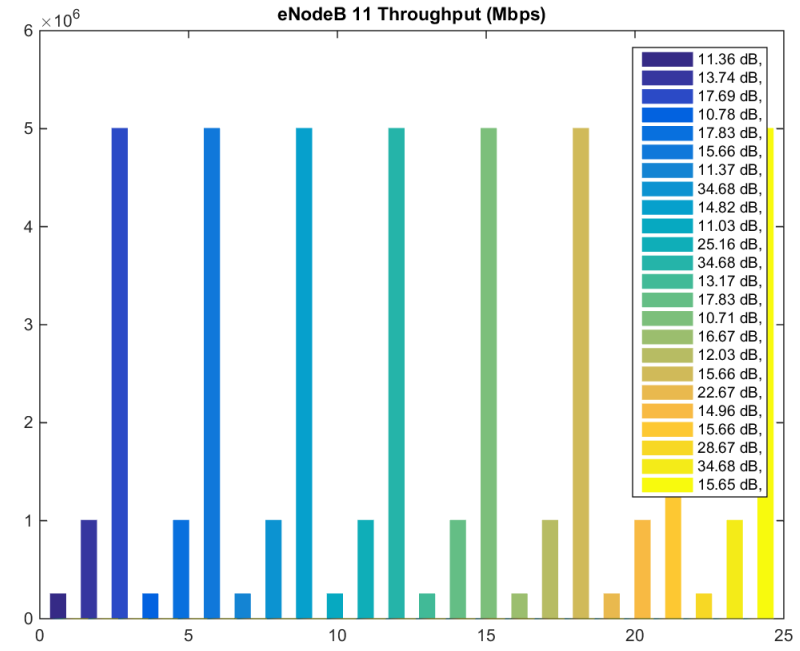


(a) eNB 10 Coverage

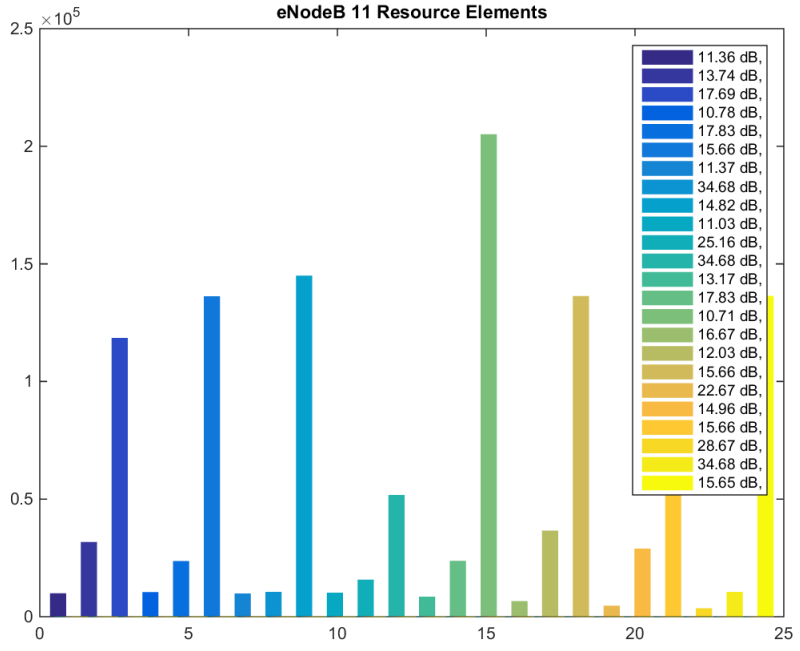


(b) eNB 10 REM

Figure 8.43: The system contains 14 UEs, each concurrently running a real-time application.



(a) Throughput of the UEs for eNB 11



(b) Resources Allocated to UEs by eNB 11

Figure 8.44: The system contains 24 UEs, each concurrently running a real-time application.

UEs 2, 5, 8, ..., and is less than the rates allocated to UEs 3, 6, 9, ... compared one to one (i.e. corresponding indices compared to each other). Also, we see that for same bit rate requirements, UEs with higher SNR are allocated less resources. For instance, UE 15 is allocated more REs since it is at the lowest SNR in the system 10.7 dB.

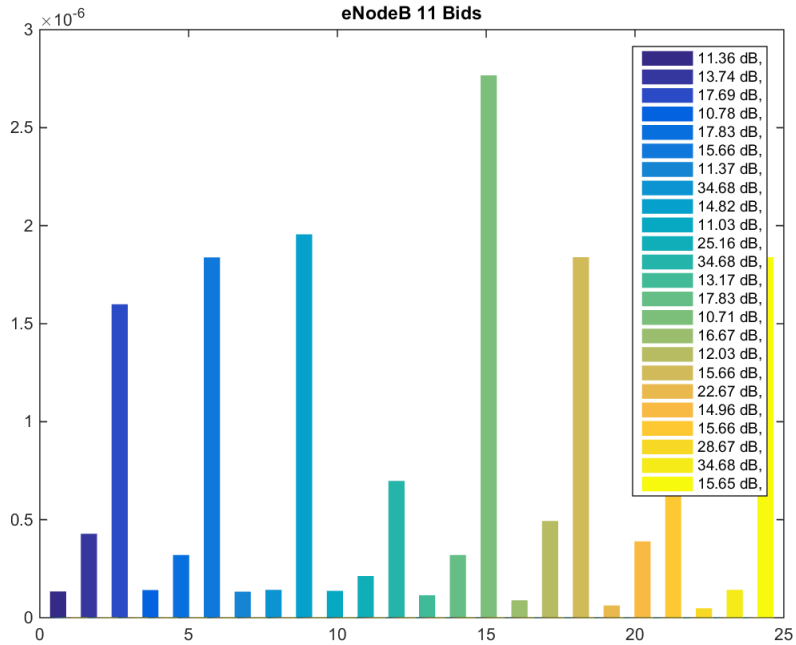


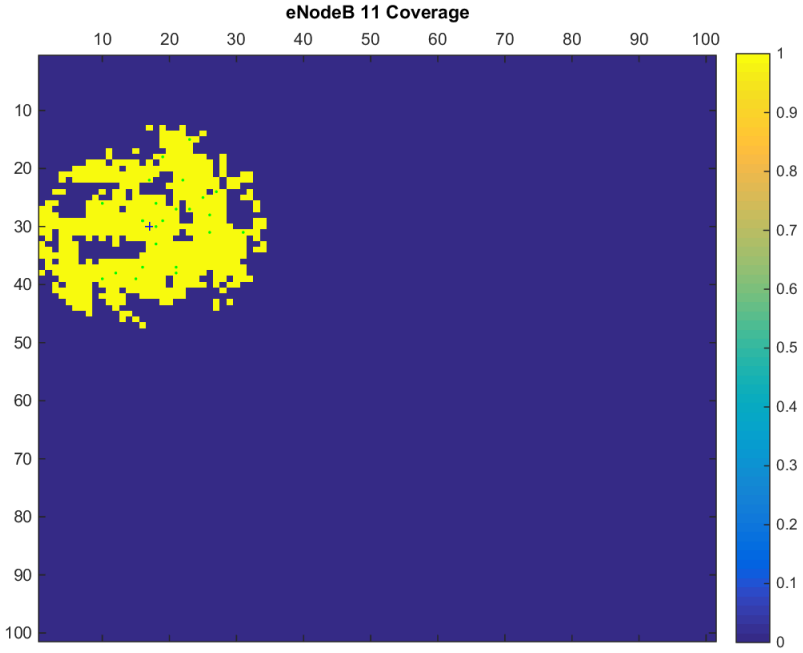
Figure 8.45: UE Bids pledged to eNB 11

Moreover, as we can see from Figure (8.45), UEs $3 + 3K$ are higher than UEs $2 + 3K$ which bid higher than UEs $1 + 3K$ due to the fact the they require more resources in view of their applications (5 Mbps vs. 1 Mbps vs 0.25 Mbps.) This plot shows the last iteration of the algorithm where the shadow prices are converged. Interestingly, UE 15 which needed high resources at good channel is bidding less than its counterparts. Furthermore, we can observe the coverage area of eNodeB 11 in Figure (8.46(a)), and as we see the coverage is not circular which is a result of REM and elevations in various directions leading to different SNRs. The black addition symbol in the middle of the coverage, shown in yellow, is the eNodeB 11 coordinates in the 101×101 grid. The side bar, shows 0 - dark blue - and 1 -

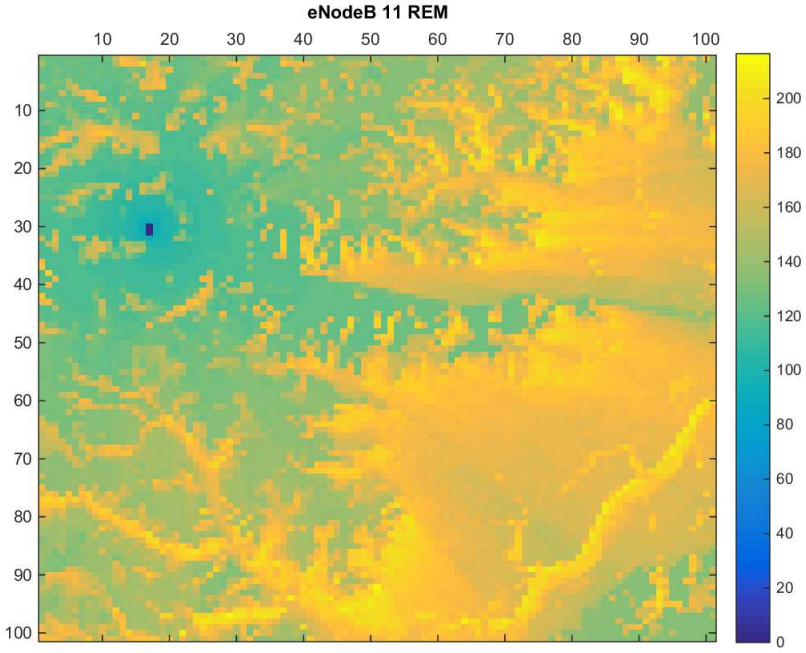
yellow - for the areas which are respectively not covered and are under coverage. The green dots in the region are the UEs scattered all over the area. As we can see, UEs are in the foot print of the eNodeBs, and had good channel conditions. These simply chose eNodeB 11 as it was the strongest signal they could ever receive. Furthermore, we can see the REM which shows that a blue dot at the eNodeB location, which is for the eNodeB 11 and the blue color from the side bar represents a 0 dB loss which is expected as we are at the eNodeB position. However, as we go further away from the eNodeB, the path loss increases, which is shown by the color spectrum light blue, green, and yellow in the order of increase. The highest pathloss in the area is about 206 dB.

For eNodeB 12, the plots of resource elements allocated to the UEs as well as the throughput in Mbps are given in Figures (8.47(b)) and (8.47(a)), respectively. As we can observe from the figures, the horizontal axis is the UE indices which indicate that there are 25 UEs being served by the eNodeB 12. The vertical axis is the number of resource elements allocated by the optimization in Algorithms (21) and (22). Each UE has only 1 application running. This assumption simplifies the simulation because the goal of this chapter is observing the effect of the channel for which channel-aware EURA is performed. Furthermore, the legend show the SNR of the UEs in dB. The bit rate requirements of the applications is according to $\{0.25, 1, 5, 0.25, 1, 5, ...\}$ Mbps.

As we can observe from Figure (8.47(b)), the UEs with low SNRs are receiving more resource elements in order to meet their bit rate requirements. On the other hand, UEs with high SNR are receiving less resources. On the other hand, (8.47(a)) shows that UE throughputs are met. This is due all UEs are at good channel conditions such that the minimum SNR was -8.67 dB. We can see these in Figure (8.47(b)) that UEs with higher bit rate needs were allocated more resources, i.e. for UE k, $R_{1+3K} < R_{2+3K} < R_{3+3K}$. This statements says that the rate allocated to UE 1, 4, 7, ..., is less than the rate allocated to UEs 2, 5, 8, ..., and is less than the rates allocated to UEs 3, 6, 9, ... compared one to one

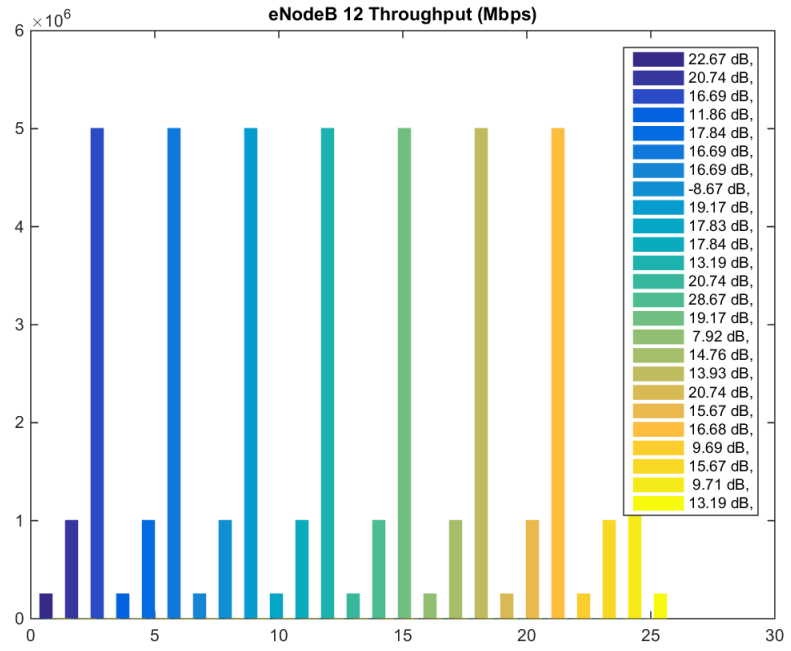


(a) eNB 11 Coverage

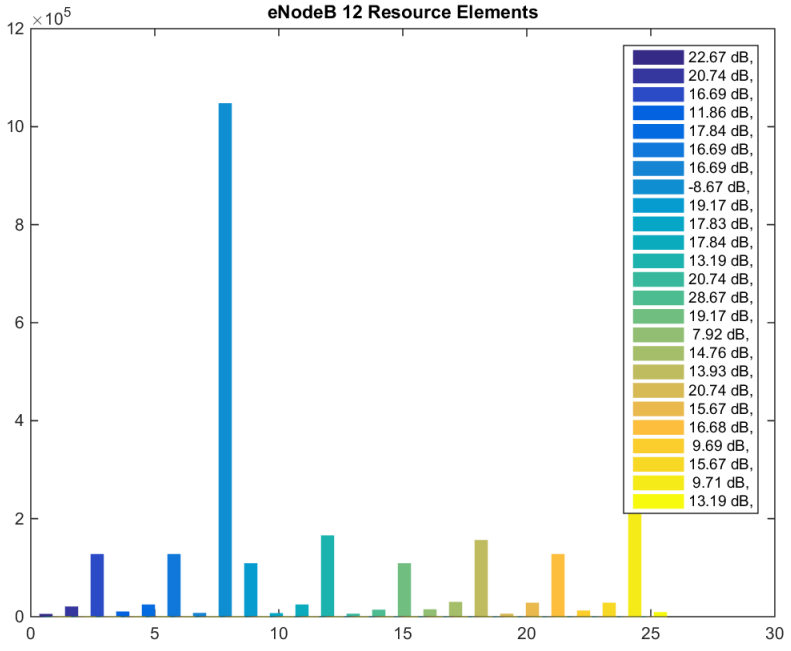


(b) eNB 11 REM

Figure 8.46: The system contains 24 UEs, each concurrently running a real-time application.



(a) Throughput of the UEs for eNB 12



(b) Resources Allocated to UEs by eNB 12

Figure 8.47: The system contains 25 UEs, each concurrently running a real-time application.

(i.e. corresponding indices compared to each other). Also, we see that for same bit rate requirements, UEs with higher SNR are allocated less resources. For instance, UEs 11 and 12 and 18 are allocated more REs since they are at lower SNRs. In particular, the spike for UE 8 is interesting esince it is at the lowest SNR situation in the system -8.67 dB.

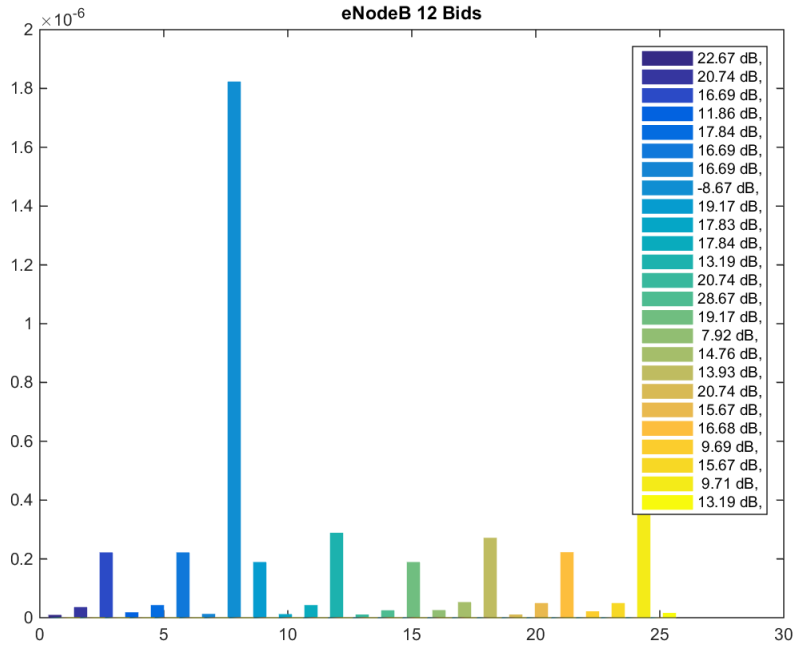


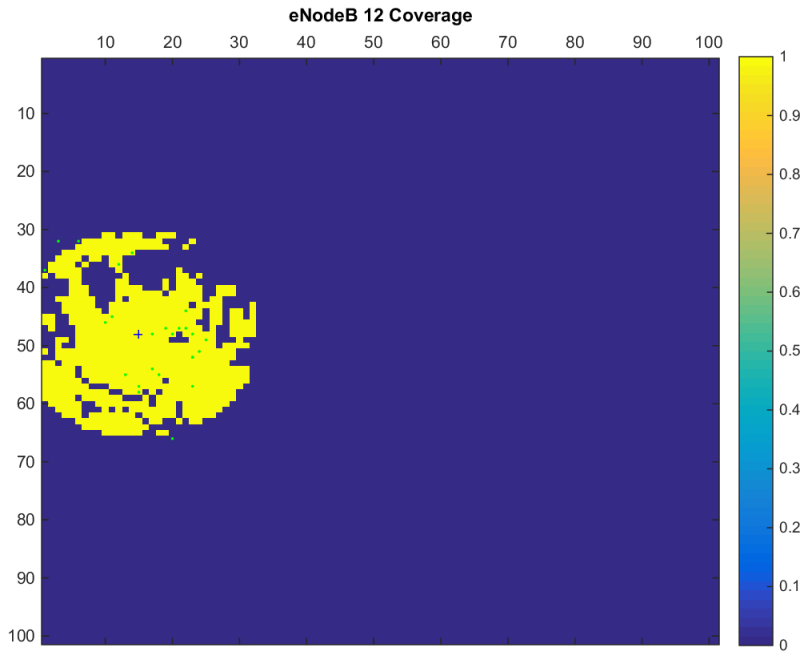
Figure 8.48: UE Bids pledged to eNB 12

Moreover, as we can see from Figure (8.48), UEs $3 + 3K$ are higher than UEs $2 + 3K$ which bid higher than UEs $1 + 3K$ due to the fact the they require more resources in view of their applications (5 Mbps vs. 1 Mbps vs 0.25 Mbps.) This plot shows the last iteration of the algorithm where the shadow prices are converged. Interestingly, UE 8 which needed high resources at bad channel is bidding higher than its counterparts. Furthermore, we can observe the coverage area of eNodeB 12 in Figure (8.49(a)), and as we see the coverage is not circular which is a result of REM and elevations in various directions leading to different SNRs. The black addition symbol in the middle of the coverage, shown in yellow, is the eNodeB 12 coordinates in the 101×101 grid. The side bar, shows 0 - dark blue - and 1 -

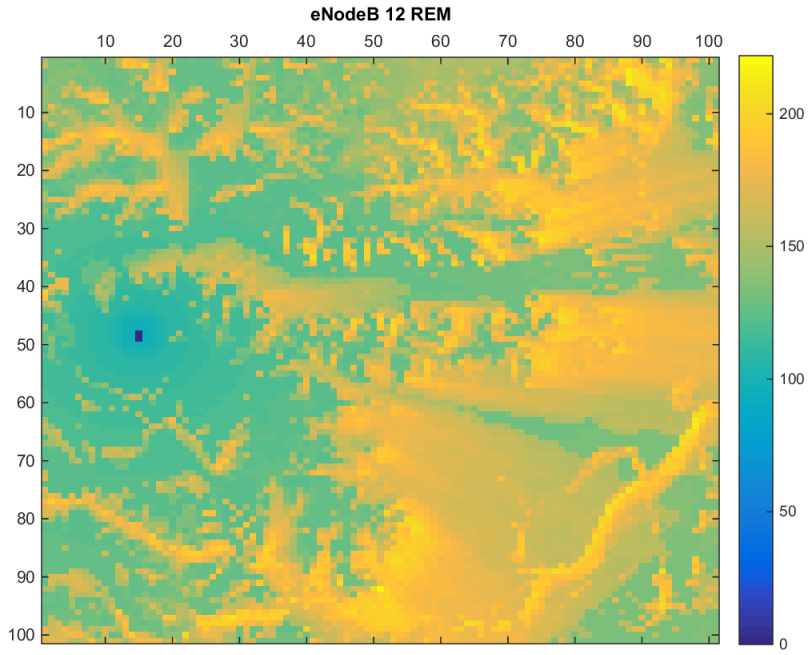
yellow - for the areas which are respectively not covered and are under coverage. The green dots in the region are the UEs scattered all over the area. As we can see, UEs are in the foot print of the eNodeBs, and had good channel conditions. These simply chose eNodeB 12 as it was the strongest signal they could ever receive. Furthermore, we can see the REM which shows that a blue dot at the eNodeB location, which is for the eNodeB 12 and the blue color from the side bar represents a 0 dB loss which is expected as we are at the eNodeB position. Furthemore, we see some UEs are outside the footprint of the eNodeB which are reflected in lower SNRs in the system. However, as we go further away from the eNodeB, the path loss increases, which is shown by the color spectrum light blue, green, and yellow in the order of increase. The highest pathloss in the area is about 205 dB.

For eNodeB 13, the plots of resource elements allocated to the UEs as well as the throughput in Mbps are given in Figures (8.50(b)) and (8.50(a)), respectively. As we can observe from the figures, the horizontal axis is the UE indices which indicate that there are 21 UEs being served by the eNodeB 13. The vertical axis is the number of resource elements allocated by the optimization in Algorithms (21) and (22). Each UE has only 1 application running. This assumption simplifies the simulation because the goal of this chapter is observing the effect of the channel for which channel-aware EURA is performed. Furthermore, the legend show the SNR of the UEs in dB. The bit rate requirements of the applications is according to $\{0.25, 1, 5, 0.25, 1, 5, ...\}$ Mbps.

As we can observe from Figure (8.50(b)), the UEs with low SNRs are receiving more resource elements in order to meet their bit rate requirements. On the other hand, UEs with high SNR are receiving less resources. On the other hand, (8.50(a)) shows that UE throughputs are met. This is due all UEs are at good channel conditions such that the minimum SNR was 13.18 dB. We can see these in Figure (8.50(b)) that UEs with higher bit rate needs were allocated more resources, i.e. for UE k, $R_{1+3K} < R_{2+3K} < R_{3+3K}$. This statements says that the rate allocated to UE 1, 4, 7, ..., is less than the rate allocated to

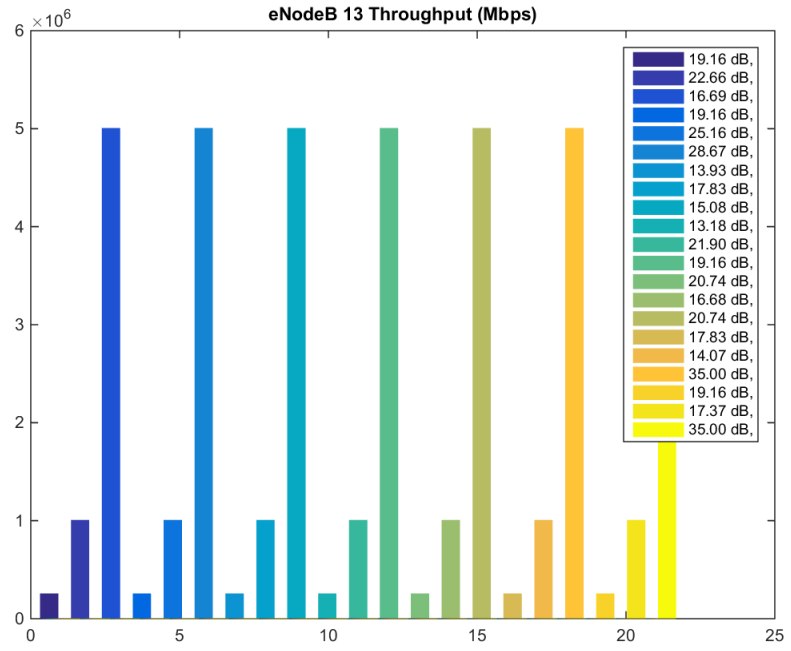


(a) eNB 12 Coverage

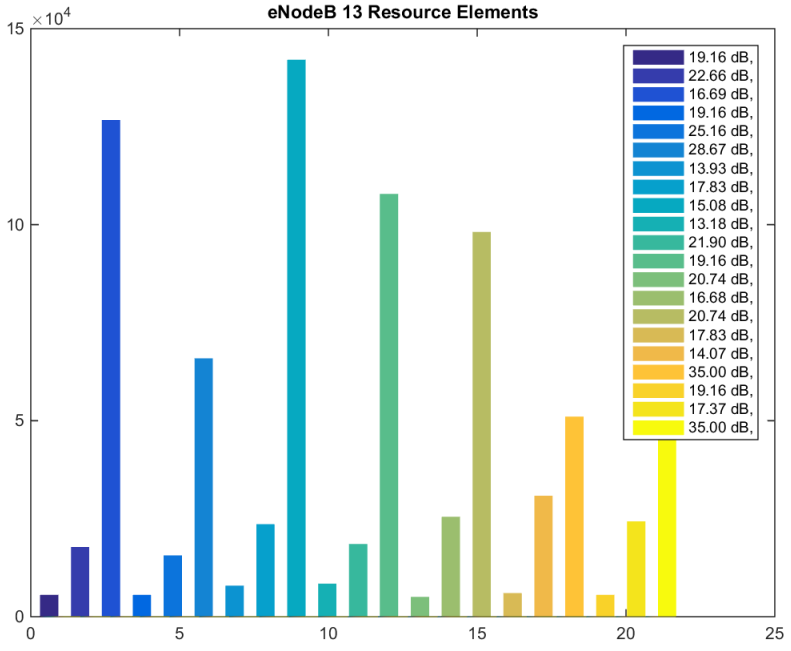


(b) eNB 12 REM

Figure 8.49: The system contains 25 UEs, each concurrently running a real-time application.



(a) Throughput of the UEs for eNB 13



(b) Resources Allocated to UEs by eNB 13

Figure 8.50: The system contains 21 UEs, each concurrently running a real-time application.

UEs 2, 5, 8, ..., and is less than the rates allocated to UEs 3, 6, 9, ... compared one to one (i.e. corresponding indices compared to each other). Also, we see that for same bit rate requirements, UEs with higher SNR are allocated less resources. For instance, UE 9 is allocated more REs since it is at lower SNRs. In particular, the spike for UE 9 is interesting because this UE is at the lowest SNR situation in the system 13.18 dB.

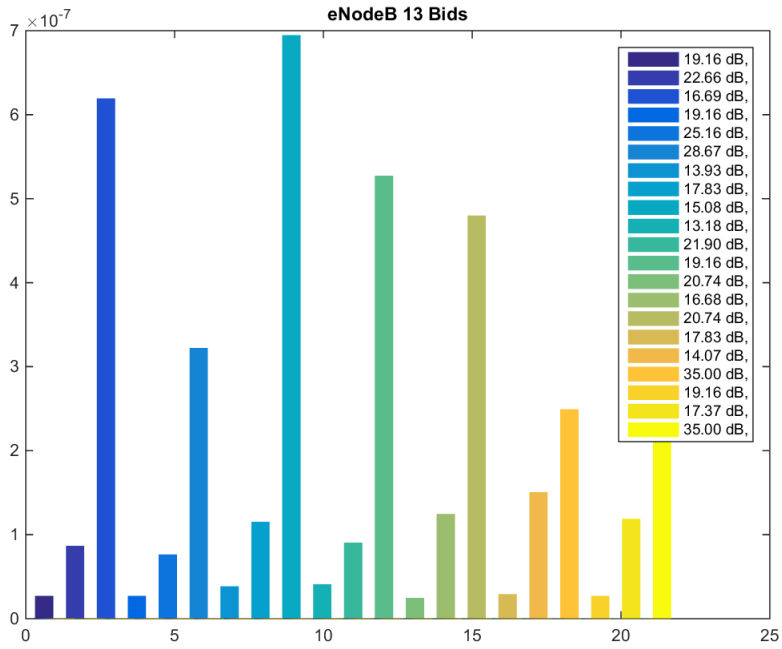


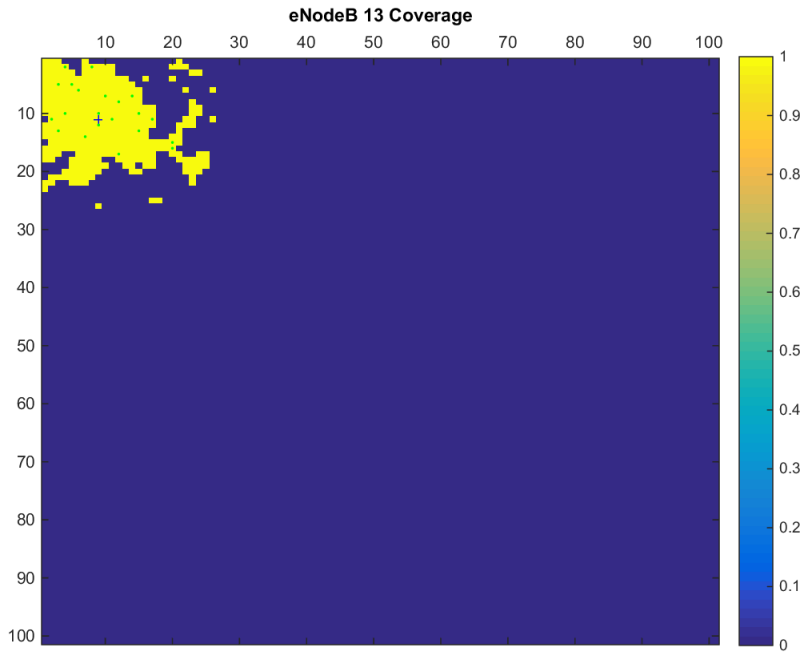
Figure 8.51: UE Bids pledged to eNB 13

Moreover, as we can see from Figure (8.51), UEs $3 + 3K$ are higher than UEs $2 + 3K$ which bid higher than UEs $1 + 3K$ due to the fact the they require more resources in view of their applications (5 Mbps vs. 1 Mbps vs 0.25 Mbps.) This plot shows the last iteration of the algorithm where the shadow prices are converged. Interestingly, UE 9 which needed high resources at good channel is bidding less than its counterparts. Furthermore, we can observe the coverage area of eNodeB 13 in Figure (8.25(a)), and as we see the coverage is not circular which is a result of REM and elevations in various directions leading to different SNRs. The black addition symbol in the middle of the coverage, shown in yellow, is the

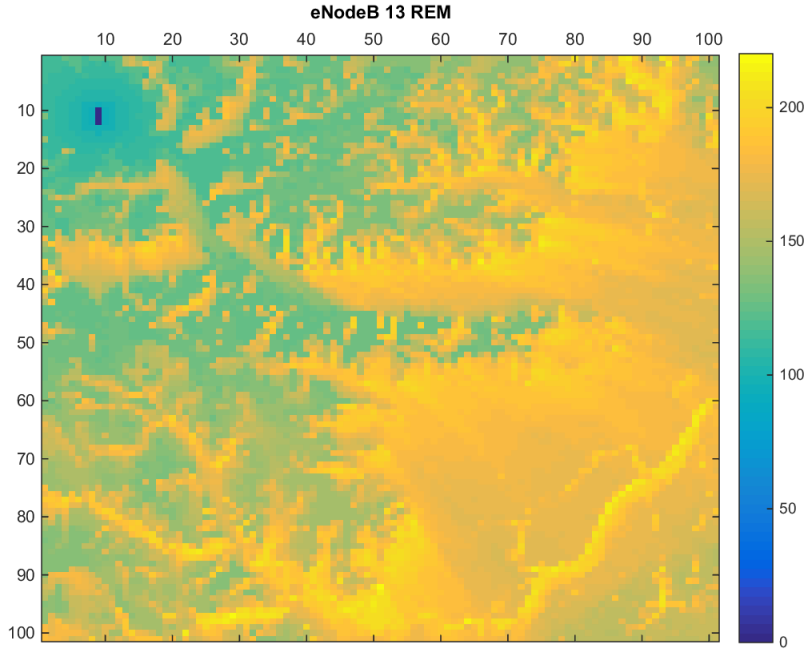
eNodeB 4 coordinates in the 101×101 grid. The side bar, shows 0 - dark blue - and 1 - yellow - for the areas which are respectively not covered and are under coverage. The green dots in the region are the UEs scattered all over the area. As we can see, UEs are in the foot print of the eNodeBs, and had good channel conditions. These simply chose eNodeB 13 as it was the strongest signal they could ever receive. Furthermore, we can see the REM which shows that a blue dot at the eNodeB location, which is for the eNodeB 13 and the blue color from the side bar represents a 0 dB loss which is expected as we are at the eNodeB position. However, as we go further away from the eNodeB, the path loss increases, which is shown by the color spectrum light blue, green, and yellow in the order of increase. The highest pathloss in the area is about 218 dB.

For eNodeB 14, the plots of resource elements allocated to the UEs as well as the throughput in Mbps are given in Figures (8.53(b)) and (8.53(a)), respectively. As we can observe from the figures, the horizontal axis is the UE indices which indicate that there are 25 UEs being served by the eNodeB 14. The vertical axis is the number of resource elements allocated by the optimization in Algorithms (21) and (22). Each UE has only 1 application running. This assumption simplifies the simulation because the goal of this chapter is observing the effect of the channel for which channel-aware EURA is performed. Furthermore, the legend show the SNR of the UEs in dB. The bit rate requirements of the applications is according to $\{0.25, 1, 5, 0.25, 1, 5, ...\}$ Mbps.

As we can observe from Figure (8.53(b)), the UEs with low SNRs are receiving more resource elements in order to meet their bit rate requirements. On the other hand, UEs with high SNR are receiving less resources. On the other hand, (8.53(a)) shows that UE throughputs are met. This is due all UEs are at good channel conditions such that the minimum SNR was 11.25 dB. We can see these in Figure (8.53(b)) that UEs with higher bit rate needs were allocated more resources, i.e. for UE k, $R_{1+3K} < R_{2+3K} < R_{3+3K}$. This statements says that the rate allocated to UE 1, 4, 7, ..., is less than the rate allocated to

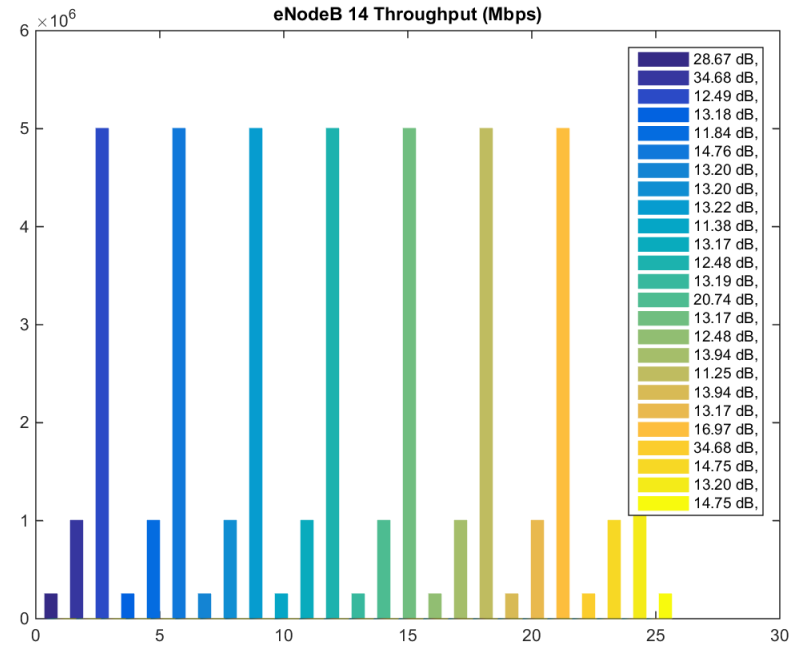


(a) eNB 13 Coverage

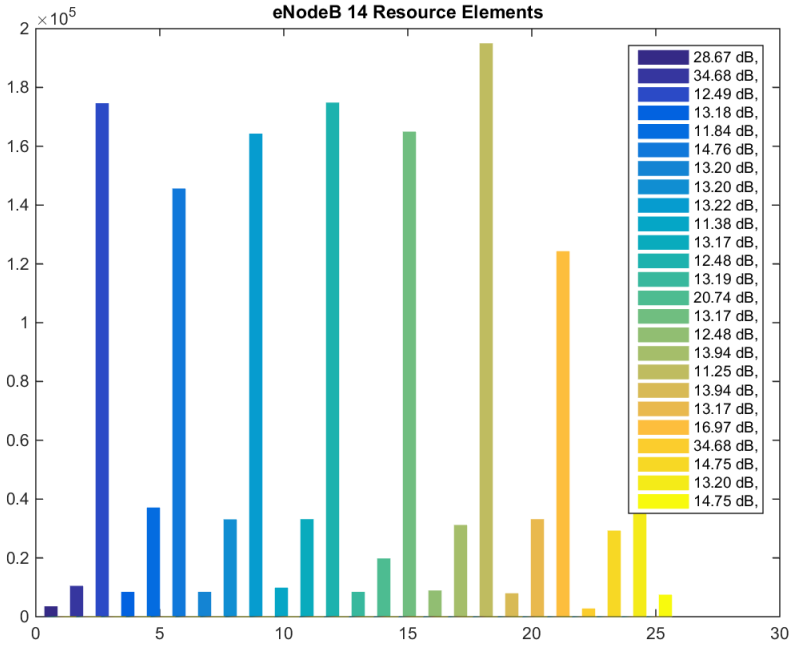


(b) eNB 13 REM

Figure 8.52: The system contains 21 UEs, each concurrently running a real-time application.



(a) Throughput of the UEs for eNB 14



(b) Resources Allocated to UEs by eNB 14

Figure 8.53: The system contains 25 UEs, each concurrently running a real-time application.

UEs 2, 5, 8, ..., and is less than the rates allocated to UEs 3, 6, 9, ... compared one to one (i.e. corresponding indices compared to each other). Also, we see that for same bit rate requirements, UEs with higher SNR are allocated less resources. For instance, UE 18 who has a high bit requirement of 5 Mbps is getting more REs than its counterparts since it is at a bad channel condition 11.25 dB.

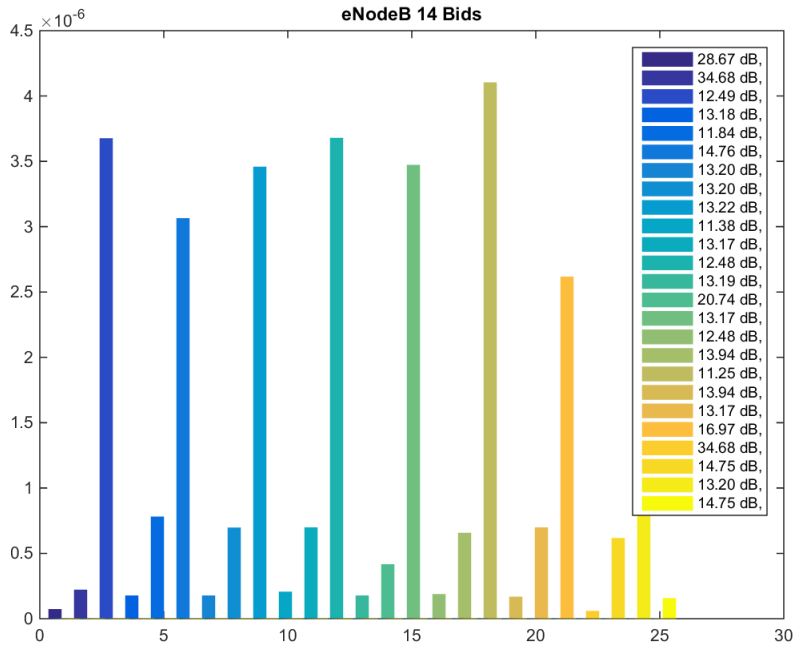


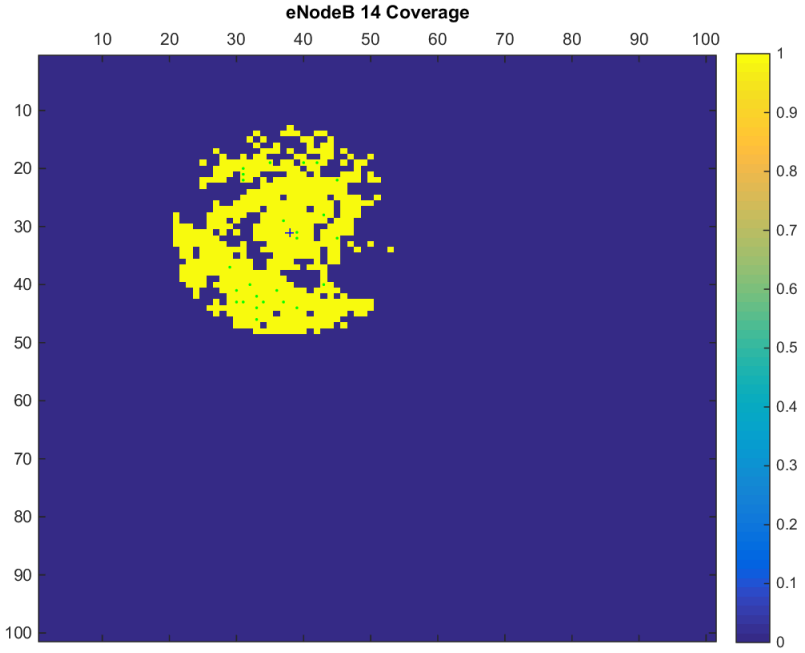
Figure 8.54: UE Bids pledged to eNB 14

Moreover, as we can see from Figure (8.54), UEs $3 + 3K$ are higher than UEs $2 + 3K$ which bid higher than UEs $1 + 3K$ due to the fact the they require more resources in view of their applications (5 Mbps vs. 1 Mbps vs 0.25 Mbps.) This plot shows the last iteration of the algorithm where the shadow prices are converged. Interestingly, UE 18 which needed high resources at good channel is bidding higher than its counterparts since it is at a bad channel condition. Furthermore, we can observe the coverage area of eNodeB 14 in Figure (8.55(a)), and as we see the coverage is not circular which is a result of REM and elevations in various directions leading to different SNRs. The black addition symbol in the middle

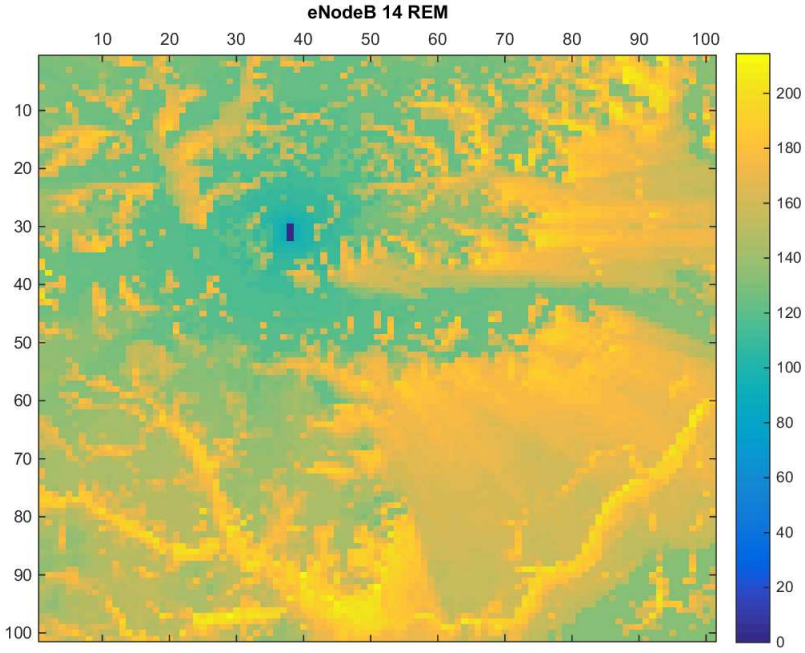
of the coverage, shown in yellow, is the eNodeB 14 coordinates in the 101×101 grid. The side bar, shows 0 - dark blue - and 1 - yellow - for the areas which are respectively not covered and are under coverage. The green dots in the region are the UEs scattered all over the area. As we can see, UEs are in the foot print of the eNodeBs, and had good channel conditions. These simply chose eNodeB 14 as it was the strongest signal they could ever receive. Furthermore, we can see the REM which shows that a blue dot at the eNodeB location, which is for the eNodeB 14 and the blue color from the side bar represents a 0 dB loss which is expected as we are at the eNodeB position. However, as we go further away from the eNodeB, the path loss increases, which is shown by the color spectrum light blue, green, and yellow in the order of increase. The highest pathloss in the area is about 209 dB.

For eNodeB 15, the plots of resource elements allocated to the UEs as well as the throughput in Mbps are given in Figures (8.56(b)) and (8.56(a)), respectively. As we can observe from the figures, the horizontal axis is the UE indices which indicate that there are 31 UEs being served by the eNodeB 15. The vertical axis is the number of resource elements allocated by the optimization in Algorithms (21) and (22). Each UE has only 1 application running. This assumption simplifies the simulation because the goal of this chapter is observing the effect of the channel for which channel-aware EURA is performed. Furthermore, the legend show the SNR of the UEs in dB. The bit rate requirements of the applications is according to $\{0.25, 1, 5, 0.25, 1, 5, ...\}$ Mbps.

As we can observe from Figure (8.56(b)), the UEs with low SNRs are receiving more resource elements in order to meet their bit rate requirements. On the other hand, UEs with high SNR are receiving less resources. On the other hand, (8.56(a)) shows that UE throughputs are met. This is due all UEs are at good channel conditions such that the minimum SNR was -2.14 dB. We can see these in Figure (8.56(b)) that UEs with higher bit rate needs were allocated more resources, i.e. for UE k, $R_{1+3K} < R_{2+3K} < R_{3+3K}$. This statements says that the rate allocated to UE 1, 4, 7, ..., is less than the rate allocated to

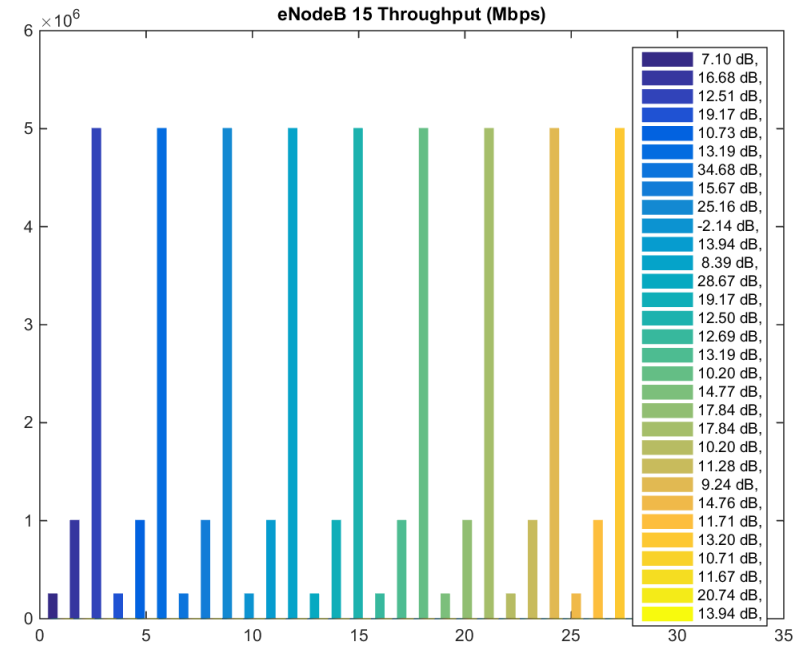


(a) eNB 14 Coverage

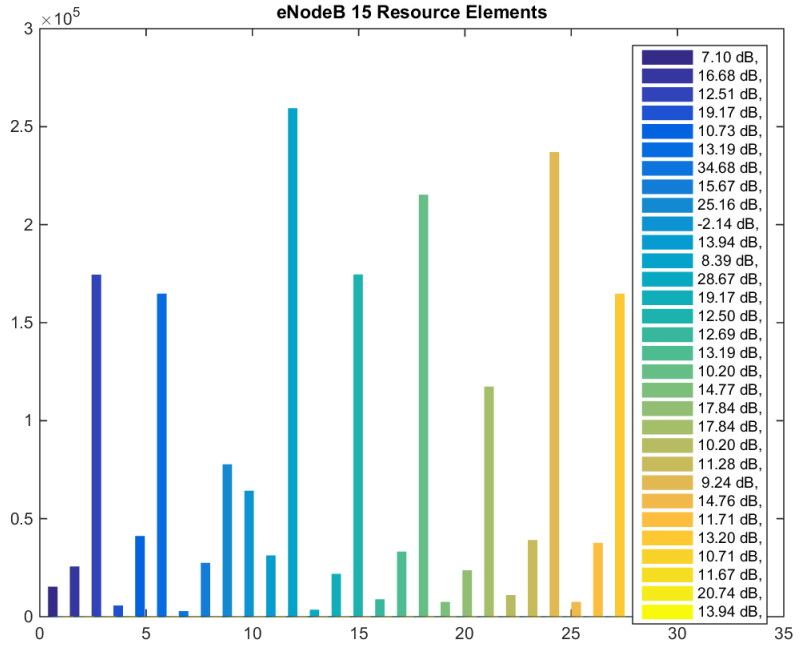


(b) eNB 14 REM

Figure 8.55: The system contains 25 UEs, each concurrently running a real-time application.



(a) Throughput of the UEs for eNB 15



(b) Resources Allocated to UEs by eNB 15

Figure 8.56: The system contains 31 UEs, each concurrently running a real-time application.

UEs 2, 5, 8, ..., and is less than the rates allocated to UEs 3, 6, 9, ... compared one to one (i.e. corresponding indices compared to each other). Also, we see that for same bit rate requirements, UEs with higher SNR are allocated less resources. For instance, UEs 12 and 24 are allocated more REs since they are at lower SNRs. In particular, the spike for UE 10 is interesting even though it has less bit needs (0.25 Mbps). The spike is because this UE is at the lowest SNR situation in the system -2.14 dB.

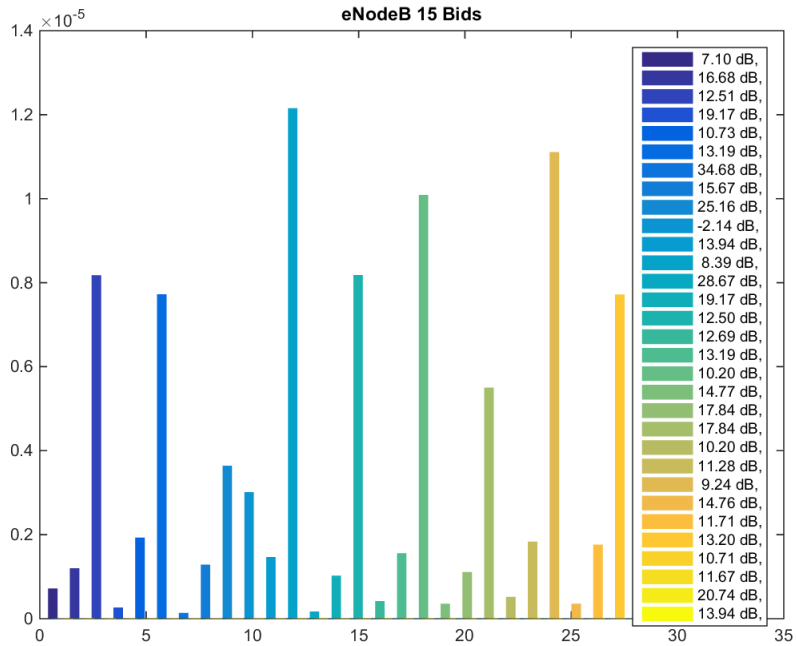


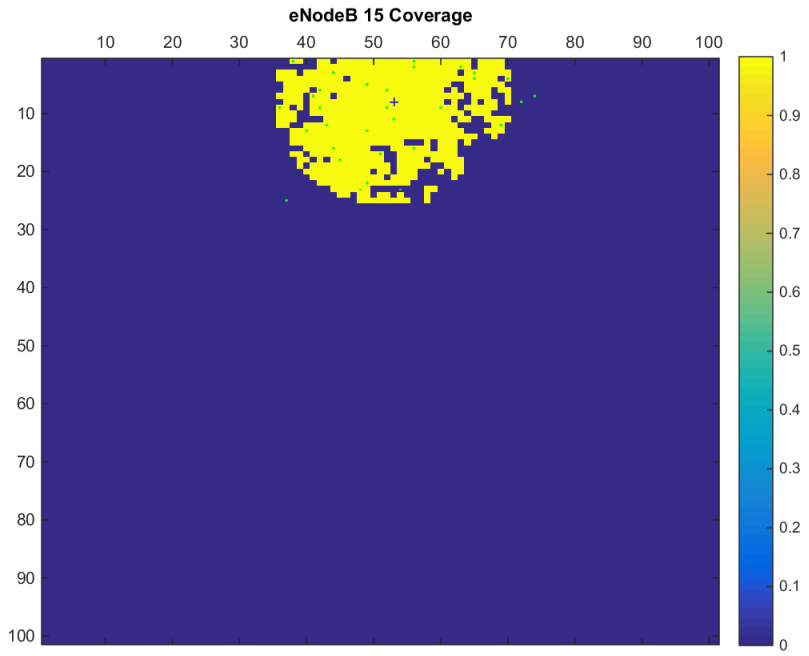
Figure 8.57: UE Bids pledged to eNB 15

Moreover, as we can see from Figure (8.57), UEs $3 + 3K$ are higher than UEs $2 + 3K$ which bid higher than UEs $1 + 3K$ due to the fact the they require more resources in view of their applications (5 Mbps vs. 1 Mbps vs 0.25 Mbps.) This plot shows the last iteration of the algorithm where the shadow prices are converged. Interestingly, UE 15 which needed high resources at good channel is bidding less than its counterparts. Furthermore, we can observe the coverage area of eNodeB 15 in Figure (8.58(a)), and as we see the coverage is not circular which is a result of REM and elevations in various directions leading to different

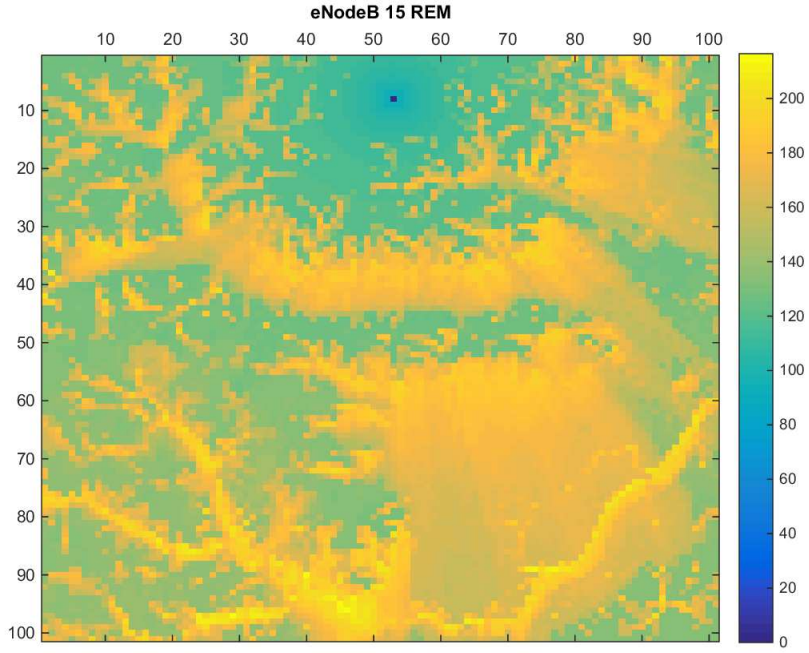
SNRs. The black addition symbol in the middle of the coverage, shown in yellow, is the eNodeB 15 coordinates in the 101×101 grid. The side bar, shows 0 - dark blue - and 1 - yellow - for the areas which are respectively not covered and are under coverage. The green dots in the region are the UEs scattered all over the area. As we can see, UEs are not just in the foot print of the eNodeBs, and these represent the lowest SNRs in the system. These simply chose eNodeB 15 as it was the strongest signal they could ever receive. Furthermore, we can see the REM which shows that a blue dot at the eNodeB location, which is for the eNodeB 15 and the blue color from the side bar represents a 0 dB loss which is expected as we are at the eNodeB position. However, as we go further away from the eNodeB, the path loss increases, which is shown by the color spectrum light blue, green, and yellow in the order of increase. The highest pathloss in the area is about 209 dB.

For eNodeB 16, the plots of resource elements allocated to the UEs as well as the throughput in Mbps are given in Figures (8.59(b)) and (8.59(a)), respectively. As we can observe from the figures, the horizontal axis is the UE indices which indicate that there are 14 UEs being served by the eNodeB 16. The vertical axis is the number of resource elements allocated by the optimization in Algorithms (21) and (22). Each UE has only 1 application running. This assumption simplifies the simulation because the goal of this chapter is observing the effect of the channel for which channel-aware EURA is performed. Furthermore, the legend show the SNR of the UEs in dB. The bit rate requirements of the applications is according to $\{0.25, 1, 5, 0.25, 1, 5, ...\}$ Mbps.

As we can observe from Figure (8.59(b)), the UEs with low SNRs are receiving more resource elements in order to meet their bit rate requirements. On the other hand, UEs with high SNR are receiving less resources. On the other hand, (8.59(a)) shows that UE throughputs are met. This is due all UEs are at good channel conditions such that the minimum SNR was -12.30 dB. We can see these in Figure (8.29(b)) that UEs with higher bit rate needs were allocated more resources, i.e. for UE k, $R_{1+3K} < R_{2+3K} < R_{3+3K}$. This

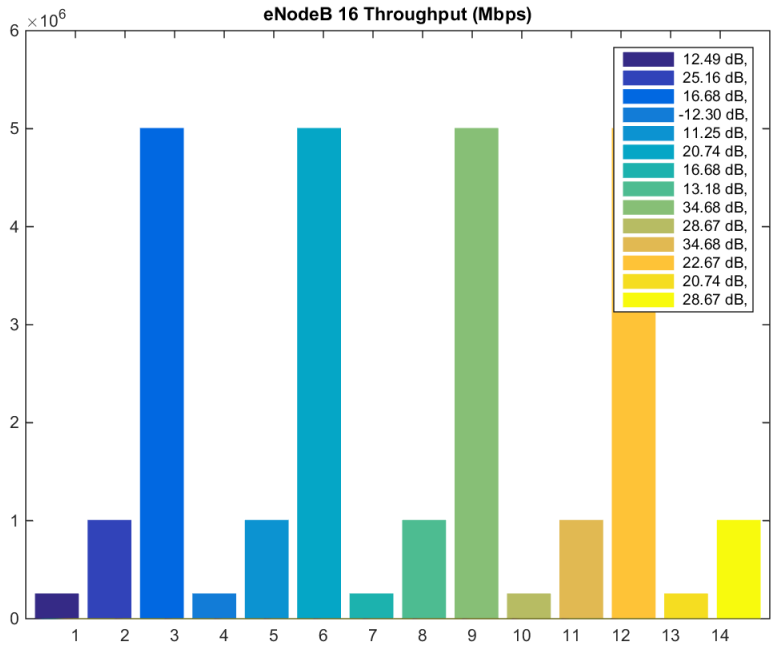


(a) eNB 15 Coverage

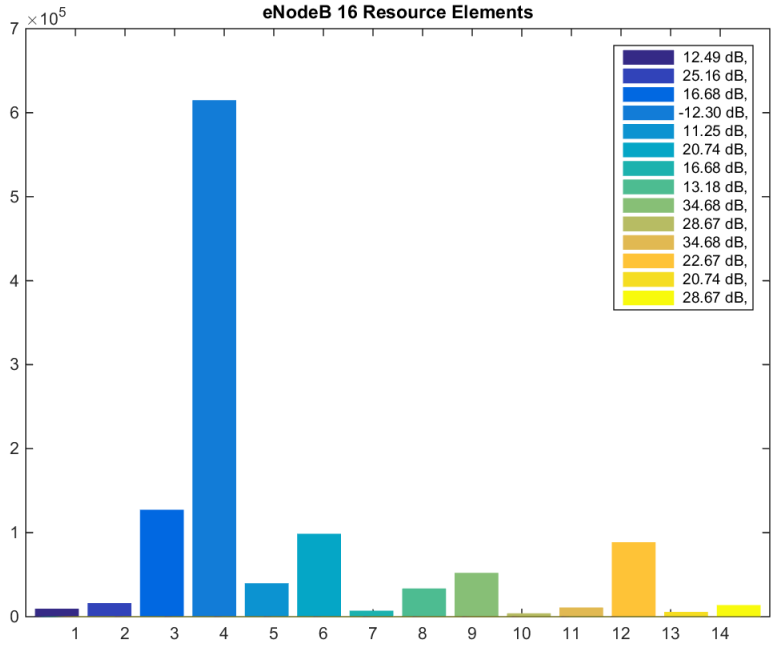


(b) eNB 15 REM

Figure 8.58: The system contains 31 UEs, each concurrently running a real-time application.



(a) Throughput of the UEs for eNB 16



(b) Resources Allocated to UEs by eNB 16

Figure 8.59: The system contains 14 UEs, each concurrently running a real-time application.

statements says that the rate allocated to UE 1, 4, 7, ..., is less than the rate allocated to UEs 2, 5, 8, ..., and is less than the rates allocated to UEs 3, 6, 9, ... compared one to one (i.e. corresponding indices compared to each other). Also, we see that for same bit rate requirements, UEs with higher SNR are allocated less resources. For instance, UE 4 is allocated more REs since it is at the lowest SNR -12.30 dB.

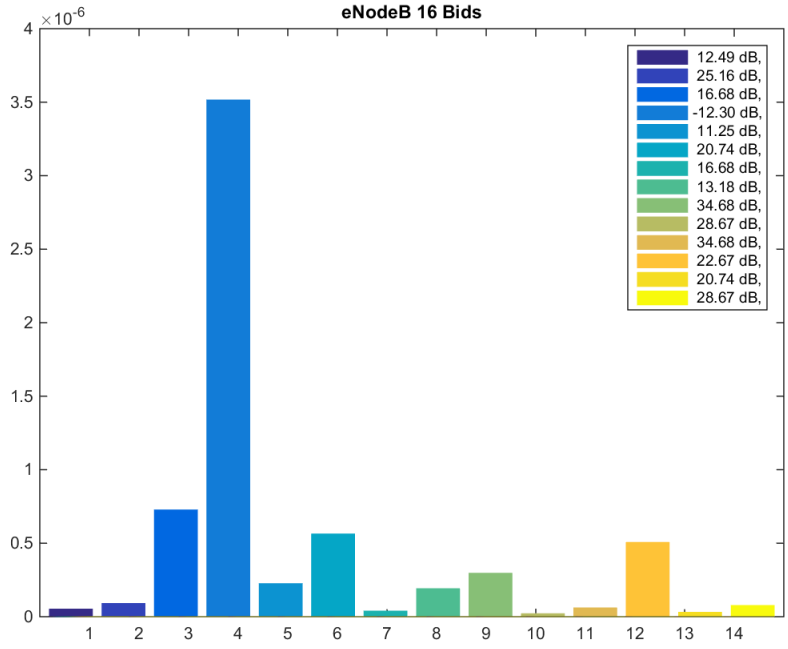


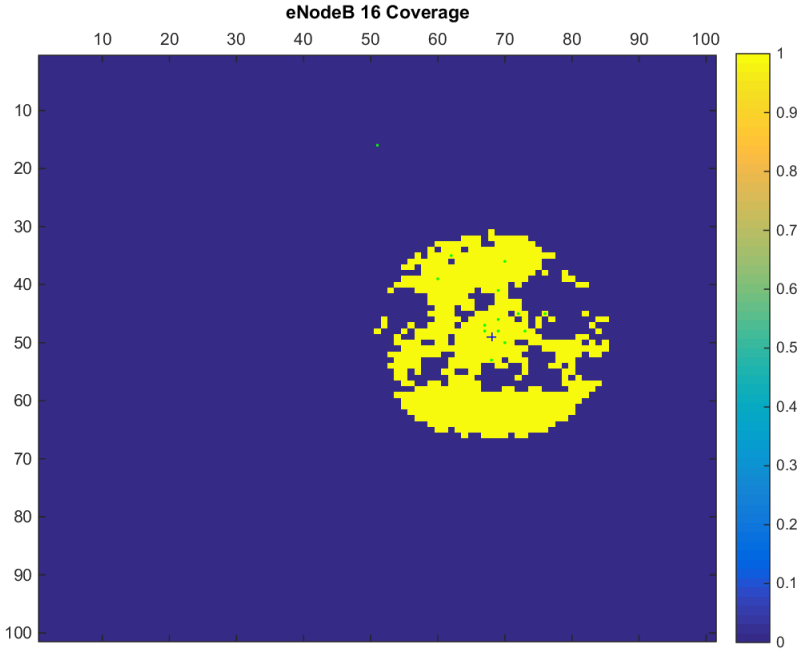
Figure 8.60: UE Bids pledged to eNB 16

Moreover, as we can see from Figure (8.60), UEs $3+3K$ are higher than UEs $2+3K$ which bid higher than UEs $1+3K$ due to the fact the they require more resources in view of their applications (5 Mbps vs. 1 Mbps vs 0.25 Mbps.) This plot shows the last iteration of the algorithm where the shadow prices are converged. Furthermore, we can observe the coverage area of eNodeB 16 in Figure (8.61(a)), and as we see the coverage is not circular which is a result of REM and elevations in various directions leading to different SNRs. The black addition symbol in the middle of the coverage, shown in yellow, is the eNodeB 16 coordinates in the 101×101 grid. The side bar, shows 0 - dark blue - and 1 - yellow - for the areas which

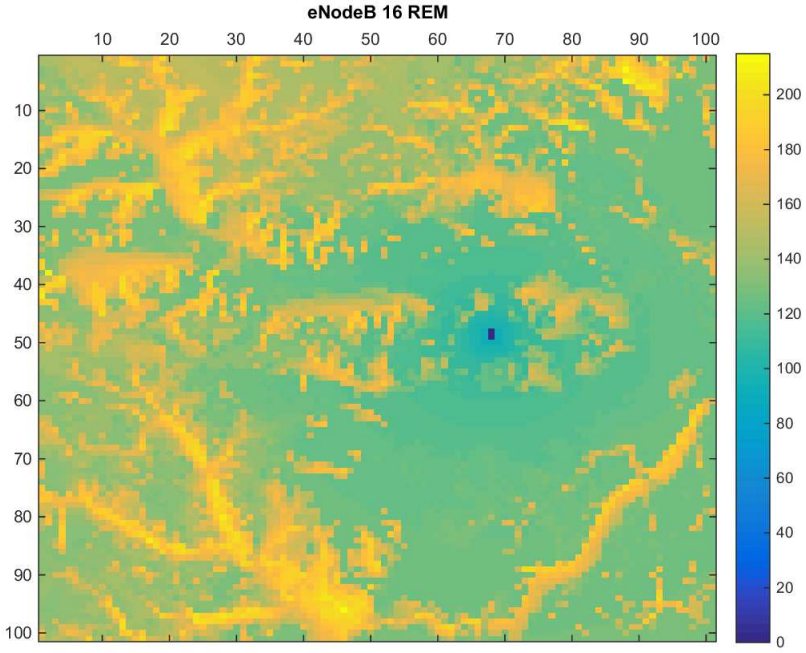
are respectively not covered and are under coverage. The green dots in the region are the UEs scattered all over the area. As we can see, UEs are in the foot print of the eNodeBs, and had good channel conditions. These simply chose eNodeB 16 as it was the strongest signal they could ever receive. There are SNRs outside the footprint of the eNodeB, which is reflected by the low SNR Furthermore, we can see the REM which shows that a blue dot at the eNodeB location, which is for the eNodeB 16 and the blue color from the side bar represents a 0 dB loss which is expected as we are at the eNodeB position. However, as we go further away from the eNodeB, the path loss increases, which is shown by the color spectrum light blue, green, and yellow in the order of increase. The highest pathloss in the area is about 205 dB.

For eNodeB 17, the plots of resource elements allocated to the UEs as well as the throughput in Mbps are given in Figures (8.62(b)) and (8.62(a)), respectively. As we can observe from the figures, the horizontal axis is the UE indices which indicate that there are 18 UEs being served by the eNodeB 18. The vertical axis is the number of resource elements allocated by the optimization in Algorithms (21) and (22). Each UE has only 1 application running. This assumption simplifies the simulation because the goal of this chapter is observing the effect of the channel for which channel-aware EURA is performed. Furthermore, the legend show the SNR of the UEs in dB. The bit rate requirements of the applications is according to $\{0.25, 1, 5, 0.25, 1, 5, ...\}$ Mbps.

As we can observe from Figure (8.62(b)), the UEs with low SNRs are receiving more resource elements in order to meet their bit rate requirements. On the other hand, UEs with high SNR are receiving less resources. On the other hand, (8.62(a)) shows that UE throughputs are met. This is due all UEs are at good channel conditions such that the minimum SNR was 8.68 dB. We can see these in Figure (8.62(b)) that UEs with higher bit rate needs were allocated more resources, i.e. for UE k, $R_{1+3K} < R_{2+3K} < R_{3+3K}$. This statements says that the rate allocated to UE 1, 4, 7, ..., is less than the rate allocated to

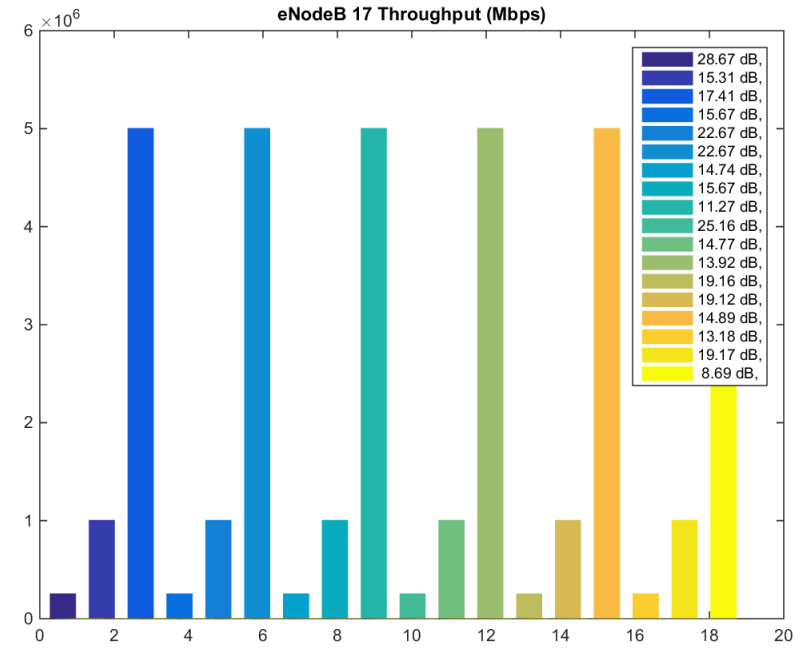


(a) eNB 16 Coverage

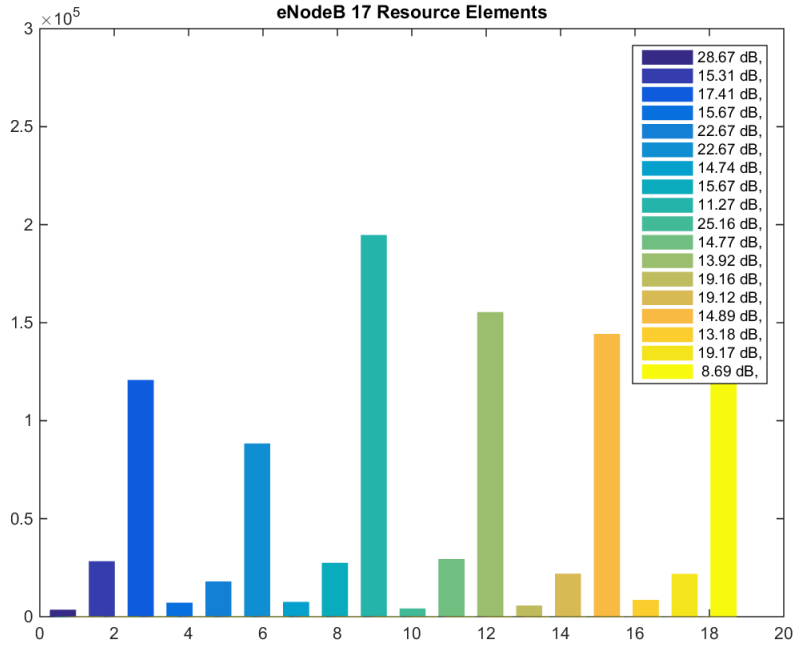


(b) eNB 16 REM

Figure 8.61: The system contains 14 UEs, each concurrently running a real-time application.



(a) Throughput of the UEs for eNB 17



(b) Resources Allocated to UEs by eNB 17

Figure 8.62: The system contains 18 UEs, each concurrently running a real-time application.

UEs 2, 5, 8, ..., and is less than the rates allocated to UEs 3, 6, 9, ... compared one to one (i.e. corresponding indices compared to each other). Also, we see that for same bit rate requirements, UEs with higher SNR are allocated less resources. For instance, UE 9 is allocated more REs since than UEs 3, 12, 15, and 18 as it is at lower SNRs.

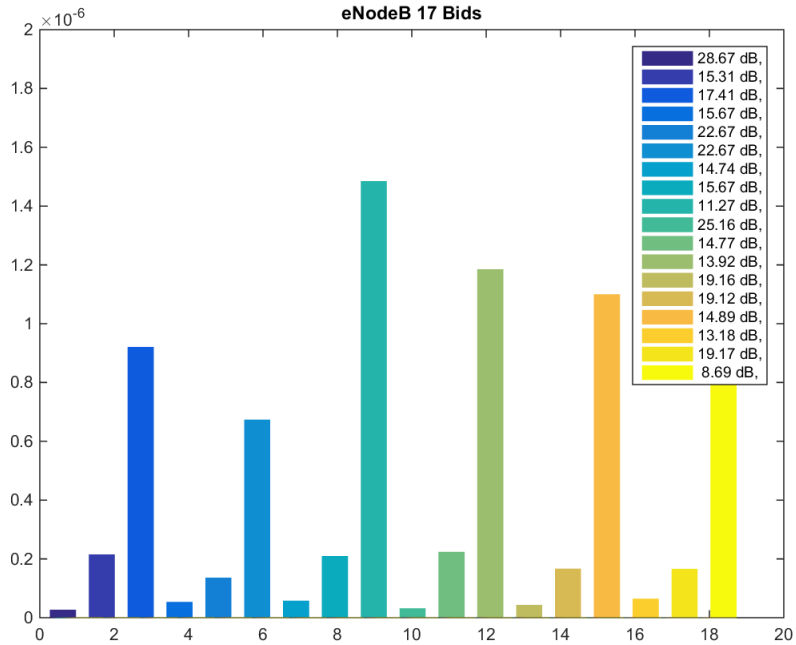


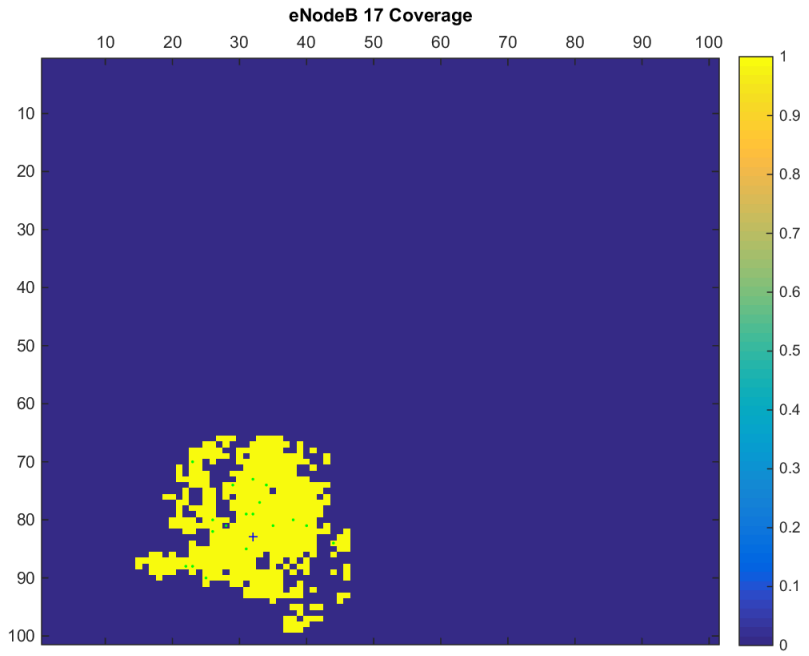
Figure 8.63: UE Bids pledged to eNB 17

Moreover, as we can see from Figure (8.63), UEs $3+3K$ are higher than UEs $2+3K$ which bid higher than UEs $1+3K$ due to the fact the they require more resources in view of their applications (5 Mbps vs. 1 Mbps vs 0.25 Mbps.) This plot shows the last iteration of the algorithm where the shadow prices are converged. Interestingly, UE 9 which needed high resources at bad channel conditions is bidding higher than its counterparts. Furthermore, we can observe the coverage area of eNodeB 17 in Figure (8.64(a)), and as we see the coverage is not circular which is a result of REM and elevations in various directions leading to different SNRs. The black addition symbol in the middle of the coverage, shown in yellow, is the eNodeB 17 coordinates in the 101×101 grid. The side bar, shows 0 - dark blue - and 1 -

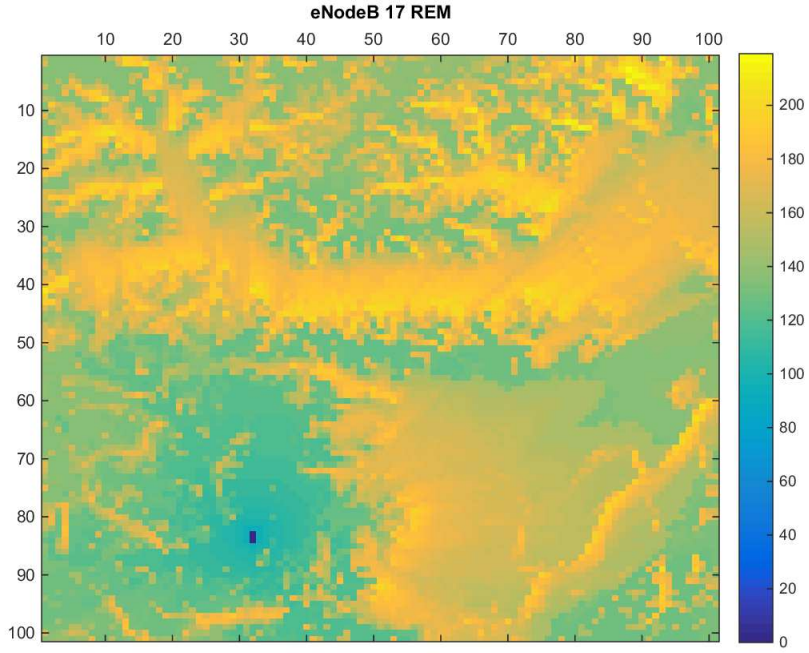
yellow - for the areas which are respectively not covered and are under coverage. The green dots in the region are the UEs scattered all over the area. As we can see, UEs are in the foot print of the eNodeBs, and had good channel conditions. These simply chose eNodeB 17 as it was the strongest signal they could ever receive. Furthermore, we can see the REM which shows that a blue dot at the eNodeB location, which is for the eNodeB 17 and the blue color from the side bar represents a 0 dB loss which is expected as we are at the eNodeB position. However, as we go further away from the eNodeB, the path loss increases, which is shown by the color spectrum light blue, green, and yellow in the order of increase. The highest pathloss in the area is about 210 dB.

For eNodeB 18, the plots of resource elements allocated to the UEs as well as the throughput in Mbps are given in Figures (8.65(b)) and (8.65(a)), respectively. As we can observe from the figures, the horizontal axis is the UE indices which indicate that there are 30 UEs being served by the eNodeB 4. The vertical axis is the number of resource elements allocated by the optimization in Algorithms (21) and (22). Each UE has only 1 application running. This assumption simplifies the simulation because the goal of this chapter is observing the effect of the channel for which channel-aware EURA is performed. Furthermore, the legend show the SNR of the UEs in dB. The bit rate requirements of the applications is according to $\{0.25, 1, 5, 0.25, 1, 5, ...\}$ Mbps.

As we can observe from Figure (8.65(b)), the UEs with low SNRs are receiving more resource elements in order to meet their bit rate requirements. On the other hand, UEs with high SNR are receiving less resources. On the other hand, (8.65(a)) shows that UE throughputs are met. This is due all UEs are at good channel conditions such that the minimum SNR was 5.64 dB. We can see these in Figure (8.65(b)) that UEs with higher bit rate needs were allocated more resources, i.e. for UE k, $R_{1+3K} < R_{2+3K} < R_{3+3K}$. This statements says that the rate allocated to UE 1, 4, 7, ..., is less than the rate allocated to UEs 2, 5, 8, ..., and is less than the rates allocated to UEs 3, 6, 9, ... compared one to one

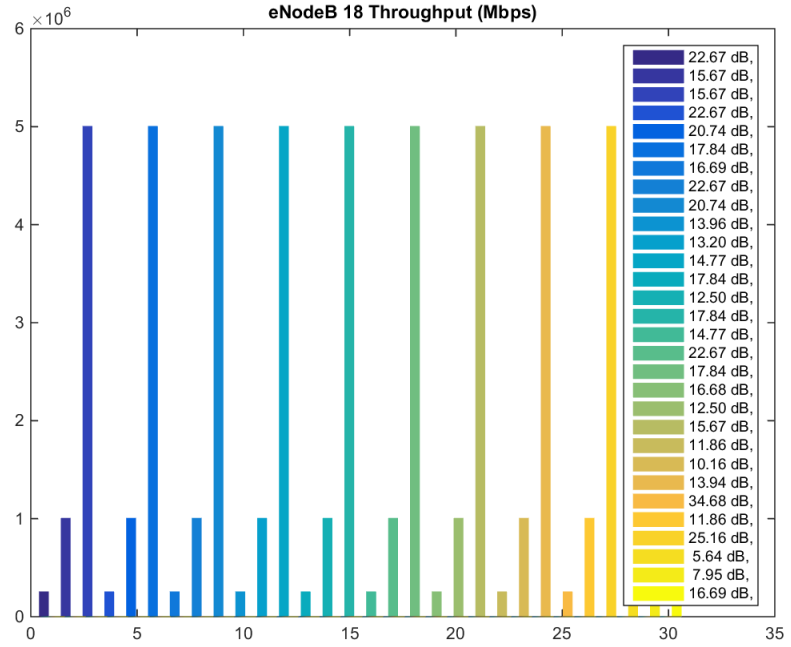


(a) eNB 17 Coverage

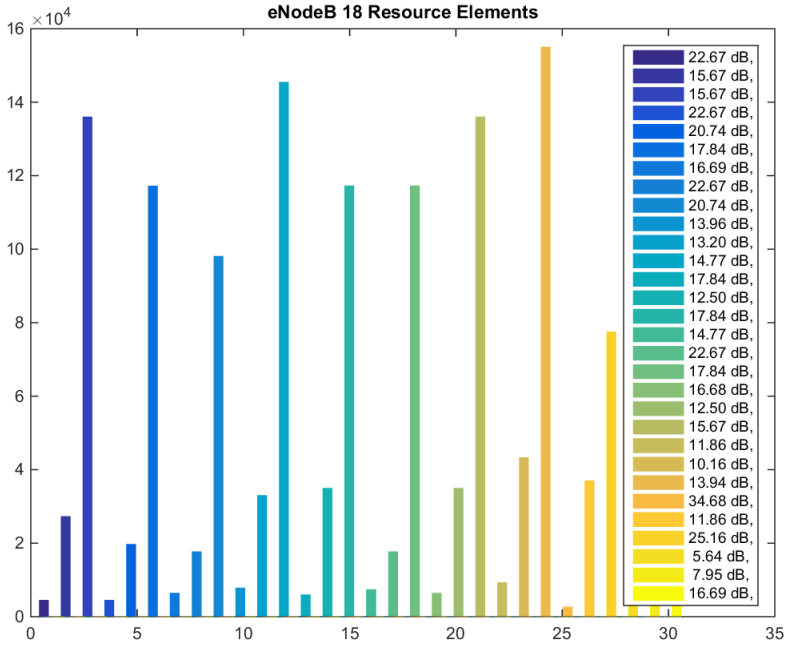


(b) eNB 17 REM

Figure 8.64: The system contains 18 UEs, each concurrently running a real-time application.



(a) Throughput of the UEs for eNB 18



(b) Resources Allocated to UEs by eNB 18

Figure 8.65: The system contains 30 UEs, each concurrently running a real-time application.

(i.e. corresponding indices compared to each other). Also, we see that for same bit rate requirements, UEs with higher SNR are allocated less resources. For instance, UEs 12, 21, and 24 and 18 are allocated more REs since they are at lower SNRs with respect to their counterparts.

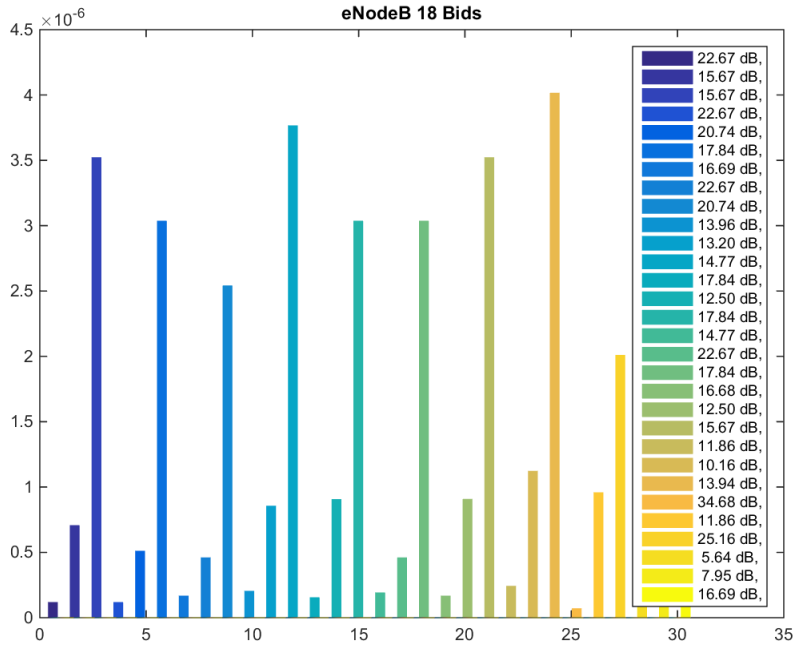


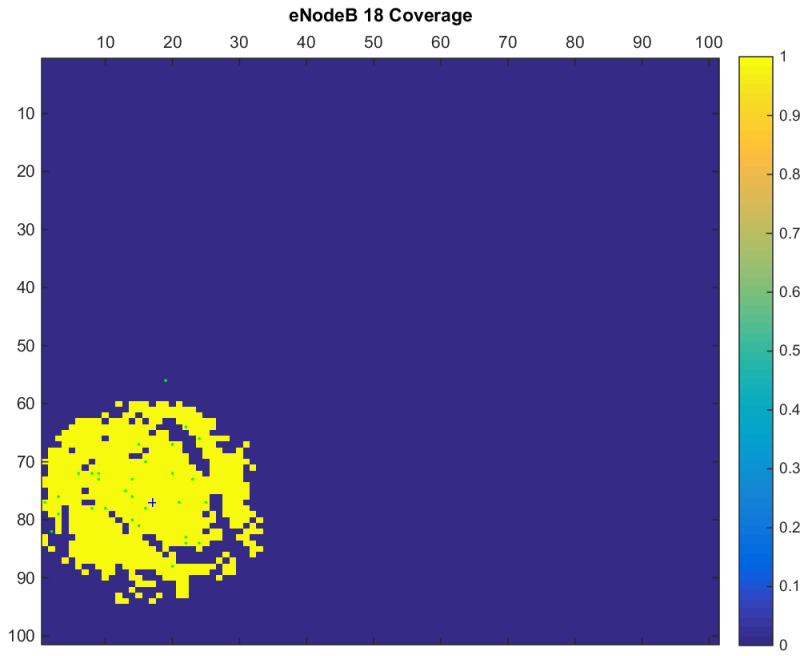
Figure 8.66: UE Bids pledged to eNB 18

Moreover, as we can see from Figure (8.66), UEs $3+3K$ are higher than UEs $2+3K$ which bid higher than UEs $1+3K$ due to the fact the they require more resources in view of their applications (5 Mbps vs. 1 Mbps vs 0.25 Mbps.) This plot shows the last iteration of the algorithm where the shadow prices are converged. Interestingly, UE 24 which needed high resources at bad channel conditions is bidding higher than its counterparts. Furthermore, we can observe the coverage area of eNodeB 18 in Figure (8.67(a)), and as we see the coverage is not circular which is a result of REM and elevations in various directions leading to different SNRs. The black addition symbol in the middle of the coverage, shown in yellow, is the eNodeB 18 coordinates in the 101×101 grid. The side bar, shows 0 - dark blue - and 1 -

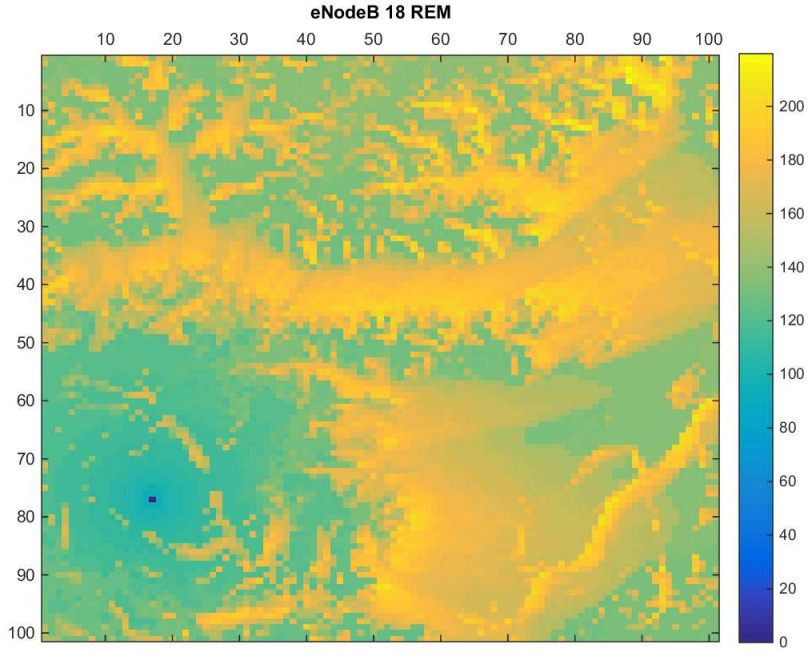
yellow - for the areas which are respectively not covered and are under coverage. The green dots in the region are the UEs scattered all over the area. As we can see, UEs are in the foot print of the eNodeBs, and had good channel conditions. These simply chose eNodeB 18 as it was the strongest signal they could ever receive. Furthermore, we can see the REM which shows that a blue dot at the eNodeB location, which is for the eNodeB 3 and the blue color from the side bar represents a 0 dB loss which is expected as we are at the eNodeB position. However, as we go further away from the eNodeB, the path loss increases, which is shown by the color spectrum light blue, green, and yellow in the order of increase. The highest pathloss in the area is about 217 dB.

For eNodeB 19, the plots of resource elements allocated to the UEs as well as the throughput in Mbps are given in Figures (8.68(b)) and (8.68(a)), respectively. As we can observe from the figures, the horizontal axis is the UE indices which indicate that there are 23 UEs being served by the eNodeB 19. The vertical axis is the number of resource elements allocated by the optimization in Algorithms (21) and (22). Each UE has only 1 application running. This assumption simplifies the simulation because the goal of this chapter is observing the effect of the channel for which channel-aware EURA is performed. Furthermore, the legend show the SNR of the UEs in dB. The bit rate requirements of the applications is according to $\{0.25, 1, 5, 0.25, 1, 5, ...\}$ Mbps.

As we can observe from Figure (8.68(b)), the UEs with low SNRs are receiving more resource elements in order to meet their bit rate requirements. On the other hand, UEs with high SNR are receiving less resources. On the other hand, (8.68(a)) shows that UE throughputs are met. This is due all UEs are at good channel conditions such that the minimum SNR was -18.95 dB. We can see these in Figure (8.68(b)) that UEs with higher bit rate needs were allocated more resources, i.e. for UE k, $R_{1+3K} < R_{2+3K} < R_{3+3K}$. This statements says that the rate allocated to UE 1, 4, 7, ..., is less than the rate allocated to UEs 2, 5, 8, ..., and is less than the rates allocated to UEs 3, 6, 9, ... compared one to one

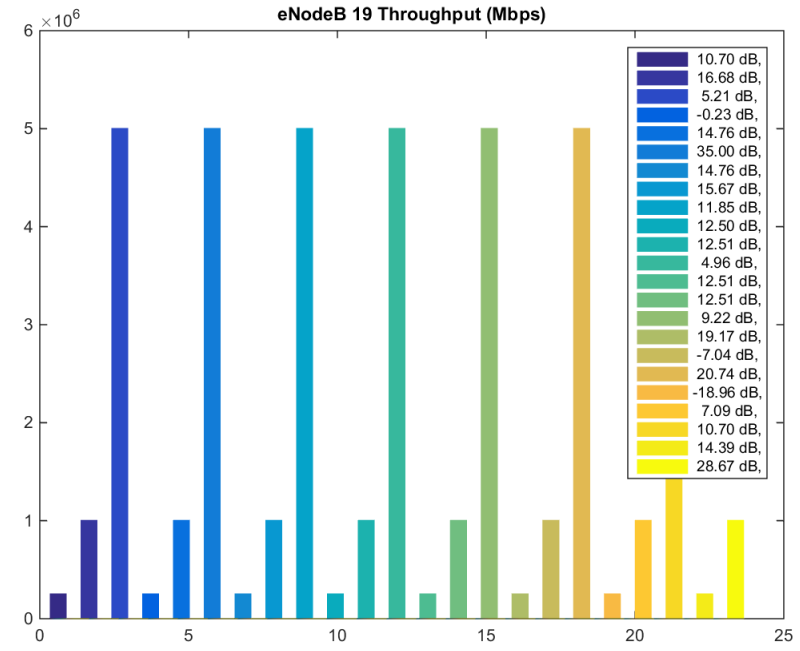


(a) eNB 18 Coverage

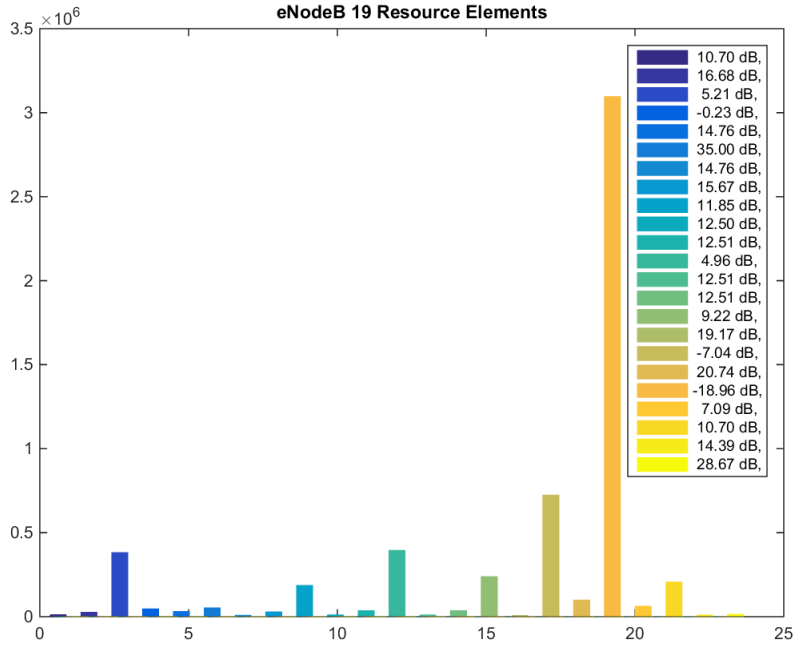


(b) eNB 18 REM

Figure 8.67: The system contains 30 UEs, each concurrently running a real-time application.



(a) Throughput of the UEs for eNB 19



(b) Resources Allocated to UEs by eNB 19

Figure 8.68: The system contains 23 UEs, each concurrently running a real-time application.

(i.e. corresponding indices compared to each other). Also, we see that for same bit rate requirements, UEs with higher SNR are allocated less resources.

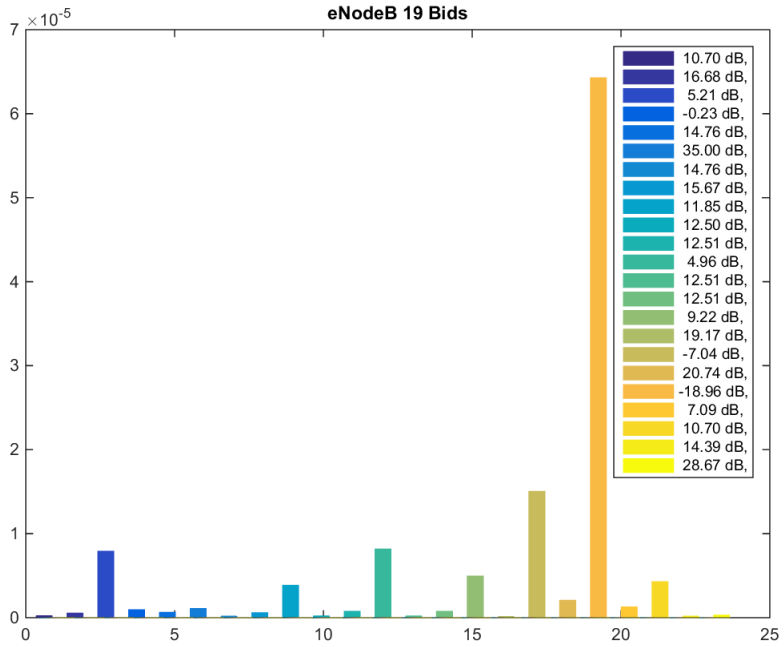
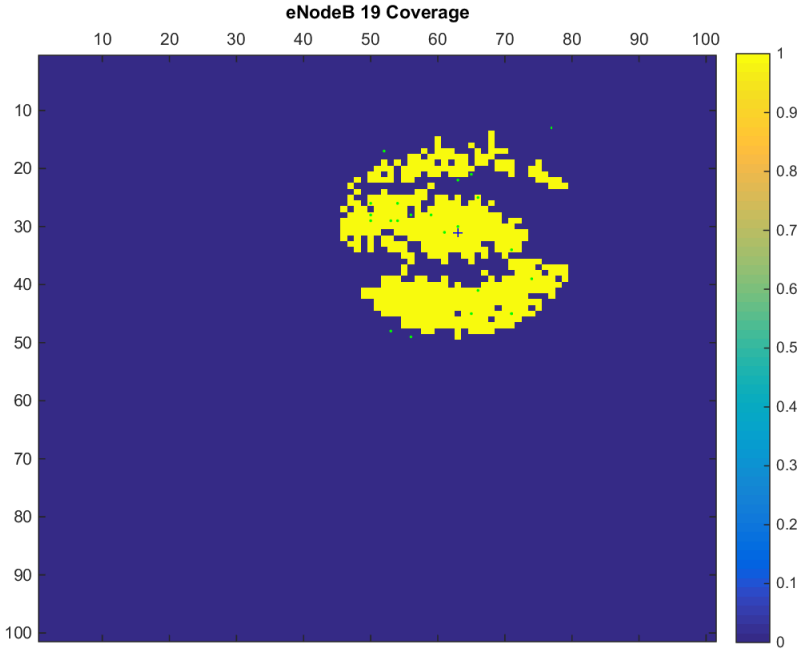
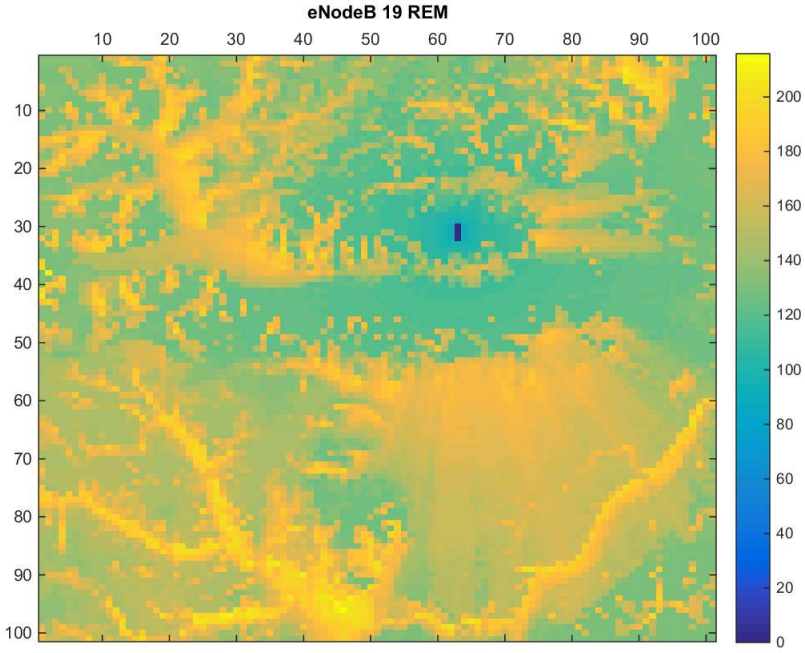


Figure 8.69: UE Bids pledged to eNB 19

Moreover, as we can see from Figure (8.69), UEs $3 + 3K$ are higher than UEs $2 + 3K$ which bid higher than UEs $1 + 3K$ due to the fact the they require more resources in view of their applications (5 Mbps vs. 1 Mbps vs 0.25 Mbps.) This plot shows the last iteration of the algorithm where the shadow prices are converged. Interestingly, UE 15 which needed high resources at good channel is bidding less than its counterparts. Furthermore, we can observe the coverage area of eNodeB 19 in Figure (8.70(a)), and as we see the coverage is not circular which is a result of REM and elevations in various directions leading to different SNRs. The black addition symbol in the middle of the coverage, shown in yellow, is the eNodeB 19 coordinates in the 101×101 grid. The side bar, shows 0 - dark blue - and 1 - yellow - for the areas which are respectively not covered and are under coverage. The green dots in the region are the UEs scattered all over the area.



(a) eNB 19 Coverage



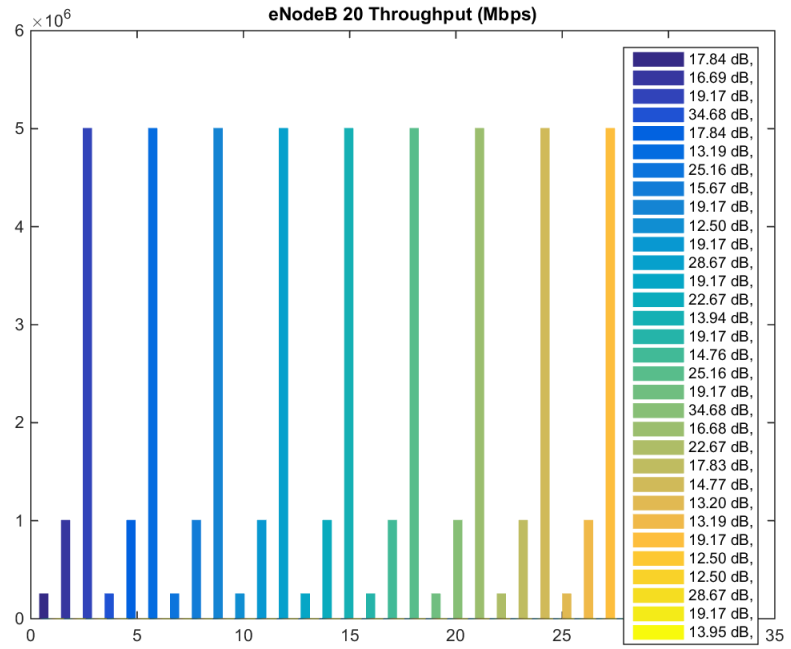
(b) eNB 19 REM

Figure 8.70: The system contains 23 UEs, each concurrently running a real-time application.

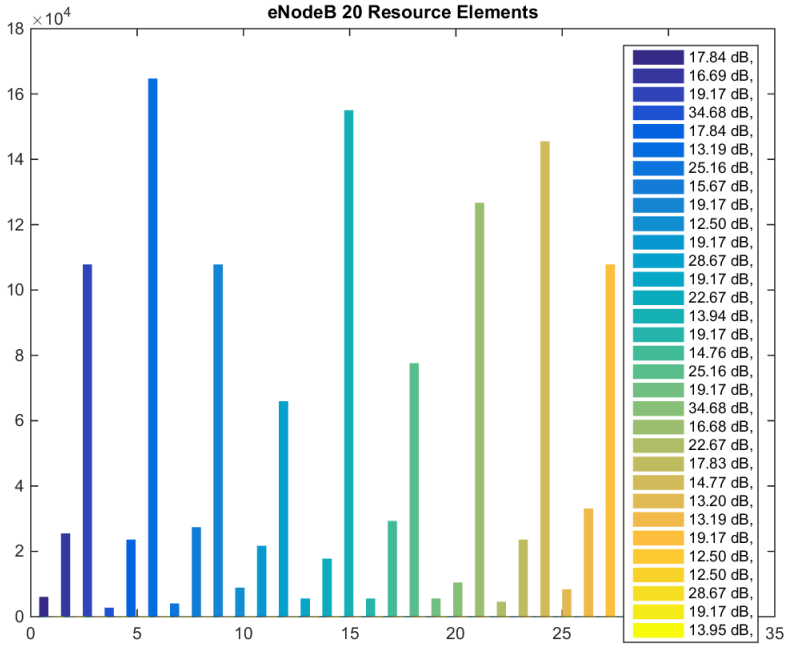
For eNodeB 20, the plots of resource elements allocated to the UEs as well as the throughput in Mbps are given in Figures (8.71(b)) and (8.71(a)), respectively. As we can observe from the figures, the horizontal axis is the UE indices which indicate that there are 32 UEs being served by the eNodeB 4. The vertical axis is the number of resource elements allocated by the optimization in Algorithms (21) and (22). Each UE has only 1 application running. This assumption simplifies the simulation because the goal of this chapter is observing the effect of the channel for which channel-aware EURA is performed. Furthermore, the legend show the SNR of the UEs in dB. The bit rate requirements of the applications is according to $\{0.25, 1, 5, 0.25, 1, 5, ...\}$ Mbps.

As we can observe from Figure (8.71(b)), the UEs with low SNRs are receiving more resource elements in order to meet their bit rate requirements. On the other hand, UEs with high SNR are receiving less resources. On the other hand, (8.71(a)) shows that UE throughputs are met. This is due all UEs are at good channel conditions such that the minimum SNR was 12.5 dB. We can see these in Figure (8.23(b)) that UEs with higher bit rate needs were allocated more resources, i.e. for UE k, $R_{1+3K} < R_{2+3K} < R_{3+3K}$. This statements says that the rate allocated to UE 1, 4, 7, ..., is less than the rate allocated to UEs 2, 5, 8, ..., and is less than the rates allocated to UEs 3, 6, 9, ... compared one to one (i.e. corresponding indices compared to each other). Also, we see that for same bit rate requirements, UEs with higher SNR are allocated less resources.

Moreover, as we can see from Figure (8.72), UEs $3 + 3K$ are higher than UEs $2 + 3K$ which bid higher than UEs $1 + 3K$ due to the fact the they require more resources in view of their applications (5 Mbps vs. 1 Mbps vs 0.25 Mbps.) This plot shows the last iteration of the algorithm where the shadow prices are converged. Furthermore, we can observe the coverage area of eNodeB 20 in Figure (8.73(a)), and as we see the coverage is not circular which is a result of REM and elevations in various directions leading to different SNRs. The black addition symbol in the middle of the coverage, shown in yellow, is the eNodeB 20



(a) Throughput of the UEs for eNB 20



(b) Resources Allocated to UEs by eNB 20

Figure 8.71: The system contains 32 UEs, each concurrently running a real-time application.

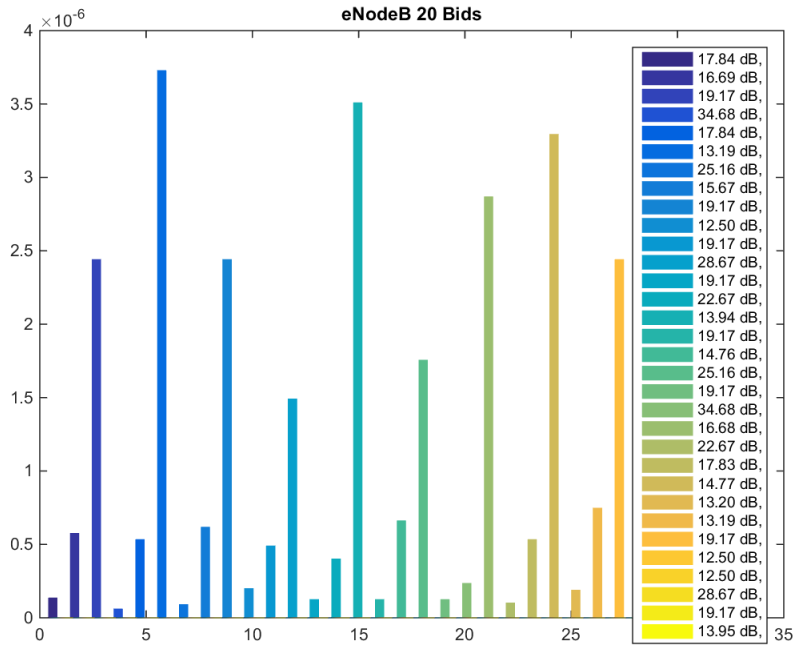
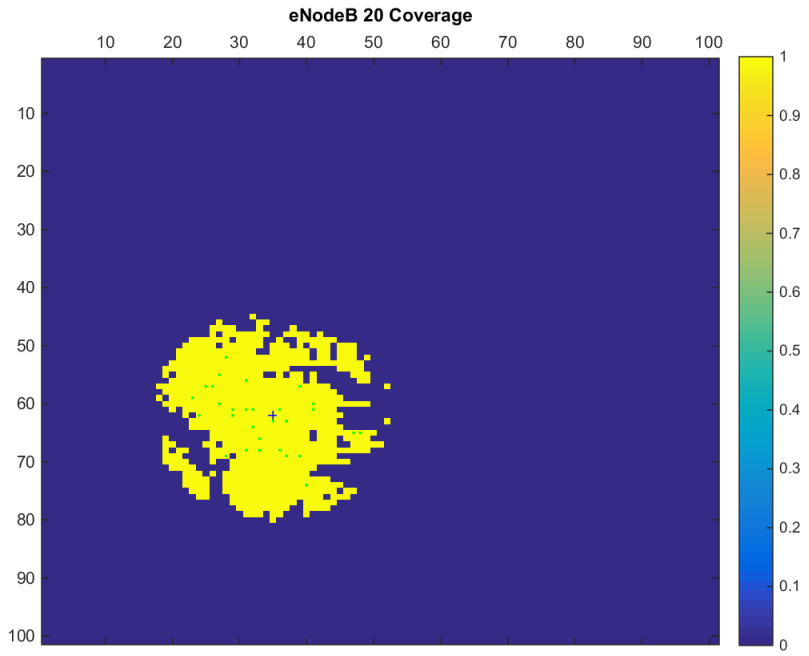


Figure 8.72: UE Bids pledged to eNB 20

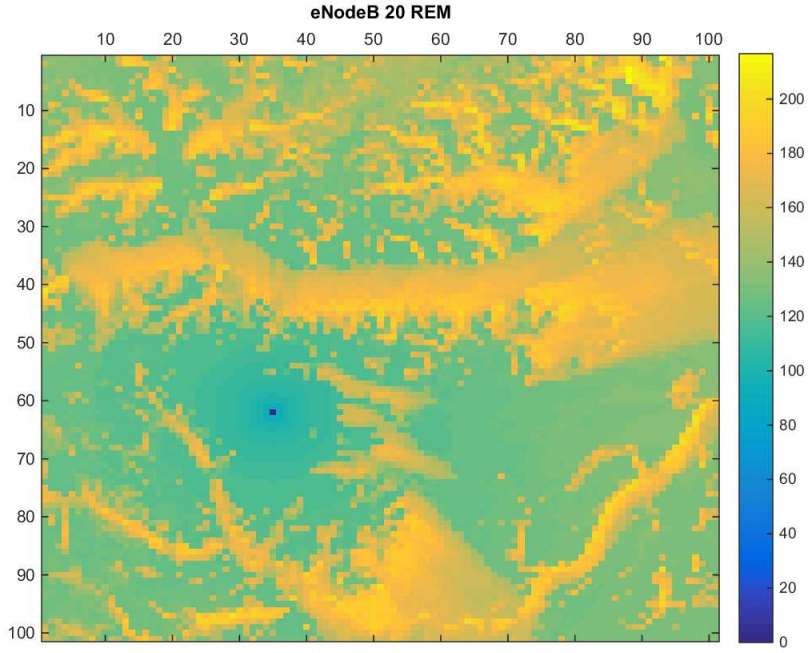
coordinates in the 101 × 101 grid. The side bar, shows 0 - dark blue - and 1 - yellow - for the areas which are respectively not covered and are under coverage. The green dots in the region are the UEs scattered all over the area. As we can see, UEs are in the foot print of the eNodeBs, and had good channel conditions. These simply chose eNodeB 20 as it was the strongest signal they could ever receive. Furthermore, we can see the REM which shows that a blue dot at the eNodeB location, which is for the eNodeB 20 and the blue color from the side bar represents a 0 dB loss which is expected as we are at the eNodeB position.

8.7 Chapter Summary

In this chapter, we developed channel-aware distributed architecture for the QoS-minded utility proportional fairness framework for resource allocation for the cells of a cellular communications system that was introduced in chapter 3. The distributed architecture was



(a) eNB 20Coverage



(b) eNB 20 REM

Figure 8.73: The system contains 32 UEs, each concurrently running a real-time application.

composed of a EURA optimization which allocated the UE rates by the eNB and an IURA optimization which assigned application rates by the UEs. Not only did we prove that the proposed distributed resource allocation architecture's EURA and IURA optimization problems are convex and solved them through the Lagrangian of their dual problem, but also we proved the optimality of the rate assignments. We showed that under abundance of resources, the resource allocation assigns more resources to the UEs with bad channel conditions, in order to meet their QoS requirements for their applications. This is in light of the fact that under a bad channel, lower modulation orders and coding schemes can be used which reduces the spectrum efficiency. On the other hand, when the resources are constrained, more resources are allocated to UEs with good channel conditions so as to meet their bit rate requirements. Ultimately, we performed simulations in MATLAB to show the application of the proposed distributed resource allocation architecture to a cellular communications system. We also performed a large scale simulation through a network planning where the channel conditions was dicatated based on the radio environment map and applied the algorithm to assign resources based on the channel conditions.

Chapter 9

Conclusions

In this chapter, we summarize the findings in this thesis work and suggest research trajectories for people interested in continuing the work presented in this thesis.

9.1 Findings

In this section, we summarize the findings in various chapters of this thesis. The results within this thesis are as below.

- In this thesis, we showed that we can leverage sigmoidal utility functions for realtime applications QoS satisfaction under a proportional fairness resource allocation framework; and proved that the optimizations are convex.
- We showed that as a result of the convexity of the optimizations under proportional fairness with sigmoidal utility the allocated rates are optimal.
- We leveraged the set of sigmoidal and logarithmic utilities to model the QoS satisfaction of the realtime and delay tolerant applications, and showed that the optimizations using

the hybrid traffic originated from sigmoidal and logarithmic utilities are convex and the solutions are optimal.

- By proving the convexity of the optimizations involving sigmoidal utilities, we proved that an NP-hard problem of realtime traffic resource allocation, which traditionally used to be solved via approximating the sigmoidal utilities to logarithmic ones can be solved with polynomial complexity.

- We incorporated application temporal usage into the resource allocation formulation and showed that higher usage of the application leads to a dynamic resource allocation according to the temporal focus of the user leading to various application usages.

- We incorporated UE priority weights into the resource allocation architecture and showed that higher weights lead to priority in the allocation of the weights so that those UEs with higher weights are assigned more resources initially.

- We provided a centralized architecture which assigned the rates to the applications in a singular stage.

- We proved the convexity of the centralized architecture when the traffic is a hybrid of realtime and delay tolerant traffic represented by sigmoidal and logarithmic utility functions.

- We derived the transmission overhead of the centralized resource allocation architecture.

- We derived the sensitivity of the centralized method to the changes in the number of UEs in the system and showed that the rates remain optimal when UEs rebid for resources in the face of changes in the number of UEs in the system.

- We analyzed the sensitivity of the centralized method to the changes in the number of UEs in the system and showed that the rates do not remain optimal when all UES

do not rebid for resources in the face of the dynamics introduced by changes in the number of UEs in the system.

- We analyzed the sensitivity of the centralized architecture to the changes in the usage percentage of the applications in the system and showed that when all UEs rebid for resources, the rates remain optimal in the face of the application usage changes.

- We analyzed the sensitivity of the centralized resource allocation to the changes in the usage percentage of the applications in the system and showed that when all UEs do not rebid for resources, the rates do not remain optimal in the face of changes in the system.

- We Introduced a distributed architecture for the resource allocation which allocated the rates to the UEs by the BSs and the rates to the applications by the UEs.

- We proved that the distributed optimizations for the network and the device are convex and lead to optimal solutions both for the UEs and the applications running inside UEs.

- We analyzed the sensitivity of the distributed approach to the changes in the number of UEs in the system and showed that when all UEs do not rebid for resources in the face of dynamics in the number of UEs, the rates do not remain optimal.

- We analyzed the sensitivity of the distributed approach to the changes in the number of UEs in the system and showed that when all UEs rebid for resources in the face of dynamics in the number of UEs, the rates remain optimal.

- We analyzed the sensitivity of the distributed approach to the changes in the application usage of the UEs in the system and showed that when all UEs do not rebid for resources in the face of dynamics in the applications usage, the rates do not remain optimal for the network optimization.

- We analyzed the sensitivity of the distributed approach to the changes in the application usage of the UEs in the system and showed that when all UEs rebid for resources in the face of dynamics in the applications usage, the rates remain optimal for the network optimization.

- We analyzed the sensitivity of the distributed approach to the changes in the application usage of the UEs in the system and showed that when all applications do not rebid for resources in the face of dynamics in the applications usage, the rates do not remain optimal for the network optimization.

- We analyzed the sensitivity of the distributed approach to the changes in the application usage of the UEs in the system and showed that when all applications rebid for resources in the face of dynamics in the applications uage, the rates remain optimal for the network optimization.

- We proved that the centralized and the distributed architectures are mathematically equivalent by proving that the UE and application rates assigned by each method is equal to those allocated by the other method.

- We analyzed the transmission overhead of the distributed resource allocation and derived lower bounds for the transmission overhead.

- We proved that the shadow price for the centralized approach converges to the optimal value for all available resources.

- We proved that the shadow price for the distributed approach converges to optimal value when there are more resources available at the BS than the addition of inflection points of all realtime applications.

- We proved that the shadow price for the distributed approach oscillates around the optimal value when there are not enough resources available at the BS as opposed

ot the addition of the inflection points of the realtime applications sigmoidal utility functions.

- We provided a mechanism to stabilize the distributed architecture for all BS available resources by means of decay functions.
- We made the resource allocation more realistic by looking at the LTE radio resource block structure and introduced a modified optimization which allocated resource blocks to the UEs.
- We showed that the resource block allocation optimization with hybrid traffic is convex.
- We presented a method to solve the resource block optimization and introduced a fast mechanism to map the optimal continuous rates to the closest optimal discrete rates.
- We proved that the proposed discrete optimization mapping mechanism reduces the search spaces complexity significantly and leads to a pool of selections for the resource blocks.
- We made the optimal resource allocation more realistic by accounting for wireless channel conditions and provided with a modified hybrid traffic channel-aware resource allocation mechanism.
- We proved that the preceding application-aware channel-aware resource allocation is convex and the allocated rate are optimal.
- We showed that the channel condition changes the slope of the sigmoidal utilities under the optimization and makes it higher or lower for respectively good or bad channel conditions.
- We shoed that the channel effect moves the inflection point of the sigmoidal utilities to lower numbers for good channel conditions and to higher for bad channel conditions to account for the modulation and coding scheme supportable by the methodology.

- We showed that under abundance of resources, the application-aware channel-aware resource allocation assigns more resources to the UEs which are at bad channel conditions in order to satisfy the QoS of their applications. This is done by accounting for the lower order modulation and coding schemes that can be used in bad channel conditions.

- We showed that under the scarcity of resources, more resources are allocated to the UEs at good channel conditions.

- We also looked at the shared-spectrum operation and evaluated the impact of radar systems into LTE system under shared-spectrum operation and showed that the old exclusion zones that were introduced by the NTIA were incorrect.

- Motivated by the preceding effort, we incorporated shared-spectrum operation into our resource allocation optimization through a carrier aggregation method.

- We provided algorithms to assign resources to various sections of the cellular system in a way to ensure fairness under shared-spectrum operation.

9.2 Future Trajectory

In this section, we suggest some trajectories for future work for researchers who are interested in pursuing the effort presented in the current thesis to a higher level.

- In this work, the UEs chose the serving BSs according to highest level of signal strength that they would observe in the downlink from the BSs. The difficulty with this method is that we may end up with a situation where a cell becomes overcrowded since its BS signal is strong at many UE positions, while another cell at a slightly lower SNR may be undercrowded and therefore, the resources will not be efficiently utilized by the

network. It would be a great research effort for the UEs to choose the serving BS by looking at the traffic situation at various cells as well in order to balance the load. For instance, UEs can choose a cell with a lower shadow price which represents a higher availability of resources.

- The implementation of application-aware channel-aware method is very cumbersome in terms of runtime. It would make a great project to look at a parallel implementation of the proposed method to increase the speed in order to pave the way for real-world deployment of the method.

- In this thesis, we assume a flat-fading and slow-fading channel for the application-aware channel-aware method. It would make a great research work to look at frequency-selective and fast-fading channels.

- It would be of immense value to map the proposed method to LTE structure by defining protocols for message exchanges needed for the resource allocation and finally moving towards defining standards for the method.

- It would make a great research study to look at the effect of the resource allocation and interference jointly. Resources which are allocated to UEs may lead to in-band and out-of-band interference and a study on intracell and intercell interference while allocating resources will be extremely valuable.

References

[1] I. Research, "Mobile voip subscribers will near 410 million by 2015; VoLTE still a long way off," Mobile Services and Subscribers: Voice, SMS/MMS, and Broadband, 2010.

[2] NTIA, "An assessment of the near-term viability of accommodating wireless broadband systems in the 1675-1710 MHz, 1755-1780 MHz, 3500-3650 MHz, 4200-4220 MHz and 4380-4400 MHz bands," U.S. Dept. of Commerce, 2010.

[3] "Ericsson mobility report," in *Ericsson Press Release*, 2013.

[4] A. AbdelHadi and C. Clancy, "Context-aware resource allocation in cellular networks," Dynamic Specttrum Access Networks (Dyspan), 2014.

[5] G. T. . V. (2010-03, "Further advancements for e-utra physical layer aspects," Release 9, 2010.

[6] S. Zarei, "Channel coding and link adaptation," Seminar LTE: Der Mobilfunk der Zukunft, 2009.

[7] "Frame structure - downlink," http://www.sharetechnote.com/, 2009.

[8] M. Richards, J. Scheer, and W. Holm, "Principles of modern radar," SciTech Publishing, 2010.

[9] G. T. . V9.0.0, "Further advancements for E-UTRA physical layer aspects," Measuring of Heterogeneous Wireless and Wired Networks, 2012.

[10] N. Solutions and Networks, "Enhance mobile networks to deliver 1000 times more capacity by 2020," White Paper, 2013.

[11] G. Intelligence, "Smartphone users spending more 'face time' on apps than voice calls or web browsing," Technical Report, 2011.

[12] N. S. Networks, "Understanding smartphone behavior in the network," White Paper, 2011.

[13] R. Braden, "Integrated services in the internet architecture: an overview," IETF RFC 1633, 1994.

[14] S. Blake, "An architecture for differentiated services," IETF RFC 2475, 1998.

[15] R. Braden, "Resource ReServation Protocol (RSVP) - version 1 functional specification," IETF RFC 2205, 1997.

[16] K. Nichols, "A two-bit differentiated services architecture for the internet," IETF RFC 2638, 1999.

[17] K. Nahrstedt, "The QoS broker," IEEE Multimedia, 1995.

[18] A. Ghosh and R. Ratasuk, "Essentials of LTE and LTE-A," The Cambridge Wireless Essentials Series, 2011.

[19] A. Larmo, M. Lindstrom, M. Meyer, G. Pelletier, J. Torsner, and H. Wiemann, "The LTE link-layer design," IEEE Communications Magazine, 2009.

[20] C. Ciochina and H. Sari, "A review of OFDMA and single-carrier FDMA," Wireless Conference (EW), 2010.

[21] G. Piro, L. Grieco, G. Boggia, and P. Camarda, "A two-level scheduling algorithm for qos support in the downlink of LTE cellular networks," Wireless Conference (EW), 2010.

[22] G. Monghal, K. Pedersen, I. Kovacs, and P. Mogensen, "Qos oriented time and frequency domain packet schedulers for the UTRAN Long Term Evolution," IEEE Vehicular Technology Conference (VTC), 2008.

[23] D. Soldani, H. X. Jun, and B. Luck, "Strategies for mobile broadband growth: Traffic segmentation for better customer experience," IEEE Vehicular Technology Conference (VTC), 2011.

[24] S. Ali and M. Zeeshan, "A delay-scheduler coupled game theoretic resource allocation scheme for LTE networks," Frontiers of Information Technology (FIT), 2011.

[25] D. Fudenberg and J. Tirole, "Nash equilibrium: multiple nash equilibria, focal points, and pareto optimality," MIT Press, 1991.

[26] P. Ranjan, K. Sokol, and H. Pan, "Settling for less - a QoS compromise mechanism for opportunistic mobile networks," SIGMETRICS Performance Evaluation, 2011.

[27] R. Johari and J. Tsitsiklis, "Parameterized supply function bidding: Equilibrium and efficiency," Operations Research, 2011.

[28] H. Y. and S. Alamouti, "OFDMA: A Broadband Wireless Access Technology," IEEE Sarnoff Symposium, 2006.

[29] L. Chung, "Energy efficiency of QoS routing in multi-hop wireless networks," IEEE International Conference on Electro/Information Technology (EIT), 2010.

[30] M. Alasti, B. Neekzad, H. Jie, and R. Vannithamby, "Quality of service in WiMAX and LTE networks [Topics in Wireless Communications]," IEEE Communications Magazine, 2010.

[31] J. Andrews, A. Ghosh, and R. Muhamed, "Fundamentals of WiMAX: Understanding broadband wireless netwroking," Prentice Hall Communications Engineering and Emerging Technologies Series, 2007.

[32] H. Ekstrom, "QoS control in the 3GPP evolved packet sysem," IEEE Communications Magazine, 2009.

[33] B. Mota, "Quality of service in wireless backhaul applications with vortiqa software for service provider equipment," FreeScale, 2010.

[34] IXIACOM, "Quality of Service (QoS) and Policy Management in Mobile Data Networks," White Paper, 2010.

[35] F. Li, "Quality of service, traffic conditioning, and resource management in universal mobile teleccomunication system (UMTS)," Doctoral Dissertation, Norwegian University of Science and Technology, 2003.

[36] G. Gorbil and I. Korpeoglu, "Supporting QoS traffic at the network layer in multi-hop wireless mobile networks," Wireless Communications and Mobile Computing Conference (IWCMC), 2011.

[37] J. Tournier, J. Babau, and V. Olive, "Qinna, a component-based qos architecture," in *Proceedings of the 8th International Conference on Component-Based Software Engineering*, 2005.

[38] N. Ahmed and H. Yan, "Access control for mpeg video applications using neural network and simulated annealing," in *Mathematical Problems in Engineering*, 2004.

[39] I. Jung, I. J., Y. Y., H. E., and H. Yeom, "Enhancing qos and energy efficiency of realtime network application on smartphone using cloud computing," in *IEEE Asia-Pacific Services Computing Conference (APSCC)*, 2011.

[40] Tellabs, "Quality of Service in the Wireless Backhaul," White Paper, 2012.

[41] W. Stallings, "Data and computer communications," in *William Stallings Books on Computer and Data Communications*, 2013.

[42] C. Dovrolis, D. Stiliadis, and P. Ramanathan, "Proportional differentiated services: delay differentiation and packet scheduling," IEEE/ACM Transactions on Networking, 2002.

[43] A. Sali, A. Widiawan, S. Thilakawardana, R. Tafazolli, and B. Evans, "Cross-layer design approach for multicast scheduling over satellite networks," in *Wireless Communication Systems, 2005. 2nd International Symposium on*, 2005.

[44] E. Lutz, D. Cygan, M. Dippold, F. Dolainsky, and W. Papke, "The land mobile satellite communication channel-recording, statistics, and channel model," IEEE Transactions on Vehicular Technology, 1991.

[45] H. Perros and K. Elsayed, "Call admission control schemes: A review," 1994.

[46] P. C. o. A. o. S. Executive Office of the President and T. (PCAST), "Realizing the full potential of government-held spectrum to spur economic growth," 2012.

[47] F. C. Commission, "Proposal to create a citizens broadband service in the 3550-3650 MHz band," FCC Docket No. 12-354, 2012.

[48] S. Wilson and T. Fischetto, "Coastline population trends in the united states: 1960 to 2008," in *U.S. Dept. of Commerce*, 2010.

[49] Z. Kbah and A. Abdelhadi, "Resource allocation in cellular systems for applications with random parameters," *CoRR*, 2015.

[50] A. Abdel-Hadi, J. Michel, A. Gerstlauer, and S. Vishwanath, "Real-time optimization of video transmission in a network of aavs," in *Vehicular Technology Conference (VTC Fall), 2011 IEEE*, pp. 1–5, Sept 2011.

[51] F. Wilson, I. Wakeman, and W. Smith, "Quality of service parameters for commercial application of video telephony," Human Facotrs Telecommunications Symposium, 1993.

[52] A. Aggarwal and T. Meng, "Minimizing the peak-to-average power ratio of ofdm signals using convex optimization," in *IEEE Transactions on Signal Processing*, 2006.

[53] H. Al-Shatri and T. Weber, "Achieving the maximum sum rate using D.C. programming in cellular networks," in *IEEE Transactions on Signal Processing*, 2012.

[54] A. Abdelhadi, A. Gerstlauer, and S. Vishwanath, "Real-time rate-distortion optimized streaming of wireless video," *CoRR*, 2014.

[55] S. Shenker, "Fundamental design issues for the future internet," IEEE Journal on Selected Areas in Communications, 1995.

[56] V. Annapureddy and V. Veeravalli, "Sum capacity of MIMO interference channels in the low interference regime," in *IEEE Transactions on Information Theory*, 2011.

[57] A. Abdelhadi and T. Clancy, "An optimal resource allocation with frequency reuse in cellular networks," 2015.

[58] A. Abdelhadi, H. Shajaiah, and C. Clancy, "A multi-tier wireless spectrum sharing system leveraging secure spectrum auctions," *CoRR*, vol. abs/1503.04899, 2015.

[59] A. Abdel-Hadi and S. Vishwanath, "On multicast interference alignment in multihop systems," in *IEEE Information Theory Workshop 2010 (ITW 2010)*, 2010.

[60] J. Jose, A. Abdel-Hadi, P. Gupta, and S. Vishwanath, "On the impact of mobility on multicast capacity of wireless networks," in *INFOCOM, 2010 Proceedings IEEE*, pp. 1–5, March 2010.

[61] J. Lee, R. Mazumdar, and N. Shroff, "Downlink power allocation for multi-class wireless systems," IEEE/ACM Transactions in Networking, 2005.

[62] Y. Wang and A. Abdelhadi, "A qos-based power allocation for cellular users with different modulations," *CoRR*, vol. abs/1507.07141, 2015.

[63] G. Tychogiorgos, A. Gkelias, and K. Leung, "Utility proportional fairness in wireless networks," IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), 2012.

[64] Y. Wang and A. Abdelhadi, "Optimal power allocation for lte users with different modulations," *CoRR*, vol. abs/1507.07159, 2015.

[65] A. Abdelhadi, A. Khawar, and T. C. Clancy, "Optimal downlink power allocation in cellular networks," *CoRR*, vol. abs/1405.6440, 2014.

[66] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, P. Whiting, and R. Vijayakumar, "Providing quality of service over a shared wireless link," Communications Magazine, IEEE, 2001.

[67] J. A. Mahal, A. Khawar, A. Abdelhadi, and T. C. Clancy, "Radar precoder design for spectral coexistence with coordinated multi-point (comp) system," *CoRR*, vol. abs/1503.04256, 2015.

[68] A. Khawar, A. Abdelhadi, and T. C. Clancy, "QPSK waveform for MIMO radar with spectrum sharing constraints," *CoRR*, vol. abs/1407.8510, 2014.

[69] M. Li, Z. Chen, and Y. Tan, "A maxmin resource allocation approach for scalable video delivery over multiuser mimo-ofdm systems," in *IEEE International Symposium on Circuits and Systems (ISCAS)*, 2011.

[70] A. Khawar, A. Abdelhadi, and T. C. Clancy, "Channel modeling between seaborne MIMO radar and MIMO cellular system," *CoRR*, vol. abs/1504.04325, 2015.

[71] A. Khawar, A. Abdelhadi, and T. C. Clancy, "3d channel modeling between seaborne MIMO radar and MIMO cellular system," *CoRR*, vol. abs/1504.04333, 2015.

[72] A. Khawar, A. Abdelhadi, and T. C. Clancy, "Coexistence Analysis between Radar and Cellular System in LoS Channel," *CoRR*, vol. abs/1506.07468, 2015.

[73] C. Shahriar, A. Abdelhadi, and T. C. Clancy, "Overlapped-mimo radar waveform design for coexistence with communication systems," *CoRR*, vol. abs/1502.04117, 2015.

[74] A. Khawar, A. Abdel-Hadi, T. Clancy, and R. McGwier, "Beampattern analysis for mimo radar and telecommunication system coexistence," in *Computing, Networking and Communications (ICNC), 2014 International Conference on*, pp. 534–539, Feb 2014.

[75] R. Kurrle, "Resource allocation for smart phones in 4G-LTE advanced carrier aggregation," Master Thesis, Virginia Tech, 2012.

[76] A. Khawar, A. Abdelhadi, and T. Clancy, "A mathematical analysis of cellular interference on the performance of s-band military radar systems," in *Wireless Telecommunications Symposium (WTS), 2014*, pp. 1–8, April 2014.

[77] M. Ariaudo, I. Fijalkow, J. Gautier, B. A. M. Brandon, and B. Milevsky, "Green radio despite dirty RF front-end," in *IEEE Transactions on Information Theory*, 2012.

[78] A. G. Bagherikaram, S. Motahari, and A. K. Khandani, "The secrecy capacity region of the Gaussian MIMO broadcast channel," in *IEEE Transactions on Information Theory*, 2009.

[79] R. Prabhu and B. Daneshrad, "An energy-efficient water-filling algorithm for ofdm systems," in *IEEE International Conference on Communications (ICC)*, 2010.

[80] V. Balakrishnan, S. Boyd, and S. Balemi, "Robust downlink beamforming based on outage probability specifications," in *International Journal on Robust and Nonlinear Control*, 1991.

[81] A. Khawar, A. Abdelhadi, and C. Clancy, "Target detection performance of spectrum sharing mimo radars," *Sensors Journal, IEEE*, vol. 15, pp. 4928–4940, Sept 2015.

[82] H. Kushner and P. Whiting, "Convergence of proportional-fair sharing algorithms under general conditions," IEEE Transactions on Wireless Communications, 2004.

[83] H. aShajaiah, A. Khawar, A. Abdel-Hadi, and T. Clancy, "Resource allocation with carrier aggregation in lte advanced cellular system sharing spectrum with s-band radar," in *Dynamic Spectrum Access Networks (DYSPAN), 2014 IEEE International Symposium on*, pp. 34–37, April 2014.

[84] F. Kelly, A. Maulloo, and D. Tan, "Rate control in communication networks: shadow prices, proportional fairness and stability," in *Journal of the Operational Research Society*, 1998.

[85] A. Parekh and R. Gallager, "A generalized processor sharing approach to flow control in integrated services networks: the single-node case," IEEE/ACM Transactions on Networking, 1993.

[86] A. Demers, S. Keshav, and S. Shenker, "Analysis and simulation of a fair queueing algorithm," SIGCOMM Comput. Commun. Rev., 1989.

[87] A. Khawar, A. Abdelhadi, and T. C. Clancy, "On The Impact of Time-Varying Interference-Channel on the Spatial Approach of Spectrum Sharing between S-band Radar and Communication System," in *Military Communications Conference (MILCOM)*, 2014.

[88] S. Boyd and L. Vandenberghe, *Introduction to convex optimization with engineering applications*. Cambridge University Press, 2004.

[89] S. Low and D. Lapsley, "Optimization flow control, i: Basic algorithm and convergence," IEEE/ACM Transactions on Networking, 1999.

[90] P. de Kerret and D. Gesbert, "The multiplexing gain of the network MIMO channel with distributed CSI," in *arXiv:1108.3742*, to apprear.

[91] M. Ghorbanzadeh, A. Abdelhadi, and C. Clancy, "A utility proportional fairness resource allocation in spectrally radar-coexistent cellular networks," in *Military Communications Conference (MILCOM)*, 2014.

[92] J. Lee, R. Mazumdar, and N. Shroff, "Non-convex optimization and rate control for multi-class services in the internet," IEEE/ACM Transactions on Networking, 2005.

[93] H. aShajaiah, A. Abdelhadi, and C. Clancy, "Spectrum sharing approach between radar and communication systems and its impact on radar's detectable target parameters," in *Vehicular Technology Conference (VTC Spring), 2015 IEEE 81st*, pp. 1–6, May 2015.

[94] A. AbdelHadi and C. Clancy, "A utility proportional fairness approach for resource allocation in 4G-LTE," in *IEEE International Conference on Computing, Networking, and Communications (ICNC), CNC Workshop*, 2014.

[95] A. Abdelhadi and C. Clancy, "A robust optimal rate allocation algorithm and pricing policy for hybrid traffic in 4G-LTE," in *IEEE nternational Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 2013.

[96] A. Abdelhadi, C. Clancy, and J. Mitola, "A resource allocation algorithm for users with multiple applications in 4G-LTE," in *ACM Workshop on Cognitive Radio Architectures for Broadband (MobiCom Workshop CRAB)*, 2013.

[97] H. Shajaiah, A. Abdelhadi, and C. Clancy, "Spectrum sharing between public safety and commercial users in 4G-LTE," IEEE International Conference on Computing, Networking and Communications (ICNC), 2014.

[98] H. Shajaiah, A. Abdelhadi, and C. Clancy, "Utility proportional fairness resource allocation with carrier aggregation in 4G-LTE," IEEE Military Communications Conference (MILCOM), 2013.

[99] M. Han, T. Yu, J. Kim, K. Kwak, S. Lee, S. Han, and D. Hong, "OFDM channel estimation with jammed pilot detector under narrow-band jamming," IEEE Transactions on Vehicular Technology, 2008.

[100] M. Ghorbanzadeh, A. Abdelhadi, and T. Clancy, "A utility proportional fairness approach for resource block allocation in cellular networks," in *IEEE IInternational Conference on Computing, Networking and Communications (ICNC)*, 2015.

[101] H. Shajaiah, A. Abdelhadi, and C. Clancy, "Multi-application resource allocation with users discrimination in cellular networks," IEEE nternational Symposium on Personal, Indoor and Mobile Radio Communications, 2014.

[102] F. Guo and W. Chenggui, "A jamming scheme based on pilot assisted channel estimation of OFDM," Journal of Electronic Warfare Technology, 2008.

[103] A. Khawar, A. Abdel-Hadi, and T. C. Clancy, "MIMO radar waveform design for coexistence with cellular systems," in *IEEE International Symposium on Dynamic Spectrum Access Networks: SSPARC Workshop (IEEE DySPAN 2014 - SSPARC Workshop)*, 2014.

[104] G. Tychogiorgos, A. Gkelias, and K. Leung, "A new distributed optimization framework for hybrid adhoc networks," in *GLOBECOM Workshops*, 2011.

[105] T. Harks, "Utility proportional fair bandwidth allocation: An optimization oriented approach," in *QoS-IP*, 2005.

[106] A. Abdelhadi and C. Clancy, "Context-aware resource allocation in cellular networks," in *arXiv*, 2014.

[107] R. Etkin, A. Parekh, and D. Tse, "Spectrum sharing for unlicensed bands," in *IEEE Journal on Selected Areas in Communications*, 2010.

[108] R. Etkin, D. Tse, and H. Wang, "Gaussian interference channel capacity to within one bit," in *IEEE Transactions on Information Theory*, 2008.

[109] T. Nandagopal, T. Kim, X. Gao, and V. Bharghavan, "Achieving mac layer fairness in wireless packet networks," in *Proceedings of the 6th annual International Conference on Mobile Computing and Networking (Mobicom)*, 2000.

[110] P. Dighe, R. Mallik, and S. Jamuar, "Analysis of transmit-receive diversity in rayleigh fading," in *IEEE Transactions on Communications*, 2012.

[111] P. Dighe, R. Mallik, and S. Jamuar, "On downlink beamforming with greedy user selection: Performance analysis and a simple new algorithm," in *IEEE Transactions on Signal Processing*, 2005.

[112] M. Ghorbanzadeh, A. Abdelhadi, and T. Clancy, "A utility proportional fairness bandwidth allocation in radar-coexistent cellular networks," in *Military Communications Conference (MILCOM)*, 2014.

[113] M. Ghorbanzadeh, A. AbdelHadi, and C. Clancy, "A utility proportional fairness approach for resource block allocation in cellular networks," in *IEEE IInternational Conference on Computing, Networking and Communications (ICNC)*, 2015.

[114] T. Erpek, A. Abdelhadi, and C. Clancy, "An optimal application-aware resource block scheduling in LTE," in *IEEE International Conference on Computing, Networking and Communications (ICNC) Worshop CCS)*, 2015.

[115] A. Abdelhadi and C. Clancy, "An optimal resource allocation with joint carrier aggregation in 4g-lte," in *Computing, Networking and Communications (ICNC), 2015 International Conference on*, pp. 138–142, Feb 2015.

[116] M. Ghorbanzadeh, A. AbdelHadi, and C. Clancy, "A utility proportional fairness bandwidth allocation in radr-coexistent cellular networks," in *Military Communications Conference (MILCOM)*, 2014.

[117] A. Khawar, A. Abdel-Hadi, and T. C. Clancy, "Spectrum sharing between S-band radar and LTE cellular system: A spatial approach," in *2014 IEEE International Symposium on Dynamic Spectrum Access Networks: SSPARC Workshop (IEEE DySPAN 2014 - SSPARC Workshop)*, 2014.

[118] M. Ghorbanzadeh, A. Abdelhadi, A. Amanna, J. Dwyer, and T. Clancy, "Implementing an optimal rate allocation tuned to the user quality of experience," in *Computing, Networking and Communications (ICNC), 2015 International Conference on*, pp. 292–297, Feb 2015.

[119] T. Jiang, L. Song, and Y. Zhang, "Orthogonal frequency division multiple access fundamentals and applications," Auerbach Publications, 2010.

[120] M. Dohler, R. Heath, A. Lozano, C. Papadias, , and R. Valenzuela, "Is the PHY layer dead?," in *IEEE Communications Magazine*, 2011.

[121] Y. Eldar and N. Merhav, "A competitive minimax approach to robust estimation of random parameters," in *IEEE Transactions on Signal Processing*, 2004.

[122] K. Eriksson, S. Shi, N. Vucic, M. Schubert, and E. Larsson, "Globally optimal resource allocation for achieving maximum weighted sum rate," in *IEEE Global Communications Conference*, 2010.

[123] M. Ghorbanzadeh, Y. Chen, and C. Clancy, "Fine-grained end-to-end network model via vector quantization and hidden markov processes," in *IEEE Conference on Communications (ICC)*, 2013.

[124] P. de Kerret and D. Gesbert, "Towards optimal csi allocation in multicell MIMO channels," in *in Proceedings of IEEE International Conference on Communications*, 2012.

[125] H. Shajaiah, A. Abdelhadi, and T. C. Clancy, "A price selective centralized algorithm for resource allocation with carrier aggregation in LTE cellular networks," *CoRR*, vol. abs/1408.4151, 2014.

[126] A. Abdelhadi and C. Clancy, "Context-aware resource allocation in cellular networks," in *arXiv*, 2014.

[127] A. Abdelhadi and T. C. Clancy, "Network MIMO with partial cooperation between radar and cellular systems," *CoRR*, vol. abs/1507.07267, 2015.

[128] H. Shajaiah, A. Abdelhadi, and T. C. Clancy, "Robust resource allocation with joint carrier aggregation for multi-carrier cellular networks," *CoRR*, vol. abs/1503.08994, 2015.

[129] H. Shajaiah, A. Abdelhadi, and T. C. Clancy, "An efficient multi-carrier resource allocation with user discrimination framework for 5g wireless systems," *CoRR*, vol. abs/1506.02448, 2015.

[130] H. Shajaiah, A. Abdelhadi, and T. Clancy, "Utility proportional fairness resource allocation with carrier aggregation in 4g-lte," in *IEEE Military Communications Conference (MILCOM)*, 2013.

[131] H. Shajaiah, A. Abdelhadi, and T. Clancy, "Multi-application resource allocation with users discrimination in cellular networks," in *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2014.

[132] H. Shajaiah, A. Abdelhadi, and T. Clancy, "Spectrum sharing between public safety and commercial users in 4g-lte," in *IEEE International Conference on Computing, Networking and Communications (ICNC)*, 2014.

[133] A. Lackpour, M. Luddy, and J. Winters, "Overview of interference mitigation techniques between wimax networks and ground based radar," Proc. 20th Annual Wireless and Optical Comms. Conf., 2011.

[134] F. Sanders, J. Carrol, G. Sanders, and R. Sole, "Effects of radar interference on LTE base station receiver performance," NTIA, U.S. Dept. of Commerce, 2013.

[135] E. Bjornson, "Multiantenna cellular communications: Channel estimation, feedback, and resource allocation," in *PhD thesis,KTH Royal Institute of Technology*, 2011.

[136] E. Bjornson, M. Bengtsson, and B. Ottersten, "Receive combining vs. multistream multiplexing in multiuser mimo systems," in *in Proceedings of IEEE Swedish-Communication Technologies Workshop*, 2001.

[137] E. Bjornson, M. Bengtsson, and B. Ottersten, "areto characterization of the multicell mimo performance region with simple receivers," in *IEEE Transactions on Signal Processing*, 2012.

[138] E. Bjornson, M. Bengtsson, and B. Ottersten, "Computational framework for optimal robust beamforming in coordinated multicell systems," in *Proceedings of IEEE Computational Advance in Multi-Sensor Adaptive Processing*, 2011.

[139] E. Bjornson, M. Bengtsson, and B. Ottersten, "Optimality properties, distributed strategies, and measurement-based evaluation of coordinated multicell ofdma transmission," in *IEEE Transactions on Signal Processing*, 2011.

[140] E. Bjornson and B. Ottersten, "A framework for training-based estimation in arbitrarily correlated Rician MIMO channels with Rician disturbance," in *IEEE Transactions on Signal Processing*, 2010.

[141] E. Bjornson, R. Zakhour, D. Gesbert, and B. Ottersten, "Cooperative multicell precoding: Rate region characterization and distributed strategies with instantaneous and statistical CSI," in *IEEE Transactions on Signal Processing*, 2010.

[142] E. Bjornson, R. Zakhour, D. Gesbert, and B. Ottersten, "Optimal coordinated beamforming in the multicell downlink with transceiver impairments," in *IEEE Global Communication Conference*, 2012.

[143] E. Bjornson, R. Zakhour, D. Gesbert, and B. Ottersten, "Characterization of convex and concave resource allocation problems in interference coupled wireless systems," in *IEEE Transactions on Signal Processing*, 2011.

[144] H. Boche and M. Schubert, "A general duality theory for uplink and downlink beamforming," in *IEEE Vehicular Technology Conference*, 2002.

[145] H. Boche and M. Schubert, "A calculus for log-convex interference functions," in *IEEE Transactions on Information Theory*, 2008.

[146] F. Bock and B. Ebstein, "Assignment of transmitter powers by linear programming," in *IEEE Transactions on Electromagnetic Compatibility*, 1964.

[147] D. Cai, T. Quek, C. Tan, and S. Low, "Max-min SINR coordinated multipoint downlink transmission duality and algorithms," in *IEEE Transactions on Signal Processing*, 2012.

[148] G. Caire, N. Jindal, M. Kobayashi, , and N. Ravindran, "Multiuser MIMO achievable rates with downlink training and channel state feedback," in *IEEE Transactions on Information Theory*, 2010.

[149] G. Caire, S. Ramprashad, and H. Papadopoulos, "Rethinking network mimo: Cost of csit, performance analysis, and architecture comparisons," in *IEEE Transactions on Information Theory*, 2010.

[150] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna gaussian broadcast channel," in *IEEE Transactions on Information Theory*, 2003.

[151] I. Csiszar and J. Korner, "Broadcast channels with confidential messages," in *IEEE Transactions on Information Theory*, 1978.

[152] H. Dahrouj and W. Yu, "Coordinated beamforming for the multicell multiantenna wireless system," in *IEEE Transactions on Wireless Communications*, 2010.

[153] D. Dardari, V. Tralli, and A. Vaccari, "A theoretical characterization of nonlinear distortion effects in OFDM systems," in *IEEE Transactions on Communications*, 2000.

[154] K. Fan, "Minimax theorems," in *Proceedings of National Academic Society*, 1953.

[155] C. Farsakh and J. Nossek, "Channel allocation and downlink beamforming in an SDMA mobile radio system," in *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, 1995.

[156] H. Ekstrom, "Qos control in the 3GPP evolved packet system," IEEE Communications Magazine, 2009.

[157] M. Ghorbanzadeh, Y. Chen, K. Ma, C. Clancy, and R. McGwier, "A neural network approach to category validation of android applications," in *IEEE Conference on Computing, Networking, and Communications (ICNC)*, 2013.

[158] Z. Kbah and A. Abdelhadi, "Resource allocation in cellular systems for applications with random parameters," *CoRR*, vol. abs/1507.07608, 2015.

[159] M. Gast, "802.11 wireless networks: The definitive guide," O'REILLY, 2005.

[160] M. Portnoy, "Virtualization essentials," WILEY, 2012.

[161] IBM, "x3250 m4," IBM Redbooks Product Guide, 2011.

[162] A. Bachmutsky, "System design for telecommunication gateways," WILEY, 2010.

[163] C. Spurgeon, "Ethernet: The definitive guide," O'REILLY, 2014.

[164] L. Chappell and G. Combs, "Wireshark network analysis: The official wireshark certified network analyst study guide," IProtocol Analysis Institute, Chappell University, 2010.

[165] A. Abdelhadi and T. Clancy, "An optimal resource allocation with frequency reuse in cellular networks," vol. abs/1507.07161, 2015.

[166] I.-H. Hou and C. S. Chen, "Self-organized resource allocation in lte systems with weighted proportional fairness," in *Communications (ICC), 2012 IEEE International Conference on*, 2012.

[167] K. Levenberg, "A method for the solution of certain non-linear problems in least squares," in *Quarterly of Applied Mathematics*, 1944.

[168] D. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," in *SIAM Journal on Applied Mathematics*, 1963.

[169] M. Yokoo, E. Durfee, T. Ishida, and K. Kuwabara, "The distributed constraint satisfaction problem: formalization and algorithms," IEEE Transactions on Knowledge and Data Engineering, 1998.

[170] K. Kiwiel, T. Larsson, and P. Lindberg, "Lagrangian relaxation via ballstep subgradient methods," Mathematics of Operations Research, 2007.

[171] Y. Wang, H. Zhou, and J. Wu, "Reference signals power allocation for uplink MU-MIMO in LTE system," Systems and Networks Communications (ICSNC), 2010.

[172] H. Shajaiah, A. Abdel-Hadi, and C. Clancy, "Utility Proportional Fairness Resource Allocation with Carrier Aggregation in 4G-LTE," MILCOM, 2013.

[173] "Enhance mobile networks to deliver 1000 times more capacity by 2020."

[174] C. M. and D. R., "Spectrum occupancy measurements of the 3550-3650 Megahertz maritime radar band near san diego, california," NTIA, U.S. Dept. of Commerce, 2014.

[175] S. Wilson and T. Fischetto, "Coastline population trends in the united states: 1960 to 2008," U.S. Dept. of Commerce, 2010.

[176] J. Reed and N. Tripathi, "Cellular communications: A comprehensive and practical guide," Wiley-IEEE Press, 2014.

[177] G. Hufford, A. Longley, and W. Kissick, "A guide to the use of the its irregular terrain model in the area prediction mode," U.S. Dept. of Commerce, 1982.

[178] V. Shumilov, A. Trushanin, and R. Maslennikov, "Design of link-to-system mapping interface for LTE-a uplink system level simulations," Mobile Services and Subscribers report, 2012.

[179] J. G. Proakis, "Digital communications," McGraw Hill, 2007.

[180] T. U. S. G. Survey, "National land cover data set," in *ftp://edcftp.cr.usgs.gov/pub/data/landcover/states*, 2005.

Index

3.5 GHz band, 183

3550 - 3650 MHz, 181

3GPP, 183

Access Point, 65

Adaptive Modulation and Coding, AMC, 205

Aggregate Utility Function, 77

antenna azimuth, 187

Antenna Boresight, 187

antenna elevation, 187

antenna pattern, 187

Application Layer QoS, 10

Application Usage Dynamics, 123

Application Utility Function, 31

area prediction mode, 191

Automatic Repeat Request, ARQ, 205

Backlobe, 185

Centralized Architecture, 54

centralized resource allocation, 90

Channel Quality Indicator,CQI, 206

Cluster, 189

cochannel, 192

composite antenna pattern, 187

computational complexity, 130

Congested/Contested Environments, 172

continuous rate, 148

Convex, 76, 143

convex optimization, 55

Cosine Antenna Pattern, 185

Council of Advisers on Science and Technology, 181

Cross Layer QoS, 11

delay-tolerant applications, 32

diffraction loss, 183

digital elevation model, DEM, 230

digital signal processing,DSP, 206

Distributed Architecture, 72

Distributed Architecture with Rebidding, 116

Distributed Architecture without Rebidding, 119

DL:UL split, 188

Downtilt, 194

Dual Problem, 80, 221

dwell time, 185

Elastic Traffic, 32

eNB, 72, 203

Ethernet, 65

exclusion zone, 183

exponential effective SNR mapping,EESM, 187

External UE Resource Allocation, EURA, 75

Federal Communications Commission, 181

File Transfer Protocol, 36

Fixed-Utility UE, 126

fluctuation decay function, 152

Forward Error Correction, FEC, 205

Frank Kelly Algorithm, 40

free space path loss, 183

FTP, 36

Geometric Interpretation, 81

Global Optimal Solution, 75

Hardware QoS, 11

HARQ, 187

horizontal beamwidth, 184

HTTP, 65

hysical Uplink Shared Channel, PUSCH, 208

IBM, 65

Implementation, 63

indoor small cell, 187

inflection point, 103

Infonetics Research, 1

Instranet, 65

intercell interference, 167

internal UE rate allocation,IURA, 74

intersite distance, 187

intra-UE computation, 130

intracell interference, 167

irregular terrain model, ITM, 183

Lagrange Multiplier , 81

Lagrangian, 91

Land Use Land Cover, LCLU, 232

Line-of-Sight, LoS, 190

Link Layer QoS, 8

link-to-system mapping, 187

log-centralized optimization, 90

Logarithmic Utility, 94

LTE, 142

LTE bandwidth, 188

LTE sub-frame, 193

LTE symbol, 193

macro cell model, 187

main beam, 185

mathematical equivalence, 88

Max-Min Resource Allocation, 38

Minimum QoS, 146
Mobile Management Entity, 174
Mobile Network Operators, 1
modulations and coding scheme, MCS, 204
multi input multi output, MIMO, 187
multimedia telephony, 142

National Telecommunications and Information Administration, NTIA, 181
Network Layer QoS, 9
NF,noise figure, 206
Non-Ideal Link Adaptation, 187
non-LoS, NLoS, 190
Nonlinear, 143
Nonnegativity, 34

out-of-band interference, 193
outdoor small cell, 193

PCAST, 181
Physical Uplink Control Channel, PUCCH, 208
Propagation Model, 191
Proportional Fairness, 39
Pulse Repetition Interval, PRI, 193
pulse repetition interval, PRI, 184

QoS Challanges, 16
QoS Requirement, 108
Quality of Service, 3

Radar, 179
radar-to-LTE interference, 183
radio environment map, REM, 229
Rate Matching Module, RMM, 206
Real-Time Application, 33
refractivity, 191
Resource Allocation Requirements, 46
Resource Block Allocation, 153
resource block allocation, 141
resource element, 234, 237, 241, 245, 249, 253, 257, 261, 265, 269, 273, 277, 281, 285, 289, 293, 297, 301, 305, 310
Resource Management, RM, 208
Robust Algorithm, 86
Router, 65

sector, 187
Sensitivity Analysis, 110
Server, 65
Shadow Price, 81, 222
shadow price, 126
shipborne, 191
sigmoidal Utility, 94
single carrier orthogonal frequency division multiple access,SC-OFDMA, 187
single-carrier, 109

SINR, 193

slope curvature function, 82

Solution Existence, 77

Sounding Reference Signals, SRS, 207

subscriber differentiation, 88

System Model, 48

Time Division Duplex, TDD, 183

Traffic Analysis, 110

Traffic Shaping, 68

traffic-dependent, 103

transmission overhead, 119

transmission time interval,TTI, 185

troposcatter loss, 183

Turbo coding, 205

turbo decoder, 187

UE Quantity Dynamics, 111

UE throughput, 193

United Stated Geological Survey, USGS, 230

Uplink, 183

User Equipment, UE, 72, 203

Utility Function Parameters, 36

video streaming, 95

VoIP, 94

WiFi, 65

WiMAX, 182

YouTube, 65