Congestion Mitigation in LTE Base Stations using Radio Resource Allocation Techniques with TCP End to End Transport

A thesis submitted in fulfilment of the requirements for the degree of Master of Engineering

Aviroop Ghosh

B.Eng, The University of Melbourne

School of Engineering
College of Science, Engineering and Health
RMIT University

November 2019
Declaration

I certify that except where due acknowledgement has been made, the work is that of the author alone; the work has not been submitted previously, in whole or in part, to qualify for any other academic award; the content of the thesis is the result of work which has been carried out since the official commencement date of the approved research program; any editorial work, paid or unpaid, carried out by a third party is acknowledged; and, ethics procedures and guidelines have been followed.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

______________________________
Aviroop Ghosh

26th November 2019
Abstract

As of 2019, Long Term Evolution (LTE) is the chosen standard for most mobile and fixed wireless data communication. The next generation of standards known as 5G will encompass the Internet of Things (IoT) which will add more wireless devices to the network. Due to an exponential increase in the number of wireless subscriptions, in the next few years there is also an expected exponential increase in data traffic. Most of these devices will use Transmission Control Protocol (TCP) which is a type of network protocol for delivering internet data to users. Due to its reliability in delivering data payload to users and congestion management, TCP is the most common type of network protocol used. However, the ability for TCP to combat network congestion has certain limitations especially in a wireless network. This is due to wireless networks not being as reliable as fixed line networks for data delivery because of the use of last mile radio interface. LTE uses various error correction techniques for reliable data delivery over the air-interface. These cause other issues such as excessive latency and queuing in the base station leading to degradation in throughput for users and congestion in the network. Traditional methods of dealing with congestion such as tail-drop can be inefficient and cumbersome. Therefore, adequate congestion mitigation mechanisms are required. The LTE standard uses a technique to pre-empt network congestion by a mechanism known as Discard Timer. Additionally, there are other algorithms such as Random Early Detection (RED) that also are used for network congestion mitigation. However, these mechanisms rely on configured parameters and only work well within certain regions of operation. If the parameters are not set correctly then the TCP links can experience congestion collapse.

In this thesis, the limitations of using existing LTE congestion mitigation mechanisms such as Discard Timer and RED have been explored. A different mechanism to analyse the effects of using control theory for congestion mitigation has been developed. Finally, congestion mitigation in LTE networks has been addresses using radio resource allocation techniques with non-cooperative game theory being an underlying mathematical framework. In doing so, two key end-to-end performance measurements considered for measuring congestion for the game theoretic models were identified which were the total end-to-end delay and the overall throughput of each individual TCP link. An end to end wireless simulator model with the radio access network using LTE and a TCP based backbone to the end server was developed using MATLAB. This simulator was used as a baseline for testing each of the congestion mitigation mechanisms. This thesis also provides a comparison and performance evaluation between the congestion mitigation models developed using existing techniques (such as Discard Timer and RED), control theory and game theory.
Acknowledgements

For most of the duration of this thesis, I was employed full time at NBN Co. I would like to thank NBN Co for allowing me to carry out research at RMIT. I would especially like to thank my former colleagues Mr Colin Rudolph and Mr Benedick Agdipa for sparking my interest in wireless networks and mentoring me during the early years of my career.

I would like to thank my supervisors Professor Kandeepan Sithamparanathan and Dr Karina Gomez Chavez for giving me the opportunity to work with them and for their supervision.

Finally, I would like to acknowledge my friends and family especially my parents for their encouragement and support.
# Table of Contents

**CHAPTER 1: INTRODUCTION** .............................................................................................................. 1
  1.1 BACKGROUND .......................................................................................................................... 1
  1.2 RESEARCH GAPS .................................................................................................................... 3
  1.3 RESEARCH QUESTIONS ......................................................................................................... 5
  1.4 RESEARCH CONTRIBUTIONS .................................................................................................. 6
  1.5 LIST OF PUBLICATIONS .......................................................................................................... 9
  1.6 THESIS ORGANIZATION ....................................................................................................... 9

**CHAPTER 2: BACKGROUND AND LITERATURE REVIEW ON CONGESTION MITIGATION MECHANISMS** .... 11
  2.1 TCP FLOW CONTROL AND QUEUE MANAGEMENT TECHNIQUES ........................................ 11
  2.2 CONGESTION MANAGEMENT IN WIRELESS NETWORKS .................................................... 12
  2.3 GAME THEORETIC ANALYSIS IN NETWORKS .......................................................................... 13
  2.4 ANALYSING LTE BASE STATION CONGESTION MANAGEMENT ........................................... 14
    2.4.1 Limitations of Existing Congestion Mitigation Mechanisms ........................................... 14
    2.4.2 Applying Game Theory to LTE Networks .......................................................................... 15
  2.5 OTHER APPROACHES TO CONGESTION MITIGATION ............................................................. 15
    2.5.1 Split TCP Connection ......................................................................................................... 15
    2.5.2 Software Defined Networking ........................................................................................... 16
    2.5.3 Machine Learning Techniques ........................................................................................... 16
  2.6 CHAPTER SUMMARY ............................................................................................................... 17

**CHAPTER 3: END TO END SIMULATED WIRELESS NETWORK MODEL DESIGN** .............................. 18
  3.1 SIMULATED MODEL DESIGN OVERVIEW ............................................................................. 19
  3.2 TRANSPORT LAYER ............................................................................................................... 22
    3.2.1 Introduction to Transmission Control Protocol ................................................................. 23
    3.2.2 Background on TCP ........................................................................................................ 23
    3.2.3 Principles of Operation ...................................................................................................... 25
    3.2.4 TCP Packet Generation ................................................................................................... 27
  3.3 NETWORK LAYER .................................................................................................................... 29
    3.3.1 Packet Processing ............................................................................................................. 30
    3.3.2 Packet Multiplexing .......................................................................................................... 30
  3.4 DATA LINK LAYER .................................................................................................................... 31
    3.4.1 LTE Protocol Stack ............................................................................................................ 32
    3.4.2 Transmission over Downlink ............................................................................................ 32
  3.5 PHYSICAL LAYER ....................................................................................................................... 33
    3.5.1 Determining Signal to Noise Ratio .................................................................................... 34
    3.5.2 Downlink Transmission ..................................................................................................... 38
    3.5.3 Physical Layer Parameter Settings .................................................................................... 43
  3.6 SIMULATED END-TO-END WIRELESS NETWORK DESIGN .................................................... 44
  3.7 CHAPTER SUMMARY ............................................................................................................... 48

**CHAPTER 4: BASE STATION CONGESTION MITIGATIONS MODELS BASED ON QUEUE MANAGEMENT** .... 50
4.1 PROPORTIONAL FAIR SCHEDULER .................................................................................................................. 51
4.1.1 Proportional Fair Model .................................................................................................................................. 51
4.1.2 Simulated Results for Proportional Fair Model ............................................................................................. 53
4.2 CONGESTION MITIGATION USING DISCARD TIMER ..................................................................................... 53
4.2.1 Discard Timer Model .......................................................................................................................................... 54
4.3 CONGESTION MITIGATION USING CONTROL THEORY .................................................................................. 54
4.3.1 Random Early Detection ................................................................................................................................... 55
4.3.2 Proportional Controller Model .......................................................................................................................... 56
4.4 RESULTS AND DISCUSSION ............................................................................................................................ 60
4.4.1 Good Radio Conditions ...................................................................................................................................... 61
4.4.2 Average Radio Conditions ................................................................................................................................. 62
4.4.3 Poor Radio Conditions ....................................................................................................................................... 63
4.5 CONCLUSION ....................................................................................................................................................... 64
4.6 CHAPTER SUMMARY ......................................................................................................................................... 65

CHAPTER 5: BASE STATION CONGESTION MITIGATION BASED ON GAME THEORY ................................. 66
5.1 DEFINING PERCEIVED THROUGHPUT AND RUNNING AVERAGE .............................................................. 67
5.1.1 Perceived Throughput ...................................................................................................................................... 67
5.1.2 Running Average ............................................................................................................................................... 68
5.2 GAME THEORETIC MODEL FOR BASE STATION CONGESTION MITIGATION: BUYER AND SELLER MODEL ................................................................................................................................. 70
5.2.1 Establishing the Price for Seller .......................................................................................................................... 71
5.2.2 Non-Cooperative Game amongst Buyers ......................................................................................................... 72
5.2.3 Analysing the Buyer and Seller Model .............................................................................................................. 73
5.3 GAME THEORETIC MODEL FOR BASE STATION CONGESTION MITIGATION: EQUILIBRIUM MODEL ..................................................................................................................................................... 75
5.3.1 Game Model ......................................................................................................................................................... 75
5.3.2 Analysing the Equilibrium Model ....................................................................................................................... 76
5.4 RESULTS AND DISCUSSION ............................................................................................................................ 79
5.4.1 Good Radio Conditions ...................................................................................................................................... 80
5.4.2 Average Radio Conditions ................................................................................................................................. 81
5.4.3 Poor Radio Conditions ....................................................................................................................................... 82
5.5 CONCLUSION ....................................................................................................................................................... 83
5.6 CHAPTER SUMMARY ......................................................................................................................................... 84

CHAPTER 6: CONCLUSIONS AND SCOPE FOR FUTURE WORK ................................................................. 87
6.1 CONCLUSIONS .................................................................................................................................................... 87
6.2 FUTURE WORK .................................................................................................................................................. 88

BIBLIOGRAPHY ................................................................................................................................................... 98
List of Figures

Figure 1-1: TCP packet loss over a wireless link ................................................................. 2
Figure 1-2: Average Throughput for a single user .................................................................. 4
Figure 1-3: Wireless End to End Model .................................................................................. 6
Figure 1-4: Different Queue-Based Congestion Control Models ........................................... 7
Figure 3-1: End-to-End Wireless Network Design ................................................................. 19
Figure 3-2: Controller Files Relationship ............................................................................... 20
Figure 3-3: Sub-file References for source_controller.m ...................................................... 21
Figure 3-4: eNodeB_scheduler_controller.m Sub-file Reference .......................................... 22
Figure 3-5: TCP Slow-Start and Congestion Avoidance ......................................................... 24
Figure 3-6: Wireless Slow-Start as a Multiplexer ................................................................... 31
Figure 3-7: LTE Protocol Stack ............................................................................................. 32
Figure 3-8: Transport Block ..................................................................................................... 33
Figure 3-9: Base Station and UE Interaction .......................................................................... 33
Figure 3-10: Wireless Channel Environment ......................................................................... 34
Figure 3-11: Wireless Propagation Environment .................................................................... 36
Figure 3-12: Probability Density Function of different Radio Conditions .............................. 37
Figure 3-13: One Physical Resource Block ............................................................................. 38
Figure 3-14: SNR and CQI ....................................................................................................... 39
Figure 3-15: CRC with Transport Block Size ......................................................................... 40
Figure 3-16: Formation of FEC .............................................................................................. 42
Figure 3-17: Transport Layer Flow Diagram ......................................................................... 45
Figure 3-18: Network Layer Flow Diagram .......................................................................... 46
Figure 3-19: Data Link and Physical Layer Flow Diagram ...................................................... 47
Figure 4-1: Base Station Congestion Controllers ................................................................. 50
Figure 4-2: Buffer Occupancy lower than Discard Timer Threshold ...................................... 54
Figure 4-3: Buffer Occupancy higher than Discard Timer Threshold ..................................... 54
Figure 4-4: AQM - RED Packet Probability Drop Behaviour ............................................... 55
Figure 4-5: N Sources in a Wireless Network ........................................................................ 57
Figure 4-6: Feedback Loop for Proportional Controller ....................................................... 59
Figure 4-7: Good Radio Conditions ...................................................................................... 62
Figure 4-8: Average Radio Conditions ................................................................................... 63
Figure 4-9: Poor Radio Conditions ......................................................................................... 64
Figure 5-1: Base Station Congestion Controllers ................................................................. 66
Figure 5-2: End-to-End Network Latency ............................................................................. 69
Figure 5-3: Willingness to Sell Index .................................................................................... 72
Figure 5-4: Results for Scenario 1 ......................................................................................... 77
Figure 5-5: Results for Scenario 2 ......................................................................................... 78
Figure 5-6: Results for Scenario 3.................................................................................................. 79
Figure 5-7: Price Metric for all Scenarios .................................................................................... 79
Figure 5-8: Good Radio Conditions .............................................................................................. 81
Figure 5-9: Average Radio Conditions ......................................................................................... 82
Figure 5-10: Poor Radio Conditions ............................................................................................. 83
Figure 6-1: Use of Numerologies in 5G ....................................................................................... 90
Figure 6-2: Example of Implementing QoS with the Equilibrium Model..................................... 93
List of Tables

Table 1-1: Parameters Affected by Congestion Mitigation Algorithms ............................................ 8
Table 1-2: Reporting metric ................................................................................................................. 9
Table 3-1: Descriptions for Sub-file Reference for source_controller.m ........................................... 21
Table 3-2: Descriptions for Sub-file Reference for eNodeB_scheduler_controller.m ...................... 22
Table 3-3: TCP Principles of Operation .............................................................................................. 26
Table 3-4: Path Loss Model Parameters ............................................................................................ 35
Table 3-5: Radio Channel Parameters ............................................................................................... 37
Table 3-6: CQI and Modulation Scheme Mapping ............................................................................. 40
Table 3-7: CQI and TBS Index Mapping .............................................................................................. 41
Table 3-8: RBs Generated with Spatial Multiplexing of 2 ................................................................. 41
Table 3-9: Physical Layer Parameters ............................................................................................... 44
Table 3-10: Sample Output for 10 Users ............................................................................................ 48
Table 3-11: Sample Output for 15 Users ............................................................................................ 49
Table 4-1: CQI and TBS Allocation for Proportional Fair Scheduler .................................................. 53
Table 5-1: Example of Seller and Buyer ............................................................................................ 71
Table 5-2: Data for Example 1 ........................................................................................................... 73
Table 5-3: Data for Example 2 ........................................................................................................... 74
Table 5-4: Fairness Amongst Users for Different Controllers ............................................................ 85
Table 6-1: Service Differentiation and Resource Allocation ............................................................... 95
# List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIMD</td>
<td>Additive Increase Multiplicative Decrease</td>
</tr>
<tr>
<td>ACK</td>
<td>Acknowledgement</td>
</tr>
<tr>
<td>AQM</td>
<td>Active Queue Management</td>
</tr>
<tr>
<td>ARED</td>
<td>Active Random Early Detection</td>
</tr>
<tr>
<td>ARQ</td>
<td>Automatic Repeat Requests</td>
</tr>
<tr>
<td>BER</td>
<td>Bit Error Rate</td>
</tr>
<tr>
<td>CQI</td>
<td>Channel Quality Index</td>
</tr>
<tr>
<td>CoDel</td>
<td>Controlled Delay</td>
</tr>
<tr>
<td>DT</td>
<td>Discard Timer</td>
</tr>
<tr>
<td>eNodeB</td>
<td>Evolved Node B</td>
</tr>
<tr>
<td>EPC</td>
<td>Evolved Packet Core</td>
</tr>
<tr>
<td>FEC</td>
<td>Forward Error Correction</td>
</tr>
<tr>
<td>FIFO</td>
<td>First In First Out</td>
</tr>
<tr>
<td>FPGA</td>
<td>Field Programmable Gate Array</td>
</tr>
<tr>
<td>FTTH</td>
<td>Fibre To The Home</td>
</tr>
<tr>
<td>FTTN</td>
<td>Fibre To The Node</td>
</tr>
<tr>
<td>HARQ</td>
<td>Hybrid Automatic Repeat Request</td>
</tr>
<tr>
<td>IoT</td>
<td>Internet of Things</td>
</tr>
<tr>
<td>IP</td>
<td>Internet Protocol</td>
</tr>
<tr>
<td>LTE</td>
<td>Long Term Evolution</td>
</tr>
<tr>
<td>MAC</td>
<td>Medium Access Control</td>
</tr>
<tr>
<td>ML</td>
<td>Machine Learning</td>
</tr>
<tr>
<td>MTU</td>
<td>Maximum Transmission Unit</td>
</tr>
<tr>
<td>NACK</td>
<td>Negative Acknowledgement</td>
</tr>
<tr>
<td>NR</td>
<td>New Radio</td>
</tr>
<tr>
<td>OSI</td>
<td>Open Systems Interconnection</td>
</tr>
<tr>
<td>PDCP</td>
<td>Packet Data Convergence Protocol</td>
</tr>
<tr>
<td>PHY</td>
<td>Physical</td>
</tr>
<tr>
<td>P</td>
<td>Proportional</td>
</tr>
<tr>
<td>PER</td>
<td>Packet Error Rate</td>
</tr>
<tr>
<td>PI</td>
<td>Proportional Integral</td>
</tr>
<tr>
<td>PRB</td>
<td>Physical Resource Block</td>
</tr>
<tr>
<td>QoS</td>
<td>Quality of Service</td>
</tr>
<tr>
<td>RAN</td>
<td>Radio Access Network</td>
</tr>
<tr>
<td>RE</td>
<td>Resource Element</td>
</tr>
<tr>
<td>RED</td>
<td>Random Early Detection</td>
</tr>
<tr>
<td>RLC</td>
<td>Radio Link Control</td>
</tr>
<tr>
<td>RTO</td>
<td>Retransmission Timeout</td>
</tr>
<tr>
<td>RTT</td>
<td>Round Trip Time</td>
</tr>
<tr>
<td>SDN</td>
<td>Software Defined Networking</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal to Noise Ratio</td>
</tr>
<tr>
<td>SUI</td>
<td>Stanford University Interim</td>
</tr>
<tr>
<td>TCP</td>
<td>Transmission Control Protocol</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>------------------------------------</td>
</tr>
<tr>
<td>TTI</td>
<td>Transmission Time Interval</td>
</tr>
<tr>
<td>UDP</td>
<td>User Datagram Protocol</td>
</tr>
<tr>
<td>UE</td>
<td>User Equipment</td>
</tr>
</tbody>
</table>
Chapter 1: Introduction

A study by Ericsson showed that as of 2016 there were 7.4 billion mobile subscriptions worldwide. Data traffic reached around 7 exabytes each month. By 2021, this number is expected to increase by 12 times [1]. There is also an expected exponential increase in traffic from Internet of Things (IoT) devices. Wireless is an excellent alternative of delivering broadband to homes and business since there are high costs involving deployment of Fibre to the Home (FTTH) and Fibre to the Node (FTTN) especially in rural and regional areas. With the advent of LTE-Advanced and 5G technologies, wireless broadband speeds will be able to match some fixed line technologies [2]. A report in 2018 by the Federal Communications Commission showed that 92.3% of all Americans have access to 25 Mbps/3 Mbps speeds however, 24 million Americans still lack that speed [3]. Wireless broadband services will only keep being a competitive option if it is able to deploy cheaper, faster and more importantly keep up with user application demands.

Users rely on transport level protocols such as Transmission Control Protocol (TCP) or User Datagram Protocol (UDP) for data delivery. Most wireless data traffic will use TCP since it is the most dominant internet transport application [4]. In a wireless network, the TCP link competes with other TCP links under fluctuating radio conditions for the Base Station or Evolved Node B (eNodeB) radio resources. Unlike fixed line services, wireless broadband services are limited by the Radio Access Network (RAN). The RAN is a shared interface that can be subjected to various signal and power degradations and serves as a bottleneck.

Globally it is a challenge for telecommunications operators as well as vendors to configure base stations such that they make optimal use of their spectrum allocations. With an increase in data traffic and/or the number of users there is an increase in network congestion. Additionally, adverse radio conditions also cause congestion in the base station resulting in poor end user throughput. To mitigate congestion, more spectrum allocation can be allocated. But this is an extremely costly solution and therefore, adequate congestion management and mitigation strategies are required.

1.1 Background

TCP has a basic form of congestion control mechanism. The TCP algorithm tries to constantly probe the network conditions. A packet drop signals the TCP source of network congestion. In case of a packet drop in the network, TCP will restrict the number of packets that can be sent. The logic behind is to ramp up and down the TCP transmission rate till the correct equilibrium for the link is reached. With the growth in internet traffic, evolution of network equipment and improvements in networks the TCP congestion mechanism is insufficient for optimal network performance [5].

There are additional problems with a TCP link especially when the last mile is a wireless connection. Wireless interfaces can be lossy, and a packet drop over the air-interface can falsely signal the TCP source of network congestion [6] as shown in Figure 1-1.
Figure 1-1: TCP packet loss over a wireless link

However, there can be improvements made at the Data Link Layer to improve the quality of transmission over the air-interface. With techniques such as Forward Error Correction (FEC) techniques and Automatic Repeat Request (ARQ) techniques like Hybrid Automatic Repeat Request (HARQ), the quality of data over the air interface can be improved. Additionally, with improvements in hardware and processing ability base station buffer sizes can be kept very large [7].

Due to the base station buffer size being large another issue known as bufferbloat arises. Bufferbloat is when the packets in the base station queue up for extended periods before they reach the User Equipment (UE) thereby inducing a greater latency [8]. This can lead to significant performance degradation especially with TCP based applications where the packets could be waiting for prolonged periods when the link is stable, and the buffer is not congested enough. On the contrary if the base station buffer is congested either the packet processing time increases, or the packets are discarded from the buffer. This results in unstable performance for TCP links and poor throughput.

Therefore, TCP’s congestion control ability is limited especially in dynamic networks such as wireless networks. These issues with TCP can be alleviated with adequate congestion management techniques. Modern day applications use queue management in network devices known as Active Queue Management (AQM) [9] whereby packets are dropped pre-emptively before the onset of network congestion. This reduces the TCP transmit rate for the link for which the packet was dropped thereby reducing congestion.

Different mechanisms have been proposed for AQM. The standardized LTE method uses a latency-based Discard Timer (DT) [10] whereby all packets are dropped which occupy the buffer longer
than a given configurable interval. Other mechanisms exist which use buffer-based control where the number of packets occupying the buffer is a factor for packet drop. One such mechanism is known as Random Early Detection (RED), which is primarily used on congestion control for routers [11]. For implementing RED, if the queue size is over a given threshold then packets are dropped probabilistically. Other AQM implementation applicable to wireless networks such Adaptive RED (ARED) and Controlled Delay (CoDel) [12] have been devised which adjust the parameter values based on the buffer load.

1.2 Research Gaps

A major drawback of implementing RED and similar mechanisms is that the algorithm requires careful parameter configuration and tuning. Given the dynamic nature of wireless networks, buffer based AQM mechanisms may prove to be cumbersome. It should be noted that AQM mechanisms like RED, ARED and CoDel are analysed mathematically using control theory [13] [14] since the TCP mechanism can be mathematically modelled as a feedback loop. However, using control theory limits the analysis in wireless networks due to the dynamism of radio interface capacity.

In addition, DT and RED require parameter setting for congestion mitigation. The complications associated with incorrect parameter tuning is not being able to maintain the most efficient operating point for the TCP links thus degrading throughput. More severe cases can result in congestion collapse whereby the packets for a given TCP link are dropped before reaching the user thereby reducing the throughput for the link [15]. An example has been shown in Figure 1-2. The output is from the MATLAB tool developed as a part of this thesis. The y-axis represents the throughput for a single user and the x-axis represents the number of users simultaneously served by the base station. Note that the number of users is progressively increased implying an increase in congestion. It can be observed that the parameters are set for optimal use when there are between 1 and 6 users. The throughput drastically deteriorates when more users are added (6 – 9 users) i.e. the networks starts getting congested. After 10 users, the link experiences congestion collapse.
Since the base station is a bottleneck and is responsible for allocation radio resources to users, congestion mitigation algorithms can be employed by the base station to improve end user throughput. The research gaps can be addressed if the LTE resource allocation mechanisms can be combined with the principles of Active Queue Management. However, a different approach mathematically would need to be explored.

Game theory is an important mathematical field and allows network congestion problems to be analysed as economic problems of supply and demand. This can be applied to LTE resource allocation strategies where the users act as independent players and try to maximize their payoff. AQM induced packet drop for congestion management can be triggered when a user is trying to monopolise over radio resources to maximise their payoff at the detriment of the other users. This study formulates a game theoretic analysis for network congestion problems for TCP networks in LTE base stations.

In this thesis, congestion mitigation mechanisms that use queue-based indicators for congestion such as DT, RED or any model that uses the base station buffer (either occupancy or amount stored) as a congestion indicator are defined as queue-based congestion control mechanisms. Those that make use of game theory are defined as game theoretic based congestion control mechanisms. To address the research gaps, existing congestion mitigation algorithms need to be explored and will have to be compared with congestion mitigation algorithms that are devised by optimizing radio resource allocation by the base station. It should be noted that this thesis analyses
users in low mobility or fixed situations. Only downstream and best effort traffic has been taken into account when analysing the base station congestion mitigation mechanisms.

1.3 Research Questions

There are three research questions that will be addressed as a part of the thesis. These have been listed below.

Research Question 1: How can congestion mitigation models be developed and analyzed with an LTE radio network and an end to end TCP backbone?

Before developing different congestion mitigation models, a suitable tool is required for carrying out development and testing as a baseline. This research question explores any tool development, method of evaluation and any reverse engineering (such a method of evaluating Channel Quality Index) that is required before exploring the research gaps.

Research Question 2: For a wireless LTE network with the base station as the bottleneck, for multiple TCP links, how can a base station congestion mitigation algorithm be derived factoring in variable radio capacity using control theory? How does this model compare with DT and RED? What are the limitations of this model?

Control theory usually forms the mathematical basis for analysing congestion management for networks using TCP. This research question examines the feasibility of using control theory in a dynamic wireless network to perform congestion control by factoring the variable radio link capacity. A comparison of the results with DT and RED will show the improvements provided by this model. Further examination should reveal the practicability of this approach in a wireless network.

Research Question 3: For a wireless LTE network with the base station as the bottleneck, for multiple TCP links, how can congestion mitigation algorithms be derived using game theory? Do these models overcome the limitations of using control theoretic and 3GPP discard timer models?

The research question is aimed at introducing non-cooperative game theory using LTE radio resource allocation and congestion management. A method by which congestion can be measured for the games to take place will need to be defined. Following this, the non-cooperative models will need to be developed and compared with the control theoretic and the DT models to determine their robustness.
1.4 Research Contributions

This section elaborates on the novel research contributions from this thesis to address the research questions and have been summarised as follows.

**Wireless End to End Tool development using MATLAB**

To address the research questions a suitable platform for testing and validation is required. As a part of this thesis 18 MATLAB scripts were developed. The scripts developed simulate the behaviour of multiple number of TCP sources in a wireless network. They also emulate the LTE radio interface and for radio base station resource allocation as well as include radio condition signalling from the User Equipment (UEs) to the base station in the form of a Channel Quality Index (CQI). In addition, each base station congestion mitigation model is designed using a script. These scripts can be integrated with the tool as plug-ins to test each of the controllers individually. The output from the tool is used to provide a comparative evaluation of the different controllers. Figure 1-3 shows the scope of the design for the wireless end to end model. In Chapter 3, the design of an end to end wireless network design tool that was simulated using MATLAB has been described in detail.

![Figure 1-3: Wireless End to End Model](image)

Developing a simulation tool instead of using an open-source or existing simulation tool provides the required flexibility to perform the investigation and analysis required for this study. The
simulation tool developed is a major contribution as an outcome of this thesis and provides the foundational platform for addressing all the research questions and addresses research question 1.

Development of a Proportional Controller Model Factoring Radio Conditions

Due to TCP congestion control mechanism, the mathematical analysis of the end-to-end network can be performed using control theory. This approach assumes the use of a fixed value for capacity. This feedback control mechanism has been shown in Figure 1-4 (a). However, in wireless networks the radio base station capacity varies. Therefore, analysing wireless networks using control theory is limited.

This thesis attempts to develop a control theoretic congestion mitigation model that factors in variable radio conditions as shown in Figure 1-4 (b). For this purpose, a proportional controller has been developed. Just like RED, the congestion indication is based on the queue size of the base station but factors in the radio conditions. It is expected that the outcome from this particular study would provide a feasibility studying if a control theoretic analysis can provide the desired outcome for congestion mitigation. Design and analysis of the proportional controller model addresses research question 2. The outcomes for this study have been presented in Chapter 4.

This research contribution addresses research question 2.

![Figure 1-4: Different Queue-Based Congestion Control Models](image)

Foundations required for Game Theoretic Models

Queue-based congestion management use the base station buffer as an indicator for congestion. However, for the game theoretic models developed in this study there is no correlation between the base station buffer size and indication of congestion. For the game theoretic models, congestion is
based on the rate of TCP transmission and the resource allocation strategies of the users. The strategies for the game depend on each payoff function of the user. Therefore, it is important to define the variables used to determine a payoff function of the user. The concepts of running average and perceived throughput have been introduced in Chapter 5 for this purpose.

This research contribution provides the technical foundation for addressing research question 3.

Non-Cooperative games for Congestion Mitigation

Two congestion mitigation models have been developed using the principles of non-cooperative game theory. The first is known as the Buyer-Seller model, where users are classified as either buyers or sellers. The users are allowed to trade radio resources with each TTI based on a game model. The second model is the Equilibrium model where users predict an action path to maximise their payoff given the action path of the other users. These have been described in Chapter 5. Novel methods for resource distribution and congestion mitigation have also been introduced in a game theoretic framework for LTE networks. At the end of the chapter, the results have been compared with the queue-based congestion models. This research contribution addresses research question 3.

Method for Analysing Data

There have been 5 different congestion mitigation algorithms implemented as a part of this study. To provide a comparative study between the different models, there needs to be certain metrics and parameter output that have to be defined. The references [16] [17] in provide a list of parameters that can be varied to measure the performance for congestion mitigation models and the metrics that can be reported.

Table 1-1 shows the list of parameters that affect the performance of congestion mitigation algorithms. In this study, these parameters were either kept constant or varied in order to test the robustness of each of the models.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traffic Pattern</td>
<td>The traffic pattern for each user is identical</td>
</tr>
<tr>
<td>Network Load</td>
<td>The network load has been varied by increasing/decreasing the number of users</td>
</tr>
<tr>
<td>Link Capacity</td>
<td>Radio link capacity has been varied and have been classified as good, average and poor</td>
</tr>
<tr>
<td>Buffer Size</td>
<td>The base station buffer size is assumed to be infinite in size except for the queue-based congestion mitigation models where an upper limit threshold value for buffer size is factored in</td>
</tr>
<tr>
<td>Round Trip Time (RTT)</td>
<td>The RTT for all the users is consistent and is mainly dependent on the buffer occupancy in the base station buffer</td>
</tr>
</tbody>
</table>
Table 1-2 shows the reporting metrics that were produced as an output for each user from the MALTAB tool that was developed. Note, to quantify the output in a presentable format the mean and standard deviation for all the users in a given scenario was considered. Using these a set of principles that is used for analysis has been defined in Chapter 4 for queue-based congestion controllers and have been reused in Chapter 5 after the introduction of the game theoretic models. This research contribution is used for addressing research questions 2 and 3.

Table 1-2: Reporting metric

<table>
<thead>
<tr>
<th>Reporting Metric</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Throughput</td>
<td>The throughput values for each user over a given period is reported</td>
</tr>
<tr>
<td>Packet Loss/Retransmission Rate</td>
<td>The number of packets that are re-transmitted by the TCP source due to no acknowledgement received</td>
</tr>
<tr>
<td>Latency</td>
<td>The round-trip delay of each TCP link which is measured by the packet transmission and the Acknowledgement or Negative-Acknowledgement sent to the TCP source</td>
</tr>
<tr>
<td>Window Size</td>
<td>The average window size measured over a given interval</td>
</tr>
</tbody>
</table>

1.5 List of Publications

Two research papers that have been developed as an outcome of this work. The first paper is based on the Equilibrium model and the second paper is based on the Buyer-Seller model. The titles of the papers are:

2. Aviroop Ghosh, Karina Gomez Chavez, Sithamparanathan Kandeepan; “Congestion Mitigation using Non-Cooperative Game Theory in LTE Networks” (based on the Buyer-Seller Model)

1.6 Thesis Organization

This section provides an overview the thesis structure and a breakdown on each of the subsequent chapters. The thesis has six chapters.

- **Chapter 2: Background and Literature Review on Congestion Mitigation Mechanisms.** This chapter provides the literature review required to identify and understand research gaps and research questions. Discussions focus on the deficiencies of TCP’s own congestion control ability, existing AQM mechanism and their limitations and existing game theory literature for networks. Based on these discussions the final section discusses how non-cooperative game theory can be introduced in LTE networks in the context of mitigation congestion.
• **Chapter 3: End to End Simulated Wireless Network Model.** This chapter discusses the wireless network model that was designed using MATLAB. The theoretical background and its corresponding practical implementation have been provided. The tool overview and structure have also been discussed in detail.

• **Chapter 4: Base Station Congestion Mitigation Models based on Queue Management.** The proportional fair algorithm is the underlying mechanism for radio resource allocation for queue-based congestion control models. This chapter introduces the unique way in which the proportional fair algorithm allocates radio resources. Following this, three congestion mitigation mechanism have been developed namely, DT, RED and a proportional controller model that factors in radio conditions has also been introduced. The results from the three controllers have been evaluated under different congestion situations and radio conditions.

• **Chapter 5: Base Station Congestion Mitigation based on Game Theory.** The concepts of running average and perceived throughputs have been defined in this chapter. On the basis of these concepts two different non-cooperative game congestion mitigation models have been developed.

• **Chapter 6: Conclusion and Future Work.** The last chapter provides a discussion on a final conclusion for all the models developed. The future work such as applicability of the studies applicability to 5th generation of cellular network technology (5G) have been discussed and implementation based on different Quality of Service (QoS) strategies.
Chapter 2: Background and Literature Review on Congestion Mitigation Mechanisms

The objective of this thesis is to develop a wireless base station congestion mitigation model in an LTE environment in low mobility conditions for users using TCP traffic. This chapter provides a review of previous literature on congestion control and mitigation in networks. Key aspects detailed are the mathematical conceptualization of TCP’s inherent congestion behavior, implementation of existing algorithms that provide network congestion mitigation and advocated approaches to congestion mitigation in wireless networks. An examination of these topics lead-up to the novel congestion mitigation techniques that have been detailed as a conclusion to this chapter.

2.1 TCP Flow Control and Queue Management Techniques

The TCP algorithm has the ability to perform congestion control by increasing or decreasing the source transmission rate in the absence of congestion or detection of congestion. The detection of congestion is based upon how the network responds to the packets [18]. If during a transmit cycle, the TCP source does not detect congestion, then the rate is increased in the consecutive transmit cycle. However, if the network is congested and there is a packet drop then the TCP source reduces the transmit rate. These concepts have been discussed further in Chapter 3.

The dynamics of TCP can be modelled by a set of differential equations known as the fluid flow model [19]. The model allows a simplified mathematical method of analysing TCP behaviour and provides a better understanding of TCP congestion control. Based on the modelling, it can be determined that the TCP throughput is inversely proportional to the Round-Trip Time and the square-root of the packet loss probability [20] [21] as shown in (1).

\[
\text{Throughput} = \frac{1.22 \times \text{MTU}}{\text{RTT} \times \sqrt{\text{Loss}}} \quad (1)
\]

Assuming that the main reason for packet loss is due to buffer overflows instead of link loss, the equation shows that packet drops are synchronous across all the links i.e. the transmission rate for multiple TCP links will be reduced concurrently. For congestion control implementation using tail drop, all the packets exceeding a queue threshold are discarded. This would result in a number of TCP links discarding packets simultaneously. This occurrence is known as TCP global synchronization [22].

Studies have shown that buffer management systems which break this synchronisation lead to a fairer resource allocation for links with unequal latencies and payload data [23] like most networks. To prevent this, Active Queue Management (AQM) techniques have been developed. AQM pre-empts network congestion by dropping packets before the TCP links experience congestion in the buffer [24]. This results in the TCP source to decrease the transmit rate thereby mitigating congestion.
One type of AQM known as Random Early Detection (RED) provides congestion mitigation by probabilistically discarding packets. The packet discard probability increases with an increasing queue size.

By analysing the TCP/AQM dynamics using the fluid flow model, congestion management can be implemented using control theory. To simplify the mathematics and control theoretic analysis, a linearized model for the TCP/AQM dynamics was proposed in [13]. Different models using control theory have been proposed. The congestion control model in [25] uses a proportional-derivative (PD) controller and a disturbance observer (DOB) to reduce the network convergence time in order to stabilise the queue. The DOB is used for compensating for non-linearity in a queue. In [5] the use of a proportional-integral (PI) and a proportional (P) controller were implemented. The results showed that both the PI and P controllers had faster response times than RED.

AQM using control theory can provide better performance than passive congestion mechanisms such as tail-drop [26] although there are certain challenges involved in implementing controllers. For example, RED requires careful parameter configuration for optimized use [27]; when the number of TCP connections exceed a certain threshold, the buffer can become unstable [28].

2.2 Congestion Management in Wireless Networks

Unlike fixed line services, the last mile of the physical layer of wireless broadband services are shared amongst multiple users. This shared interface is known as the Radio Access Network (RAN). TCP encounters various problems in wireless networks such as high link error rates which is caused due to propagation over the air interface. Additional issues such as high link latency and large delay variations happen due to radio resource scheduling over the air interface. Base stations constantly adapt to radio conditions and this causes unstable latencies [29].

The challenge of high link error rates has been greatly mitigated by improving the quality of data going over the radio interface. Methods such as Forward Error Correction (FEC) and Automatic Repeat Request (ARQ) techniques like Hybrid Automatic Repeat Request (HARQ) have reduced the probability of data corruption over the air-interface. Due to employing such techniques to minimize link error rates and improvements in the base station’s packet processing ability, the base station buffer sizes are large for the additional processing needed [7]. Consequently, an issue known as bufferbloat arises. This happens when the packets in the base station queue up for extended periods before they reach the User Equipment (UE) hence inducing a greater latency [30]. This further leads to high link latency and large delay variations for TCP connections.

There have been different AQM implementations proposed to combat bufferbloat. One such implementation is known as Adaptive RED (ARED). With ARED, the target queue size is defined by the operator and the probability of packet drop adjusts depending upon this parameter [31]. However,
the mathematics that is involved in defining ARED assumes a fixed capacity. In a wireless network, the
capacity of the base station varies depending upon the radio conditions.

A different AQM implementation know Controlled Delay (CoDel) has also been proposed. CoDel does not require queue size as a congestion indicator like AQM models such as RED [12] since it uses the localised delay in the bottleneck as an indicator for link congestion. This ensures scalability, for example when the link capacity is high it expected that a larger flow of traffic can be tolerated. The CoDel algorithm sets a target queue period for a link by monitoring when the packets entered the queue and left the queue. If the period exceeds the minimum period, then CoDel drops a packet from the tail of the queue. The algorithm then decreases the next drop interval, which has an inverse relationship to the number of packets dropped [12] [32]. In studies such as [32], [33] and [34] variations of CoDel have been developed either based on the delay timer as in [32] and [33] or adjusting based on traffic type (such as low-rate traffic) as developed in [34]. For implementation in a wireless network, it is important to factor in wireless parameters such as radio channel quality to improve the performance of wireless systems as was done in [35].

The standardized LTE method uses a latency-based discard timer [10] whereby all packets which occupy the buffer longer than a given configurable interval are dropped. The discard timer mechanism is effectively tail drop in time domain [36]. The studies from [36] conclude that neither discard timer nor RED implementation in LTE networks is suitable for managing base station PDCP buffers for internet traffic to users.

2.3 Game Theoretic Analysis in Networks

In this thesis, game theory has been applied for the purposes of congestion mitigation and resource allocation. Game theory is an active research topic and finds applicability in various network related problems. Previous studies using game theory for wireless networks have been for power control [37] [38], network selection and admission control [39], wireless security [40] [41] and subcarrier and spectrum allocation [42] [43].

For congestion management, a good analogy is with the economic concepts of supply and demand which factor in user/network utility and use factors such as price and cost to act as control parameters to deal with network congestion. Game theory provides an ideal mathematical framework to perform such an analysis. Within the context of game theory, noncooperative game theoretic models and the concept of Nash Equilibrium finds useful application in analysing congestion management.

Several studies focus on the noncooperative aspects of game theory in particular, when analysing networks. Users in a network can be considered players with the game defining each player’s objective. For a congestion mitigation game, the objective for a user is to minimize cost (or maximize payoff) given the strategy of all the other players where cost for a user is a function of the congestion
in the network and/or the network resource usage required to satisfy the utility requirement of the user. This approach to network congestion control has been used in [44] [45] [46]. The concept of pricing has been expanded further to include service providers trying to maximize their revenue and dictating their pricing strategies to the users [47] [48]. In [48] [49] a Stackelberg game (leader-follower) is used with the base station as the leader and the users as the followers. The leader sets the price and the followers are allocated resources based on a noncooperative game. This allows a service provider to determine the price setting and when to provide additional capacity during congestion scenarios.

2.4 Analysing LTE Base Station Congestion Management

Based on the literature review in previous sections, this section discusses the limitations of using existing congestion mitigation techniques such as DT and RED (Section 2.4.1) in wireless networks. The following section (2.4.2) discusses the applicability of game theoretic models for congestion mitigation in LTE base stations.

2.4.1 Limitations of Existing Congestion Mitigation Mechanisms

There are three main aspects considered in this study where existing congestion mitigation are limited for applicability in LTE networks.

**Configurable Parameters**

Control theoretic models such as RED or Proportional control and LTE implementation of Discard Timer need static configurable parameters as indicators for congestion hence require certain regions of operation for the system to be stable. If the congestion threshold is set too high, then the packets will queue in the base station buffer thereby inducing latency. If the threshold is set too low, then the packets will be discarded before TCP is able to fully scale leading to congestion collapse. Both these factors result in throughput degradation or unstable performance.

**Mathematical Formulation of using the Fluid Flow Model**

The mathematical formulation using the fluid flow model requires constant parameters. For example, the study in [50] shows that modelling RTT variations using fluid flow model results in skewed conclusions regarding the window flow control mechanisms. Wireless channel capacity depends upon the performance over the radio interface which can fluctuate. Therefore, incorrect analysis and conclusions can be drawn by analysing congestion control using the fluid flow model for wireless networks.
Using Wireless Network Specific Congestion Mitigation Mechanisms

Specific implementation to cater to wireless networks such as ARED and CoDel have been developed. ARED also assumes a fixed capacity in its analysis [51]. LTE based simulations have shown that Reno achieves twice the throughput when CoDel is not used [51].

2.4.2 Applying Game Theory to LTE Networks

Game theory is a viable alternative for congestion mitigation in LTE networks. In order to develop a non-cooperative game, there are a few principles that need to be followed:

- Defining an objective function for each user
- Objective function will factor in user’s utility and a cost of using radio resources by the base station
- Objective of the user will be to minimize cost (or maximize payoff) while maximizing its utility given the strategy of other players

These principles can be adopted when considering the mechanism by which radio resource allocation is performed by the base station. An objective function will determine if the user’s requirements are met given the amount of radio resources it receives. In addition, the price parameter can be used as an indicator for base station congestion. Based on these a non-cooperative game can determine the specific actions the users need to undertake in order to satisfy their respective objective function.

Congestion mitigation models using control theory or queue management are limited to a certain operating region and it is hard to extend mathematical analysis to variable parameters. Given a proper mathematical framework, game theoretic models will not have such limitations. Therefore, congestion mitigation in LTE base station for TCP flows will be more robust when using a game theoretic approach.

2.5 Other approaches to Congestion Mitigation

In this section, some of the other recent developments and active research areas of congestion mitigation in networks have been discussed. These approaches have been provided to complement this chapter but have bearing to the later developments and contributions in this thesis.

2.5.1 Split TCP Connection

With a split TCP connection, the end to end TCP mechanism is broken at the base station or the wireless gateway into two separate connections [52] [53] [54]. A key advantage is that it enables to separate the reliable TCP backbone connection with the lossy TCP connection over the air interface.
This implies that the backbone TCP connection does not reduce the transmission rate due to mistakenly identifying an air-interface drop as congestion.

There are many potential pitfalls with this approach. It increases the complexity of the network by requiring the base station or the gateway to maintain the per-connection state [54]. Also, if the wireless link is lossy then a split TCP connection cannot do much to improve it. As previously mentioned in Section 2.2 with improvements over the air-interface, the utility of a split-TCP connection is negated.

### 2.5.2 Software Defined Networking

Software Defined Networking (SDN) is the next evolution in networks with a gradual shift from traditional legacy networks. Traditional networks store control plane information such as routing and switching decisions in each node across as network. For example, in a leaf-spine network topology, a Layer 3 node will have a network map based on the IP table while a Layer 2 node will have a network map based on the MAC table. SDN architecture will allow for a centralised control plane management using a centralised SDN controller [55]. SDN will allows streamlining and increase efficiency of network management. Although most of SDN research efforts have been towards the control plane, there have been recent attempts to include SDN to the data plane.

Greater congestion mitigation at a bottleneck node can be achieved if the AQM scheme is applied to each TCP connection based on the application that is running. The study in [56] proposes a method to implement this by extending SDN principle to the data plane. However, dynamically adapting AQM is not currently possible for switches and routers due to the logic required being implemented in the hardware [55]. [56] uses a Field Programmable Gate Array (FPGA) to circumvent this problem. The FPGAs can be reconfigured by an SDN controller. There are limitations that have been identified in this approach such as the inability for applications to signal their objectives (throughput, low delay, power, flow completion, transaction completion, tail completion) to the controller.

### 2.5.3 Machine Learning Techniques

A radically different approach to congestion control has been proposed in [57]. The overall end to end congestion control scheme is designed using Machine Learning (ML) techniques by a computer algorithm. The algorithm known as Remy, is fundamentally different from all the congestion control algorithms.

Remy first uses an objective function that is to be minimized on the basis of average throughput or average round-trip time. Remy then keeps track of the network conditions that is updated each time an acknowledgement (ACK) is received by the source. Remy only keeps track of three state variables. Remy maps regions of this state space to an action. This mapping of a state space
to one particular action is known as a rule. This rule is then split by considering alternative actions which are then further broken into smaller regions by considering a further set of alternative actions. On convergence, Remy splits the most popular rule and repeats the process thereby fine tuning the congestion control process. The results from Remy show that it performs much better than traditional TCP base congestion control algorithms. However, due to its complexity and the reasons for its superior performance, Remy is still being understood and investigated [58].

2.6 Chapter Summary

This chapter began with explaining on how the TCP algorithm’s congestion control mechanism can be mathematically analysed using the fluid flow model. Relying on this model for congestion control can result in TCP synchronisation where concurrent TCP links drop packets simultaneously. To overcome this, it was explained that network device-initiated congestion mitigation mechanisms such as RED have been introduced. The congestion mitigation mechanisms use the queue size as a reference for packet drops and can be mathematically explained using control theory. LTE networks use a discard timer mechanism which drops packets after a given configured period of buffer occupancy.

Wireless congestion control can be challenging since wireless networks require additional queueing to run various techniques in order not to compromise data quality when transmitting over the air-interface. This causes excessive queuing and a problem known as Bufferbloat. Therefore, careful congestion management is required in wireless networks. However, the fundamental problem of analysing wireless networks using control theory is that they do not cater for variable parameters such as radio channel capacity in their mathematical analysis.

Based on existing literature and implementation of game theory, it can be postulated that game theory can provide an innovative approach to overcome the constraints of existing congestion mitigation mechanisms such as RED or discard timer for congestion management in wireless networks.
Chapter 3: End to End Simulated Wireless Network Model Design

The objective of this thesis is to validate and analyse different base station congestion mitigation controllers. Each of these controllers will have contrasting mathematical and theoretical foundations. To provide a proper framework for validating the effectiveness of the controllers, an end-to-end wireless network model has been designed as a baseline for testing. The wireless network model was designed and created using scripts from the mathematical software tool MATLAB. In the overarching wireless network model, different base station congestion controller models were scripted to act as plug-ins.

Figure 3-1 shows the end-to-end wireless model that was designed. The sources 1 to N are TCP packet generating sources. Each source is assumed to be running the same TCP application type. The packets from the sources aggregate at the wireless gateway. This then forwards the aggregated traffic to the base station over a common fibre link. From a physical layer perspective, the sources have been designed to be equidistant from the wireless gateway.

At the base station, the radio interface transfers the packets to the User Equipment (UE) devices over the radio channel. Based on the type of scheduler and the radio conditions, the base station will determine the physical layer Protocol Data Unit (PDU) for the UEs. The scheduling will follow standard LTE based scheduling. The various controllers will have unique ways of managing the base station buffer and the number of radio connections for a user.

Note that the model has been designed such that there is a 1:1 ratio between the sources and the UEs with each source having a corresponding UE. Additionally, the backhaul network has been dimensioned assuming that there is fibre connectivity. The type of traffic in the network is Best Effort (BE) only.

The user-generated inputs are the number of sources specified, the type of base station scheduler to be executed and the duration of the program execution (in seconds). Additionally, the radio conditions, Maximum Transmission Unit (MTU), antenna spatial multiplexing configuration, LTE radio band, radio antenna transmit power, base station gain, and UE gain can be pre-configured before program execution. The outputs that will be generated are the throughput (in bits/s), latency (in seconds), percentage of packets retransmitted (in %) and average window size (in packets). By varying the input parameters and collecting the respective outputs, the results can be used to validate the robustness of the base station congestion controller and give a comparison of the performance.
The next section will discuss the model architecture and how the scripts in MATLAB were designed to simulate the end to end wireless network. In the subsequent sections, the discussions will consider the use of the Open Systems Interconnection (OSI) model as a reference. The aspects of the Transport, Network, Datalink and Physical layers that are modelled will be discussed in detail using theoretical foundations and modelling parameters used.

3.1 Simulated Model Design Overview

There were three main control m-files created using MATLAB. These three control functions interact with each other to simulate the wireless network. The control function interactions have been shown in Figure 3-2. The user defined inputs, namely those of the number of sources, the duration of the simulation and the eNodeB scheduler type will be received by server_controller.m. The file server_controller.m captures the behaviour of the sources and the TCP algorithm. This has been described in Transport Layer (Section 3.2). The data generated from the server_controller.m file is then further processed by the transport_controller.m file. The transport_controller.m file is used for
modelling the network layer functionalities such as packet processing and the packet gateway behaviour and is subject to discussion in Network Layer (Section 3.3). The data generated from the server_controller.m and processed by the transport_controller.m files are received by the eNodeB_scheduler.m file. The eNodeB_scheduler_controller.m has more complexities involved since it deals with the eNodeB congestion controllers created as a part of this study. Aside from the congestion control models, the m-file contains the simulators fundamental behaviour of the LTE radio resource scheduling and the radio interface that is generic across all congestion control models. Data Link Layer (Section 3.4) and Physical Layer (Section 3.5) discuss the theory and modelling in further detail. The results from the eNodeB_scheduler_controller.m file are sent back to transport_controller.m and server_controller.m. The loop continues until the run-time has elapsed. The server_controller.m then generates the usable output. The process and output received have been elaborated in Simulated End-to-End Wireless Network Design (Section 3.6) and Chapter Summary (Section 3.7) respectively.

The global_parameters.m file is an overarching file that contains a list of configurable parameters (such as MTU size used, radio frequency band) and is referred by all the controller files. This is why the global_parameters.m is represented by a dotted line in Figure 3-2.

To efficiently design and update the controller files of source_controller.m and eNodeB_scheduler_controller.m, there have been other m-files created which are referred within the controller file by their function reference. These files act as plug-ins to the controller files.

The references within the source_controller.m file have been shown in Figure 3-3. The dotted lines between the files show that there is a relationship between these files. For example, the output from source_determine_bitrate.m is used as an input by source_packet_generator and source_window_size.m. Additionally the outputs from source_determine_link_capacity.m and source_RTO_estimation.m can influence the source_window_size.m. This has been further elaborated in Principles of Operation (Section 3.2.3). The descriptions of each sub-file and the references to the theoretical and simulated results have been shown in Table 3-1.
Table 3-1: Descriptions for Sub-file Reference for source_controller.m

<table>
<thead>
<tr>
<th>M-File</th>
<th>Description</th>
<th>Section Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>source_window_size.m</td>
<td>Determines the Window size of each source per iteration</td>
<td>Principles of Operation</td>
</tr>
<tr>
<td>source_determine_link_capacity.m</td>
<td>Determines the maximum number of packets that can be generated by each source without an acknowledgement (maximum allowable window size)</td>
<td>Principles of Operation</td>
</tr>
<tr>
<td>source_determine_bitrate.m</td>
<td>Used for calculating the estimated achievable bitrate</td>
<td>TCP Packet Generation</td>
</tr>
<tr>
<td>source_packet_generator.m</td>
<td>Generates the number of packets based on the achievable bitrate</td>
<td>TCP Packet Generation</td>
</tr>
<tr>
<td>source_RTO_estimation.m</td>
<td>Calculates the TCP Timeout value for each source</td>
<td>TCP Retransmission Time Out</td>
</tr>
</tbody>
</table>

Similar to the source_controller.m, the eNodeB_scheduler_type.m file has four sub-files which are used as functions within the controller. This has been shown in Figure 3-4. Table 3-2 refers to the description and the section references for each file.
Table 3-2: Descriptions for Sub-file Reference for eNodeB_scheduler_controller.m

<table>
<thead>
<tr>
<th>M-File</th>
<th>Description</th>
<th>Section Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>eNodeB_Pathloss.m</td>
<td>Used for calculating the path-loss based on the radio</td>
<td>Path Loss Model</td>
</tr>
<tr>
<td></td>
<td>environment</td>
<td></td>
</tr>
<tr>
<td>eNodeB_CQI_Inputs.m</td>
<td>Generates the CQI values</td>
<td>Channel Quality Indicator</td>
</tr>
<tr>
<td>eNodeB_radio_parameters.m</td>
<td>The radio parameters measured by the UE are sent to</td>
<td>Transport Block Size Formation</td>
</tr>
<tr>
<td></td>
<td>the base station to calculate the radio resource to be</td>
<td></td>
</tr>
<tr>
<td></td>
<td>allocated for each UE</td>
<td></td>
</tr>
<tr>
<td>eNodeB_packet_prob_drop.m</td>
<td>The probability of packet drops for each transmission</td>
<td>Packet Drop Estimation</td>
</tr>
<tr>
<td></td>
<td>to a given UE over the air-interface</td>
<td></td>
</tr>
</tbody>
</table>

3.2 Transport Layer

The transport layer is responsible for providing reliable and efficient data transmission service between the two end points. Unlike network layer, mac or physical layer packet transmission, which have localised functionality (for example localised to the radio interface or communication between routers), the transport layer provides an overarching framework for traffic management between a source and a destination. From Figure 3-1, the transport layer connectivity will be maintained between each source and the corresponding user.

TCP is one of the most common transport layer mechanisms and is subject to further discussions in the next sections.
### 3.2.1 Introduction to Transmission Control Protocol

The choice of traffic that is used for in the network model is TCP, which is a dominant internet transport application [4]. TCP uses an algorithm that tries to maximise the throughput, and, at the same time, probes the network conditions. TCP is able to increase its rate of sending packets steadily in order to ensure congestion detection. In case of adverse network conditions such as link loss or congestion, TCP has a self-control mechanism that will restrict the number of packets that can be sent. This prevents from further packet losses and maintains a steady flow of packets. TCP additionally has an acknowledgement mechanism; when packets are not acknowledged, they are retransmitted. TCP therefore is useful for applications, which are sensitive to packet loss, congestion and require reliable data transfer such as Netflix, which uses TCP [59] as a transport protocol.

There are different versions of TCP like TCP Reno [60], TCP Cubic [61], TCP Vegas [62], of which TCP Reno is the most commonly used and fundamental TCP version; all other versions are built based on TCP Reno [63]. For this study, TCP Reno has been used.

Of particular importance in the simulation and modelling context is the TCP congestion control mechanism and the TCP timer control. TCP connection establishment and TCP connection release have not been discussed in detail since those aspects were not required for modelling. The simulation performs a basic latency calculation required for TCP connection establishment and the connection remains established unless there is a TCP timeout, following which the TCP connection needs to be re-established. Additionally, the modelling does not delve into the intricacies of TCP segment structure.

### 3.2.2 Background on TCP

In the following sections, the appropriate backgrounds for performing modelling for TCP congestion control mechanism and TCP re-transmission mechanisms have been provided.

#### TCP Congestion Control

As highlighted earlier, TCP has a unique feedback mechanism by which it is able to gauge the network conditions. The maximum number of unacknowledged packets the TCP application is allowed to inject into the network is defined as the window size. If, after a given period, acknowledgements are received for all the packets sent, then the TCP window size is increased. If there is a negative acknowledgement or the acknowledgement timer times out, then the window size is reduced. The key problem with employing this mechanism is to find the ideal operating point of the TCP rate.

TCP uses a concept known as Additive Increase Multiplicative Decrease (AIMD) to address this problem. AIMD is used such that the link converges to a fair state regardless of the initial state of the network. The size of the TCP window is increased by an additive factor when congestion is not detected, or packets are not lost, i.e., all packets are successfully transmitted over the network (all
acknowledgements received). If there is a negative acknowledgement or congestion detection, then the window size is reduced by a multiplicative factor. For example, for TCP Reno, the additive factor is 1 and the multiplicative factor is $1/2$ the window size. This has been defined in (1). It is important to note that the window size changes every Round-Trip Time (RTT), where RTT is defined as the time it takes for the source to send all the packets and receive all the acknowledgements.

$$W_{RTT+1} = \begin{cases} W_{RTT} + 1 & \text{if positive ACK} \\ \frac{W_{RTT}}{2} & \text{on packet drop} \end{cases}$$  \hfill (1)

The problem with AIMD is that it takes the network a long time to reach a suitable operating point if the packet size is increased by a small additive number (like 1 in the case of TCP Reno). The converse will also be true if the packet size is increased by a large additive number. To overcome this problem, TCP connections start with a phase known as slow start whereby the packets sent increase by a factor of 2 each RTT. The slow start phase continues until a packet drop is detected or a certain threshold is reached. Following the slow start phase, the AIMD phase is instated which shows a sawtooth like behaviour. This has been shown pictorially in Figure 3-5.

![Figure 3-5: TCP Slow-Start and Congestion Avoidance](image)

In addition to slow start and AIMD, TCP uses the Fast Retransmit/Fast Recovery mechanism to retransmit packets which have been lost. These packets are detected when the TCP sender receives duplicate acknowledgements for the same packet. This packet is the same as the last packet that was successfully received by the TCP receiver. Since this study is interested in TCP properties for packet
generation and behaviour but not the information carried by the packet itself, Fast Retransmit/Fast Recovery have not been modelled.

Different TCP types have different congestion mechanisms, for example TCP versions like Vegas use a delay-based mechanism in which the window size is pre-emptively reduced. However, to simulate AQM mechanisms it was judged that a loss-based mechanism was the best choice. TCP Reno was used for modelling and analysis.

TCP Retransmission Time Out

Each time a TCP packet is sent out, a timer is started that provides an estimate of when the TCP packet will be acknowledged. If the packet is not acknowledged within that particular timeframe then the packet is re-transmitted. The parameter estimation for TCP Retransmission Time Out (RTO) has been a well-researched area since predicting if packets have reached their respective destinations is not trivial. A dynamic algorithm has been developed in [64]. A variable parameter known as Smoothed Round-Trip Time (SRTT) maintains an estimate of the current RTT session and the equation is given in (2).

\[
\text{SRTT} = (1 - \alpha) \times \text{SRTT} + \alpha \times \text{R}
\]  

(2)

\(\alpha\) is a smoothing factor and \(R\) is the measured RTT. Along with SRTT, another parameter known as Round Trip Time VARIation (RTTVAR) is calculated in order to set RTO. As, packet behaviour in networks can be probabilistic, RTTVAR is used to absorb any variances in the round-trip times for packets and is given by (3).

\[
\text{RTTVAR} = \beta \times \text{RTTVAR} + (1 - \beta) \times |\text{SRTT} - \text{R}|
\]

(3)

The Retransmission Time Out value is then given by (4).

\[
\text{RTO} = \text{SRTT} + \max (G, K \times \text{RTTVAR})
\]

(4)

where, \(G\) is the clock granularity and \(K = 4\). Assuming that the computation time is negligible compared to the RTTVAR value, this results in (5), which is the preferred equation for the model.

\[
\text{RTO} = \text{SRTT} + K \times \text{RTTVAR}
\]

(5)

Estimating RTO is essential since it gives an indication of the data quality and the latency in the network.

3.2.3 Principles of Operation

Following from the previous introductory section on TCP, the operating principles of TCP using the model will be demonstrated in this section. Table 3-3 below shows a list of conditions that will be simulated using different combinations of the input variables. The model will show the output in the
form of the change in the window size. A value of 1 will represent that the variable as TRUE, a value of 0 represents the variable as FALSE, and X signifies that the same output is generated irrespective of the input parameter value.

Table 3-3: TCP Principles of Operation

<table>
<thead>
<tr>
<th>Condition</th>
<th>Slow Phase</th>
<th>Start</th>
<th>AIMD Phase</th>
<th>Timeout</th>
<th>Congestion</th>
<th>Window Limit</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition 1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Condition 2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Condition 3</td>
<td>X</td>
<td>X</td>
<td>0</td>
<td>X</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Condition 4</td>
<td>X</td>
<td>X</td>
<td>0</td>
<td>X</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Condition 5</td>
<td>X</td>
<td>X</td>
<td>0</td>
<td>X</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**Condition 1**

As shown in Table 3-3, the TCP source is in the slow start phase. It is important to note that the slow start and AIMD phase are mutually exclusive.

If the previous cycle showed a window size = 10 and measured RTT = 30 ms, then the model generates an updated window = 33.

**Condition 2**

Here, the slow start phase is set to 0, therefore the AIMD phase is 1. The rest of the variables are 0. For an input window size = 10 and RTT = 30 ms, the updated window size = 11.

This is consistent with the TCP Reno principles of increasing the window size by 1 for successful acknowledgements.

**Condition 3**

As discussed previously, TCP timeout happens when the packets are not acknowledged within a certain time threshold. In the case of the model, it usually implies that the latency is above certain tolerable threshold. Due to timeout, the window size value is reset to 1 (as per the operating principles of TCP Reno). The TCP connection is entering the slow start phase.

**Condition 4**

In this instance, the far end of notifies of a congestion in the link. In this study, the far end is the base station. If there is congestion detected, then the window size is halved.

For example, the window size = 10, AIMD = 1 (Slow-Start = 0), Timeout = 0 and Window Size Limit = 0. Then the updated window size = 5. If the application is in the slow start phase, then it enters the AIMD phase after halving the window size.
Condition 5

Depending upon the source, the window size will be limited for the TCP application the source is running. This limit on the window size has been factored in. If the window size limit is exceeded, then the window size is halved.

Just like Condition 4, if the application was in the slow start phase then it enters the AIMD phase.

Consider for example, the window size = 72, Slow-Start = 1 (AIMD = 0), Timeout = 0 and Congestion = X. Then the updated window size = 36. This is as per TCP Reno’s Fast Retransmit principle [58].

The following pseudo-code is generated to summarise the implementation of the TCP mechanism.

```plaintext
Enter Window Size
Enter RTT
Program Execution
  IF (TIMEOUT == 1) THEN
    Updated Window Size = 1 //Window Size is set to 1
    Slow Start = 1 //The TCP application enters Slow Start
  ELSEIF (Congestion == 1) OR (Window Size Limit == 1)
    Updated Window Size = Window Size/2 //Window Size is halved
    IF (Slow Start == 1) THEN
      AIMD = 1 //Enter AIMD
    END
  END
  ELSEIF (Congestion != 1) AND (Window Size Limit != 1)
    IF (Slow Start == 1) THEN
      Updated Window Size = Window Size + \log(2)/RTT
    END
    IF (AIMD == 1)
      Updated Window Size = Update Window Size + 1
    END
  END

3.2.4 TCP Packet Generation

The TCP window size determines the number of unacknowledged packets that can be included in the network at a given time. Calculation of the TCP window size was discussed in the previous section, TCP Principles of Operation. The time it takes for the packets to be acknowledged is the RTT. In this section, how the source application generates packets using the window size and RTT have been discussed.
The window size and RTT determine the achievable bitrate for an application the source is running. In this model, the achievable bitrate is grouped in bins \{1, 2, 3..., 19, 20\} with the unit as Mbps. Therefore, the smallest achievable bitrate is 1 Mbps and the highest is 20 Mbps. The achievable bitrate is given by (6).

\[
\text{Achievable Bitrate} = \frac{\text{WindowSize} \times \text{mtu}}{\text{RTT}}
\]  

where, \text{mtu} is the maximum transmission unit for an IP packet. \text{RTT}_{\text{max}} is the maximum allowable RTT in order to achieve the bitrate. The \text{mtu} chosen for the model is 1500 bytes or 120000 bits. The packets are generated based on a lognormal distribution. Consider as an example, the achievable is 3Mbits/s, the window size is 10 packets and the maximum allowable RTT is 30 ms. The model will generate the following sequence of packets, [12000, 12000, 12000, 7000, 12000, 12000, 12000, 10119].

Using the examples below, the packet generated by the model has been shown for different source conditions:

**Example 1: No Congestion**

The source measures an RTT of 15 ms and has a window size of 15 packets. From 6, the achievable bitrate = \(12 \times 10^6\) bps

For the input parameters listed above, the sequences of packets generated are,

<table>
<thead>
<tr>
<th>12000</th>
<th>12000</th>
<th>12000</th>
<th>11359</th>
<th>12000</th>
<th>12000</th>
<th>12000</th>
<th>11152</th>
<th>12000</th>
<th>12000</th>
</tr>
</thead>
<tbody>
<tr>
<td>12000</td>
<td>12000</td>
<td>2743</td>
<td>12000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note that the size of each packet is shown by the number of bits it contains.

For the succeeding cycle, assuming that there was no congestion, the window size is increased by 1 to cater for 16 packets. Also assume that the measured RTT increases to 25 ms. The achievable bitrate = \(7 \times 10^6\) bps.

Based on the input parameters of a window size of 16 packets, RTT of 25 ms and achievable bitrate of 7 Mbps, the following sequence of packets are generated:

<table>
<thead>
<tr>
<th>12000</th>
<th>8176</th>
<th>12000</th>
<th>7721</th>
<th>12000</th>
<th>7808</th>
<th>12000</th>
<th>7908</th>
<th>12000</th>
<th>8055</th>
<th>12000</th>
</tr>
</thead>
<tbody>
<tr>
<td>7800</td>
<td>12000</td>
<td>7422</td>
<td>12000</td>
<td>7989</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Example 2: Congestion

In this example, the same starting parameters as the first example will be used. However, it is assumed that the packets will face a packet drop and thereby reduction of window size. As listed previously, for a window size of 15 packets, RTT of 15 ms and achievable bitrate of 12 Mbps, the following packets are generated:

<table>
<thead>
<tr>
<th>12000</th>
<th>12000</th>
<th>12000</th>
<th>11359</th>
<th>12000</th>
<th>12000</th>
<th>12000</th>
<th>11152</th>
<th>12000</th>
<th>12000</th>
</tr>
</thead>
<tbody>
<tr>
<td>12000</td>
<td>12000</td>
<td>2743</td>
<td>12000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

After this cycle, the congestion is marked by the far end (user end). The window size is reduced by half to 7 packets. Assuming that the measured RTT is 10 ms, the achievable bitrate = $8 \times 10^6$ bps. The following packets are generated:

| 12000 | 12000 | 12000 | 12000 | 1184  | 12000 | 12000 |

Example 3: Low Latency

In this example, the consequences of low core latencies will be explored. Assume that source measures an RTT of 8 ms with window size of 20 packets. The achievable bitrate = 19 Mbps. The following sequence of packets are generated:

<table>
<thead>
<tr>
<th>12000</th>
<th>12000</th>
<th>12000</th>
<th>10208</th>
<th>12000</th>
<th>12000</th>
<th>12000</th>
<th>1784</th>
<th>12000</th>
</tr>
</thead>
<tbody>
<tr>
<td>12000</td>
<td>12000</td>
<td>4546</td>
<td>12000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

It is important to note that only 14 packets were generated even though the window size is 20. Therefore, the number of packets generated is dependent on the achievable bitrate. The window size represents the upper limit of the number of packets that can be generated.

Comparing the results of Example 1 and Example 2, it is obvious that having a higher window size does not necessarily increase the bitrate. Latency plays an important factor as well. The achievable bitrate of 7 Mbps from Example 1 is lower than the achievable bitrate of 8 Mbps from Example 2 since the RTT is 25 ms from Example 1 and 10 Mbps from Example 2 even though the window size is 16 packets for Example 1 and 7 packets for Example 2.

3.3 Network Layer

The network layer is defined as the layer where Internet Protocol (IP) packets are forwarded through routers. Although this might be deemed a simple functionality of the network layer, the overall functionality of packet forwarding, and routing is very intricate. In itself, the network layer has capabilities of adaptively forwarding packets over the optimal routing path, performing congestion control and maintaining Quality of Service (QoS). However, for modelling the network layer for this
study, the interest lies in packet processing and forwarding capability and the time taken for each of these activities.

3.3.1 Packet Processing

The packet processing time at the source and wireless gateway will be required for latency calculations. The downlink layer 3 network has packet forwarding capabilities for the source and packet processing and forwarding capabilities for the wireless gateway. The time required for packet generation at the source is considered negligible. To simplify uplink latency estimations, it is assumed that only one packet per link is used for acknowledgement/negative acknowledgement irrespective of the number of packets that were forwarded on the downlink. Therefore, per link the uplink latency comprises of forwarding and processing that acknowledgement/negative acknowledgement of that packet.

The model does not use a specific router to emulate the source or the wireless gateway functionality and does not delve into the line-card and network processor unit architecture. It is estimated that the time required to process/forward each packet is 5 micro-seconds which has been leveraged from the study in [66]. The essence of the wireless gateway is to emulate the network devices in the data plane in the Evolved Packet Core (EPC).

3.3.2 Packet Multiplexing

The wireless gateway acts as a multiplexer and packets from each individual source are queued on a First-in-First-Out (FIFO) basis. Packets are also not differentiated based on priorities. Figure 3-6 shows how packets from different sources (in downstream) are multiplexed and then queued before being forwarded via the egress port. The model has an internal mechanism to determine the inter-arrival time between each packet. The packets are queued and processed individually irrespective of the source they arrive from.
3.4 Data Link Layer

The data link layer is the logical layer between the network and the physical layers. In this study, the layer 2 data link layer is between the wireless gateway and the base station and the processing done at the base station. The base station receives these IP encapsulated packets known as frames.

The base station receives the frames. At the data link layer, the base station has the following functionalities:

- Buffer management
- Implementing congestion control
- Gauging radio conditions and connection with each UE
- Transmitting data units over the air-interface
- Maintaining data integrity of the data transferred over the physical layer
3.4.1 LTE Protocol Stack

Once the frames arrive at the base station, they are queued for scheduling. The scheduling mechanism in LTE networks happen at every Transmission Time Interval (TTI). For 4G networks, the TTI is a 1 millisecond interval. At each TTI, the base station allocates radio resources to each UE, which are then propagated via the air-interface. Before being scheduled, the packets arriving at the base station are queued, arranged and sized before being transferred over the air-interface. There are various layers in the LTE protocol stack that are important for understanding base station and UE interactions. Each IP packet encapsulated as frames being transported over the fibre backhaul stacks in the Layer 2 LTE rank before being passed over the physical (PHY) layer to the UE. The LTE protocol stack has been shown in Figure 3-7.

The Packet Data Convergence Protocol (PDCP) layer is the top sublayer within layer 2. The PDCP layer can be configured to discard data packets in case the IP packets timeout, i.e. wait in the buffer for a longer configured period [67] [68]. The Radio Link Control (RLC) layer is responsible for controlling the error correcting ARQ behaviour. The RLC holds the packets in the RLC buffer for retransmission. The RLC is also used to concatenate or segment packets received from the PDCP [69]. The final layer before the physical layer is the Medium Access Control (MAC) layer. The MAC is used for Hybrid Automatic Repeat Requests (HARQ) parameter setting as well as transporting packets over the air interface.

The LTE protocol stack plays an important role in congestion control and will be detailed in subsequent chapters.

3.4.2 Transmission over Downlink

At each TTI, the base station uses transport-blocks to transfer MAC layer Packet Data Units to the user via the physical layer. The transport-block that is sent to the UE is shown in Figure 3-8. The payload is a segment of the data frames that will be sent to the UE. At each TTI, the payload undergoes compression, segmentation and multiplexing before being passed on to the physical interface.
Transport block generation requires an understanding of how the physical layer behaves. This has been discussed in the next section.

### 3.5 Physical Layer

The interaction between the base station and the UE over the radio interface forms the basis of the physical layer. Analysing radio conditions between the base station and UE plays a pivotal role to determine the number of radio resources each UE is allocated. Modelling the base station and UE interactions is shown comprehensively in Figure 3-9.

The radio communication path in the form of electromagnetic waves between the base station and UE is affected by different natural factors such as reflection, refraction, absorption etc. The Signal to Noise Ratio (SNR) gives an assessment of the quality of the signal sent from the base station with respect to noise in the environment. Note that, in this study it is assumed that the UEs do not have any interference. A set of standard propagation models have been devised, which factor in variables such as UE antenna height, distance of UE to base station to determine the path-loss. Using one these models, the Signal to Noise Ratio (SNR) is calculated. Using the SNR, the Channel Quality Indicator (CQI) is calculated by the UE and the index is sent to the base station. The CQI determines the quality of the signal and plays a central role in radio resource allocation by the base station for the UE. The base station arranges radio resource information to the UE in time and frequency blocks known as Physical Resource Blocks (PRBs). The PRBs are sent to the UE’s and if decoded correctly by the UE, an acknowledgement is sent back to the base station. If the information is corrupted or lost while being transmitted over the air-interface, then the PRBs are retransmitted.
3.5.1 Determining Signal to Noise Ratio

Before detailing how CQI is calculated and used, it is important to determine how SNR is calculated since the CQI value depends on the SNR. However, in order to calculate and simulate SNR, details on the radio interface model will need to be discussed. This is done in this section.

3.5.2.1 Path Loss Model

Path loss is defined as the ratio of transmitted power to received power and is expressed in decibels. Figure 3-10 below shows a typical wireless channel environment. Note that even though all the UEs get the signal from the same base station, their radio paths are different. For example, UE 1 has direct line of sight while UE 3 is obstructed by foliage. Therefore, for UE 3 there should be reduction of power from the obstruction. This is one of the causes for path loss. Other reasons could be due to distance, cable losses, reflection, refraction and diffraction.

There are various standard empirical models to evaluate path loss such as Stanford University Interim (SUI) model, the COST-231 Hata model and the ECC-23 model. Each of these models have been developed by different standardisation bodies based on a set of experimental data. However, most of the models cater for urban, suburban or rural or all of the areas [70].

For the development of the path loss model, the Stanford University Interim (SUI) model has been considered. This model is suitable for a suburban environment. Three types of terrain parameters are used with the SUI model. Type A has the highest path loss with hilly and high foliage density, Type B is characterised as having a fairly flat terrain with medium to high level of foliage and Type C is defined as having minimum path loss with flat terrain and low foliage. Type B has been chosen as an appropriate condition for modelling. The path loss equation has been given by (7).

\[
P_L = A + 10\gamma \log_{10} \left( \frac{d}{d_0} \right) + X_f + X_h + s \quad \text{for } d > d_0
\]  

\[(7)\]
where,

d: Distance between the base station and the UE

s: Shadowing component, a zero mean Gaussian distributed factor with a variance of $\sigma^2$, used for measuring the fading caused by foliage

$d_0$: 10 m

The other parameters for are listed below.

\[
A = 20 \log_{10} \left( \frac{4\pi d_0}{\lambda} \right) \quad (8)
\]
\[
\gamma = a - bh_b + \frac{c}{h_b} \quad (9)
\]
\[
X_f = 6 \log_{10} \left( \frac{f}{2000} \right) \quad (10)
\]
\[
X_h = -10.8 \log_{10} \left( \frac{h_r}{2000} \right) \quad (11)
\]

where, $h_b$ is the base station height (in meters), $f$ is the frequency (in MHz) and $h_r$ is the UE antenna height (in meters). The parameter values for $a$, $b$ and $c$ are provided in Table 3-4.

<table>
<thead>
<tr>
<th>Model Parameter</th>
<th>Terrain A</th>
<th>Terrain B</th>
<th>Terrain C</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a$</td>
<td>4.6</td>
<td>4.0</td>
<td>3.6</td>
</tr>
<tr>
<td>$b$</td>
<td>0.0075</td>
<td>0.0065</td>
<td>0.005</td>
</tr>
<tr>
<td>$c$</td>
<td>12.6</td>
<td>17.1</td>
<td>20</td>
</tr>
</tbody>
</table>

3.5.2.1 Wireless Propagation Environment

In order to model the radio environment between the base station and the UEs, an omnidirectional antenna was simulated. All the UEs were placed within a radius of 1.2 km from the base station. This is shown in Figure 3-11. The UEs are represented by the blue cross hairs. Using this model as a baseline the SNR for each UE was calculated. This has been described in the following section.
Figure 3-11: Wireless Propagation Environment

3.5.2.1 Signal to Noise Ratio

The Signal to Noise Ratio (in dB) is given by (12).

$$SNR = P + G_{BS} + G_{UE} - PL - c_{loss} - P_N$$  \hfill (12)

where,

- $P$: Base station transmit power (in dBm)
- $G_{BS}$: Base station antenna gain (in dB)
- $G_{UE}$: UE antenna gain (in dB)
- $PL$: Path loss (in dB)
- $c_{loss}$: Factors in for any cable losses (in dB)
- $P_N$: Noise Power (dBm)

For modelling, the parameters used are shown in Table 3-5.
Table 3-5: Radio Channel Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base Station Transmit Power</td>
<td>46 dBm</td>
</tr>
<tr>
<td>Cable Losses</td>
<td>3 dB</td>
</tr>
<tr>
<td>Base Station Gain</td>
<td>20 dB</td>
</tr>
<tr>
<td>UE Gain</td>
<td>3 dB</td>
</tr>
<tr>
<td>Noise Power</td>
<td>-70 dBm</td>
</tr>
<tr>
<td>Cell radius</td>
<td>1200 m</td>
</tr>
</tbody>
</table>

The SNR parameters for each UE vary based on the shadowing standard deviation. In this thesis there are 3 different path loss exponent and shadowing standard deviation values chosen to generate the results in Chapter 4 and Chapter 5. Based on these the radio conditions are classified as Good, Average and Poor radio conditions. The probability density functions for the different radio conditions have been shown Figure 3-12.

![Figure 3-12: Probability Density Function of different Radio Conditions](image-url)
3.5.2 Downlink Transmission

Following discussions in the previous sections on the LTE protocol stack and calculating SNR, the dynamics of the LTE protocol stack and SNR can be used to explain how the frames are arranged in a physical layer service data unit before being transmitted over the air-interface. It should also be noted that while transmitting over the air-interface there is a certain probability of the data that is, getting corrupted due to external influence during transmission. This probability reduces with each successive retransmission. This section will describe how the model simulates the downlink transmission environment by providing the theoretical background and results from the simulation.

3.5.2.1 Physical Resource Blocks

The base station allocates radio resources to the users at every 1 millisecond interval known as the Transmission Time Interval (TTI). These TTIs alternate between downlink and uplink transmission for Time Domain Duplex (TDD) but not for Frequency Division Duplex (FDD). For downlink transmission, there is a maximum of 100 PRBs which are produced by the base station every TTI. PRBs are resources allocated in time and frequency blocks. Each RB contains 84 Resource Elements (RE), some of which are used as reference signals. In Figure 3-13, this has been highlighted with a bold ‘R’.

![Figure 3-13: One Physical Resource Block](image)

The base station scheduling mechanism allocates these PRBs every TTI and the PRBs are distributed amongst the UEs. Details on how the PRBs allocation is distributed amongst the UEs will not be covered in this chapter since it is the foundation for the studies involving base station congestion management that is covered extensively in Chapter 4. For modelling 100 PRBs are allocated per TTI.

3.5.2.3 Channel Quality Indicator

The Channel Quality Indicator (CQI) is an important parameter used to measure the reported radio conditions by the UE. CQI values range from 1 to 15 with 15 being the highest. CQIs are sent
periodically by the UE to the base station. The base station then uses CQI as one of the parameters to allocate resources. The UE estimates the CQI based on various parameters such as SNR, multipath delay and Block Error Rate (BLER) [71]. The algorithms for CQI estimation are proprietary to network operators and vendors. To estimate CQI values in this model, an algorithm has been developed which only uses SNR. Figure 3-14 shows a range of SNR values {1 dB, 35 dB}. For each SNR value a range of 100 CQI values was calculated with mode of the CQI values displayed in the Y-axis.

In LTE there is a mapping between CQI, modulation scheme and code rate. Table 3-6 shown below shows this mapping for modulation ranges from QPSK to 256QAM and is taken from [72]. The coding rate indicates the number of real bits of data present out of 1024 and the efficiency provides the number of information bits per modulation symbol. For example, with a CQI index of 2 and code rate of 193/1024 and modulation QPSK, the Efficiency = $2 \times \frac{193}{1024} = 0.3770$. This modulation scheme and code rate is required to determine the Transport Block and has been discussed in the next section.
Table 3-6: CQI and Modulation Scheme Mapping

<table>
<thead>
<tr>
<th>CQI index</th>
<th>Modulation</th>
<th>Code rate x 1024</th>
<th>Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td>out of range</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>QPSK</td>
<td>78</td>
<td>0.1523</td>
</tr>
<tr>
<td>2</td>
<td>QPSK</td>
<td>193</td>
<td>0.3770</td>
</tr>
<tr>
<td>3</td>
<td>QPSK</td>
<td>449</td>
<td>0.8770</td>
</tr>
<tr>
<td>4</td>
<td>16QAM</td>
<td>378</td>
<td>1.4766</td>
</tr>
<tr>
<td>5</td>
<td>16QAM</td>
<td>490</td>
<td>1.9141</td>
</tr>
<tr>
<td>6</td>
<td>16QAM</td>
<td>616</td>
<td>2.4063</td>
</tr>
<tr>
<td>7</td>
<td>64QAM</td>
<td>466</td>
<td>2.7305</td>
</tr>
<tr>
<td>8</td>
<td>64QAM</td>
<td>567</td>
<td>3.3223</td>
</tr>
<tr>
<td>9</td>
<td>64QAM</td>
<td>666</td>
<td>3.9023</td>
</tr>
<tr>
<td>10</td>
<td>64QAM</td>
<td>772</td>
<td>4.5234</td>
</tr>
<tr>
<td>11</td>
<td>64QAM</td>
<td>873</td>
<td>5.1152</td>
</tr>
<tr>
<td>12</td>
<td>256QAM</td>
<td>711</td>
<td>5.5547</td>
</tr>
<tr>
<td>13</td>
<td>256QAM</td>
<td>797</td>
<td>6.2266</td>
</tr>
<tr>
<td>14</td>
<td>256QAM</td>
<td>885</td>
<td>6.9141</td>
</tr>
<tr>
<td>15</td>
<td>256QAM</td>
<td>948</td>
<td>7.4063</td>
</tr>
</tbody>
</table>

3.5.2.3 Transport Block Size Formation

In LTE Transport Block Size (TBS) is the information passed from the MAC layer to the physical layer per TTI. The physical layer will add a Cyclic Redundancy Check (CRC) before transmitting to the UEs. This has been shown in Figure 3-15.

![Figure 3-15: CRC with Transport Block Size](image)

The TBS will depend on the number of resource blocks allocated per TTI and the modulation. Based on the CQI, the Modulation and Coding Scheme (MCS) index is selected using the table in [66]. The MCS index is then mapped to a Transport Block Size (TBS) index, from which the Transport Block Size is calculated. The mapping used for this model has been shown in Table 3-7.
Table 3-7: CQI and TBS Index Mapping

<table>
<thead>
<tr>
<th>CQI index</th>
<th>Modulation</th>
<th>TBS Index ($I_{TBS}$)</th>
<th>MCS ($I_{MCS}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>QPSK</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>QPSK</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>QPSK</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>16QAM</td>
<td>14</td>
<td>9</td>
</tr>
<tr>
<td>5</td>
<td>16QAM</td>
<td>16</td>
<td>10</td>
</tr>
<tr>
<td>6</td>
<td>16QAM</td>
<td>19</td>
<td>13</td>
</tr>
<tr>
<td>7</td>
<td>64QAM</td>
<td>21</td>
<td>15</td>
</tr>
<tr>
<td>8</td>
<td>64QAM</td>
<td>22</td>
<td>16</td>
</tr>
<tr>
<td>9</td>
<td>64QAM</td>
<td>24</td>
<td>18</td>
</tr>
<tr>
<td>10</td>
<td>64QAM</td>
<td>25</td>
<td>19</td>
</tr>
<tr>
<td>11</td>
<td>64QAM</td>
<td>29</td>
<td>22</td>
</tr>
<tr>
<td>12</td>
<td>256QAM</td>
<td>30</td>
<td>23</td>
</tr>
<tr>
<td>13</td>
<td>256QAM</td>
<td>31</td>
<td>24</td>
</tr>
<tr>
<td>14</td>
<td>256QAM</td>
<td>32</td>
<td>25</td>
</tr>
<tr>
<td>15</td>
<td>256QAM</td>
<td>33</td>
<td>27</td>
</tr>
</tbody>
</table>

Based on the number of resource blocks allocated and the TBS Index specified, the Transport Block Size is determined. For simulation, the indexes from Table 3-7 was used in MATLAB’s `lteTBS(number of resource blocks, $I_{TBS}$, spatial multiplexing)` function. An additional parameter that of spatial multiplexing can be specified. In this study, spatial multiplexing values of 2 was used. Table 3-8 shows a range of resource blocks from 20 to 27 using a spatial multiplexing of 2, the CQI and the corresponding TBS index have been listed in the left most columns.

Table 3-8: RBs Generated with Spatial Multiplexing of 2

<table>
<thead>
<tr>
<th>CQI</th>
<th>$I_{TBS}$</th>
<th>RB 20</th>
<th>RB 21</th>
<th>RB 22</th>
<th>RB 23</th>
<th>RB 24</th>
<th>RB 25</th>
<th>RB 26</th>
<th>RB 27</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6</td>
<td>4136</td>
<td>4392</td>
<td>4776</td>
<td>5352</td>
<td>6200</td>
<td>7224</td>
<td>8504</td>
<td>9912</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>5544</td>
<td>5992</td>
<td>6456</td>
<td>7224</td>
<td>8504</td>
<td>9912</td>
<td>11448</td>
<td>13536</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>6968</td>
<td>7480</td>
<td>7992</td>
<td>9144</td>
<td>10680</td>
<td>12216</td>
<td>14688</td>
<td>16992</td>
</tr>
<tr>
<td>4</td>
<td>14</td>
<td>11448</td>
<td>12216</td>
<td>12960</td>
<td>14688</td>
<td>16992</td>
<td>19848</td>
<td>23688</td>
<td>27376</td>
</tr>
<tr>
<td>5</td>
<td>16</td>
<td>12960</td>
<td>13536</td>
<td>14688</td>
<td>16992</td>
<td>19848</td>
<td>22920</td>
<td>26416</td>
<td>31704</td>
</tr>
<tr>
<td>6</td>
<td>19</td>
<td>16992</td>
<td>18336</td>
<td>19848</td>
<td>22152</td>
<td>25456</td>
<td>30576</td>
<td>35160</td>
<td>40576</td>
</tr>
<tr>
<td>7</td>
<td>21</td>
<td>19848</td>
<td>21384</td>
<td>22920</td>
<td>26416</td>
<td>30576</td>
<td>35160</td>
<td>40576</td>
<td>48936</td>
</tr>
<tr>
<td>8</td>
<td>22</td>
<td>21384</td>
<td>22920</td>
<td>24496</td>
<td>28336</td>
<td>32856</td>
<td>37888</td>
<td>43816</td>
<td>51024</td>
</tr>
<tr>
<td>9</td>
<td>24</td>
<td>24496</td>
<td>25456</td>
<td>28336</td>
<td>31704</td>
<td>36696</td>
<td>42368</td>
<td>51024</td>
<td>59256</td>
</tr>
<tr>
<td>10</td>
<td>25</td>
<td>25456</td>
<td>26416</td>
<td>29296</td>
<td>32856</td>
<td>37888</td>
<td>43816</td>
<td>52752</td>
<td>61664</td>
</tr>
<tr>
<td>11</td>
<td>29</td>
<td>29296</td>
<td>31704</td>
<td>34008</td>
<td>39232</td>
<td>45352</td>
<td>52752</td>
<td>61664</td>
<td>71112</td>
</tr>
</tbody>
</table>
3.5.2.4 Packet Drop Estimation

Wireless transmission links over the air-interface are susceptible to noise and interference. This could cause unnecessary frame loss and low throughput. There are techniques where transmission over the air-interface can be improved. These have been used in LTE and wireless systems and have been modelled as a part of this study.

Once such mechanism is known as Forward Error Correction (FEC) where a few redundant bits are added to each data unit sent. FEC allows the receiver to correct the errors received over the air-interface without the need to retransmission. Another method is to retransmit the failed messages, known as Automatic Repeat Request (ARQ). In LTE, a combination of these are used to transmit messages over the air known as Hybrid Automatic Repeat Request (HARQ). However, there can be certain problems associated with using HARQ. The first is the redundant bits may not be sufficient and the second the ARQ process might induce higher latencies. Therefore, an optimal set of parameters are required to adjust the amount of FEC bits are added as well as the number of times the packets are re-transmitted before being dropped [74].

Figure 3-16 summarises the HARQ operation. The transport block generated at the RLC layer further undergoes a FEC overhead in the MAC layer before being transported over the air interface. If the TBS is dropped when being transported over the air interface, then the RLC retransmits the transport block.
The packet error ratio (PER) can be calculated using (13).

\[
PER = 1 - (1 - BER)^n
\]

where, \( n \) is the number of bits that the UE receives for the transport block. Note by packet, it is implied the MAC layer Protocol Data Unit (PDU). BER is the Bit Error Rate for the chosen modulation scheme.

The PER stated in (13) is the raw packet error rate. However as stated previously, LTE uses FEC techniques to reduce the packet error. Following the analysis performed in [74], the packet error with FEC is given by (14).

\[
PER_{FEC} = \left( \sum_{i=n-k+1}^{n} \binom{n}{i} PER^i (1 - PER)^{n-i} \right)
\]

where, \( k \) is the number of units for data transmission, \( n-k \) is the redundant units added to the frame being transported and \( k/n \) is the rate.

If a PDU is dropped, then it is retransmitted. Assuming that each retransmission is independent of each other, the net packet drop with ARQ is given by (15).

\[
PER_{ARQ} = 1 - (1 - PER_{FEC})^X
\]

where \( X \) is the number of retransmissions per TTI which is set to 4 in this model. After a maximum of 4 re-transmission a packet drop is induced.

### 3.5.3 Physical Layer Parameter Settings

The preceding sections provided a detailed discussion on the physical layer aspects of the modelling. Table 3-9 provides the relevant LTE physical layer parameter settings that were used when modelling the physical layer.
### Table 3-9: Physical Layer Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spatial Multiplexing</td>
<td>2</td>
</tr>
<tr>
<td>Transmission Method</td>
<td>FDD</td>
</tr>
<tr>
<td>Band</td>
<td>32</td>
</tr>
<tr>
<td>Downlink Transmission Frequency</td>
<td>1.475 GHz</td>
</tr>
<tr>
<td>Channel Bandwidth</td>
<td>20 MHz</td>
</tr>
<tr>
<td>Number of Resource Blocks per TTI</td>
<td>100</td>
</tr>
<tr>
<td>Maximum Modulation</td>
<td>256 QAM</td>
</tr>
<tr>
<td>TTI Interval Duration</td>
<td>1 milli-seconds</td>
</tr>
<tr>
<td>Maximum number of retransmissions</td>
<td>4</td>
</tr>
</tbody>
</table>

---

3.6 Simulated End-to-End Wireless Network Design

As explained in the introductory section, the execution of the MATLAB tool relies on a three-variable input, namely the number of users, the base station controller type and the runtime (specified in seconds). The outputs generated form this model and the base station controller type will be discussed in subsequent chapters. In this section, the working process/logic of the model will be described.

The behaviour of the transport layer with the intricacies of the TCP protocol were discussed in Section 3.2. Figure 3-17 below shows the logical flow for a single connection (i.e. between one source and the corresponding UE). Following the establishment of the TCP connection and subsequent slow start phase, the TCP packets are generated.

A sub-process is shown in the form of the network layer. The TCP algorithm will use the information provided from the network layer to determine if there was a TCP timeout or congestion. Additionally, latency information provided by the network layer will be used to determine the change in window size. The run-time is determined by the number of RTTs, i.e., the amount of time the packet has been in the network. Until the total sum of RTTs exceeds the run-time, packets are re-generated. Once the total sum of RTTs exceed the run-time, the program stops.
Figure 3-17: Transport Layer Flow Diagram
Figure 3-18 shows the sub-process of the network layer. From the simulation perspective, the network layer is for measuring the amount of latency due to packet processing and multiplexing as was discussed in the section detailing the network layer. The sub-process within the network layer is the data link layer. The delays in the data link layer also factored in. The network layer reports on the latency and congestion experienced in the network. This information is required by the TCP algorithm.

![Network Layer Flow Diagram](image)

The final diagram in Figure 3-19 shows the Data Link Layer and the sub-processes in the physical layer. For the purposes of modelling, the unit of measurement for processing in the base station is in bits taken from the layer 2 frames. The time measurement is in TTIs. This subsection of the simulation ends when there is congestion detected in the base station buffer or the number of HARQ re-transmits exceeds the allocated number or all the bits in the base station buffer are scheduled. The success of packet transfer (congestion) and the amount of time the layer 2 frames spend in the base station buffer are reported back to the network layer.
Figure 3-19: Data Link and Physical Layer Flow Diagram
3.7 Chapter Summary

In this chapter, the modelling has been explained. Modelling is crucial since it provides a baseline for testing the wireless congestion controllers which is the scope for the thesis. To summarise, the model executes the program based on user-generated inputs of the number of sources specified, the type of scheduler the base station will employ and the duration of the program execution (in seconds). Following which, the model will simulate an end to end wireless network, which entails simulating the TCP behaviour for a source, the source generating packets, packet processing in a simulated network and scheduling over the wireless interface. The radio conditions, Maximum Transmission Unit (MTU), antenna spatial multiplexing configuration, LTE radio band, radio antenna transmit power, base station gain, and UE gain have been pre-configured. The outputs generated are the throughput (in bits/s), latency (in seconds), percentage of packets retransmitted (in %) and average window size (in packets).

A sample output has been shown in Table 3-10 where the following parameters are specified:

- **Maximum Window Size:** The maximum allowable window size for the user
- **Average Window Size:** The average window size reported through the run-time
- **Median bitrate:** The median bitrate measured through the run-time
- **Retransmission Percentage:** The percentage of packets that have to be re-transmitted
- **Throughput:** The throughput measured during the run-time
- **Latency:** The average amount of latency experienced by the packets
- **Scheduler:** The congestion controller implemented

<table>
<thead>
<tr>
<th>User</th>
<th>Link Capacity (packets)</th>
<th>Average Window Size (packets)</th>
<th>Median Bitrate (bits/s)</th>
<th>Retransmission Percentage (%)</th>
<th>Throughput (bits/s)</th>
<th>Latency (s)</th>
<th>Scheduler</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>46</td>
<td>41</td>
<td>3.00E+06</td>
<td>0</td>
<td>3.14E+06</td>
<td>0.1225</td>
<td>Discard Timer</td>
</tr>
<tr>
<td>2</td>
<td>36</td>
<td>23</td>
<td>2.00E+06</td>
<td>0</td>
<td>2.76E+06</td>
<td>0.11174</td>
<td>Discard Timer</td>
</tr>
<tr>
<td>3</td>
<td>40</td>
<td>29</td>
<td>1.00E+06</td>
<td>0</td>
<td>3.03E+06</td>
<td>0.11933</td>
<td>Discard Timer</td>
</tr>
<tr>
<td>4</td>
<td>44</td>
<td>38</td>
<td>3.00E+06</td>
<td>0</td>
<td>3.09E+06</td>
<td>0.11935</td>
<td>Discard Timer</td>
</tr>
<tr>
<td>5</td>
<td>41</td>
<td>32</td>
<td>3.00E+06</td>
<td>0</td>
<td>3.08E+06</td>
<td>0.12019</td>
<td>Discard Timer</td>
</tr>
<tr>
<td>6</td>
<td>44</td>
<td>38</td>
<td>3.00E+06</td>
<td>0</td>
<td>3.11E+06</td>
<td>0.12015</td>
<td>Discard Timer</td>
</tr>
</tbody>
</table>

Table 3-10: Sample Output for 10 Users
A sample output for a 15-user scenario is displayed in Table 3-11.

<table>
<thead>
<tr>
<th>User</th>
<th>Link Capacity (packets)</th>
<th>Average Window Size (packets)</th>
<th>Median Bitrate (bits/s)</th>
<th>Retransmission Percentage (%)</th>
<th>Throughput (bits/s)</th>
<th>Latency (s)</th>
<th>Scheduler</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>40</td>
<td>23</td>
<td>1.00E+06</td>
<td>18.666</td>
<td>1.66E+06</td>
<td>0.11573</td>
<td>AQM-RED</td>
</tr>
<tr>
<td>2</td>
<td>46</td>
<td>32</td>
<td>2.00E+06</td>
<td>34.956</td>
<td>1.51E+06</td>
<td>0.10073</td>
<td>AQM-RED</td>
</tr>
<tr>
<td>3</td>
<td>45</td>
<td>26</td>
<td>1.00E+06</td>
<td>17.48</td>
<td>1.75E+06</td>
<td>0.11908</td>
<td>AQM-RED</td>
</tr>
<tr>
<td>4</td>
<td>41</td>
<td>35</td>
<td>2.00E+06</td>
<td>17.864</td>
<td>1.82E+06</td>
<td>0.11914</td>
<td>AQM-RED</td>
</tr>
<tr>
<td>5</td>
<td>39</td>
<td>32</td>
<td>1.00E+06</td>
<td>0</td>
<td>2.14E+06</td>
<td>0.1359</td>
<td>AQM-RED</td>
</tr>
<tr>
<td>6</td>
<td>48</td>
<td>38</td>
<td>2.00E+06</td>
<td>0</td>
<td>2.18E+06</td>
<td>0.14033</td>
<td>AQM-RED</td>
</tr>
<tr>
<td>7</td>
<td>39</td>
<td>23</td>
<td>1.00E+06</td>
<td>18.772</td>
<td>1.66E+06</td>
<td>0.11872</td>
<td>AQM-RED</td>
</tr>
<tr>
<td>8</td>
<td>41</td>
<td>38</td>
<td>2.00E+06</td>
<td>0</td>
<td>2.12E+06</td>
<td>0.14025</td>
<td>AQM-RED</td>
</tr>
<tr>
<td>9</td>
<td>36</td>
<td>26</td>
<td>1.00E+06</td>
<td>0</td>
<td>1.89E+06</td>
<td>0.13304</td>
<td>AQM-RED</td>
</tr>
<tr>
<td>10</td>
<td>42</td>
<td>37</td>
<td>2.00E+06</td>
<td>0</td>
<td>2.03E+06</td>
<td>0.14005</td>
<td>AQM-RED</td>
</tr>
<tr>
<td>11</td>
<td>57</td>
<td>11</td>
<td>1.00E+06</td>
<td>66.883</td>
<td>7.51E+05</td>
<td>0.066488</td>
<td>AQM-RED</td>
</tr>
<tr>
<td>12</td>
<td>40</td>
<td>37</td>
<td>2.00E+06</td>
<td>0</td>
<td>2.06E+06</td>
<td>0.14316</td>
<td>AQM-RED</td>
</tr>
<tr>
<td>13</td>
<td>42</td>
<td>37</td>
<td>2.00E+06</td>
<td>0</td>
<td>2.02E+06</td>
<td>0.14403</td>
<td>AQM-RED</td>
</tr>
<tr>
<td>14</td>
<td>43</td>
<td>38</td>
<td>2.00E+06</td>
<td>0</td>
<td>2.11E+06</td>
<td>0.14256</td>
<td>AQM-RED</td>
</tr>
<tr>
<td>15</td>
<td>34</td>
<td>22</td>
<td>1.00E+06</td>
<td>0</td>
<td>1.96E+06</td>
<td>0.13666</td>
<td>AQM-RED</td>
</tr>
</tbody>
</table>

It should be noted from Table 3-11 that the retransmission percentage is primary due to AQM based congestion drops. In subsequent chapters the reasoning behind this as well as how different schedulers impact the retransmission percentage with respect to the number of users and throughput will be explained.
Chapter 4: Base Station Congestion Mitigations Models based on Queue Management

The base station congestion mechanisms designed for this thesis are based on two different mathematical frameworks. The first technique uses queue management mechanisms to mitigate congestion in the base stations. The second technique uses the principles of non-cooperative game theory for congestion mitigation. Figure 4-1 shows the two different techniques. In this chapter, congestion mitigation mechanisms designed based on managing the base station queue have been discussed.

As shown in Figure 4-1, three different models, the Discard Timer Model, the Random Early Detection (RED) Controller Model and the Proportional Controller Model were designed. The Discard Timer model is a standardised LTE based implementation [10] whereby all packets which occupy the buffer longer than a given configurable interval are dropped. The other mechanism known as Random Early Detection (RED) is primarily used for congestion mitigation in routers [75]. RED has the ability to discard packets probabilistically based on the buffer size. In this chapter, the principles of control theory have been applied for congestion mitigation to develop two controllers, namely, RED and the Proportional controller. Both these congestion mitigation mechanisms use the base station buffer size as a reference for packet drop. The key difference between RED and Proportional Controller is that the latter factors in the variation of radio conditions in estimating the proportional value used to determine the probability for a packet drop.

It is important to note that these base station congestion mitigation models only manage congestion but do not provide a mechanism for radio resource allocation. The underlying mechanism
for radio resource allocation used is the proportional fair scheduler. The implementation of the proportional fair scheduler will be the first section (Section 4.1) of discussion in this chapter.

4.1 Proportional Fair Scheduler

In a wireless environment the proportional fair algorithm provides a method of distributing radio resources (Physical Resource Blocks) in an equitable manner (as opposed to equal). This implies that all users are guaranteed the allocation of the same or nearly the same number of bits per TTI despite being in varying conditions (such as distance from base station or radio conditions). This makes it an ideal scheduler for UEs which are in low mobility and fixed scenarios.

Telecommunication vendors have their proprietary method of implementing the proportional fair scheduler. In this study, a unique method of the scheduler has been devised. This section also describes the framework for a proportional fair scheduler and how these principles have been applied to design an LTE base station model for resource block allocation.

4.1.1 Proportional Fair Model

The concept of utility functions is used for the mathematical formulation of a proportional fair scheduler. The utility function is known as a measure of the satisfaction a user gets from consuming network resources, in this case radio resources. The utility function (denoted by $U$) depends on various parameters such as congestion, latency and the type of application being used. The utility is a function of $x$, where $x$ is the resources allocated to the user per TTI.

In order to increase their utility, a user needs to pay a certain price for the service (denoted by $\omega$). The user will try to maximise the utility while minimizing the price it pays. This is denoted by $\max\left[U_i(x_i) - \omega_i\right]$.

In the context of LTE resource allocation, $U_i$ is the utility of the $i^{th}$ user for being allocated $x_i$ resource blocks allocated in the given TTI. For this, user $i$ has to pay a price per TTI which is denoted by $\omega_i$ and $\omega_i \geq 0$. Note, in this context price does not have monetary significance but is used as an indicator required for resource allocation. The utility function and price metric are used to develop the proportional fair scheduling mechanism.

To maximize the utility amongst the users, the base station will have the maximization problem $\sum_{i=1}^{N} U_i(x_i)$ for $N$ number of users with condition $\sum_{i=1}^{N} x_i \leq R$ [76] where $R$ is the maximum number of PRBs available per TTI. In order for the utility function to maximize proportional fairness in the network it will need to be as show in (1) [77].

$$U_i(x_i) = \omega_i \log x_i$$

(1)
Therefore, solving \( \max \left( \sum_{i=1}^{N} \omega_i \log x_i \right) \) subject to \( \sum_{i=1}^{N} x_i \leq R \) maximizes proportional fairness in the network. This optimization problem can be solved using the method of Lagrange multipliers. The Lagrangian is given by (2). \( \lambda \) is used to denote the rate of change of the parameter being optimized, in this case \( x \).

\[
L(x, \lambda) = \sum_{i=1}^{N} \omega_i \log x_i + \lambda (R - \sum_{i=1}^{N} x_i)
\]

(2)

To get the optimal values of (2), the differential of the Lagrangian equation is set to zero as shown in (3).

\[
\nabla_{x, \lambda} L(x, \lambda) = 0
\]

(3)

The solution for (3) is given in (4) and (5).

\[
\sum_{i=1}^{N} x_i = R
\]

(4)

\[
\lambda = \frac{\sum_{i=1}^{N} \omega_i}{\sum_{i=1}^{N} x_i} = \frac{\sum_{i=1}^{N} \omega_i}{R}
\]

(5)

From (4) and (5) it is clear that \( \omega_i \) will be a function of \( x_i \), and is determined by the CQI. Therefore, \( \omega_i \) will be a function of the CQI parameters for all the users being served by the base station. Hence \( \omega_i \) will also be linked with the reported radio conditions of the UEs. Since \( \omega_i \) is a price metric, user \( i \) will be charged with respect to all other uses being served by the base station. Also, as per the proportional controller implementation, all users will be provided an equitable share of radio resources irrespective of the CQI value reported to the base station. Therefore, users with poor radio conditions (lower CQI) will have to pay a greater price in order to match the service quality of users with good radio conditions (higher CQI).

Define \( \omega_i \) for user \( i \) as shown in (6) where for \( N \) users with a range of CQI given by \( \{\text{CQI}_1, \text{CQI}_2, \ldots, \text{CQI}_{\text{fu}}\} \).

\[
\omega_i = \min\left(\frac{\text{CQI}_{i[v]}}{\text{CQI}_i}\right)
\]

(6)

Equation (6) shows that the users with the lowest CQI pay the highest price. The prices for other users are set relative to the minimum CQI value. From (5) and (6), for each user the relation between the resource allocated \( (x_i) \) and price \( (\omega_i) \) is shown in (7). Since \( x_i \) is a positive integer value, the floor values are used.
4.1.2 Simulated Results for Proportional Fair Model

The simulated results shown in this section are from the base station model where the TBS is displayed as an output. Table 4-1 shows a sample of measurements taken in a single TTI for six different users being served by the same base station. The algorithm notes the CQI and derives the price metric for the users as shown in the third row. After obtaining the price metric, the number of resource blocks allocated for each user is determined. The TBS is calculated accordingly.

Table 4-1: CQI and TBS Allocation for Proportional Fair Scheduler

<table>
<thead>
<tr>
<th>Parameters</th>
<th>User 1</th>
<th>User 2</th>
<th>User 3</th>
<th>User 4</th>
<th>User 5</th>
<th>User 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>CQI</td>
<td>15</td>
<td>13</td>
<td>7</td>
<td>10</td>
<td>6</td>
<td>12</td>
</tr>
<tr>
<td>Price Metric</td>
<td>0.4000</td>
<td>0.4615</td>
<td>0.8571</td>
<td>0.6000</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td>Resource Blocks Allocated</td>
<td>10</td>
<td>12</td>
<td>22</td>
<td>15</td>
<td>26</td>
<td>13</td>
</tr>
<tr>
<td>TBS (in bits)</td>
<td>19848</td>
<td>19848</td>
<td>22152</td>
<td>19080</td>
<td>22152</td>
<td>20616</td>
</tr>
</tbody>
</table>

Although the TBS indicates that the allocation is all not equal in value, they can be considered close enough to each other in value hence signifying the implementation of the proportional fair scheduler.

4.2 Congestion Mitigation using Discard Timer

In this section, the standardized LTE method for congestion mitigation has been discussed. LTE base stations use a latency-based discard timer [10] whereby all packets are dropped which occupy the buffer longer than a given configurable interval. There are various configurable intervals that can be set. The Discard Timer can also be disabled.
4.2.1 Discard Timer Model

From Chapter 3, the LTE protocol stack demonstrated the functionality of the PDCP buffer in the eNodeB. For each Radio Access Bearer (RAB) connection to a UE there is a dedicated PDCP buffer. Note that in practice the PDCP buffer this is irrespective of the number of applications that the user may be concurrently using. However, in this model the user does not have multiple concurrent applications. If the discard timer threshold is breached for the particular PDCP, all the packets for the RAB would be flushed out. For a TCP based application if the discard timer is triggered and a packet drop induced, TCP would reduce the packet send rate by adjusting the window size.

Suppose that the Discard Timer is set to 100 milliseconds. If the buffer occupancy of any of the IP packet data units in the PDCP lower than 100 milliseconds, then the Discard Time does not take effect. This is shown in Figure 4-2. If the buffer occupancy of any IP packet data unit is greater than 100 milliseconds, then all the packets in the PDCP buffer is flushed out as highlighted in Figure 4-3.

![Figure 4-2: Buffer Occupancy lower than Discard Timer Threshold](image1)

![Figure 4-3: Buffer Occupancy higher than Discard Timer Threshold](image2)

In this study, the discard timer threshold is set to 100 milliseconds.

4.3 Congestion Mitigation using Control Theory

Two congestion controllers have been designed based on control theory and have been applied to an LTE environment. Both these controllers implement AQM by using a probabilistic packet discard mechanism. As opposed to tail-drop mechanisms where all packets are dropped over a certain buffer threshold, the probabilistic packet discard mechanism drop packets probabilistically between certain thresholds. This provides a more gradual approach of reducing congestion for TCP links as opposed to Discard Timer.
There are two controllers that have been discussed in this section, the RED and the proportional controller. Though similar in principle, the RED controller does not factor in variations in the base station capacity while the proportional controller does.

4.3.1 Random Early Detection

The probabilistic approach to drop packets is known as Random Early Detection [78]. As discussed in Chapter 3, the base station buffer is partitioned for each Radio Bearer which has a dedicated PDCP buffer in the base station. In order to implement RED, the base station buffer is considered as a whole instead of individual PDCP buffers reserved for UEs. Between certain thresholds of buffer occupancy for the base station buffer packets are dropped randomly. The probability of this random drop increases as the buffer occupancy of all the packets in the base station increases. If a packet has to be dropped, then the link that is consuming the higher buffer occupancy of the base station will be removed from the base station buffer. However, if the base station buffer exceeds the upper limit of the threshold then all the packets in the base station buffer will be discarded irrespective of the PDCP buffer occupancy levels for each RAB.

The equation for RED is shown in (8). If the buffer occupancy $b$ in the base station buffer is below a certain limit, then there is no packet drop. This limit has been represented as $B_{\text{min}}$. The probability of randomly dropping packets in the base station increases when the buffer occupancy is between $B_{\text{min}}$ and $B_{\text{max}}$. The probability of dropping all the packets in the buffer becomes 1 if the buffer occupancy exceeds the threshold. The implementation of RED has been represented pictorially in Figure 4-4.

\[
p_{\text{RED}} = \begin{cases} 
0 & \text{if } b < B_{\text{min}} \\
K(b-B_{\text{min}}) & \text{if } B_{\text{min}} < b < B_{\text{max}} \\
1 & \text{if } b > B_{\text{max}} 
\end{cases} \tag{8}
\]

![Figure 4-4: AQM - RED Packet Probability Drop Behaviour](image-url)
To implement RED in this study, the assessment for base station occupancy levels are made every TTI. The advantage of this arrangement is the base station buffer status is constantly monitored. If the base station buffer is within $B_{\text{min}}$ and $B_{\text{max}}$ and a packet drop takes place, then the probability of a packet drop happening is reduced in the next TTI. However, if the subsequent TTI admits a larger number of packets then the probability is increased again.

4.3.2 Proportional Controller Model

TCP has its own congestion control mechanism which depends upon the successful acknowledgement of packets. Based on this the window size is updated every RTT. The congestion mechanism of TCP can be mathematically modelled using the fluid flow model [13] which for TCP traffic is based on the AIMD behaviour. The fluid flow model is a set of differential equations that are used to describe the behaviour of the evolution of the TCP window size and the base station buffer size.

Assuming the number of links between the source and the UEs do not vary, the expected TCP window size and the buffer occupancy are given by (9) and (10) respectively [13] [79]. These equations are non-linear and capture the dynamic model of TCP behaviour.

\[
\frac{dW}{dt} = \frac{1}{\tau(t)} - \frac{W(t)W(t-\tau(t))}{2\tau(t-\tau(t))} - p(t-\tau(t)) \tag{9}
\]

\[
\frac{db}{dt} = \frac{W(t)N(t)}{\tau(t)} - C(t) \tag{10}
\]

Where,

$W$ - Expected TCP window size

$b$ - Buffer occupancy at the base station

$\tau$ - Round Trip Time

$C$ - Radio interface capacity

$N$ - Number of TCP links in the base station

$p$ - Probability of packet drop/congestion measure

As shown in Figure 4-5, $N$ TCP sources enter the network. The base station serves as the congestion point. The base station has a buffer limit represented by $b$ and capacity $C$. The congestion measure for each UE denoted by $p$ is sent via the backhaul in the form of AQM induced packet drops.
In studies such as [80], the fluid flow model has been adjusted for a wireless network. Additional variables have been introduced to cater for probability of packet drop over the downlink radio channel and the time taken for the effects of the packet drop to take place. However, in this thesis, these are assumed to be negligible and have not been factored in any calculations.

For further analysis of (9) and (10), the equations must be linearized around an operating point $x = x_0 + \Delta x$. The variables $W$, $b$, $C$ and $p$ will be linearized about their equilibrium points ($W_0$, $B_0$, $C_0$ and $P_0$). Using the perturbation formula, the linearized equations for (9) and (10) can be obtained which are (11) and (12) respectively.

$$\frac{dW_0(t)}{dt} = -\frac{\beta}{\tau_0} W_0(t) - \frac{\alpha}{\tau_0} p_0(t - \tau_0)$$ (11)

$$\frac{db_0(t)}{dt} = \frac{N}{\tau_0} W_0(t) - \frac{1}{\tau_0} b_0(t)$$ (12)
Two constants have been introduced in (11), \( \alpha \) and \( \beta \). Based on the model designed, the values of \( \alpha \) and \( \beta \) selected are
\[
\alpha = \frac{(10^{-3})^2}{N^2} \quad \text{and} \quad \beta = N / 10^{-3} C.
\]
\( C \) is the radio link capacity for all the users in the given TTI. Therefore, the values of \( \alpha \) and \( \beta \) will be influenced by the number of links in the system \( N \) and the radio capacity \( C \).

Using Laplace transforms on (11) and (12) the transfer functions obtained are shown in (13) and (14). The TCP dynamics is represented by \( G_{\text{tcp}} \) and the base station queue dynamics is represented by \( G_{\text{queue}} \).

\[
G_{\text{tcp}}(s) = \frac{W_0(s)}{P_0(s)} = \frac{\alpha/\tau_0}{s + \beta/\tau_0}
\]
(13)

\[
G_{\text{queue}}(s) = \frac{b_0(s)}{W_0(s)} = \frac{N_0/\tau_0}{s + 1/\tau_0}
\]
(14)

Define \( G(s) \) with the equation shown in (15) with \( e^{-\tau_0 s} \) representing the delay parameter.

\[
G(s) = G_{\text{tcp}}(s)G_{\text{queue}}(s)e^{-\tau_0 s}
\]
(15)

Replacing (15) by Padé’s approximation, (16) is obtained.

\[
G(s) = \frac{N_0\alpha/\tau_0^2}{(s + \beta/\tau_0)(s + 1/\tau_0)}
\]
(16)

On expanding (16), (17) is obtained.

\[
G(s) = \frac{N_0(2 - \tau_0 s)\alpha/\tau_0^2}{\tau_0 s^3 + (\beta + 3)s^2 + \left(\frac{3\beta + 2}{\tau_0}\right)s + \frac{2\beta}{\tau_0^2}}
\]
(17)
With the transfer functions now established, a control feedback loop can be derived. Figure 4-6 shows the feedback loop. Note that the radio link capacity ($C(s)$) is represented in the form of a disturbance that influences the buffer size. Based on the outcome, the controller (in the base station) sets a certain probability packet drop or congestion measure (represented by $p_0$). This congestion measure takes effect after a delay or the round-trip time and influences the TCP and queue dynamics modelling.

![Figure 4-6: Feedback Loop for Proportional Controller](image)

The AQM controller is a proportional controller ($K_p$) with the transfer function represented by $G_C(s)$. The open loop transfer function for the feedback loop is $G_{open}(s) = K_p G(s)$.

$$G_{open}(s) = K_p \frac{\left( \frac{N_o \alpha}{\tau_0^2} \right) \left( 2 - s \tau_0 \right)}{\tau_0 s^3 + (3 + \beta) s^2 + \left( \frac{3 \beta + 2}{\tau_0} \right) s + \frac{2 \beta}{\tau_0^2}} \quad (18)$$

Define $\gamma = \left( \frac{N_o \alpha}{\tau_0^2} \right)$. The characteristic equation for (18) is given by (19).

$$1 + K_p G(s) = \tau_0 s^3 + (\beta + 3) s^2 + \left( -K_p \gamma \tau_0 + \left( \frac{3 \beta + 2}{\tau_0} \right) \right) s + \left( 2 \gamma K_p + \frac{2 \beta}{\tau_0^2} \right) = 0 \quad (19)$$

A necessary condition for the closed-loop system stability is that all the coefficients must be strictly positive. Thus, the $K_p$ value should be as shown in (20).

$$K_p < \frac{3 \beta + 2}{\gamma \tau_0^2} = \frac{3 \beta + 2}{N_o \alpha} \quad (20)$$
For a sufficient condition for stability, the Routh-Hurwitz criterion needs to be employed. This has been shown below,

\[
\begin{array}{ccc}
   s^3 & \tau_0 & \left(-K_p\gamma\tau_0 + \frac{3\beta + 2}{\tau_0}\right) \\
   s^2 & (\beta + 3) & \left(2\gamma K_p + \frac{2\beta}{\tau_0^2}\right) \\
   s^1 & \lambda & \left(2\gamma K_p + \frac{2\beta}{\tau_0^2}\right) \\
   s^0 & & \\
\end{array}
\]

where,

\[
\lambda = \frac{(\beta + 3)\left(\frac{3\beta + 2}{\tau_0} - K_p\gamma\tau_0\right) - \tau_0\left(\frac{3\beta + 2}{\tau_0} - K_p\gamma\tau_0\right)}{(\beta + 3)} \tag{21}
\]

Based on the Routh-Hurwitz criterion, all the variables in the first column need to be positive. Since \( \lambda > 0 \), solving (21), (22) is obtained.

\[
K_p < \frac{(\beta + 3)(3\beta + 2) - 2\beta}{N_0\alpha(\beta + 3) + 2N_0\alpha} \tag{22}
\]

Define \( K_p \) as (23). A gain parameter of 10 has been included in (23) to calibrate the probability of packet drop (after some trial and error). (23) satisfies the condition stated in (20) and (22).

\[
K_p = \frac{10\beta}{N_0\alpha} \tag{23}
\]

4.4 Results and Discussion

In this section the performance of DT, RED and Proportional controllers are compared. In order to evaluate the merits of each controller a proper evaluation method is required. The key methodology for evaluation has been provided below:

1. A simulation run consists of setting a fixed number of users in a given radio condition with a given controller model (for example, DT or RED or Proportional) over a period of 1 second.
2. Using the same number of users and fixed radio conditions the simulation run was repeated for the rest of the controllers.
3. The simulation run was repeated 50 times for each of the controllers.
4. The average values and standard deviation of the throughput, latency, window size and percentage of TCP packets retransmitted for all the links were calculated for each controller. Under the same radio conditions, these steps were repeated for a network with 10, 20, 30, 40 and 50 users.

There were three different scenarios for radio conditions as listed below:

- Good radio conditions: path loss exponent = 2.2 and shadowing standard deviation = 2 dB
- Average radio conditions: path loss exponent = 2.3 and shadowing standard deviation = 4 dB
- Poor radio conditions: path loss exponent = 2.8 and shadowing standard deviation = 7 dB

The results for each of the controllers have been represented as bar plots for comparison in the following sections. The performance reporting metrics have been defined in Table 1-2.

4.4.1 Good Radio Conditions

The results for good radio conditions have been presented in this section. The x-axis in Figure 4-7 (a) – (d) represents the number of users and the y-axis represents the throughput (Mbps), latency (ms), window size (number of packets) and percentage of packet retransmission (%) for Figure 4-7 (a), (b), (c) and (d) respectively.

When the base station faces low congestion with 10 and 20 users, it is observed that the throughputs for all the three controllers have a similar mean value and standard deviation. With 20 users, the Proportional controller has lower window size, but lower latency therefore is able to have near equivalent throughput with the other models.

When the base station is serving 30 users, RED demonstrates a substantial increase in the percentage of packets being retransmitted. This demonstrates that a sudden burst of traffic (with number of users increased from 20 to 30) the RED controller is not able to cope and has drop packets to reduce congestion. This results in lower throughput for RED as compared to DT and Proportional controller. Consequently, the latencies and the average window sizes of all the links decrease. The Proportional controller has the highest throughput when there are 30 users and the standard deviation for percentage of packets retransmitted is close to zero.

When there are 40 and 50 users, the throughput values for all the controllers plummet. The latency values for the DT reach its threshold of 100 ms and the congestion mitigation action begins to take effect. This is evident with the higher packet retransmission percentages. For 50 users, the average throughput is below 1 Mbps therefore unusable. The RED and Proportional controllers experience congestion collapse. This is evident from the near 100% packet retransmission percentages and close to 0 Mbps of average throughput. This suggests that all packets are discarded as soon as they queue in the base station buffer due to breaching the congestion thresholds set by the RED and
Proportional controllers. The latency values decrease due to the negative acknowledgement (NACK) sent to the TCP sources by the base station as soon as the packets are discarded. Subsequent transmissions cause the base station to receive all packets simultaneously and also breach the threshold simultaneously. This phenomenon of TCP global synchronisation is observed.

4.4.2 Average Radio Conditions

The results for the average radio conditions have been shown in Figure 4-8. Due to the lower quality of radio conditions for all the controllers it is expected that the average throughput values should be lower, and the latency values should be higher as compared good radio conditions. This is evident when comparing Figure 4-7 and Figure 4-8. However, the overall trend of the results is similar.

The average throughput values for 10 and 20 users are similar for all the controllers. It should be noted that the average latency has not breached the threshold for DT. Therefore, DT does not have to induce packet drop for congestion control. This is evident from the retransmission percentages. The RED and Proportional controllers perform some congestion mitigation as evident from the retransmission percentage of packets. The values of the latencies are lower than DT, but the throughput values are comparable.
For 30 users the DT and RED controllers become unstable as determined by the high variation in the window size and retransmission percentage values. The Proportional controller exhibits its robustness by maintaining consistent retransmission percentages but higher throughput.

When there are 40 and 50 users (high congestion) all the throughput for all the controllers deteriorate. This is similar to the behaviour with good radio conditions. RED and Proportional controllers face congestion collapse. DT experiences significantly degraded throughput due to excessive queuing in the PDCP buffers.

![Figure 4-8: Average Radio Conditions](image)

4.4.3 Poor Radio Conditions

The final scenario shows the results for poor radio conditions. Due to poorer radio conditions, the outcomes for this scenario are lower throughputs and higher latencies as verified from Figure 4-9. Due to increased latency it is also expected that the DT does not perform as well compared to the other radio conditions. This can be verified from the results in Figure 4-9 (a). When the number of users is 20, the throughput results for the DT are lower than RED or Proportional. In addition, the system is not stable since the percentage of retransmissions has a high standard deviation. It should also be noted that unlike DT, RED and Proportional controller do not have any threshold on the amount of time packets spend in the queue and therefore are able to exceed the 100 ms threshold.
With further increase in the number of users to 40 and 50, the RED and Proportional controller are not able to cope since all the TCP links face congestion collapse.

**Figure 4-9: Poor Radio Conditions**

### 4.5 Conclusion

The purpose of increasing the number of users was to increase the load in the base station buffer. When there are 10 or 20 users then the base station does not experience any congestion for any of the different radio scenarios. Depending upon the radio conditions, the average throughputs are similar amongst all the congestion controllers. The only anomaly to this is when there are 20 users in poor radio conditions. The average throughput from the Discard Timer is close to 1 Mbps lower comparatively. This is due to the mechanism of DT. Therefore, with varying radio conditions DT is not the most optimal controller in terms of performance.

This chapter has shown that analysing congestion mitigation mechanisms in a wireless network model by factoring the variable radio capacity as an input parameter in the form of a disturbance is possible. This implies that unlike RED, a variable proportionality factor is used for inducing packet drop each TTI. In a given scenario, $K_p$ increases inducing more packet drops when majority of the users are in poor radio conditions. RED congestion control starts becoming inconsistent for all the radio scenarios when there are 30 users in the buffer as observed by the variance in throughput and percentage of retransmissions from Figure 4-7, Figure 4-8 and Figure 4-9. However, only in one scenario does the Proportional controller provide similar inconsistent results (poor radio
conditions, Figure 4-9). This demonstrates that the parameter tuning for the gain parameter is more accurate for Proportional controller than RED.

On comparing the throughput values between the Discard Timer and Proportional controller when the number of users is 30 from Figure 4-7, Figure 4-8 and Figure 4-9, it is observed that the throughput values for the Discard Timer is lower but the latency values are higher. This demonstrates that higher latency is detrimental to throughput.

When the congestion in the base station is high (40 and 50 users), RED and Proportional controller experience congestion collapse. The maximum threshold of the queue size in the base station buffer is breached and all the incoming packets for all the links are dropped.

The limitations of using queue-based congestion management is that careful parameter tuning is required depending upon the network setting. If the network conditions sway, then performance is affected as seen in the case of DT for 20 users in different radio conditions. Therefore, queue-based congestion management need to be selected carefully and the network conditions need to be analysed before setting the parameters.

4.6 Chapter Summary

In this chapter, the radio resource allocation method using proportional fair was presented and the radio resource allocation methods were shown.

The three different congestion mitigation mechanisms based on queue management were presented. The Discard Timer model implementation is used in LTE base stations as per the 3GPP standards. RED was implemented for an LTE base station. Congestion control models based on fluid flow approximation and control theory like RED have a limitation whereby the variable parameters like radio capacity are not factored. To overcome this, the radio capacity has been modelled as a disturbance in the network feedback loop and a proportional control for managing congestion was developed.

The results of the three controllers were compared under varied radio conditions. Key findings showed that congestion mitigation for queue-based congestion management systems work well in certain operating regions. Below or beyond these regions the models will require parameter adjustment since they experience congestion collapse.

To maintain stable operability and not experience congestion collapse, game theoretic models have been proposed. These have been discussed in the subsequent chapter.
Chapter 5: Base Station Congestion Mitigation Based on Game Theory

Chapter 4 discussed the principles of base station congestion mitigation by managing the base station queue. In this chapter the congestion mitigation using the principles of non-cooperative game theory will be discussed. As shown in Figure 5-1, there have been two such congestion controller models that have been developed and presented in this thesis namely the Buyer-Seller Model and the Equilibrium Model.

The queue-based congestion mitigation models require parameter configuration for the mechanism to work. However, for game theoretic models, parameter tuning is not required since the approach used for congestion mitigation is fundamentally different to the queue-based mechanisms.

The algorithms proposed here make use of the concept of utility functions. Each TCP stream requires certain standards of user satisfaction to be met, which is mathematically represented as utility functions [81] [82]. Additionally, each user will try to maximize its utility while minimizing the cost i.e. the amount paid per unit time for resources (in this case, radio resources). Cost or unit price is set for each radio resource or Physical Resource Block (PRB) is relative to the number of users requiring a given number of PRBs for their application. The situation where the user maximizes the utility while minimizing the cost is defined as the payoff function. Over a period, each user will try to maximize its payoff that is derived from its payoff function. To achieve this there are certain strategies that can be employed by the user. The goal would be to achieve Nash Equilibrium whereby no user will be able to increase their payoff by deviating from their chosen strategy [83].

![Figure 5-1: Base Station Congestion Controllers](image-url)
In the Buyer-Seller model the underlying resource allocation method for users is the proportional fair method. Every TTI each user will have a certain PRB allocation requirement that will need to be matched. Based on the difference between the PRBs allocated by the proportional fair method and the PRB allocation requirement users can be classified as having an excess or deficit of radio resources. Users then enter a radio resource buying/selling process between each other similar to a buyer and seller in a market. Congestion mitigation takes place when users are unable to secure their required number of resource blocks and a packet drop causes the source to reduce their window size.

In the Equilibrium model each user will have a payoff function. The user has an action path to maximize their payoff function. At each TTI, the user predicts its desired action path to maximize their payoff function given the action paths of the other users. The analysis continues until all the users reach a certain equilibrium for their desired action path. This determines the PRB allocation for the user. Congestion mitigation in the form of AQM induced packet drops occurs when a user is unable to fulfil their resource requirement or under specific conditions.

In both these models there is an underlying radio resource requirement that needs to be fulfilled by each user. This forms the basis of the game models. In this chapter the concepts of perceived throughput and running average have been introduced which are used to determine the radio resource requirement for each user.

5.1 Defining Perceived Throughput and Running Average

This section explains the concepts of perceived throughput and then that of running average. The perceived throughput is a parameter used to gauge the desired rate that the source is able to maintain while sending packets to the user. The running average is then derived from the perceived throughput. The running average is the number of PRBs each user needs to maintain per TTI.

5.1.1 Perceived Throughput

As mentioned in previous chapters, the TCP congestion control algorithm relies on packet drops to reduce the number of packets sent. This is independent of the latencies in the network. Packets which have a large waiting time in the base station buffer will increase in the latency period, but this is not enough to cause a congestion drop if all the packets sent have been successfully acknowledged. The window size would increase albeit reducing the rate (throughput) due to increasing latency.

The throughput value factors in number of packets (or bits generated) and latency. For a link \( i \), the throughput \( \gamma_i \) is given by (1)

\[ \gamma_i = \frac{A_i}{RTT_i} \]  

(1)
where, \( i \) is \( i^{th} \) link with \( 1 \leq i \leq L \) out of a total \( L \) links. \( \gamma_i \) and \( A_i \) are respectively the throughput and number of bits generated for the \( i^{th} \) link. RTTi is the measured RTT for link \( i \).

The perceived throughput values are a discrete set of values. For this study, the range of perceived throughput values for any given link is chosen from the range \( Y = \{1, 2, 3, ..., 16, 18\} \) with the unit as Mbps. The throughput value calculated in (1) is compared with \( Y \). The value higher than \( \gamma_i \) is chosen as the perceived throughput value. The value for the link \( i \) is \( \gamma_{\text{perceived},i} \) and is given by (2).

\[
\gamma_{\text{perceived}} = \lceil \gamma_i \rceil
\]  

(2)

where, \( \gamma_{\text{perceived},i} \in Y \).

By employing perceived throughput in the calculation, the base station will use a dynamic rate based AQM congestion mitigation mechanism rather than using a delay timer or buffer capacity as an indicator. Therefore, the UE should maintain its level of perceived throughput for each interval or have the rate reduced.

### 5.1.2 Running Average

The running average relies on the timing delays in the network. Figure 5-2 shows the timing delays associated with the packets generated traversing through the network. For link \( i \), define the following:

- \( \tau_{\text{source},i} \): The time taken to generate packets and process acknowledgements
- \( \tau_{a,i} \): The propagation delay for the packets to travel downlink and uplink to and from Wireless Gateway
- \( \tau_{\text{PGW},i} \): Wireless Gateway processing and queuing time for downlink and uplink packets
- \( \tau_{b,i} \): The propagation delay for the packets to travel downlink and uplink to and from Base Station
- \( \tau_{\text{eNB},i} \): Queuing delay in the base station (eNodeB)
- \( \tau_{\text{radio},i} \): Radio propagation delay
The total RTT for link $i$ is shown in (3).

$$RTT_i = \tau_{source,i} + \tau_{\alpha} + \tau_{PGW,i} + \tau_{\beta} + \tau_{eNB,i} + \tau_{radio,i}$$

(3)

The delay in the base station ($\tau_{eNB,i}$) is the only delay parameter that the base station has the ability to influence. The network delay can be roughly approximated to be a given parameter defined as $\tau_{delay,i}$. Therefore, the RTT can be re-written as (4),

$$RTT_i = \tau_{eNB,i} + \tau_{delay,i}$$

(4)

On rearranging, the amount of time spent in the base station buffer is given by (5).

$$\tau_{eNB,i} = A_i \gamma_i - \tau_{delay,i}$$

(5)

The base station schedules radio resources every TTI. Define $N$ as (6) where $N \in Z$ and $N > 0$,

$$N = \frac{\tau_{eNB,i}}{TTI}$$

(6)
Define the running average for user \( i \) as shown in (7).

\[
    r_i = \frac{A_i}{N}
\]  

(7)

If the Base Station schedules \( R \) number of PRBs every TTI for all the users (in this thesis, 100 PRBs), then user \( i \) is allocated a certain number of PRBs every TTI out of \( R \) radio resources. Define \( x_{n,i} \) as the number of PRBs that are allocated to the user \( i \) each TTI over a period of \( N_i \) where, \( 0 \leq x_{n,i} \leq R \), \( x_{n,i} \in \mathbb{Z} \). By calculating the average allocation of PRBs the user needs to maintain the running average \( r_i \). Therefore, the it can also be defined as shown in (8).

\[
    r_i = \frac{1}{N_i} \sum_{n=1}^{N_i} x_{n,i}
\]

(8)

This would mean for user \( i \), the running average need not be satisfied every TTI. The running average value can still be achieved as long as the average can be maintained over \( N_i \) TTIs irrespective of the PRBs allocated per TTI. This information is used for developing strategies for resource allocation.

5.2 Game Theoretic Model for Base Station Congestion Mitigation: Buyer and Seller Model

This section describes the buyer and seller model for the base station congestion mitigation mechanism using game theory. In order to proceed into further discussions, it is important to provide the definitions of a Seller and a Buyer. Users are classified as Sellers and Buyers each TTI, each time the game is executed.

- **Seller**: Any user allocated more resources than their respective running average in a given TTI is classified as a Seller.
- **Buyer**: Any user allocated less resources than their respective running average in a given TTI is classified as a Buyer.

Before the buyers and sellers exchange radio resources, all users are allocated radio resources in a proportional fair manner. A proportional fair allocation does not imply that the user’s throughput requirement is satisfied with the share resources it gets. Instead proportional fair resource allocation is done to equitably allocate radio resources amongst users based on the radio conditions of all active users. The objective for the model instead is to satisfy radio resource allocation based on the user’s throughput requirements, which is determined by the user’s perceived throughput. This is done by the base station assigning each seller to share their excess resources with the Buyers at a certain price of the seller’s choosing. The Buyers then use non-cooperative game theory mechanisms to determine the cost per user to purchase those resources sold by the Seller. An example of Seller/Buyer classification has been provided in Table 5-1.
Table 5-1: Example of Seller and Buyer

<table>
<thead>
<tr>
<th>User</th>
<th>User 1</th>
<th>User 2</th>
<th>User 3</th>
<th>User 4</th>
<th>User 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRBs Allocated</td>
<td>4</td>
<td>6</td>
<td>3</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>PRBs Required</td>
<td>6</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Classification</td>
<td>Buyer</td>
<td>Seller</td>
<td>NA</td>
<td>Seller</td>
<td>Buyer</td>
</tr>
</tbody>
</table>

5.2.1 Establishing the Price for Seller

The Seller has the ability to set the price that it will charge the Buyers. The price setting is done on the basis of an index that is defined as the willingness to sell where an index value of 0 is the lowest and 1 is the highest. The price charged by the Seller has an inverse relationship to the willingness to sell index. This index is defined as a function of the number of PRBs that can be allocated by the base station. After a certain PRB threshold, the willingness to sell index ($W_s$) becomes 1. $W_s$ for a Seller is given by (9).

$$W_s = \frac{1}{\gamma} \exp(\beta)$$  \hspace{1cm} (9)

where $\gamma > 0$ and $\beta$ is given by (10).

$$\beta = \frac{\alpha}{1 + e^p e^{-\alpha}}$$  \hspace{1cm} (10)

where $\alpha > 0$, $p$ is the price charged by the Seller and $n$ is the number of PRBs defined as the PRB threshold.

As an example, Figure 5-3 shows a plot of the Willingness to Sell index against the number of PRBs. Note that in the example, the PRB threshold is approximately 36 resource blocks after which the Seller’s Willingness to Sell index is 1. The value of $p$ can be extrapolated by setting $W_s = 1$ at $n = 27$. 


By the model design, the Sellers with lower perceived throughputs have a higher PRB threshold requirement. This shall be explained in subsequent sections.

5.2.2 Non-Cooperative Game amongst Buyers

In this section, the Buyer’s game model has been described. Define a function for a Buyer \( i \) for resource \( x_i \) by (11),

\[
Q_i(x_i) = \gamma \log x_i 
\]

where \( \gamma > 0 \) and \( x_i \) is the number of PRBs required for user \( i \) to achieve its desired running average \( r_i \).

The price function for all the Buyers is given by (12),

\[
P_i(\sum x_i) = \frac{z_j}{R - \sum x_j} 
\]

where \( R \) is the Total Number of PRBs available, \( \sum x_i \) is the total number of resources required by the Buyers and \( z_j \) is the spare number of PRBs Seller \( j \) allocates to the Buyers. \( z_j > 0 \) as per the definition of what constitutes a Seller. The reference [84] classifies the price function proportional to the aggregate delay for a given user. However, in this case, the price function is a congestion indicator with respect to the number of spare resources the Seller allocates for the Buyers.

Define the cost function \( (F_i) \) for user \( i \) by (13).
where $\kappa \in \mathbb{Z}$ is used for normalizing the parameters to compare with the Seller’s price.

\[ F_i = \kappa \left( P_i \left( \sum x_i \right) - \gamma \log(x_i + 1) \right) \]  

(13)

$F_i$ is the objective function of the Buyer with $P_i \left( \sum x_i \right)$ as the price function and $Q_i(x_i)$ as the Buyer’s utility function. Based on the principles provided in [84], (13) has a unique Nash Equilibrium. Therefore, the Buyer will try to minimize their cost function every TTI given the strategies of the other players.

### 5.2.3 Analysing the Buyer and Seller Model

For determining congestion in the network, there has to be exchange of PRBs from a Seller to a respective Buyer. This can only happen if the cost that a Buyer is willing to pay to buy resources from a Seller is greater than the price the Seller is willing to sell them for. If prices cannot be matched by the Buyer, then there is an impasse and the Buyer will have no choice but to drop off from the transaction thereby inducing a packet drop in the base station buffer and reducing the TCP window size.

As a rule, a packet drop can only be induced if there are users who can be classified as Sellers in the system. If there are no resources to share (i.e. the Sellers have exhausted their extra PRBs or there are no Sellers), then the Buyers resort to the proportional fair allocation. This concept has been explained using the examples below.

**Example 1**

Consider the following set of values for a group of users in a given TTI as listed in Table 5-2. In this table, the PRBs allocated are by the base station using a proportional fair scheduling method. The PRBs required are calculated from the running average, the delta PRBs are the difference between the PRBs allocated and the PRBs required and the classification is done based on that difference.

<table>
<thead>
<tr>
<th>User</th>
<th>User 1</th>
<th>User 2</th>
<th>User 3</th>
<th>User 4</th>
<th>User 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRBs Allocated</td>
<td>20</td>
<td>25</td>
<td>18</td>
<td>7</td>
<td>20</td>
</tr>
<tr>
<td>PRBs Required</td>
<td>17</td>
<td>20</td>
<td>25</td>
<td>10</td>
<td>25</td>
</tr>
<tr>
<td>Delta PRBs</td>
<td>3</td>
<td>5</td>
<td>-7</td>
<td>-3</td>
<td>-5</td>
</tr>
<tr>
<td>Classification</td>
<td>Seller</td>
<td>Seller</td>
<td>Buyer</td>
<td>Buyer</td>
<td>Buyer</td>
</tr>
</tbody>
</table>

*Table 5-2: Data for Example 1*
From Table 5-2 the following classifications can be made: Sellers $\in \{\text{User 1, User 2}\}$ and Buyers $\in \{\text{User 3, User 4, User 5}\}$. Assume that the prices set by the Sellers are, $(\text{User 1, User 2}) = (25 \text{ units, } 20 \text{ units})$.

User 2 has the lowest price setting that all the Buyers can match. By choosing appropriate parameters for $\kappa$ and $\gamma$, the cost for the Buyers are $(\text{User 3, User 4, User 5}) = (47.8 \text{ units, } 42.7 \text{ units, } 39.4 \text{ units})$. Therefore, a transaction of the Sellers’ excess PRBs can take place. Note that User 3 is willing to pay the most for the resources. The Delta PRBs after the transaction are $(\text{User 1, User 2, User 3, User 4, User 5}) = (3, 0, -7, 0, -2)$.

Now the only remaining Seller is User 1 and the Buyers are Users 3 and 5. For a price of 30 units, the cost for the Buyers are $(\text{User 3, User 5}) = (26 \text{ units, } 13.5 \text{ units})$. The transaction between User 1 and User 3 can take place. The Delta PRBs are, $(\text{User 1, User 2, User 3, User 4, User 5}) = (0, 0, -7, 0, 0)$.

User 3 remains with a lower number of resources than required and uses the PRBs that were allocated as a result of proportional fair scheduling. Since there are no Buyers left, a packet drop is not induced by the base station. This simply means that User 3 will not satisfy $x_3$ for the TTI.

**Example 2**

This example will demonstrate when an actual packet drop is enforced. Consider that there are 5 users, as shown in Table 5-3.

<table>
<thead>
<tr>
<th>User</th>
<th>User 1</th>
<th>User 2</th>
<th>User 3</th>
<th>User 4</th>
<th>User 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRBs Allocated</td>
<td>20</td>
<td>25</td>
<td>10</td>
<td>28</td>
<td>17</td>
</tr>
<tr>
<td>PRBs Required</td>
<td>16</td>
<td>22</td>
<td>8</td>
<td>20</td>
<td>15</td>
</tr>
<tr>
<td>Delta PRBs</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>-8</td>
<td>-2</td>
</tr>
<tr>
<td>Classification</td>
<td>Seller</td>
<td>Seller</td>
<td>Seller</td>
<td>Buyer</td>
<td>Buyer</td>
</tr>
</tbody>
</table>

From Table 5-3, Sellers $\in \{\text{User 1, User 2, User 3}\}$ and Buyers $\in \{\text{User 4, User 5}\}$. Assume that the prices set by $(\text{User 1, User 2, User 3}) = (25 \text{ units, } 15 \text{ units, } 35 \text{ units})$.

User 2 has the lowest price setting. Therefore, the cost of using User 2’s resources, $(\text{User 4, User 5}) = (11.8 \text{ units, } 15.3 \text{ units})$. Only User 5 and User 2 get to make the transaction since User 5 is able to match User 2’s price. The Delta PRBs are, $(\text{User 1, User 2, User 3, User 4, User 5}) = (5, 1, 2, -8, 0)$.  

74
User 4 is unable to match the prices set by the users. The cost User 4 is willing to pay for using PRBs from Users 1, 2 and 3 are 22.68, 11.81 and 0.94 respectively. Therefore, based on the rule set by the game, User 4 will have to drop from the base station buffer.

5.3 Game Theoretic Model for Base Station Congestion Mitigation: Equilibrium Model

This section details how the second congestion controller has been developed based on game theoretic principles. Unlike the Buyer and Seller model where radio resource allocation is performed in a proportional fair manner first, in the Equilibrium model resource allocation is done on the basis of a payoff function that each user tries to maximize.

5.3.1 Game Model

A key assumption when it comes to scheduling is that the base station will schedule radio resources per TTI such that the number of PRBs allocated is maximized. This could result in users that have a high PRB requirement and can starve users who demand a lower number of PRBs from the resource allocation process. To prevent the monopoly of resource intensive users, a rule in resource allocation is set such that the radio resources are allocated in a fair manner.

For resource distribution amongst the users, every TTI, each user has two choices to make bid and hold. The action drop is enforced by the base station. These three have been elaborated below:

- Bid signifies the user will receive the desired radio resources in the given TTI i.e. user \( i \) receives \( x_{n,i} \) resources.
- Hold either signifies the user will not receive radio resources either due to it being unable to bid or the user purposely chooses not to bid (explained subsequently). For bidding, \( x_{n,i} = 0 \) and \( x_{n+1,i} = \beta \times r \) where \( \beta \) represents the number of times consecutive TTIs the user has had to play hold, \( \beta > 0 \) and \( \beta \in \mathbb{Z} \).
- Drop implies that the user has to drop out of the game. This triggers the base station’s AQM functionality by inducing packet drop to reduce the TCP rate.

Over a period of \( N \), the payoff function for the user \( i \) is to maximize its payoff \( J_i \) given by (14).

\[
J_i = \sum_{n=1}^{N} \left( \frac{x_{n+1,i}}{r_i} e^\beta - p_i x_{n,i} \right)
\]  

(14)
where, \( p_n \) is the price per resource block, 
\[ p_n = \frac{\sum_{i=1}^{L} x_{n,i}}{R} \] 
and \( \frac{x_{n,i}}{r} \) states the utility requirement for user i.

The maximization of the payoff function will result in a Nash Equilibrium.

**Proof:** According to [85], if the maximization of (14) with respect to \( x_{n,i} \) gives a unique solution of \( x_{n,i}^* \) and the second partial derivative of (14) is negative definite then \( x_{n,i}^* \) is in Nash Equilibrium.

**Step 1:** Maximization of \( J_i \) gives (15)

\[ x_{n,i}^* = \frac{R e^i}{2r_i} \]  

From (15), it is clear that \( x_{n,i}^* \) is a stationary point. Therefore, maximizing \( J_i \) produces a unique solution.

**Step 2:** The second order partial derivative of \( J_i \) should be negative definite, i.e.,

\[ \frac{\partial^2 J_i}{\partial x_{n,i}^2} = \frac{-2}{R} < 0 \]  

Therefore, the second order partial derivative of \( J_i \) is negative definite. Hence the maximization of the payoff function in (14) results in a Nash equilibrium.

For a given TTI, each player would try to maximize their objective function given the strategies of the remaining players. The algorithm predicts this and gives each user its prospective action path. Once the action path is predicted, the user will choose an action. A user that predicts that it is better off bidding for resources at a later TTI since it maximizes its payoff will do so by playing Hold. A user who does not see any benefit from bidding later or cannot (if subsequent bids will result in \( x > R \)) will play Bid. The user will play drop if they are not getting enough radio resources since they have the highest demand, or the user is starving other users due their high demand. The following section highlights these strategies and the concept of Nash Equilibrium with respect to this game.

### 5.3.2 Analysing the Equilibrium Model

In this section, the results from the game theoretic model is analysed. Consider that there are three users with the running averages \( r_i = \{10, 15, 75\} \) and run-time (in TTIs) for each user is \( N_i = \{8, 8, 5\} \). The initial value of \( x \) for each user is \( x_i = \{10, 15, 75\} \). The maximum resource block capacity is \( R = 100 \). The users will have to maximize their payoff by predicting their action path from \( n = 1 \).
There are three scenarios that for resource allocation. The first scenario is employed by schedulers such as RED and Discard Timer where the users are scheduled at each TTI (i.e. they Bid). In the second scenario, the users can make choices based on their desired action path to achieve Nash Equilibrium. In the third scenario, the effects are shown when the system deviates from Nash Equilibrium.

Scenario 1 is shown in Figure 5-4. The x-axis represents the interval (N) and for Figure 5-4 (a) – (c) the y-axis represents the payoff for that interval as a percentage of the cumulative payoff of a user. Figure 5-7 (a) shows the price per RB for each interval for scenario 1. At the 6th interval, User 3 does not require any further radio resources, following which there are only 2 active users. Therefore, the price per RB decreases due to low congestion. The total payoff for User 1 for this scenario is 176154.23 units.

Scenario 2 is shown in Figure 5-5. The x-axis and y-axis values are the same as scenario 1. User 1 plays Hold till the 5th interval and at the 6th interval beings to Bid for radio resources. Hence the payoff is 75% of the total payoff at the 6th interval. This also coincides with there being two users in the system since User 3 does not need any further radio resources.

The total payoff for User 1 is 176161.73 units, which is greater than scenario 1. Further comparing both the scenarios, the average price is the same but the price per interval between the 1st and the 5th intervals are lower in scenario 2 as demonstrated in Figure 5-7 (b). Therefore, the payoff for User 1 remains the same as scenario 1 while the payoff for User 3 is larger.
The strategies of all users in a Nash Equilibrium are the best responses to each other. This does not imply, that no user can do better by not adhering to Nash Equilibrium. Consider, scenario 3 where User 1 decides to Hold at the first interval deviating from the Nash Equilibrium as shown in Figure 5-6 (a). At the second interval, \( x_2 = \{20, 15, 75\} \), the sum of which is 110 > R. User 3 cannot Hold and since it is the highest resource utilizer, it cannot Bid as per the rule stated. Therefore, User 3 is forced to Drop from the system (Figure 5-6 (c)). User 1 and User 2 now can share R amongst themselves in a low congested environment. User 1 and 2 with User 1 having a payoff of 176189.7264 units benefit from this strategy at the expense of User 3.
5.4 Results and Discussion

In this section the results from five mechanisms namely, Discard Timer, RED, Proportional Controller, Buyer-Seller and Equilibrium models are compared. The result evaluation method remained
unchanged from Chapter 4. Each of the models were subjected to three different radio conditions which have been listed below:

- Good radio conditions: path loss exponent = 2.2 and shadowing standard deviation = 2 dB
- Average radio conditions: path loss exponent = 2.3 and shadowing standard deviation = 4 dB
- Poor radio conditions: path loss exponent = 2.8 and shadowing standard deviation = 7 dB

5.4.1 Good Radio Conditions

The results for good radio conditions for the five controllers have been shown in Figure 5-8. The x-axis represents the number of users while the y-axis represents Throughput (Mbps), Latency, average Window Size (packets) and percentage of packets retransmitted (%) for Figure 5-8 (a), (b), (c) and (d) respectively.

The 10-user and 20-user scenarios signify low congestion scenarios. From Figure 5-8 (a), it is observed that the throughput values for all 5 controllers are similar with a maximum of 1 Mbps difference between them. The latency values for the game theoretic models when compared to DT or RED are significantly lower but the window sizes are also smaller. Therefore, DT and RED allow a higher threshold queuing in the base station buffer before being dropped. This can be observed by the retransmission percentage which is close to zero for DT and RED but between 15% and 25% for the game theoretic models. This demonstrates that the congestion mitigation mechanism takes place even if there is low perceived network congestion.

The 30-user scenario is an important measurement since it pushes the operating regions of congestion mechanisms based on queue management. This is observed with over 80% packet retransmission for RED (Figure 5-8 (c)). This demonstrates the beginning of congestion collapse for RED. The game theoretic models are uniformly higher than the three controllers which use base station queue management.

The 40-user and 50-user scenarios demonstrate the results when the congestion is increased. As detailed in Chapter 4 the queue-based congestion controllers face congestion collapse. It should be noted that the game theoretic controllers maintain a useable throughput despite high congestions in the base station. The links also do not face network congestion collapse. This is evident from the throughput and latency values; the latter increases with increased congestion. The average packet retransmission rate for Buyer Seller and Equilibrium models are around 20% and 40% respectively which is capable of delivering close to 2 Mbps of throughput. It should be noted that the latency values for RED and Proportional controller decrease while the latency values for the other controllers increase with an increase in congestion. This is due to packets close to zero buffer occupancy at the base station since all the packets are discarded simultaneously (resulting in TCP global synchronisation) due to the high volume of traffic.
5.4.2 Average Radio Conditions

The results for average radio conditions have been shown in Figure 5-9. The overall trends remain the same as that of good radio conditions, but the throughput values are lower, while the latencies are higher.

When there are 10 and 20 users, the average throughput values are approximately the same for each of the controllers. From Figure 5-9 (d) it can be observed that unlike Discard Timer or RED, the game theoretic models provide congestion mitigation.

DT and RED work well within certain operating regions. When there are 30 users, the system closes in on those regions. This is observed with the high standard deviation values for window size and retransmission percentage. However, this is not the case with the game theoretic models. There is no deviation from the standard congestion mitigation behaviour.

With high congestion in the base station as observed in the 40 and 50 user scenarios, with average radio conditions, there is congestion collapse for the links using queue-based congestion management while the game theoretic models maintain a useable throughput by adjusting to the network settings.
5.4.3 Poor Radio Conditions

As expected for poor radio conditions, the average throughput and window sizes are significantly lower, and the average latency is higher than the good and average radio conditions. The results have been displayed in Figure 5-10. Due to higher latencies, DT based congestion mitigation mechanism begins to falter when there are 20 users in the base station queue. The throughput values for RED, Proportional Controller, Buyer-Seller and Equilibrium are similar for 10, 20 and 30 users.

When there are 30 users, the standard deviation values for throughput and latency for RED and Proportional controller are significantly higher than that of the game theoretic models. This demonstrates that even with poor radio conditions where amount of traffic is expected to be lower the queue-based controllers begin to demonstrate signs of not being able to cope with traffic demand whereas the game theoretic models remain more stable.

High congestion situations with 40 and 50 users the queue-based congestion mechanisms still experience congestion collapse. As shown in Figure 5-10 (a) the average user throughput is close to zero and the packet retransmission rates are 80% for DT and close to 100% for RED and Proportional controller. The game theoretic models are able to maintain a useable throughput value and have reasonable packet retransmission values.
5.5 Conclusion

Queue-based models localise the congestion decision to variable parameters with reference to the base station buffer (latency or queue size). If the chosen parameters are not within a certain operating region then proper congestion mitigation is not achieved. In wireless networks these parameters can be hard to determine due to varying conditions such as radio and active number of users resulting in inconsistent performance in congestion control. For example, 30 active users in system give different results for RED under different radio conditions. When the radio conditions are good the performance degrades as indicated by the above 80% packets retransmitted in Figure 5-8. However, when the radio conditions are poor, congestion control is more efficient as indicated by the average 20% packets retransmitted in Figure 5-10. Another example is from the Discard Timer based congestion mitigation mechanism in good and poor radio conditions. In good radio conditions with 20 users, DT does not need to perform any congestion mitigation with 0% packets retransmitted. In average radio conditions this value increases to below 10%. With poor radio conditions this increases to 50% with a large standard deviation value. These examples demonstrate that for ideal performance of congestion mitigation using queue-based mechanisms, parameter tuning will be required based upon the network conditions.

The approach of game theoretic models for congestion control is fundamentally different. Congestion mitigation is based on a strategy to optimise radio resource allocation by the base station for each user/link. The optimal value is achieved by referencing to a certain bitrate and factoring in
parameters such as the TCP link RTT and amount of data being transmitted in the given RTT. Depending on the radio conditions and the base station congestion (number of users), the desired resource allocation per TTI is determined to match the required bitrate. Each TTI, the independent users are involved in a non-cooperative game. The game allows the users to take specific actions such as trading radio resources (Buyer-Seller model) or not partaking in the resource allocation process (Hold action in the Equilibrium model). The base station is a go-between the users and only reacts by implementing congestion control when a user begins to monopolise the game due to a large discrepancy on its desired radio resource allocation and what it can be allocated using the game.

There are two variable network settings used to evaluate the models. The first is increasing congestion by increasing the number of users and the second is varying radio conditions. With the former method, the game theoretic models do not cause the links to experience any congestion collapse as demonstrated from the results. Although the throughput progressively deteriorates with increased congestion, the game theoretic models are still able to provide a useable throughput. With the second method of evaluation, the game theoretic models apply similar levels of congestion mitigation irrespective of the radio conditions. As highlighted earlier, the degree of congestion control varies for the DT based on the radio conditions when there are 20 users. Comparing the percentage of packet retransmitted for the game theoretic models in Figure 5-8 (d), Figure 5-9 (d) and Figure 5-10 (d), it can be observed that the outcomes are congruent. These results prove that game theoretic models do not need certain operating regions for providing congestion mitigation. Therefore, there is efficient use of radio resources irrespective of the network setting.

5.6 Chapter Summary

This chapter has demonstrated how congestion mitigation can be performed using the principles of non-cooperative game theory using PRB based radio resource allocation. The concepts of perceived throughput and running average were introduced which forms the foundations of the non-cooperative game. Two different controllers were implemented which use the principles of non-cooperative game theory. The Buyer-Seller model classifies users as either buyers or sellers based on the radio resources allocated and the radio resources needed. The buyers and sellers then divulge into a non-cooperative game. In the Equilibrium model, each user creates their own action path based on the predicted action path of the other users. The base station acts as a facilitator of the game and only takes congestion mitigation action if a user begins to monopolise the game.

The results from the game theoretic models were compared with the queue-based congestion control models (Discard Timer, RED and Proportional). The results demonstrated that the game theoretic models provided a more robust mechanism of congestion mitigation.

A study of fairness of resource allocation amongst all the controllers has been provided in the Appendix A of this Chapter.
Appendix A: Fairness amongst Users for the Different Controllers

The underlying method for radio resource allocation per TTI for all the controllers except the Equilibrium model was the proportional fair method as described in Chapter 4. Amongst the controllers that used proportional fair allocation, the Buyer-Seller model deviates significantly in terms of resource allocation due to the introduction of perceived throughput and running average and the exchange of resources as a part of the game model. Therefore, the issue of fairness especially for the game theoretic controllers needs deeper insight. The study in this section investigates fairness for all the controllers.

The proportional fair algorithm distributes radio resources equitably amongst all users. However, in this context fairness would mean how the bottleneck (the Base Station) provides all the users an equal allocation of the bottleneck link capacity. Therefore, each controller model would have different levels of fairness and this fairness measurement would also be determined by the throughput of each link. The Jain’s fairness index is used for measuring fairness based on the aforementioned concept [86] and is given by (17),

$$J(\lambda_1, \lambda_2, \ldots, \lambda_L) = \frac{(\sum_{i} \lambda_i)^2}{L(\sum_{i} \lambda_i^2)}$$

The index is ranges from 1 to $\frac{1}{L}$ where a totally fair allocation is close to 1 and an unfair allocation is close to $\frac{1}{L}$.

Table 5-4 below shows the mean and standard deviations of 10 values of the Jain’s fairness index for all the controllers for different number of users. The radio conditions used were Average radio conditions.

<table>
<thead>
<tr>
<th>Controller Type</th>
<th>10 Users</th>
<th>20 Users</th>
<th>30 Users</th>
<th>40 Users</th>
<th>50 Users</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discard Timer</td>
<td>0.994 (0.003)</td>
<td>0.995 (0.002)</td>
<td>0.991 (0.011)</td>
<td>0.933 (0.053)</td>
<td>0.878 (0.065)</td>
</tr>
<tr>
<td>RED</td>
<td>0.994 (0.002)</td>
<td>0.932 (0.018)</td>
<td>0.839 (0.041)</td>
<td>0.961 (0.021)</td>
<td>0.977 (0.015)</td>
</tr>
<tr>
<td>Proportional</td>
<td>0.976 (0.013)</td>
<td>0.762 (0.016)</td>
<td>0.722 (0.032)</td>
<td>0.969 (0.033)</td>
<td>0.971 (0.023)</td>
</tr>
<tr>
<td>Buyer-Seller</td>
<td>0.986 (0.015)</td>
<td>0.965 (0.025)</td>
<td>0.981 (0.019)</td>
<td>0.948 (0.041)</td>
<td>0.937 (0.043)</td>
</tr>
<tr>
<td>Equilibrium</td>
<td>0.986 (0.006)</td>
<td>0.979 (0.008)</td>
<td>0.941 (0.023)</td>
<td>0.932 (0.021)</td>
<td>0.911 (0.015)</td>
</tr>
</tbody>
</table>

With 10 users and low congestion the distribution of throughput for all the users is fair for all the controllers since the index is close 1. With an increase in the number of users to 20 and 30 users, the fairness index for the RED and Proportional controllers show that the allocation is unfair. This shows that the random packet discard to lower the link’s TCP transmit rate causes throughput degradation as compared to the other links and eventually resulting in an unfair allocation. The DT
and the Buyer-Seller controller do not show any variation in fairness with increased congestion. However, the Equilibrium controller shows a slight progressive decrease in fairness. Therefore, in terms of fairness, DT and game theoretic models are fairer at distributing radio resources.

For high congestion scenarios with 40 and 50 users, the results for RED and Proportional are a misnomer and should be discarded. This is because all the links experience congestion collapse. From Figure 5-9 (a) it can be observed that the average throughput is close to zero. Therefore, all users are receiving the same amount of throughput which is close to zero hence the high fairness index. The DT also shows decrease in fairness. Figure 5-9 (a) suggests that some links have a high and useable throughput but most of the links do not. Hence, the average throughput value is not zero and the decrease in the fairness index explains why this is the case. The game theoretic controllers while maintaining a useable throughput also have fairly consistent distribution of throughput for each link.

In conclusion, the overall fairness for the game theoretic models are higher than the queue-based congestion control models. The latter start to falter which an increase in congestion both in terms of throughput and fairness for resource allocation. The game theoretic models maintain a useable throughput and fairness since the game allows for good resource distribution across each link.
Chapter 6: Conclusions and Scope for Future Work

6.1 Conclusions

This thesis explores congestion mitigation in wireless networks for TCP traffic. In Chapter 1 the need for congestion mitigation in wireless networks was stated. TCP has a mechanism whereby the protocol congestion in the network can be detected and a reactionary measure is taken by the TCP source. However, the TCP protocol is not sophisticated enough to detect latency in the network. Since the last mile of wireless networks (radio interface) can be lossy, appropriate actions are taken by the base stations to prevent these losses. Additionally, a base station has limited capacity. These cause high latencies due to higher buffer occupancy by packets in the base station buffer. Therefore, the base station is usually the main bottleneck for any wireless networks. These can be addressed by appropriate congestion mitigation actions taken by the base station.

The main focus of this thesis is to investigate existing congestion mitigation techniques in LTE networks and provide appropriate improvements. In order to study and validate the effectiveness of congestion mitigation techniques a simulation tool was developed using MATLAB. This is a major contribution towards the research objectives since the tool allows a common platform to test each congestion mitigation models developed. Chapter 3 has been dedicated providing details of the theory and tool development.

In Chapter 4 there were two congestion mitigation models developed using existing congestion control principles, namely the Discard Timer and RED. These models were analysed mathematically using control theory. A limitation such models is the assumption of using fixed radio capacity. To address this, a congestion mitigation model using proportional control factoring variable radio had been developed. While analysing the performance of the three controllers it was observed that the controllers were effective within certain operating regions. The proportional control with variable radio capacity is more effective than RED at congestion mitigation only for a set amount of capacity in the base station. Thus, the conclusion is, to implement these congestion mitigation models for optimal performance an operator will need to understand the specifics of the network such as radio conditions and the expected number of users. If the parameter setting is not performed correctly or are beyond the suitable operating regions, then the TCP connections to this base station will experience congestion collapse as demonstrated by the results. If the conditions change then the parameter tuning will be required.

To address these limitations, a different approach for congestion mitigation was developed in Chapter 5. The models factor in LTE resource allocation for congestion mitigation hence requires a different mathematical framework to detect congestion. This was developed by estimating the TCP transmission rate and determining a base station buffer occupancy time to match the desired rate. Using this occupancy period, the number of radio resources required by each user is determined.
These parameters are used for congestion mitigation using the principles of non-cooperative game theory. The results showed that the game theoretic method implements robust congestion mitigation and is not prone to the limitations of using queue-based congestion mechanisms. Unlike the queue-based congestion mitigation models, the game theoretic models do not require any parameter tuning or adjustments. This makes the game theoretic models more scalable and suitable for a variety of conditions (such as different radio conditions and number of users) that are typically experienced in wireless networks.

6.2 Future Work
There are different focus areas that the thesis could serve as a foundation for further investigation. These have been discussed below:

1. Implementation with 5th Generation Cellular Network Technology: The studies performed in this thesis especially in Chapter 5 are in the context of congestion mitigation using radio resource allocation in 4G LTE networks. Radio resource block allocation and requirement is a cornerstone required for congestion detection and referencing in the radio base station. 5th generation of cellular network technology (5G) also use resource blocks for data transfer to UEs [87]. Therefore, the game theoretic models find applicability in 5G New Radio (NR) technology for congestion mitigation. Further details have been provided in Appendix A of this chapter.

2. Implementing Quality of Service: The data traffic type for simulation used in this thesis have been considered to be Best Effort (BE) and therefore have not been subjected to any Quality of Service (QoS) treatment. In networks there can be diverse QoS requirements which is managed end to end across each network element [88]. Therefore, QoS policies will also have to be implemented in the base station. The proposed required modifications to the theoretical foundations set in this thesis with QoS treatment have been suggested in Appendix B of this chapter.

3. Fixed Buffer Size: As stated previously, due to improved hardware processing ability, base station buffers can be kept very large [7]. Therefore, an infinite buffer has been assumed for formulating the game theoretic models. For practical considerations a finite buffer can be factored in. Therefore, the game theoretic models will need adapted to include a finite nuffer size. Directions on how this can be achieved has been detailed in Appendix C of this chapter.

4. Resource Allocation Algorithms: This thesis has addressed the problems associated with congestion mitigation in wireless networks. Radio resource allocation forms a key for developing the congestion mitigation techniques. Alternative methods for resource allocation can be explored within the framework defined in Chapter 5. The Stable Marriage is one such algorithm can provide an alternate method of resource allocation. Using the notion of Buyers and Sellers the pairings required for the Stable Marriage algorithm can be achieved [89] [90].
5. Using a Control Theoretic Approach for Non-Cooperative Games: A control theoretic approach for non-cooperative game design would allow a target price to be maintained for use of a resource block. Using a control system to influence a game has been explored in [91]. This approach would enable a closer control of resource allocation and therefore congestion mitigation.

6. Using different TCP variants: There are multiple TCP variants such as TCP New Reno, Tahoe, TCP Vegas and TCP CUBIC. The simulation tool can be used as a foundational platform to test the different TCP variants and their effects in congestion in wireless networks.

7. Mobility: A next stage of evolution to the work in this thesis would be the involvement of high mobility in users.
Appendix A: Implementing Game Theoretic Models for 5th Generation Cellular Network Technology

There are some relevant differences between 5G NR and LTE in the physical layer have been summarised below:

- **Resource Blocks:** In LTE one resource block consists of 14 symbols in time domain and 12 sub-carriers in frequency domain with a fixed bandwidth of 180 kHz and sub-carrier spacing of 15 kHz. However, in NR the resource blocks bandwidth is not fixed and is based on **numerologies** which is a multiple of 12 subcarriers and a variable subcarrier spacing (minimum is 15 kHz and maximum is 240 kHz) [87] [92]. The maximum number of Resource Block varies depending upon the numerology value. Figure 6-1 shows the difference with LTE where a 15 kHz numerology consists of 14 OFDM symbols and a 30 kHz numerology consists of 28 OFDM symbols [93].

- **Channel Efficiency:** In LTE there were 4 cell specific reference signals transmitted every millisecond. In NR there are no cell specific reference signals. This enables power savings in the base station [94].

![Figure 6-1: Use of Numerologies in 5G](image-url)
As demonstrated in Chapter 5, the running average parameter is central for the development of the non-cooperative games to take place. The running average stated in (1) below. Note that $x_{n,i}$ and $N$ are variables where $x_{n,i}$ is defined as the number of PRBs that are allocated to the user $i$ each TTI over a period of $N$ and $0 \leq x_{n,i} \leq R$ where $R$ is the total number of PRBs. Therefore, there are no changes required to the theory between LTE and 5G.

$$r_i = \frac{1}{N} \sum_{n=1}^{N} x_{n,i}$$  \hspace{1cm} (1)
Appendix B: Implementing Quality of Service in User Traffic

In this Appendix section the implications of applying QoS for base stations will be investigated within the context of the material developed in Chapter 5. There have been two cases explored, the first case looks at differentiation based on service type and the second case presents the consequences of some users receiving a guaranteed bandwidth.

Previous studies in wireless networks for implementing QoS in base stations have achieved QoS based network control by managing the scheduling algorithm [95] [96], providing admission control [97] and using network coordination [98].

Case 1: Service Differentiation

Buyer-Seller Model

With service differentiation the base station is able to differentiate user service based on different class or type of traffic. These traffic classes will be treated differently by the base station. Consider for example there are two classes served by the base station, Class A and Class B with Class A having a higher access to higher bandwidth than Class B. Using the game theoretic models, prioritisation for Class A users can be done in multiple ways. One such method is by subsidising (reducing) the price function which is applicable in the Buyer-Seller model. This will enable Class A users who are Buyers having a reduced cost function, hence have access to more radio resources from the Sellers. The cost function \( F_i \) for user \( i \) can be defined by (2) for Class A users.

\[
F_i = \kappa (\lambda \times P_i (\sum x_i) - \gamma \log (x_i + 1))
\]

Where a factor \( \lambda \) has been introduced for reducing the cost for the users and \( 0 < \lambda < 1 \).

Equilibrium Model

Another method is to provide higher weighting to running average requirement for Class A users. This can be used applied in the Equilibrium model. As defined in Chapter 5, for the Equilibrium model, the payoff \( J_i \) is given by (3).

\[
J_i = \sum_{n=1}^{N} \left( \frac{x_{n,i}}{r_i} e^{r_j} - p_n x_{n,i} \right)
\]

Where \( x_{n,i} \) signifies the radio resources user \( i \) receives over \( n \) TTIs, \( p_n \) is the price per resource block and \( r_i \) is the running average.

With Class A and Class B users, define (4)
Where $\alpha > 1$. Consequently, the Payoff for the users is defined by (5),

$$J_i = \sum_{n=1}^{N} \left( \frac{x_{n,i}}{r_i^*} - p_n x_{n,i} \right)$$

Consider the same example from Chapter 5 for the Equilibrium model. However, now User 1 is a Class A user. The running averages and the run-time are $r_i = \{10, 15, 75\}$ and $N_i = \{8, 8, 5\}$ respectively. Assuming $\alpha = 2$ then $r_i^* = \{20, 15, 75\}$. The maximum resource block capacity is $R = 100$. The comparison of $r_i$ and $r_i^*$ show the user 1 is given access to a larger share of radio resources. In addition, to maintain the weighted running average $r_i^*$, User 3 is forced to drop off from the resource allocation game as per the rule set for the Equilibrium model. As shown in Figure 6-2 (c), the payoff percentage over time is zero for User 3. Additionally, User 1 does not require any radio resources beyond $N = 4$ since $r_i^* = 2 \times r_i$. By using this method, User 1 faces lower latency. Consequently, due to the shorter resource allocation period, User 1 receives a greater amount of throughput when compared with the example in Chapter 5.

![Figure 6-2: Example of Implementing QoS with the Equilibrium Model](image-url)
Service Differentiation during Resource Allocation

Service differentiation can also be applied during resource allocation. The proportional fair resource allocation method developed in Chapter 4 can be modified to cater for different classes of traffic. The price metric \( \omega_i \) for \( L \) users can be modified to (6).

\[
\omega_i = \beta \times \frac{\min(CQI_{i, sl1})}{CQI_i}
\]  

(6)

where \( \beta < 1 \) and is used to adjust the price value for users depending upon their class. Consequently, the Transport Block allocation per TTI \( (x_i) \) is also adjusted by \( \beta \) as shown in (7).

\[
x_i = \frac{\omega_i}{\sum_{i=1}^{L} \omega_i/R}
\]  

(7)

Consider an example with 6 users, the first three users are Class A users and the other three users are Class B users. Class A users should receive twice as much bandwidth as Class B users. If \( \beta = 0.5 \) for Class B users, then the price is reduced by half as shown in Table 6-1. The radio resources allocated, and the corresponding transport block is almost half for Class B users when compared to Class A users.
Case 2: Service Prioritisation

Another common reason for applying QoS is to prioritise certain types of service and provide a guaranteed bandwidth. This would require allocation of certain number of PRBs each TTI to be reserved for the services. This impact on the game theoretic models would be with the price function and the price per resource block for the Buyer-Seller and the Equilibrium models respectively. If $X$ resources are reserved per TTI with $X < R$, then $R - X$ resource remain for Best Effort or non-guaranteed traffic. The price function ($P_j(\sum x_i)$) for the Buyer-Seller model is adjusted to be

$$P_j(\sum x_i) = \frac{z_j}{R - X - \sum x_i},$$

the price per resource block ($p_n$) for the Equilibrium model is

$$p_n = \frac{\sum x_i^j}{R - X}.$$
Appendix C: Implementing Game Theoretic Models factoring in Finite Buffers

Finite buffer provides limitations for formulation of algorithms such as the Buyer-Seller or the Equilibrium model, but it also can provide an opportunity increase efficiency of the algorithms. For example, a finite buffer can curb excessive queuing in the base station buffer and improve efficiencies in the radio resource allocation and throughput.

There have been three approaches that have been proposed where the game theoretic models can be adapted for factoring in finite buffers. These have been explained below with discussions on their advantages and disadvantages.

**Case 1: Drop on exceeding Buffer Threshold**

In this case, a buffer threshold per link $i$ is set denoted by $b_i, max$. If a greater number of packets arrive than $b_i, max$ then the packets are simply discarded. The advantage of this approach is easy implementation. However, this approach will have the same problems such as those experienced by RED or the Proportional controller when congestion is high. In addition, during high congestion the effectiveness of the game theoretic models will be limited since discarding packets on exceeding the buffer threshold will take precedence over the resource allocation game.

**Case 2: Modifications to Running Average**

A key factor in determining the outcome of the game theoretic models is the running average. In this case, the base station buffer occupancy is modified which in turn alters the running average value. From Chapter 5 the running average for a user $i$ is given by (8),

$$\tau_i = \frac{A_i}{N}$$  \hspace{1cm} (8)

Where $N = \frac{\tau_{eNB,i} / TTI}{TTI}$ and $A_i$ is the number of bits for link $i$. $\tau_{eNB,i}$ is amount of time the packets are queued in the base station buffer. Base station buffer occupancy can be represented as shown in Chapter 5 (9),

$$\tau_{eNB,i}^1 = \frac{A_i}{\gamma_i} - \tau_{delay,i}$$  \hspace{1cm} (9)

$\gamma_i$ is the perceived throughput. Factoring the base station buffer size, the maximum buffer occupancy can be given as (10),

$$\tau_{eNB,i}^2 = \frac{b_i, max}{R_i, max}$$  \hspace{1cm} (10)

$R_i, max$ is the maximum number of PRBs that can be allocated per TTI. Then $\tau_{eNB,i}$ will be determined by the perceived throughput be represented by (11),
\[
\tau_{eNB,i} = \min \left( \tau_{eNB,i}^1, \tau_{eNB,i}^2 \right)
\]  

(11)

This approach allows for the game theoretic models to function based on the principles laid out in Chapter 5. However, \( R_{i,\text{max}} \) will need to be determined appropriately on the basis of parameters such as the number of concurrent users, the radio conditions and the achievable throughput. This could be challenging especially in a dynamic wireless environment.

**Case 3: Factoring in Buffer size in the game theoretic models**

Previous work of defining objective functions that factor in bottleneck buffer size have been explored in [99] [100]. As shown in [100] the evolution of the buffer queue size for a link is a function of the rate and the capacity. The objective function \( J_i(x, t) \) has been defined as the difference between the buffer occupancy and the utility function shown in (12),

\[
J_i(x, t) = \alpha_i D_i(t) x_i - U_i(x_i)
\]

(12)

Where, \( D_i \) is the total queuing delay, \( x_i \) is the link flow and \( U_i \) is the utility. The payoff function for the Equilibrium model can be adapted in-line with (12). However, key considerations in a wireless network need to be examined. For example, a high buffer occupancy is not desirable for any user but for a user in poor radio conditions (such as in a cell edge) high buffer occupancy cannot be avoided. In an LTE context, the running average and the buffer allocation can be used to determine \( D_i \) and \( \alpha_i \) can be as a regulator so that users in poor radio conditions are not penalised. Using the principles stated the users can make the decisions for Bid and Hold each TTI.

Factoring the buffer size into the game theoretic models is an interesting future research area and is advantageous to limit the latencies in the network thereby increasing the efficiencies. However, if the base station buffer size is very large (as is the case in LTE base stations) then it will be a negligible factor in determining the outcome of the game. In such cases, the base station buffer occupancy as determined by the running average in Chapter 5 will play a much greater role. Therefore, in this thesis base station buffer size is not a key factor in defining the objective functions.
Bibliography


[87] 3GPP, 3GPP TS 38.211 V15.7.0, 2019-09.


