A QoS-Aware Resource Allocation for M2M Communication with Balanced System Performance in LTE/LTE-A Network

THESIS

Submitted in partial fulfilment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

by

UPENDRA SINGH
(2017PHXF0428P)

Under the Supervision of

DR. AMIT DUA

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI

2023

Declaration of Authorship

I, Upendra Singh, declare that this thesis titled, 'A QoS-Aware Resource Allocation for M2M Communication with Balanced System Performance in LTE/LTE-A Network' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Date

____ / ____ / ____

UPENDRA SINGH

2017PHXF0428P

BITS Pilani, Pilani Campus

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI

(RAJASTHAN)

Certificate

Certified that Upendra Singh (2017PHXF0428P) has carried out the research work presented in this thesis entitled "A QoS-Aware Resource Allocation for M2M Communication with Balanced System Performance in LTE/LTE-A Network"

Date

____ / ____ / ____

____

DR. AMIT DUA

Associate Professor

BITS Pilani, Pilani Campus

Acknowledgements

"The proactive approach to a mistake is to acknowledge it instantly, correct and learn from it"

Stephen Covey

I want to express heartfelt and earnest gratitude to my Ph.D. supervisor, Dr. Amit Dua, Associate Professor, Department of Computer Science and Information Systems, BITS Pilani, India, for their unwavering guidance, encouragement, patience, and support throughout the tenure of my Ph.D. His vast research knowledge and experience in the field, deep understanding, and flexible and welcoming attitude have helped me throughout my research. It is truly an honor to be his student, and working under his mentorship has facilitated optimal growth of research and academics. I would also like to take this opportunity to express my profound gratefulness to the members of my Doctoral Advisory Committee (DAC), Dr. V. S. Shekhawat and Dr. Ashutosh Bhatia, for their continuous feedback, time, patience, and effort in reading my thesis and providing valuable feedback and suggestions.

I am forever grateful to the administration of BITS, Pilani, Pilani Campus. My sincere regards to Prof V Ramgopal Rao (Vice-Chancellor, BITS Pilani), Prof. Sudhir Kumar Barai (Director, BITS, Pilani), Col Soumyabrata Chakraborty (Retd) (Registrar) for facilitating me to pursue my research work at such an esteemed university. I also sincerely thank Prof. Shamik Chakraborty (Associate Dean, Academic – Graduate Studies & Research) for providing valuable support throughout my doctoral tenure.

I am grateful to be a part of the diverse and inclusive Department of Computer Science and Information Systems, BITS, Pilani. I extend my profound thanks to Prof.

Navneet Goyal (Head of the Department) for facilitating a conducive and healthy environment to conduct this research with great enthusiasm and encouragement. I would also like to thank all the department faculties for encouraging me by discussing their valuable and expert views during my mid-semester and end-semester presentations to the department. I thank the Department of Computer Science and Information Systems office staff, BITS Library, and other academic departments, who patiently welcomed my queries.

I cannot do without thanking my friends at the university, Mr. Victor Saha, Mr. Mayuresh Bhise, Mr. Ankit Agrawal, Mr. Nitin Singh Rajput, Mr. Shailesh Kumar, and Ms. Shail Saharan. They made the entire duration of my doctorate not only supporting and encouraging but also fun and adventurous. With their consistent support, the span of my doctorate seems to have flown away in no time.

I am blessed to have such amazingly supportive and loving family members: my mother, Mrs Basanti Devi; my brothers Pushpendra Singh and Hemendra Singh; my beloved wife, Mrs Pooja Singh; my princess Varnika Singh; and my best friend, Prashant Kumar Baheti. I wouldn't have come this far without their emotional and tangible support. Their patience and consistent love have helped me achieve such incredible feats.

Last but not least, every step of my journey would have been incomplete and inconsequential if it were not for the Almighty God, who made me capable of this work. Thank you, Lord, for all your Blessings and Mercy throughout this journey.

(Upendra Singh)

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI

(RAJASTHAN)

Abstract

DOCTOR OF PHILOSOPHY

A QoS-Aware Resource Allocation for M2M Communication with Balanced System Performance in LTE/LTE-A Network

By:

UPENDRA SINGH

M2M communication is autonomous communication among devices without human intervention. For example, the motion sensor detects the person moving toward the door and opens the door automatically. M2M systems were vendor-specific and designed for working in a restricted coverage area. These systems lack interoperability, wide-area communication, security, reliability, and availability. With the immense growth in the number of connected devices and cellular technology, M2M communication is gaining the attention of academia and R&Ds. Today's M2M system has a presence in vast domains, from emergency services to home entertainment, i.e., industry 4.0, smart homes, e-health, and smart-city applications. Thus, the M2M system requires reliable, robust, and high-speed communication.

Long-term evolution (LTE) is a wireless communication technology standard that supports many devices' high-speed, reliable, secure, and wide-area communication. The development of 3GPP's LTE cellular network drives the M2M communication in the LTE network. Various standardization bodies like 3GPP, ETSI, and oneM2M worked to standardize the M2M communication under the LTE/LTE-Advance network. The M2M communication differs from conventional cellular communication in terms of packet size, traffic pattern, many devices, a vast range of applications, and

QoS requirements. Therefore, using LTE for M2M communication is challenging as the LTE network is optimized for traditional cellular communication (or human-to-human (H2H) communication).

Through the surveyed literature and experimentation, we found that focusing on channel quality in resource allocation provides poor application priority and delay-budget support in a heterogeneous environment. Focusing on the application priority degrades the system performance regarding throughput and radio resource utilization. There is a trade-off between system performance and QoS support. As the M2M devices generally have zero or limited mobility, it restricts the changes in channel conditions. The QoS class identifier (QCI) priority is also application-dependent and static. Thus, devices with poor channel conditions and lower priority can face starvation. This also leads to unfair resource sharing. Moreover, we observed that malfunctioning devices in the M2M system can destabilize the system. Therefore, there is a requirement for a scheduling mechanism that provides a stable system with balanced system performance and QoS support. A resource allocation mechanism needs to consider the priority, channel quality, number of devices, and environment heterogeneity to provide balanced performance between system performance and QoS support with resource-sharing fairness. We worked on the QoS-aware uplink radio resource allocation methodologies for M2M communication in the LTE/LTE-A network through the thesis, which provides balanced system performance and QoS support in a large and heterogeneous environment.

This thesis provides a foundation for machine-to-machine communication. We found that M2M communication systems have an extensive range of QoS classes. We explore the LTE radio resource architecture, packet scheduler, and scheduling mechanisms and constraints. In addition to this, we worked on the classification of scheduling mechanisms and research gaps from the current literature. Several scheduling proposals were presented in the literature to allocate radio resources using metrics with combination or individual like channel quality, application priority, delay preferences, etc. The exploration of the literature ends with problem formulation and research objectives for the thesis. We adopt the smart-building scenario where many different applications are contending for scheduling grants. M2M communication has many device deployments; some can malfunction due to arbitrary faults.

We divided the research work into two stages as per our research methodology. We validated the trade-off problem at the first stage, and later, we proposed a scalable priority-based, weighted priority proportional fair (WPPF) scheduling to provide a stable communication system in a heterogeneous environment. We have constructed an auction game model at the first stage. The constructed auction game model is integrated with an application priority-based resource allocation mechanism to validate the system performance and QoS support trade-off. The auction game model charges malfunctioning devices with total transmission opportunities. Performance evaluation of proposed scheduling performed against different system performance parameters and QoS parameters. The priority-based solution provides better QCI priority support. However, it degrades the radio resource utilization while using QCI priority only as an allocation metric.

We found that only priority-based allocation is insufficient as it degrades system performance by observing the results of the first stage. The application's priority-based solutions cause unfair distribution of resources and starvation for the devices having poor channel conditions and lower application priority. Moreover, it is unsuitable for different application priorities regarding delay-budget satisfaction. In the later stage, we proposed an improved uplink radio resource allocation mechanism based on the scalable priority jointly optimized with channel quality.

The proposed weighted priority proportional fair (WPPF) mechanism improves throughput and resource utilization using joint optimization of channel quality and application priority. The scaling of application priority concerning the number of scheduling grants, devices, and QCIs improves resource-sharing fairness and minimizes starvation. The proposed scheduling mechanism improves delay-budget satisfaction for

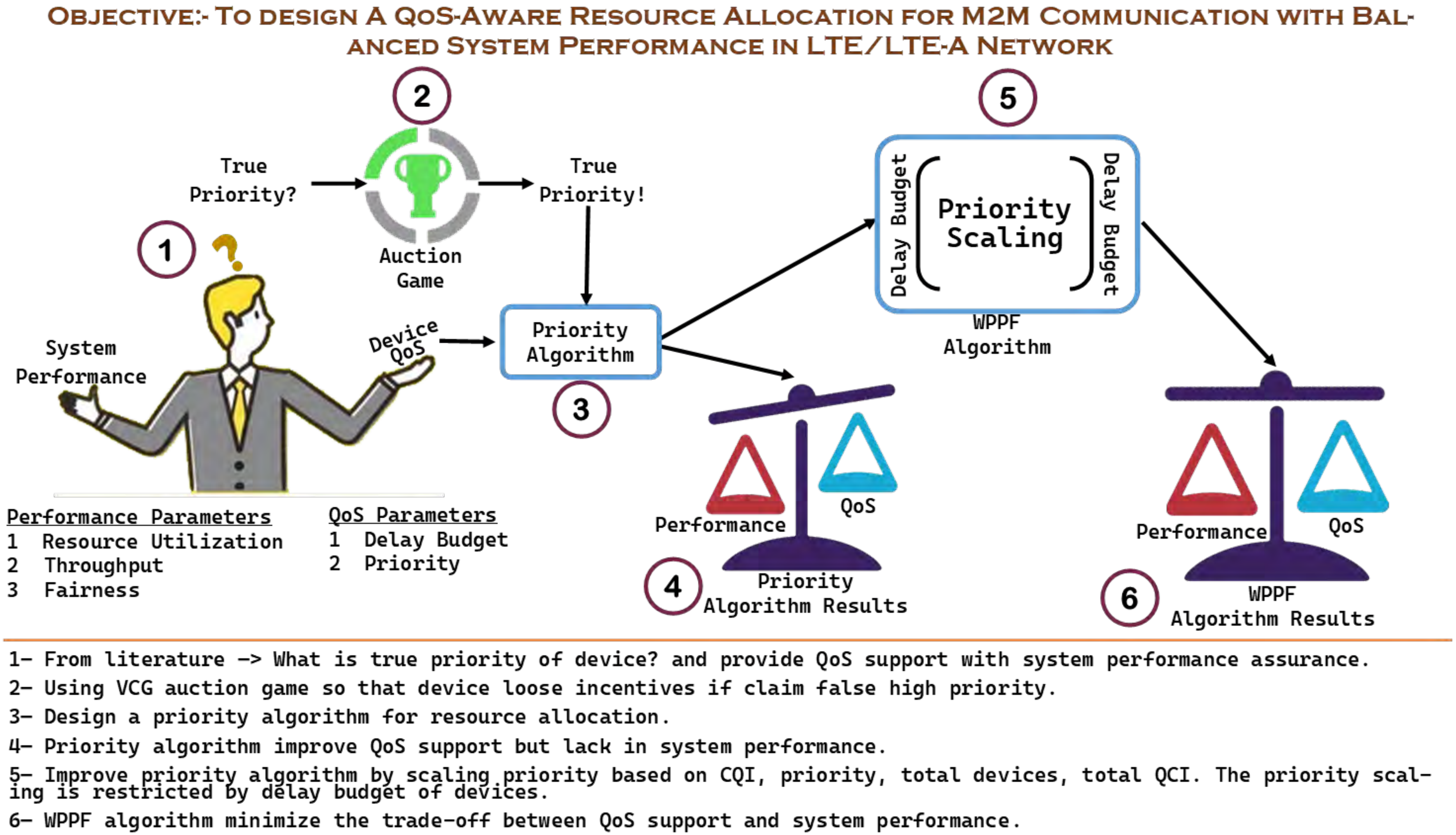


Figure 1: Graphical abstract of the thesis.

many QCIs using a delay-budget constraint in the allocation mechanism. The proposed WPPF approach provides a balanced performance regarding resource utilization, throughput, fairness, application priority, and delay-budget satisfaction. The simulation of proposed solutions has been performed on MATLAB R2020b with LTE and the 5G NR tool and improves state-of-the-art schedulers.

We have selected resource-sharing fairness, throughput, and resource utilization as system performance evaluation metrics, communication delay, and priority preferences violation as QoS support evaluation metrics to evaluate balanced system performance and QoS support. The results show that the proposed algorithm performs best in delay budget satisfaction, fairness, and better QCI priority support. The channel quality and QCI priority combination improves the proposed algorithm's resource utilization and average cell throughput.

Keywords: M2M communication, LTE, resource allocation, scheduling, Quality-of-Service, resource utilization, virtual QCI, game theory.

Abbreviations

3GPP	3rd Generation Partnership Project
Ack	Acknowledgement
BSR	Buffer Status Report
CP	Cyclic Prefix
CQI	Channel Quality Indicator
E2E	End-to-End
eNB	Evolved Node B
ETSI	European Telecommunications Standards Institute
FDPS	Frequency Domain Packet Scheduler
GBR	Guaranteed Bit Rate
H2H	Human-to-Human
HARQ	Hybrid Automatic Repeat Request
LTE	Long-Term Evolution
M2M	Machine-to-Machine
MCS	Modulation and Coding Scheme
MIMO	Multiple Input Multiple Output
MTC	Machine Type Communication
MTCD	Machine Type Communication Device
MTCG	Machine Type Communication Gateway
N-GBR	Non Guaranteed Bit Rate
Nack	Negative Acknowledgement
NOMA	Non Orthogonal Multiple Access
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
OMA	Orthogonal Multiple Access

PRB	Physical Resource Block
QCI	QoS Class Identifier
QoE	Quality-of-Experience
QoS	Quality-of-Service
RSRP	Reference Signal Received Power
SC-FDMA	Single Carrier - Frequency Division Multiple Access
SINR	Signal-to-Interference plus Noise Ratio
TBS	Transport Block Size
TDPS	Time Domain Packet Scheduler
TTI	Transmission Time Interval
UE	User Equipment
vQCI	virtual QoS Class Identifier

Contents

List of Figures

List of Tables

Dedicated to my father
Late Shri Mukhtyar Singh Faujdar

Chapter 1

Introduction to M2M

The content of this chapter is partially published in the following articles.

U. Singh, A. Dua, S. Tanwar, N. Kumar and M. Alazab, "A Survey on LTE/LTE-A Radio Resource Allocation Techniques for Machine-to-Machine Communication for B5G Networks," in IEEE Access, vol. 9, pp. 107976-107997, 2021, doi: 10.1109/ACCESS.2021.3100541. **[Q1, SCIe, IF 3.367]**
https://ieeexplore.ieee.org/document/9497084

U. Singh, A. Ramaswamy, A. Dua, N. Kumar, S. Tanwar, G. Sharma, I. E. Davidson, R. Sharma, "Coalition Games for Performance Evaluation in 5G and Beyond Networks: A Survey," in IEEE Access, doi: 10.1109/ACCESS.2022.3146158. **[Q1, SCIe, IF 3.367]**
https://ieeexplore.ieee.org/document/9691369

1 | Introduction to M2M

"Nothing in life is to be feared, it is only to be understood. Now is the time to understand more so that we may fear less."

Marie Curie

Key Points:

- Data communication involving one or more entities without human involvement is known as machine-to-machine communication.
- New generation network technologies like 4G and 5G and single chip systems accelerate the growth of the machine-to-machine (M2M) systems from industry to a smart home.
- For the scalable development of M2M communication, standardization bodies felt the requirement for a generic horizontal M2M architecture as a common platform for M2M application development.
- The M2M communication system consists of three domains: Machine type communication devices (MTCDs) Domain, Network Domain, and Application Domain.
- M2M systems have many applications like transportation, smart home, smart city, e-Health, smart metering, etc.
- QoS of MTCD depends on various parameters such as delay budget, throughput, and application priority, and these parameters have an extensive range of values.
- Resource allocation for the M2M system in new generation networks is challenging due to the vast QoS requirement of the M2M system.

1.1 Introduction

Machine-to-machine(M2M) communication is a paradigm in which the devices can communicate autonomously without the intervention of humans or with minimal intervention of humans. For example, switching on a light bulb by a human is a kind of human-to-machine (H2M) communication, whereas the detection of the human by a motion sensor and switching on the lights automatically is a Machine-to-Machine communication [1, 2, 3]. Here are some definitions of M2M communication from the literature.

"Machine-to-machine (M2M) communications are used for automated data transmission and measurement between mechanical or electronic devices. The key components of an M2M system are Field-deployed wireless devices with embedded sensors or RFID-Wireless communication networks with complementary wireline access including, but not limited to, cellular communication, Wi-Fi, ZigBee, WiMAX, wireless LAN (WLAN), generic DSL (xDSL) and fiber to the x (FTTx)." ***Gartner, Inc [4]***

"M2M communications refer to automated applications that involve machines or devices communicating through a network without human intervention. Sensors and communication modules are embedded within M2M devices, enabling data to be transmitted from one device to another device through wired and wireless communications networks." ***Department of Telecommunications, GoI [5]***

"Machine-to-machine (M2M) communications are the communication between two or more entities that do not necessarily need any direct human intervention. M2M services intend to automate decision and communication processes." ***European Telecommunications Standards Institute (ETSI) [6]***

"Machine-type communication is a form of data communication that involves one or more entities that do not necessarily need human interaction." ***3rd Generation Partnership Project (3GPP) [7]***

The rapid development of a variety of smart machines, i.e., communication devices,

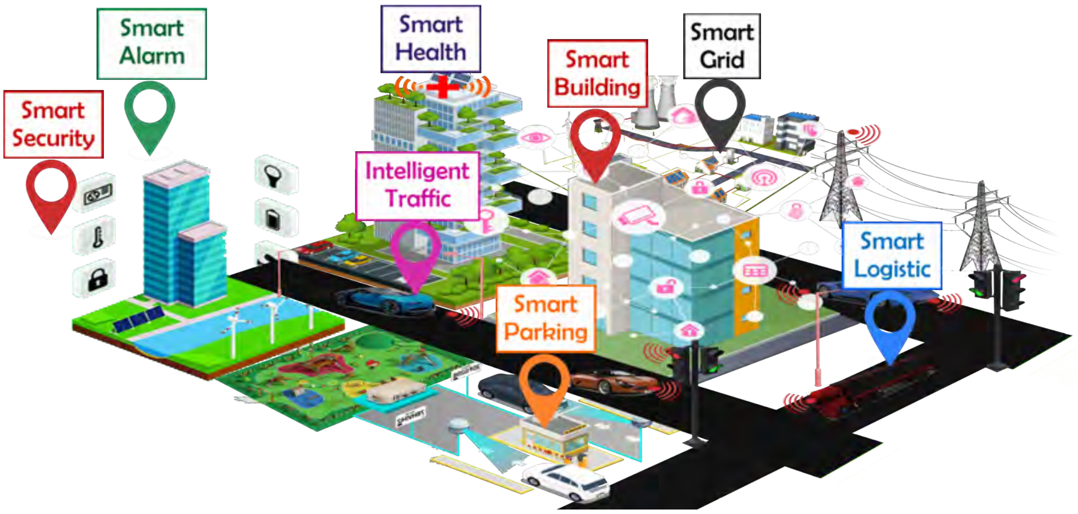


Figure 1.1: M2M communication in smart city environment

home appliances, vehicles, industrial equipment, security systems, and several applications like the infotainment system, entertainment services, surveillance, etc., make M2M communication a dominant system for ease of living and work for humans [8, 9]. So, it's likely to adopt the new communication technologies, services, and standardization of the M2M communication system for better performance, security, stability, and scalability. Figure 1.1 shows M2M communication in the smart city environment.

Through the interfacing of the M2M system with wireless sensor networks (WSN), a wide range of information can be collected by sensors, and machines can also take action through the collected information. M2M systems can be upgraded to CPS (Cyber-Physical systems). A CPS is composed of computational and physical capabilities. It's designed to act like a network of multiple variables with both physical input and output – rather than a standalone technology. CPS is an evolution of M2M with the capabilities of decision-making and automated control under the Internet of Things (IoT) [10].

The M2M communication differs from traditional H2H communication in terms of the packet size, traffic pattern, limited capacity of the device, delay bound, and larger application domain [11, 12, 13]. M2M devices are usually tasked to gather information from their surroundings and forward it to a server/computer for further processing. Thus, most of their communication is towards the uplink, which enforces the continuous allocation of radio resources for an individual device. Some devices deployed for critical information sensing may be delay-bound, such as intruder detection or disaster alarm. To preserve specific QoS for such devices or to send data before it becomes obsolete, the data should be transmitted within the specified delay budget [14, 15]. M2M communication has specific characteristics as follows.

- A massive number of connected devices generate massive data.
- Periodic or event-driven packet generation.
- Small packet size but the frequent generation of packets.
- Wide range of delay and throughput requirements.
- A Vast variety of applications promulgate a more extensive range of service requirements.

Traditional mobile communication differs from M2M communication against the above specific requirements [16].

1.1.1 Evolution of M2M communication

The growing adaptation of advanced communication technology like 4G and 5G and advancement in device connectivity encourages the research community to take advantage of the LTE network for M2M communication. Security and privacy issues, device capacity enhancement, and high-end application development direct attention toward M2M communication [1, 2, 3].

The term M2M communication is not new. It started in the year 1845 with the invention of the Russian Military's information exchange system. This was an elementary wired system for data transfer. A duplex radio communication network for data transmission in the 1900s followed it. Wired communication was used to exchange information among devices in the early 20th century. Later in the 20th

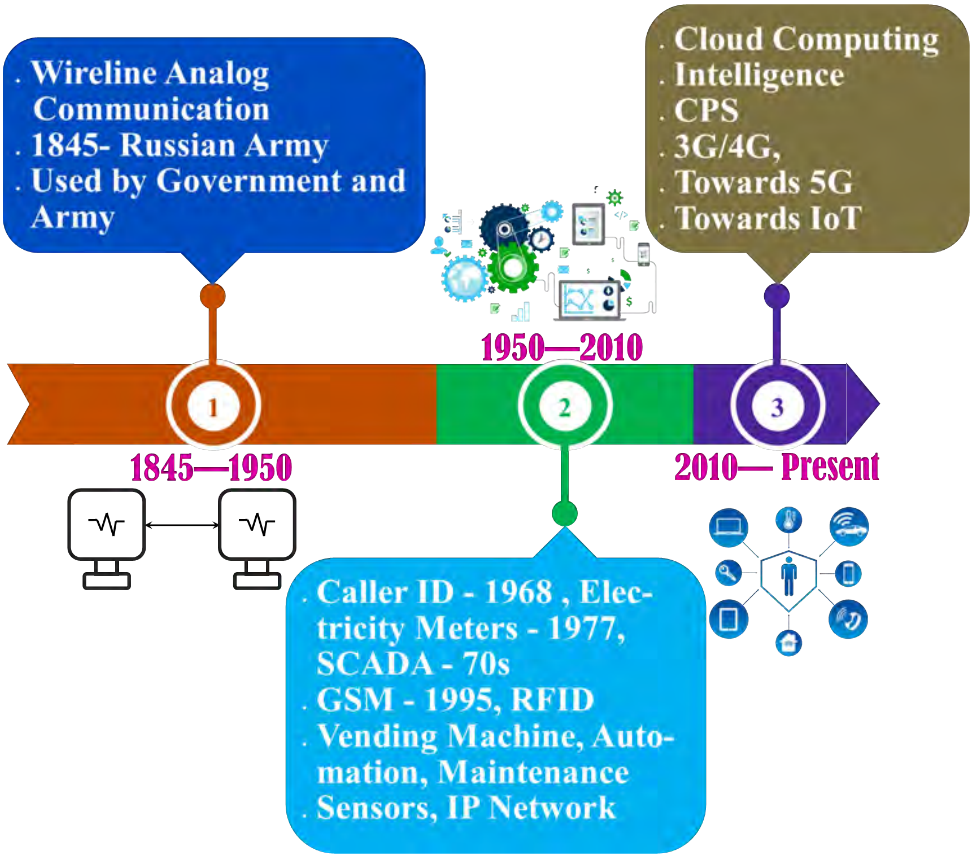


Figure 1.2: Evolution of M2M communication

century, M2M communication became more sophisticated with the advancement in computer networking and cellular communication. M2M communication expanded to applications like industrial automation, telemetry, Supervisory Control and Data Acquisition (SCADA), and many more. SCADA combines software and hardware for automating and monitoring industrial processes. In the early age of M2M communication, it was implemented through the wire-line channel. However, after the invention of the Global System for Mobile Communications (GSM), the 2nd Generation of cellular network technology in 1995, it became mature and grounds in countless applications [17, 18, 19].

At the beginning of the 21st century, cellular communication technology advanced and proposed new communication technologies named 3G and 4G LTE, which started to provide high-speed and secure data transmissions at a lower cost per bit [20]. With the advancement in cellular technology, the Internet and single-chip systems witnessed a great surge in the growth of M2M communication towards the Internet of Things (IoT) [10, 21, 22]. Both the number of connected devices and the market grew bigger. Figure 1.3 shows the number of connected devices growth. Research advisory firms Statista and CISCO have predicted that the number of connected devices will grow to 75.44 billion by 2025, which was 15.41 billion in 2015. Market advisory firm Mordor Intelligence forecasts that the M2M market will rise to USD 26.52 billion by 2025, which was USD 19.18 billion in 2019 [22, 23, 24].

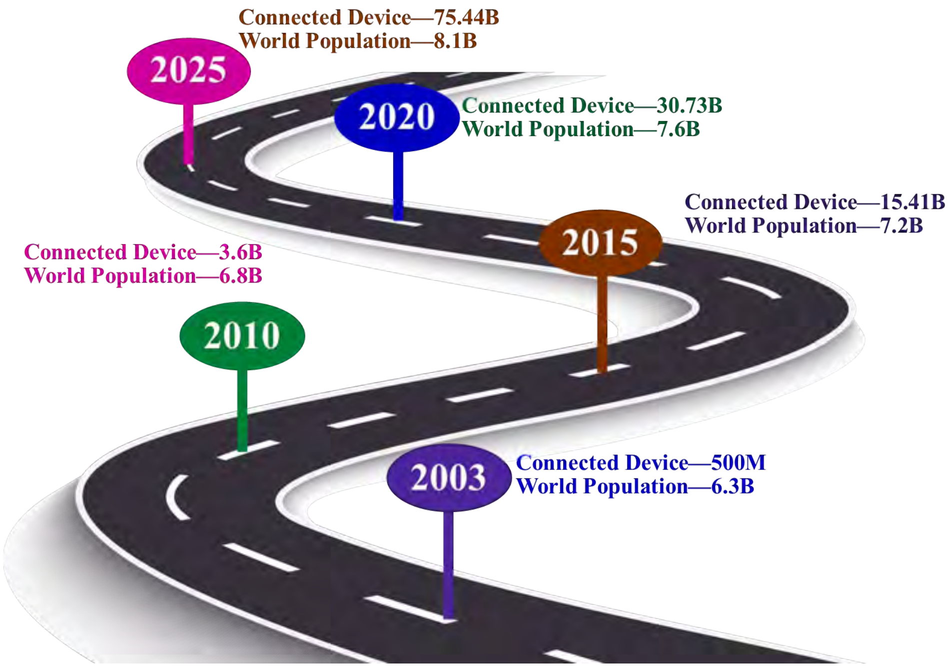


Figure 1.3: Growth in the number of connected devices

1.1.2 Milestones in M2M evolution

Important advancements in scientific innovations paved the way for M2M communication as we know it today. Following are the few marks that advanced the M2M communication as of now.

1. **Process Monitoring:** In the 1970s, software applications were developed to control industrial processes, gathering data in real-time from remote locations to control equipment and conditions. For example, SCADA (supervisory control and data acquisition) provides organizations with the tools to make and deploy data-driven decisions regarding their industrial processes.

2. **Remote Connectivity:** M2M systems had local connectivity. The development of GSM and IP networks in the 1990s, including GSM and IP networks in the M2M system, enabled remote monitoring of industrial processes.

3. **High Processing Capacity:** After 2010, M2M communication started adding information processing on its own, i.e., Cloud Computing, which helps to store extensive data and provide complex processing facilities for the M2M communication system. It is anticipated as a new Cyber-Physical System (CPS) area with intelligence integration into M2M communication [10].

4. **Robust and High Data-Rate Connectivity:** To overcome the limitation of wired networks (single point of failure) and the requirement of high data rate due to massively connected devices generating the requirement of integrating M2M communication with LTE-A / 5th Generation-New Radio (5G-NR) Network to support diverse requirements, including traffic, the number of devices, etc. Figure 1.2 shows the evolution of the M2M communication system.

1.2 Requirements of M2M Standardization

In the past, many telecommunication operators proposed commercial solutions for M2M communication in different parts of the world. These solutions were application-specific and were introduced as vertical (vendor-specific full stack development) M2M architecture or tools for a particular M2M application. For the scalable development of M2M communication, standardization bodies felt the requirement for a generic horizontal M2M architecture as a common platform for M2M application development. The standardization bodies broadly focus on the following requirements.

- **Interoperability:** Vendor-specific M2M systems have limitations in the interconnection of the devices developed by different vendors due to different communication protocols and mechanisms. For example, one device from vendor

"X" may not connect to another from vendor "Y" in group communication.

- **Sustainability:** For a sustainable M2M system, components of the system should operate properly and for the optimum duration. For example, Machine-type communication devices(MTCDs) should be power efficient to work for a long duration with a small battery.

- **Scalability:** As the number of connected devices is increasing rapidly, when additional devices are added to the M2M system at different geographical locations, it can cause a requirement for WAN connectivity instead of local connection and increased traffic requires a high data rate. The M2M system should be scalable to support the system's growth. For example, process automation and monitoring components are locally connected inside an automobile company premises. Another unit of the same company is newly started at a remote location. The connection media should be scalable to interconnect both units for process coordination and to monitor centrally.

- **Robust and Wide Area Coverage:** The M2M system is everywhere, from industrial automation to smart homes in the current scenario. To support this, robust and wide area connectivity is required, i.e., LTE/ 5G NR.

- **Privacy and Security:** Data privacy and security are key requirements for a stable M2M system.

1.3 M2M Standardization Initiatives

Various connectivity solutions are used in M2M in wired and wireless modes. Compared to wired solutions, wireless solutions offer greater ease of deployment and improved resilience against single points of failure. Cellular connectivity in the M2M system provides various advantages, including broader coverage and interoperability [25].

Different standardization bodies like the 3rd-Generation Partnership Project (3GPP) and European Telecommunications Standards Institute (ETSI) devote their efforts and propose reference architecture to support M2M communication in 4G and 5G networks to achieve robust communication. The ETSI aims to provide an end-to-end view of M2M standardization. In Aug 2010, ETSI published the first listing of general provisions for M2M service, followed by the functional requirements for M2M communication [26].

The 3GPP focuses on the recommendation for communication of MTCDs. The work has been carried out on optimizing access and core network infrastructure, allowing efficient delivery of M2M services. The 3GPP releases 8 and 9 included the first standard to support M2M communication. In 3GPP, M2M communication is called MTC (Machine type communication) [24]. More specific methods have been proposed in 3GPP release 12 for M2M communication, such as privacy, power control, group management, and service maintenance [27].

A cooperative effort of seven standardization institutes worldwide created a unique partnership for the standardization of M2M in 2012 called oneM2M. OneM2M's standard specifications enable an ecosystem to support various services and applications, including connected cars, smart grids, connected cities, automated homes, public safety, and health. In the oneM2M-standardized architecture, an IoT Service Layer is defined as a vendor-neutral software Middleware that sits between processing and communication hardware and IoT applications, offering a set of functions that IoT applications frequently require. The oneM2M Service Layer offers independent use-case functions [28].

1.4 M2M Reference Architecture

The M2M communication system consists of three domains: MTCDs Domain, Network Domain, and Application Domain [29, 30].

MTCDs Domain: MTCDs domain consists of devices, i.e., sensors, actuators, metering devices, Machine Type Communication Gateways (MTCG), etc. [29, 30]. For communication among devices, the M2M area network provides low-range connectivity among the MTCDs and MTCG using communication technologies like Smart-BLE, ZigBee, WiFi, Ultra-Wide-Band (UWB), etc. BLE (Bluetooth Low Energy, also known as Bluetooth Smart) is a power-conserving variant of Bluetooth Personal Area Network (PAN) technology. BLE is designed to use less power and cost less money than Classic Bluetooth. Ultra-wide-band (also known as UWB or ultra band) is a radio technology that enables short-range, high-bandwidth communications over a significant chunk of the radio spectrum using very little energy [31, 32, 33].

Network Domain: The network domain consists of the core network and the access network. A core network is a fixed, high-speed, intensively used communications network. It is somewhat analogous to a network of motorways and major trunk roads. Core networks often interconnect with other core networks. For example, all the mobile operators' core networks interconnect with the PSTN(Public Switched Telephone Network) core network. An access network links the end-user's equipment to the core network via a local exchange or local radio node. The access network is analogous to the minor roads that provide access to motorways and other trunk routes.

M2M devices can connect to an access network in three ways: direct, indirect, and hybrid, as shown in figure 1.4. The MTCD can communicate directly with the evolved NodeB (eNB). In indirect connection, the Machine Type Communication (MTC) gateway or cluster head/coordinator is responsible for transmission between eNB and UE (User Equipment). In contrast, the rest of the devices in the cluster communicate with that MTC gateway or cluster head/coordinator [31, 32, 33].

In hybrid communication, the device can communicate with the eNB directly and through a gateway. The network domain provides the communication services between the MTCG and the application server (Indirect Connect) or between the MTCD and the application server (Direct Connect) by using any wired or wireless WLAN

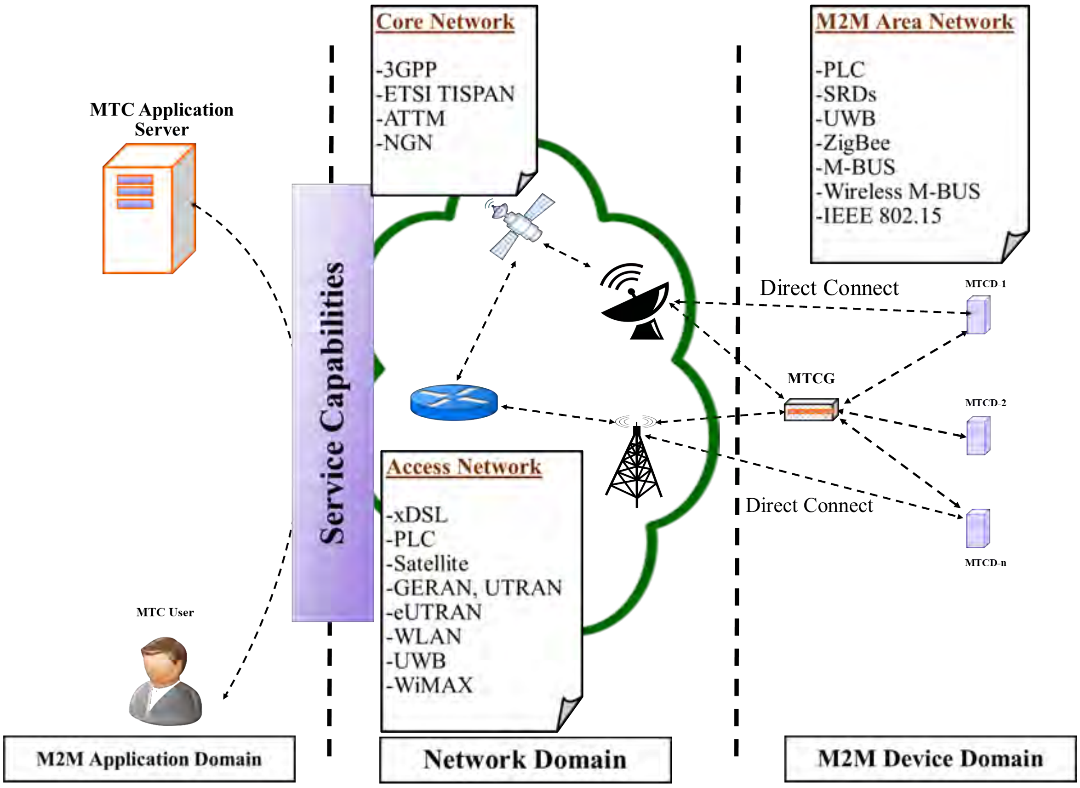


Figure 1.4: M2M communication system architecture [34]

network technology, e.g., Satellite, eUTRAN, WiMAX, etc [35, 34, 36].

Application Domain: The application domain provides a facility for users to access the information gathered by the MTCDs. The service capability layer provides access between the user and the application server. Thus, the M2M communication system enables end-to-end connectivity between the MTCD and the application server. Figure 1.4 shows the M2M communication system architecture as specified by the European Telecommunications Standards Institute (ETSI) [31, 32, 33].

1.5 M2M Application Domain

M2M systems are continually developing and covering more application areas. The automotive sector is gaining more attention, leading to the emergence of several applications [18, 6, 37]. The applications found in the literature are categorized into five groups according to the application area: transportation, smart home, smart city, e-Health, and smart metering, as shown in figure 1.5.

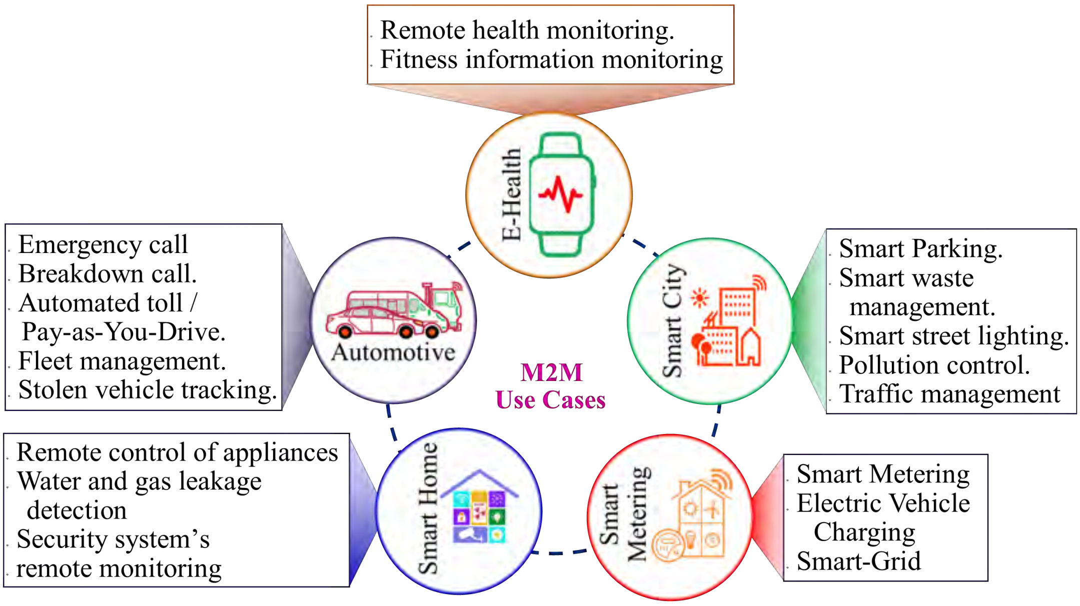


Figure 1.5: M2M applications and use case

1.5.1 Transportation

This category of M2M application encompasses all the applications related to vehicles or intelligent transportation systems. For the implementation, each vehicle has some communication modules, such as GPS and single-chip systems, enabling communication with remote servers. Some of the main applications in this domain are emergency calls, breakdown calls, automated toll / pay-as-you-drive, fleet management, and Stolen vehicle tracking [6, 38, 39].

1.5.2 e-Health

This type of application is useful in monitoring a person's health remotely. Usually, a person wears sensors like a smartwatch, which monitors a person's health, such as blood pressure monitoring, heart rate monitoring, etc.; due to limited resources and processing power, these sensors send the gathered data to the MTCG, usually a smartphone. The MTCG collects the data and sends it to the remote server of the e-Health service via eNB, where healthcare professionals analyze the data and act accordingly. These applications work like Body Sensor Network except the communication is bidirectional [40, 41, 42]. Examples of applications in this category are Remote health monitoring and fitness information monitoring.

1.5.3 Utility applications

M2M applications in this category are responsible for efficient water, electricity, and gas use through smart metering devices. These meters are an essential part of the smart grid. The major applications in this category are smart meters, automated electrical vehicle (EV) charging points, and smart grids [43, 44].

1.5.4 Smart city

This category of M2M applications belongs to the applications developed and deployed to make easy and smooth access to service to the citizens and save energy and cost. Sensors are deployed across the city, and MTCD will send data to the gateway, which is then forwarded to the MTC server for further analysis [41, 42, 45]. Some significant applications are smart vehicle parking, smart city-waste management, smart streetlights, pollution control, and smart traffic management.

1.5.5 Smart home

These applications are developed to provide comfortable living to home users with remote detection and execution of particular tasks such as Smart WiFi Plug [46, 41, 42]. Some of the significant M2M applications for the smart home are remote control of appliances, water and gas leakage detection, and security system remote monitoring.

The applications mentioned above are not the only applications of M2M. The application domain of M2M is far more significant and vaster than specified in this section.

1.6 End-to-End QoS for M2M Communication

Quality-of-service (QoS) ensures the subscribers' communication experience and helps the network work efficiently when many users access the network simultaneously. So, building a QoS architecture for M2M communications has become essential. However, it is challenging to define a single QoS class for varying M2M services [47]. The varying QoS requirements of MTCDs affect the radio resource allocation. QoS of MTCD depends on various parameters such as delay budget, throughput, and application priority, and these parameters have an extensive range of values. For example, delay budgets range from 100ms to 1 day [48]. The various applications and services' QoS requirements must be quantified in terms of criteria that indicate desired performance levels to be satisfied. Some examples of these metrics are throughput, latency, jitter, and packet loss [49, 48]. The following are the key quantitative metrics listed by 3GPP.

Throughput: Characterized through the guaranteed bit rate. After bearer establishment or modification, the guaranteed bit rate (GBR) and allotted fixed network resources remain unchanged. As a result, this data flow is a guaranteed service.

Delay: The 3GPP categorizes delays, with 75 ms being the lowest and 1 second being the highest. For delay-tolerant applications, the latter value is chosen. Packet loss: Defined as the packet error loss rate, similar to the packet delay budget.

Priority: Allocation and retention of the service data flow are prioritized according to the allocation/retention priority (ARP) parameter, which is utilized to specify this. If there are conflicts in demand for network resources, the ARP determines whether a bearer establishment or modification request can be approved or denied.

3GPP defines QoS Class Identifier (QCIs) for M2M communications over LTE, as shown in table 1.1.

Table 1.1: 3GPP QCI for M2M over LTE (Taken from [48])

QCI	Bearer Class	Priority	Delay Budget	GBR	Example Service
1	GBR	2	1s	25 kbps	Delay and Throughput sensitive bandwidth applications - Actuators.
2	GBR	4	1s	10 kbps	Delay sensitive bandwidth applications - Sensors.
3	GBR	3	10s	50 kbps	Higher bandwidth applications - Infotainment Applications.
4	GBR	5	10s	25 kbps	High bandwidth applications - e-Health.
60	GBR	0.7	75 ms	25 kbps	Mission Critical event triggered applications.
65	GBR	2	100ms	25 kbps	Non-Mission-Critical vent triggered applications.
5	Non-GBR	1	100ms	—	Time sensitive application - Instant presence sensors.
6	Non-GBR	6	1 minute	—	Low demanding applications - Temperature monitoring .
7	Non-GBR	7	10 minutes	—	Low demanding applications - non-critical monitoring.
8	Non-GBR	8	1 hour	—	Low demanding applications - smart metering.
9	Non-GBR	9	1 day	—	Low demanding applications - Utility applications.
69	Non-GBR	0.5	60 ms	—	Mission Critical delay sensitive signaling (e.g., MC- PTT signaling).
70	Non-GBR	5.5	200 ms	—	Mission Critical Data transfer.

1.7 Contribution of the Thesis

Radio resource scheduling for M2M in LTE has scalability and stability issues regarding the massive number of connected devices, standard compatible scheduling, and a trade-off between system performance and QoS priority support. We have presented a novel formulation of the priority decision problem as an action game to stabilize the M2M system in a heterogeneous environment. This game-theoretic model is integrated with the priority scheduling mechanism to validate the M2M system performance and QoS support trade-offs. The proposed game-theoretic model facilitates getting true application priority for communicating M2M devices in the presence of malfunctioning or intruder devices in the network.

Through the simulation of the proposed scheduling scheme, we absorbed that the resource utilization for all implemented algorithms observes a slack saturation after 300 devices and shows a logarithmic behavior. The algorithms that consider channel quality as an allocation metric provide high throughput. The average cell throughput improves with the number of devices but is bounded by Shannon's capacity. QCI priority violation increases rapidly for all the algorithms other than the priority-based algorithm. The simulation results show that the proposed priority algorithm performs poorly in throughput and resource utilization. It gives average results in resource-sharing fairness. The priority-based algorithms perform better than others in priority support and delay budget violations.

After validating the scheduling issues with an application priority-based scheduling scheme, we proposed an improved scalable priority-based 3GPP compatible resource scheduling methodology to overcome scalability and trade-off issues. The proposed scheduling scheme provides balanced system performance and QoS priority support to the M2M system in a heterogeneous environment. The proposed scheduling methodology uses the weighted priority to improve system performance and scale the priority concerning the number of devices and QoS classes to minimize the low resource utilization and starvation issues.

Using a combination of channel quality and QCI priority improves the proposed algorithm's resource utilization and average cell throughput performance. The proposed WPPF algorithm provides better QCI priority and delay-budget violation results. The proposed WPPF algorithm performs more than average in throughput, resource utilization, and priority support. The WPPF algorithm performs best in fairness and delay-budget violation. The improved resource-sharing fairness also minimizes the cases of starvation. The proposed WPPF algorithm perfectly balanced the system performance and QoS support for M2M communication in a heterogeneous environment.

1.7.1 Thesis organization

This thesis has four key sections: M2M communication background, LTE resource scheduling mechanism, QoS aware resource scheduling, and summary, which are spread over five chapters and organized as follows.

- Chapter 1 ***Introduction to M2M*** provides foundation for M2M communication from evolution to standardization. The chapter explains the challenges and QoS requirements for M2M in LTE networks.
- Chapter 2 ***Background Information and Literature Survey*** expounds LTE network architecture and radio resource scheduling concepts.
- Chapter 3 ***Decision on QoS Class Identifier(QCI): An Auction Model*** provides a QoS aware resource scheduling scheme for M2M communication in LTE network.
- Chapter 4 ***Balancing Performance and QCI support for M2M*** provide weighted priority proportion fair resource scheduling scheme based on the virtual QoS class identifier.

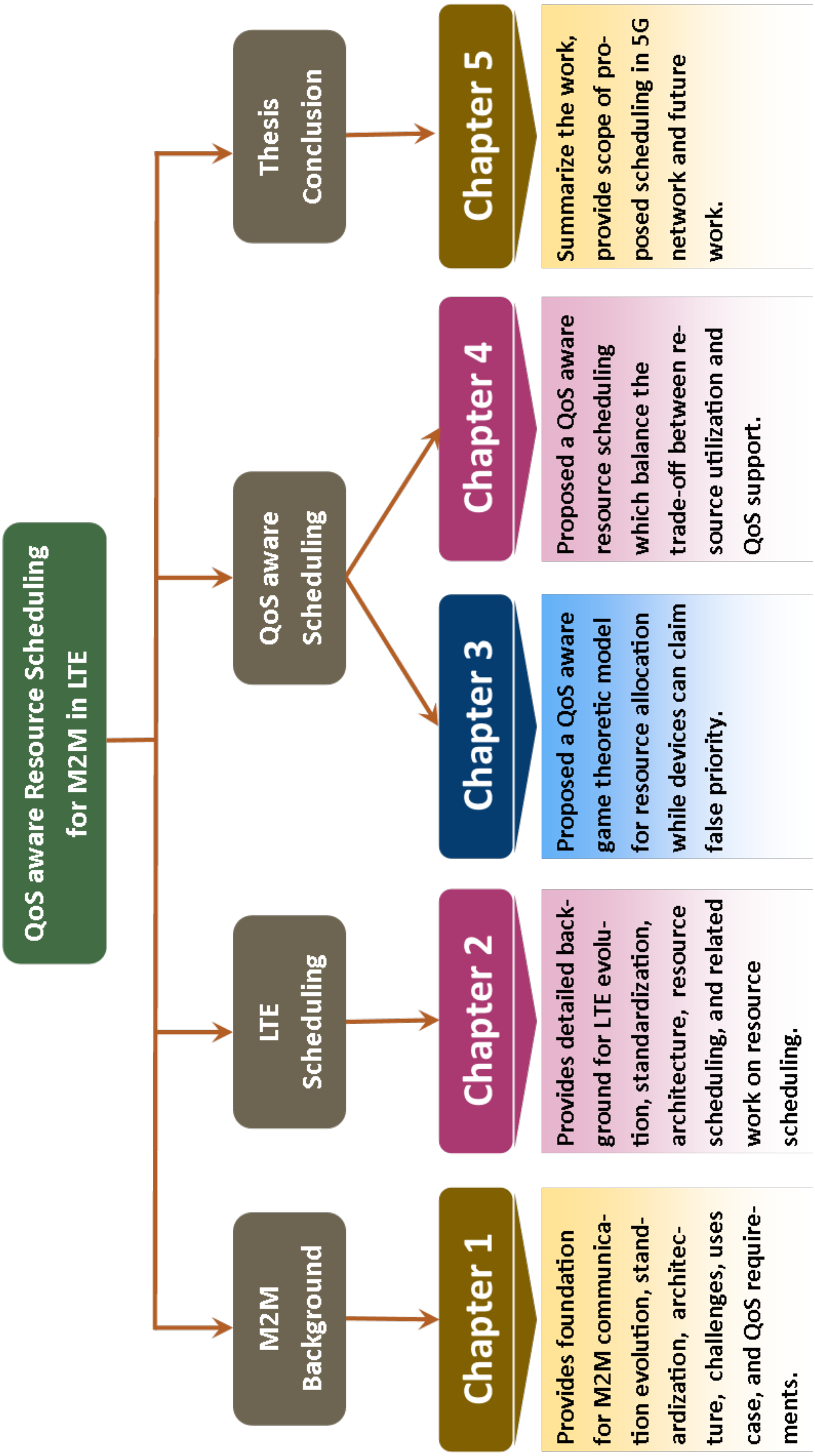


Figure 1.6: Thesis organization.

- Chapter 5 ***Thesis Summary*** summarize the work through this thesis. Moreover, this chapter provides future research direction and scope of the proposed scheduling schemes in 5G NR.

Figure 1.6 shows the organization of the thesis.

1.8 Chapter Summary

- With the emergence of applications such as online video streaming in HD and online gaming, which require a high-speed information exchange over the mobile network, high-speed communications have become an essential part of daily life.
- Machine-to-machine (M2M) communication is rapidly developing to include many devices/machines/terminals, including mobile phones, personal computers, laptops, TVs, speakers, lights, and electronic appliances.
- With the dramatic penetration of embedded devices, M2M communications have become a dominant communication paradigm.
- Applications like video on demand, surveillance, monitoring systems, and emergency communication impose the need to expand existing H2H (human-to-human) communication to H2M or M2M communication.

Chapter 2

Background Information

and

Literature Survey

The content of this chapter is partially published in the following articles.

U. Singh, A. Dua, S. Tanwar, N. Kumar and M. Alazab, "A Survey on LTE/LTE-A Radio Resource Allocation Techniques for Machine-to-Machine Communication for B5G Networks," in IEEE Access, vol. 9, pp. 107976-107997, 2021, doi: 10.1109/ACCESS.2021.3100541. **[Q1, SCIe, IF 3.367]**
https://ieeexplore.ieee.org/document/9497084

U. Singh, A. Dua, N. Kumar and M. Guizani, "QoS Aware Uplink Scheduling for M2M Communication in LTE / LTE-A Network: A Game Theoretic Approach," in IEEE Transactions on Vehicular Technology, doi: 10.1109/TVT.2021.3132535. **[Q1, SCI, IF-5.987]**
https://ieeexplore.ieee.org/document/9635654

U. Singh, A. Dua, N. Kumar, S. Tanwar, R Iqbal, M. Hijji, R. Sharma, "Scalable Priority-based Resource Allocation Scheme for M2M Communication in LTE/LTE-A Network," in Computers and Electrical Engineering,2022 Oct 1;103:108321 **[Q1, SCIe, IF-4.15]**
https://www.sciencedirect.com/science/article/pii/S0045790622005432

2 | Background Information and Literature Survey

"If you know you are on the right track if you have this inner knowledge, then nobody can turn you off... no matter what they say."

Barbara McClintock

Key Points:

- Recent developments in communication technology make Long Term Evolution (LTE)/Long Term Evolution-Advance (LTE-A) a promising technology for supporting M2M communication.
- LTE can support the diverse characteristics of M2M communication due to its complete IP connectivity, coverage area, and scalability.
- We present a survey on the classification of LTE / LTE-A scheduling methodologies from the perspective of M2M communication.
- We classify the schedulers based on their objectives, such as energy efficiency, spectrum efficiency, group-based scheduling, and Quality-of-Service (QoS) support for Machine Type Communication Devices (MTCDs).
- Through the literature survey, we find the research gaps for balanced performance concerning the system's performance and QoS support.

2.1 3GPP's Long Term Evolution

We have been witnessing immense growth in mobile data and the number of connected devices. As per the Ericsson mobility report data and forecasts(2022–2028), the number of connected mobile devices will be four times more than the world population by 2028. Half of the connected devices will be Machine Type Communication (MTC). Connected home appliances will have the highest share, and connected cars will grow rapidly. The total number of internet users globally is 5.3 billion in 2023, which was 3.9 billion in 2018. That is 66% of the world's population in 2023. Ericsson predicted that the average global mobile data will grow up to 164 Exabytes (EB) per month by 2025, which was 33 EB per month by 2019. 60 percent of cellular M2M connections are forecast to use 4G/ 5G cellular networks by the end of 2028. Recently, unprecedented growth in connected devices and a high volume of mobile data traffic have driven a demand for communication technology to fulfill communication's future requirements. Fig. 2.1 shows the evolution of the generation of wireless communication. The 3GPP's LTE can meet the demand for high-speed data transfer, massive connectivity, and low latency. The LTE is being standardized to aim for a Fiber-Like mobile broadband experience with more than 100 Mbps of data transfer speed and high mobility support [50, 51].

LTE is a cellular communication standard for mobile devices. It provides efficient high-speed transmission of up to 50 Mbps data rate in the uplink direction and 100 Mbps data rate in the downlink direction at a reduced cost per bit. The LTE communication is based on Orthogonal Frequency Division Multiplexing (OFDM) technology. It uses an Orthogonal Frequency Division Multiple Access (OFDMA) modulation scheme for downlink transmission and a Single Carrier - Frequency Division Multiple Access (SC-FDMA) modulation scheme for uplink transmission. Moreover, LTE provides better resource sharing and lower interference than the previous generations of cellular communication [52, 53]. The LTE enables several new

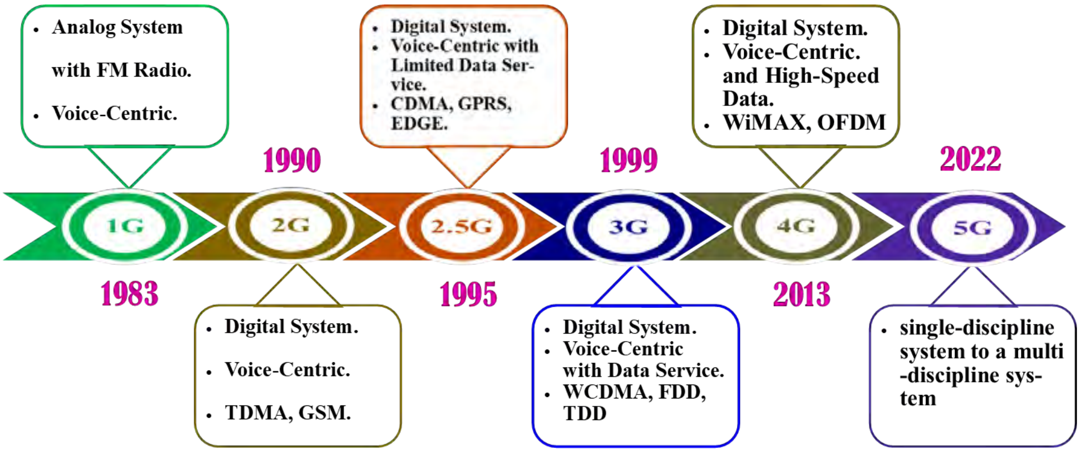


Figure 2.1: An illustration of the evolution of telecommunication technologies

application domains like augmented reality / virtual reality (AR / VR), Internet-of-Things (IoT), Internet-of-Vehicles (IoV), Device-to-Device (D2D) communication, and machine-type communication (MTC). These new application domains are leading to a rapid increase in data rate and massive device connectivity. Application domains like autonomous vehicles, AR / VR, and drone communication require low latency [54, 55]. In our test case, smart-building includes heterogeneous devices such as infotainment systems, security systems, augmented and virtual reality (AR/VR), remote surgery, surveillance cameras, fire or earthquake alert sensors, and many smart utility applications that have a heterogeneous requirement in context to delay-budget, data-rate, and massive connectivity.

2.1.1 LTE network architecture

The LTE network architecture consists of evolved Node Base (eNodeB or eNB), UEs, and core-network, called System Architecture Evolution (SAE), as described in figure 2.2. The SAE primarily consists of the following components: the Mobility Management Entity (MME), the Serving-Gateway (S-GW), the Packet Data Network Gateway (PDN-GW/P-GW), and the Home Subscriber Server (HSS). The SAE core, Evolved Packet Core (EPC), provides multiple services like authentication, mobility management, setting up of bearers, and control of QoS parameters [56, 57]. The

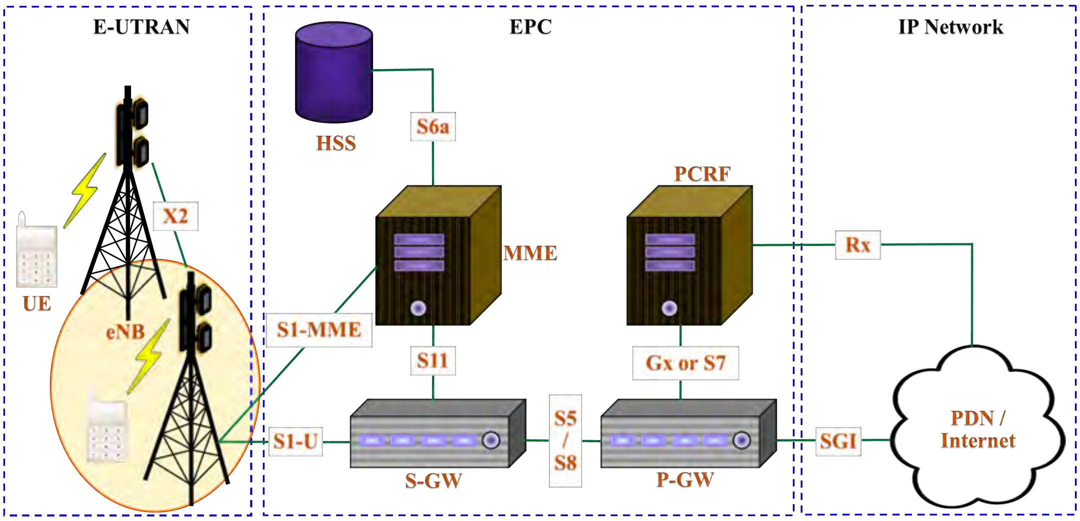


Figure 2.2: LTE communication network architecture [61]

eNodeB is responsible for connecting the UEs to this core network. UE or user equipment is a user or a machine that connects to the eNodeB to access the network [58, 59, 60].

The primary duties of an eNB are radio resource management, encryption and compression of IP data packets, and selection of an MME. The eNB performs radio resource scheduling at its MAC. The MME is part of the core network that deals with user authentication, session management, and mobility management. This entity keeps track of the user's device, and only one MME can be connected to a device at a time. The S-GW handles the data packet routing and forwarding and manages mobility between LTE and other networks [62, 63, 64]. This component also allows for the replication of user data for lawful interception. The P-GW provides the facility to connect the UE with an external network, i.e., the Internet. The P-GW also enforces the charging policy and allows packet analysis or interception. The HSS is the master database for all users and is typically stored in a single node. This component helps authenticate and authorize users for the services offered by the network [58, 35].

In this thesis, we work on the uplink communication between an M2M device and eNB shown as a shaded circle in figure 2.2.

2.1.2 LTE packet scheduler structure

In any cellular communication technology such as LTE, multiple devices contend for limited resources the network operator's infrastructure offers. The distribution of these resources to numerous devices over the radio channel involves assigning time slots and frequency channels to these devices. The process of assigning time and frequency to devices is called radio resource scheduling. The algorithms used for this allocation are instrumental in providing optimal services to end-users and applications [52, 53].

The packet scheduler performs packet scheduling at eNB. Figure 2.3 shows the functional diagram of the LTE packet scheduler. The scheduler performs the scheduling task in both the time and frequency domains. In the first phase, the Time Domain Packet Scheduler (TDPS) scheduler selects sufficient devices that can be assigned the resources [65, 66]. The selection of the devices by the TDPS scheduler is based on criteria such as device priority, channel quality, Buffer Status Report (BSR), etc. After selecting eligible UEs, TDPS passes the list of Radio Network Temporary Identifiers (RNTIs) of selected UEs to the Frequency Domain Packet Scheduler (FDPS) for the further resource allocation process. The FDPS assigns the physical resources to UEs as per the device, channel status, and device requirements [58, 53, 67].

LTE packet scheduler also performs the task related to the Hybrid Automatic Repeat Request (HARQ) management for failed packet transmission, link adaptation based on the packet's feedback (ACKs/NACKs) and Channel Quality Index (CQI) to adjust the transmission rate, transmission power level, Modulation and Coding Schemes (MCS) for error-free transmission. Packet scheduler also receives inputs about QoS, BSR, and Medium Access Control (MAC) & Radio Link Control (RLC) information to perform efficient scheduling decisions [58, 53]. This thesis's proposed resource scheduling scheme is designed to work in time domain packet scheduling (TDPS). We have used ACK/NACK, QoS attributes, buffer status reports (BSR), and MTCD's MAC information as feedback to perform resource scheduling in the

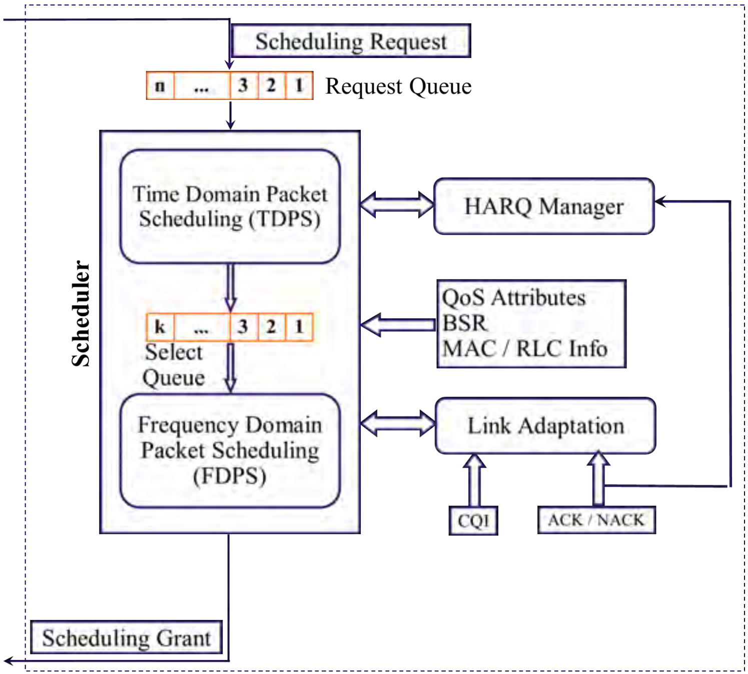


Figure 2.3: Block diagram of LTE packet scheduler [59]

proposed work.

2.1.3 LTE radio frame structure

LTE communication uses OFDMA in the downlink and SC-FDMA in the uplink channel. The uplink and downlink data is transmitted as frames of 10ms duration, as shown in figure 2.4. Each frame is further divided into ten subframes of length 1ms each. The duration of the subframe is known as the transmission time interval (TTI), and each such subframe is further divided into two slots of 0.5ms duration each. The resource units are allocated in slots of 0.5ms long in the time domain and 180KHz bandwidth in the frequency domain [68, 69, 70]. The block of 0.5 ms long in the time domain and 180KHz wide in the frequency domain is called Physical Resource Block (PRB). The PRB is the minimum unit to be allocated to a UE, and resources are allocated in multiple PRBs. Each PRB is a grid of 12*6 or 12*7 Resource Element (RE) comprised of 12 subcarriers of 15KHz each in the frequency domain and 6 (extended CP) or 7 (Normal CP) symbols in the time domain [58, 53, 71]. The cyclic prefix refers to an additional piece of each OFDM symbol that the transmitter duplicates from its end to its beginning before sending the entire signal.

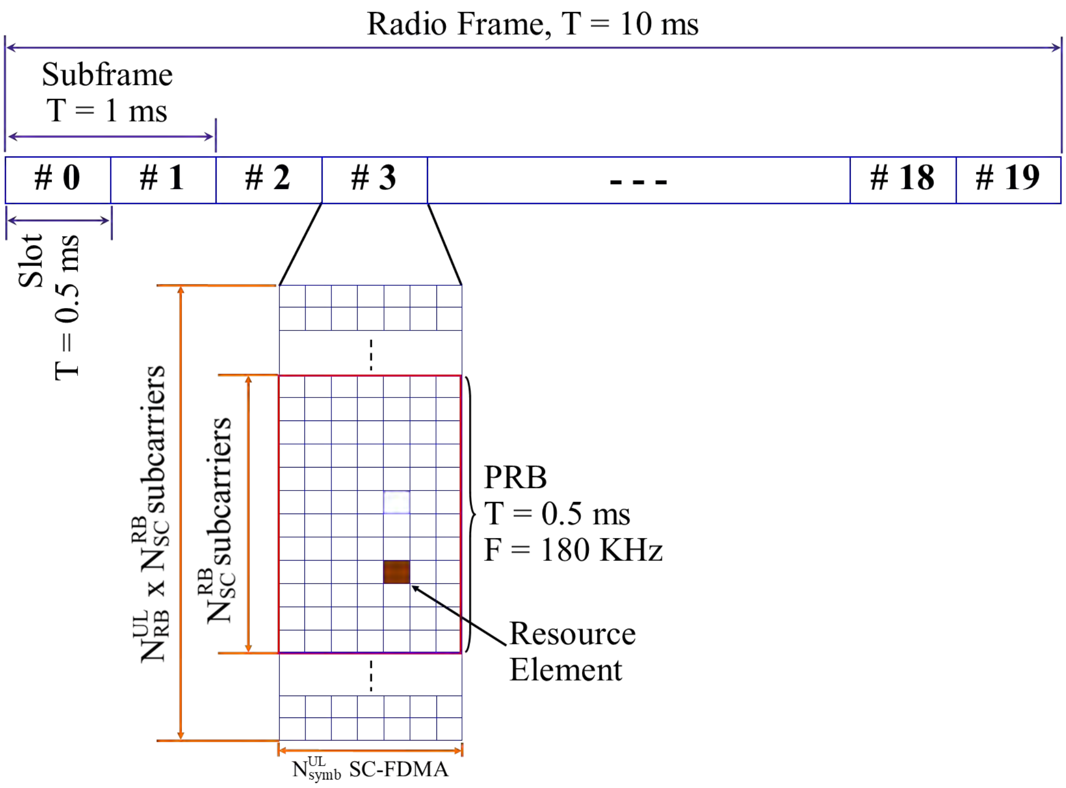


Figure 2.4: LTE radio frame type-2 structure [59]

Depending on the bandwidth, the number of PRBs in the uplink ranges from 6 to 100. LTE provides a facility of flexible bandwidth from 1.4 MHz to 20 MHz. Thus, the bandwidth of 1.4 MHz provides 6 PRBs of 0.5ms*180KHz, and the bandwidth of 20 MHz provides 100 PRBs of 0.5ms*180KHz. The most basic modulation unit is a resource element, a single block of 12*7 grid of a PRB, and contains one symbol of 15khz. Each resource element may contain two or more bits depending upon the modulation and coding scheme (2 bits in QPSK, 4 bits in 16QAM) [72, 32, 73]. There are two radio frame types, Type-1 and Type-2. Type-1 uses FDD, and Type-2 uses TDD mode. In type 2, there is 7 configuration from 0 to 6. We have used radio frame type-2 and configuration 0 with a DL and UL switching periodicity of 5ms. We have chosen a bandwidth of 5 MHz with 25 PRBs in our simulation environment, and other time domain and frequency domain configurations are as shown in figure 2.4.

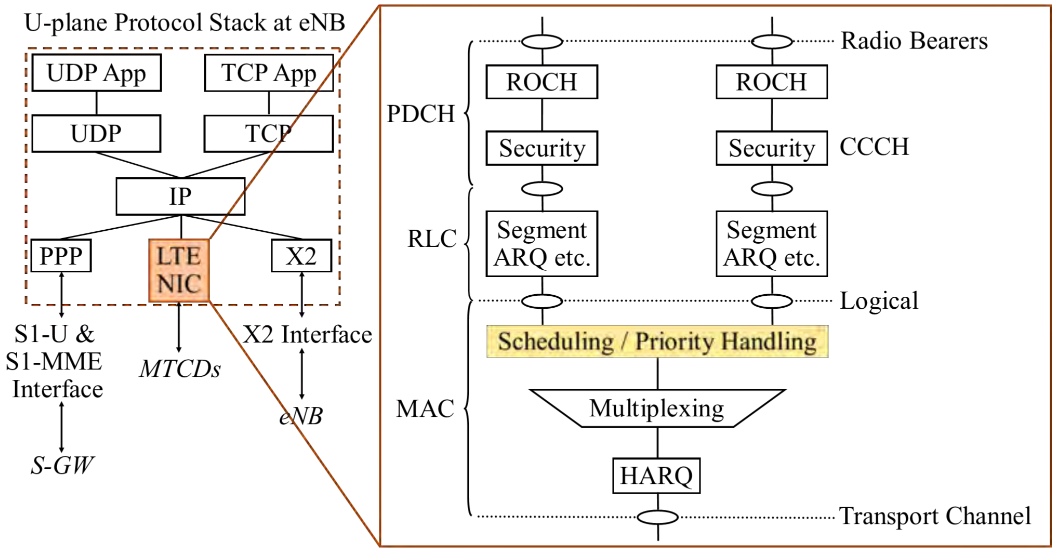


Figure 2.5: Schematic diagram of LTE scheduling process [77]

2.1.4 LTE resource scheduling

The LTE packet scheduler is a MAC layer functionality of eNB, which is responsible for packet scheduling and physical resource-sharing decisions. When a MTCD sends a scheduling request (SR) to eNB, the packet scheduler assigns the required resources to that UE based on the received information from UE and network, such as BSR, a sounding reference signal (SRS), and available resources. Physical resource scheduling decisions depend on various attributes such as QoS attributes, CQI, fairness, energy efficiency, and spectral efficiency per the scheduling objective. Figure 2.5 shows schematic diagram of LTE scheduling process [74, 75, 76].

Whenever a UE or a machine has data in its buffer for transmission, it sends a BSR packet as an uplink resource scheduling request to the eNB over the Physical Uplink Control Channel (PUCCH). BSR reporting interval can be configured to send periodically or based on data availability in UE's buffer [78, 79, 80]. Upon receiving scheduling requests over PUCCH from UE, the scheduler allocates m available physical resources to n requesting UEs, using a specified algorithm according to received reference signals and algorithm's objective and sends scheduling grant information in Downlink Control Information-0 (DCI0) format to UE. The scheduling grant contains information about MCS, frame number, transmit power, etc. The scheduling task is performed per subframe of 1 millisecond as shown in figure 2.4. Suppose a UE_i received a scheduling grant in n^{nt} subframe. In that case, the UE_i can send data in or after $(n+3)^{th}$ subframe over Physical Uplink Shared Channel (PUSCH) [81, 82, 83, 84, 85]. For example, if a UE received scheduling grant in subframe 1 consisting of slots 1 & 2 as shown in figure 2.4, then UE is allowed to send data in subframe four consisting of slots 7 & 8. PUCCH and PUSCH are logical channels in the MAC layer, as shown in figure 2.5.

LTE resource scheduling is divided into two categories. The first is dynamic or channel-dependent, and the other is static or channel-independent scheduling. Dynamic scheduling considers the channel quality between UE and eNB in scheduling decisions. In contrast, static scheduling does not consider CQI in scheduling decisions. In LTE, 15 CQIs are defined by 3GPP, ranging from 1 to 15. CQI values are used to decide the transmission's modulation and coding rate to achieve a lower Bit Error Rate (BER). Table 2.1 shows the 4-bit CQI table for LTE [86, 87].

Table 2.1: Channel Quality Indicator Table [58]

CQI	Modulation	Code Rate (x1204)	Efficiency
0	Out of range		
1	QPSK	75	0.1523
2	QPSK	120	0.2344
3	QPSK	193	0.3770
4	QPSK	308	0.6016
5	QPSK	449	0.8770
6	QPSK	602	1.1758
7	16QAM	378	1.4766
8	16QAM	490	1.9141
9	16QAM	616	2.4063
10	64QAM	466	2.7305
11	64QAM	567	3.3223
12	64QAM	666	3.9023
13	64QAM	772	4.5234
14	64QAM	873	5.1152
15	64QAM	948	5.5547

Quadrature Phase Shift Keying (QPSK),Quadrature Amplitude Modulation(QAM)

2.1.5 M2M Communication Challenges in LTE/LTE-A

M2M communications are widely used in various applications, including security, surveillance, intelligent transportation systems, emergency alerts, healthcare, and smart metering [67]. Many devices are deployed to gather information and send it via a network to processing devices. Due to its intrinsic IP connectivity and ability to serve many devices across vast distances, LTE cellular networks are a serious candidate for M2M communications. The needs of M2M communications differ depending on the supported service, making it difficult to integrate M2M services into LTE/LTE-A networks. These traits and difficulties are covered in [88] and can be summed up as follows:

- **Energy Management:** Since a majority of M2M devices function using batteries in general, lowering power consumption is one of the main issues in M2M communications [48].
- **Access Priority:** When competing for network access in particular applications, M2M devices must be given preference over other network nodes. Emergency scenarios and military support missions call for priority access [48].
- **Group Control:** The system should be capable of handling many M2M devices (massive deployments).
- **Small Burst Transmissions:** The size of the transmitted data bursts is typically minimal. Therefore, it is important to enable low traffic overhead [59].
- **Time-controlled Operation:** M2M communications need to allow time-controlled operation, in which M2M devices send or receive data at specific pre-defined times based on the type of application. Event-driven applications are the sole exception [79].

- **Delay-tolerant Operation:** Delay-tolerant M2M devices should have a lower access priority from the system, which would delay their data transmission following their application needs [78].

- **Extremely Low Latency:** For M2M devices that are not delay-tolerant, this entails lowering network access latency and data transmission latency. In the event of emergencies, this becomes critical [48].

- **Infrequent Traffic:** M2M communications are typically infrequent and have long inter-arrival times. Therefore, establishing effective scheduling approaches requires modeling the M2M traffic characteristics, similar to the study in [89].

The air interface presents the most serious difficulties in considering the needs mentioned earlier. These difficulties generally concern random access and resource allocation with provisioning for QoS.

2.1.6 Importance of resource scheduling for M2M in LTE

LTE is a cellular communication standard that offers high bandwidth and the flexibility to accommodate the varying requirements of UEs. M2M communication typically consists of low bandwidth bursts of data with different QoS requirements from their Human-2-Human(H2H) counterparts [53]. M2M communication is predominantly uplink-based and contends with the uplink H2H traffic. Various mission-critical applications of M2M communication need to be prioritized over H2H, and this contention is an issue that the uplink scheduler has to manage [90, 91].

As the M2M communication system has many devices with varying QoS requirements, designing scheduling schemes that support machine-type communication over the LTE network is challenging [54]. Many MTCDs are infrequently sending varying sizes of data packets; the LTE bandwidth offers a limited number of physical resources and is optimized for H2H communication [89, 92].

Therefore, it is required to design a solution for M2M communication that optimizes the available physical resources while satisfying the unique QoS requirements of M2M communication.

2.1.7 Scheduling metrics

Scheduling metrics are parameters related to the network and UE. The scheduler uses these metrics to optimize the system utility and to fulfill the scheduling objectives. This subsection introduces the most common scheduling metrics used in the existing schedulers.

Channel quality identifier

The 3GPP standard describes the channel quality between eNB and UE as channel quality identifiers (CQI). CQI values range from 0 (very poor channel) to 15 (very good channel). In LTE, signal-to-noise plus interference ratio (SNIR) describes the channel quality. The eNB uses reference signal received power (RSRP), noise level, and interference level to estimate the CQI for a UE-eNB link. The eNB estimates CQI for each resource block for all UEs. UE computes effective SNIR $SNIR_{eff}$ based on the received SNIR samples over multiple OFDM symbols [93]. The effective SNIR $SNIR_{eff}$ is defined as follows.

$$SNIR_{eff} = \alpha_1 I^{-1}\left(\frac{1}{N}\sum_{k=1}^{M}\left(\frac{SNIR_k}{\alpha_2}\right)\right) \tag{2.1}$$

Where N is the number of SNIR samples. α_1 and α_2 are parameters to adapt different modulation and coding. The maximum achievable sum rate is constrained by Shannon capacity and depends on SNIR [93]. Shannon's capacity is defined as follows.

$$C_S = log_2\left(1 + SNIR\right) \tag{2.2}$$

Thus, a high data rate depends on channel quality. UE can send significant data using higher modulation and coding schemes (MCS) over a channel with higher channel quality as the transport block size (TBS) depends on the MCS. The modulation constraint Shannon capacity is defined as follows.

$$C_S = log_2(M) + \frac{1}{2\pi M}\sum_{m=1}^{m-1}\int e^{-\gamma(y-x_m)^2} log_2\left(\frac{e^{-\gamma(y-x_m)^2}}{\sum_{k=0}^{M-1} e^{-\gamma(y-x_m)^2}}\right) \quad (2.3)$$

Where M is the size of the modulation alphabet, γ is SNIR, and x_m is the modulation symbol. Channel quality should support significant MCS to achieve a high data rate. In the 3GPP standard for LTE, CQI is defined as a $N \times R$ matrix, where N is the number of UEs and R is the number of available resources N_{RB}^{av}.

$$CQI_{i,j} = \begin{bmatrix} CQI_{11} & CQI_{12} & CQI_{13} & \dots & CQI_{1R} \\ CQI_{21} & CQI_{22} & CQI_{23} & \dots & CQI_{2R} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ CQI_{N1} & CQI_{N2} & CQI_{N3} & \dots & CQI_{NR} \end{bmatrix}$$

The scheduler fetches the $CQI_{i,j}$ for the UE i over resource j. The throughput-hungry scheduling policies consider CQI as the primary allocation metric.

QoS class identifier

3GPP defines QoS class identifiers (QCI) for different types of applications corresponding to application requirements. A QCI encompasses parameters like required bit rate, type of bearer (GBR/N-GBR), delay budgets, and application's priory, depending on the type of application. The LTE scheduler considers QCI priority to give preference for an UE over another in the time domain selection [48]. Other parameters are used to decide on frequency domain resource allocation. Table 1.1 shows standard QCIs for M2M communication. Let a UE i have priority p and allocated resource set R_i^p. Another UE j has priority q and allocated resource set R_j^q. Then, the resource allocation algorithm fails to satisfy the QCI if.

$$R_i^p = \phi \; and \; R_j^q \neq \phi \; if \; q \leq p \tag{2.4}$$

Equation (2.4) states that the MTCD j with lower priority gets resource even if MTCD i has higher priority than MTCD j.

Delay

Network traffic is categorized into two types: delay bound and delay tolerant. Delay-bound traffic has a strict deadline [79]. If the deadline is missed, the packet is not significant and hence is considered a dropped or lost packet. However, the delay-tolerant packet can be queued for some time. Delay-aware packet scheduling policies use the delay budget of the application packet as the primary metric for resource allocation [94]. If a UE i has a delay budget of D_b^i and a delay in the network access D_i, then a delay-aware packet scheduling algorithm tries to satisfy the following condition.

$$\frac{D_i}{D_b^i} \leq 1 \; \forall i \in I \tag{2.5}$$

Buffer status

UE sends periodic updates about their buffer level as a buffer status report (BSR) to eNB. The buffer status of a UE represents the amount of data available in the UE's buffer ready for transmission. A higher buffer level increases the probability of queue overflow and information loss. Buffer-aware resource allocation algorithms prioritize UEs with higher buffer levels to provide more transmission opportunities than those with lower buffer levels [53].

Traffic pattern

Network traffic can be classified based on traffic characteristics like time-driven (TD), event-driven (ED), and delay-bound or delay-tolerant. Generally, traffic characteristics are used to prioritize a packet over others, data aggregation on gateway devices, and clustering of devices. The data aggregation mechanism enables communication in a dense network [79].

Round robin

Round robin scheme allocates equal resources to all UEs irrespective of any characteristic or requirement of UEs. The round-robin approach aims to give all the devices a fair chance [53].

The mentioned allocation metrics can be used individually or combined with multiple metrics to optimize the allocation. The selection of allocation metrics depends on the scheduling objective. We have used channel quality, QoS class identifier, and delay as selection metrics to optimize the resource scheduling in this thesis.

2.1.8 Scheduling objectives

The scheduling objective is an expected outcome of the allocation strategy. The scheduling strategy optimizes the scheduling metrics subject to scheduling constraints to achieve scheduling objectives. The primary scheduling objectives identified in the literature are as follows.

Energy efficiency

Energy efficiency is a primary objective when the connected devices have power constraints, i.e., battery-operated devices. The primary allocation metric is channel quality between eNB and the device to achieve energy efficiency [48]. Moreover, some researchers also minimize the interference to increase energy efficiency. Generally, energy efficiency is measured as energy consumed per bit of information transfer by the communicating device. The objective is to minimize overall energy consumption $e_{i,j}$ for the devices $i \in I$ with the resource block $j \in J$.

$$Minimize \sum_i \sum_j e_{i,j} \tag{2.6}$$

Resource utilization

It is a ratio of resources utilized to total available resources. A combination of multiple allocation metrics is used to optimize resource utilization. The allocation strategies with straight allocation mechanisms, i.e., Round-Robin, provide better resource utilization [48, 84]. Resource utilization U_{RB} is defined as follows.

$$U_{RB} = \frac{\sum_{i=1}^{I} RB_i^u}{RB_{Total}} \tag{2.7}$$

Where RB_t^u is the resources the devices utilize in a time interval.

QoS support

Different devices have different QoS requirements like delay budgets, throughput required, or priority. Generally, multiple allocation metrics, like a priority, buffer level, or CQI, are jointly optimized to provide QoS support. However, there is a trade-off between resource utilization and QoS [95].

Massive connectivity

With growing connectivity to the network, one of the objectives of scheduling strategies is to provide connectivity for many devices. Traffic patterns and QoS class are standard metrics jointly used for traffic aggregation to support massive connectivity [89].

Fairness

Fairness is applied when the resources to be shared are limited. Fairness refers to the likely equal sharing of resources among the communicating devices. Fairness in sharing resources is challenging due to the trade-off between fairness and efficiency [83]. The resource fairness index I_{RB} shows whether the resources are fairly shared among the devices. The resource fairness index varies in the interval [0 1], 0 represents entirely unfair, and 1 represents fairness entirely [96]. The fairness index is defined as follows.

$$I_{RB} = \left(\frac{1}{1+\hat{x}} \right) \tag{2.8}$$

$$\hat{x} = \sqrt{\frac{\sum (RB_i - RB_{Avg})^2}{N_{UEs}}}$$

Where RB_i is the resource allocated to device i and $\hat{x}$ is the standard deviation in resource share among devices.

Throughput

Dedicated LTE bearer classes can be guaranteed bit rate (GBR) or Non-GBR. 3GPP specifies GBR for a specific application. The LTE scheduler uses GBR to estimate the minimum required resources for a particular UE [97]. The achieved throughput R_i^{Ache} for UE i should be more than or equal to GBR and less than or equal to maximum achievable throughput R_{max}. Throughput-aware resource allocation algorithms satisfy the following conditions.

$$GBR_i \leq R_i^{Ache} \leq R_{max} \; \forall i \in I \tag{2.9}$$

$$R_i^{Ache} = \frac{N_{RB} * N_{RE}^{RB} * B_{symb}}{TTI}$$

Where N_{RB} is the number of resource blocks allocated to UE i, N_{RE}^{RB} is the number of resource elements per RB, and B_{symb} represent bits per symbol and depends on MCS used.

The performance of an allocation strategy is the percentage of fulfillment of the scheduling objective by optimizing the scheduling metrics. The limit of optimization of scheduling metrics depends on the scheduling constraints. We have considered resource utilization, QoS support, fairness, and throughput as scheduling objectives in the proposed scheduling scheme. Throughput and resource utilization are selected as system performance objectives, and others are considered as QoS objectives.

2.1.9 Scheduling constraints and flags

Scheduling constraints limit the optimization of scheduling metrics. During the packet scheduling process, eNB satisfies the QoS of devices and efficiently allocates resources among the devices for the best utilization of radio resources (i.e., maximizes cell capacity). To allocate radio resources efficiently among MTCDs, and consider many factors such as CQI, SNR, BSR report, HARQ, etc. [53]. Many other limitations should be considered in the packet scheduling process. Some of these are -

1. ***Allocation Flag:*** If a resource R_j is allocated to an MTC device i, it can not be allocated to any other device. The allocation flag can be defined as follows.

$$\alpha_{i,j} = \begin{cases} 1 & if\, R_j\, is\, allocated\, to\, a\, MTCD \\ 0 & Otherwise \end{cases} \tag{2.10}$$

   Where $\alpha_{i,j}$ is a flag such that resource R_j is allocated to the device i. A single resource (PRB) can not be assigned to more than one MTC device [58].

$$R_{i,j} \cap R_{i',j} = \phi; \quad \forall\, i \in I,\; j \in J \tag{2.11}$$

   Where $R_{i,j}$ are resources allocated to device i and $R_{i',j}$ are resources allocated to the device i'

2. ***Hybrid Automatic Repeat Request (HARQ) Flag:*** If transmission of a packet has failed, then packet retransmission is initiated by eNB. The eNB provides high priority for the packets in the retransmission phase. LTE scheduler uses a mechanism to re-transmit a packet called Hybrid Automatic Repeat Request (HARQ) [84]. If the HARQ flag is enabled for a device i, it should be prioritized over other devices. HARQ flag is defined as follows.

$$\beta_i = \begin{cases} 1 & if\, reported\, by\, HARQ \\ 0 & Otherwise \end{cases} \tag{2.12}$$

Where β_i is a flag such that device i has a packet for retransmission.

3. ***Delay Constraint:*** The maximum delay for a device i, dl_i, should not be more than the maximum allowable delay budget (DB_i)

$$dl_i \leq DB_i$$

$$\frac{dl_i}{DB_i} \leq 1; \quad \forall\, i \in I \tag{2.13}$$

4. ***Throughput Constraint:*** Throughput for a device i should be greater than the minimum required Guaranteed Bit Rate (GBR) GBR_i and should be less than the maximum allowable throughput T [58].

$$GBR_i \leq \frac{N_{RB} * N_{RE}^{RB} * B_{symb}}{TTI} \leq T \tag{2.14}$$

Where n_{RE} is the number of resource elements, and b_{sym} is the number of bits per symbol.

5. ***Maximum Resource Constraint:*** The maximum number of resources (PRBs) allocated to devices in a single Transmission Time Interval (TTI)for all selected devices (scheduled for transmission) should be less than or equal to available

resources in that TTI [84].

$$| R_{i,1} \cup R_{i,2} \cup \cdots \cup R_{i,j} | \leq | R_{av} |; \quad \bigcup_{i,j=0}^{m} R_{i,j} \leq |R_{av}| \tag{2.15}$$

$R_{i,j}$ refers to a resource j allocated to device i and R_{av} is the number of available resources in a TTI.

6. ***Continuity Constraint:*** Multiple resources for a single device can only be allocated continuously in the frequency domain [27]. For example, device A needs three PRBs, device B needs one PRB, and available resources are four PRBs. Now, the PRB allocation pattern should be (A-123)-(B-4) or (A-234)-(B-1). It can not be (A-134)-(B-2).

$$R_{i,j} = \phi \quad \forall j \geq m+1 \; if \; R_{i,m} = \phi \tag{2.16}$$

Where $R_{i,j}$ refers to resource j allocated to device i.

7. ***Power Constraint:*** For a device, i maximum transmission power can not exceed the maximum allowable limit [27] -

$$0 \leq \varepsilon_{i,j} \leq \varepsilon_{max}; \forall i \in I, \; j \in J \tag{2.17}$$

Where $\varepsilon_{i,j}$ is transmission power used by the device for j^{th} resource.
A device i should have an equal power level for all PRBs allocated to the device in a TTI -

$$\varepsilon_{i,j} = \varepsilon_{i,j'} \quad \forall i \in I, \; j \in J \tag{2.18}$$

Before designing a scheduling mechanism, constraints and metrics should be identified to fulfill the scheduling objective. We have used the constraints based on feedback received from the MTCDs in the proposed scheduling scheme. From the above-listed constraints, delay and throughput are used as MTCD's constraints,

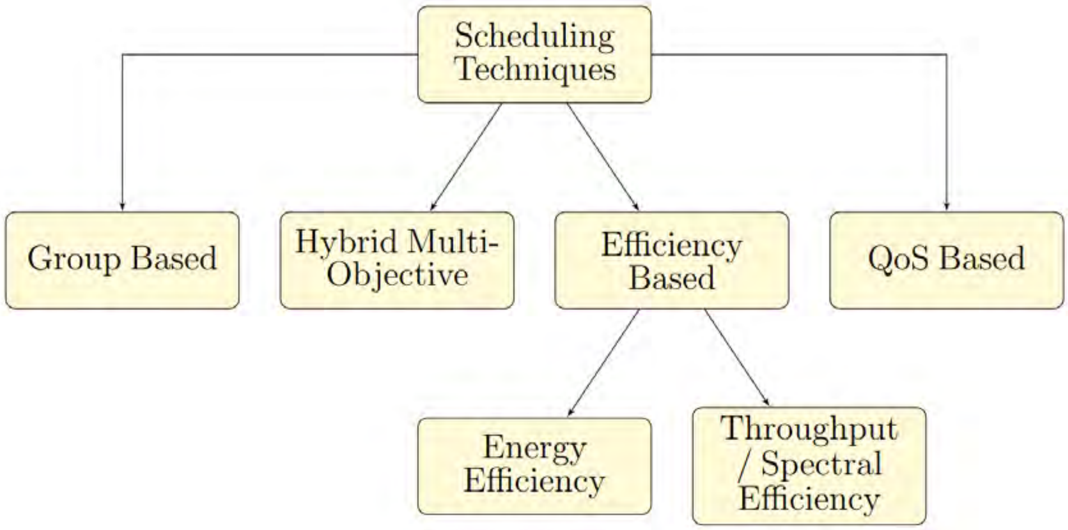


Figure 2.6: Classification of Scheduling Techniques

while continuity and maximum resource are used as eNB constraints. The HARQ and power constraints are applicable for both MTCD and eNB.

2.2 Classification of Scheduling Techniques

This section discusses selected physical resource scheduling techniques for M2M and H2H communication in the LTE network. All the scheduling techniques are classified depending on the focused objective of the scheduling, as shown in figure 2.6.

2.2.1 Efficiency-based techniques

M2M devices have low processing capabilities and limited battery life. The majority of these devices are deployed with fixed batteries. Some of these devices also serve real-time mission-critical applications. Due to the nature of M2M communications, suitable radio resource allocation techniques are required to provide these devices' energy efficiency and throughput. These can be classified into two categories based on their optimization: energy-efficient and spectral/throughput efficient resource scheduling techniques [98, 99, 100, 101].

Energy efficient techniques

These algorithms aim to minimize energy consumption in MTCDs individually or for the whole network. One approach is to give preference in allocating resource blocks to the MTCD with the best channel quality to increase throughput and decrease power consumption. The transmission rate depends on the channel quality between the UE and eNB [102, 82, 103]. For the poor channel condition, a sufficient data rate can be achieved using high transmit power (P_{tx}) and MCS such as 16QAM/64QAM. Thus, the transmit power P_{tx} of UE can be seen as one of the link adaptation schemes. Some applications can work efficiently with flexible data rates. For such cases, the energy efficiency of a UE can be increased with a lower data rate. In practice, the data rate of radio-link between the UE and eNB is controlled by Modulation and Coding rate [104, 85, 105]. Figure 2.7 shows the relationship among the transmit power P_{tx}, data rate T_{tx}, and transmission channel quality.

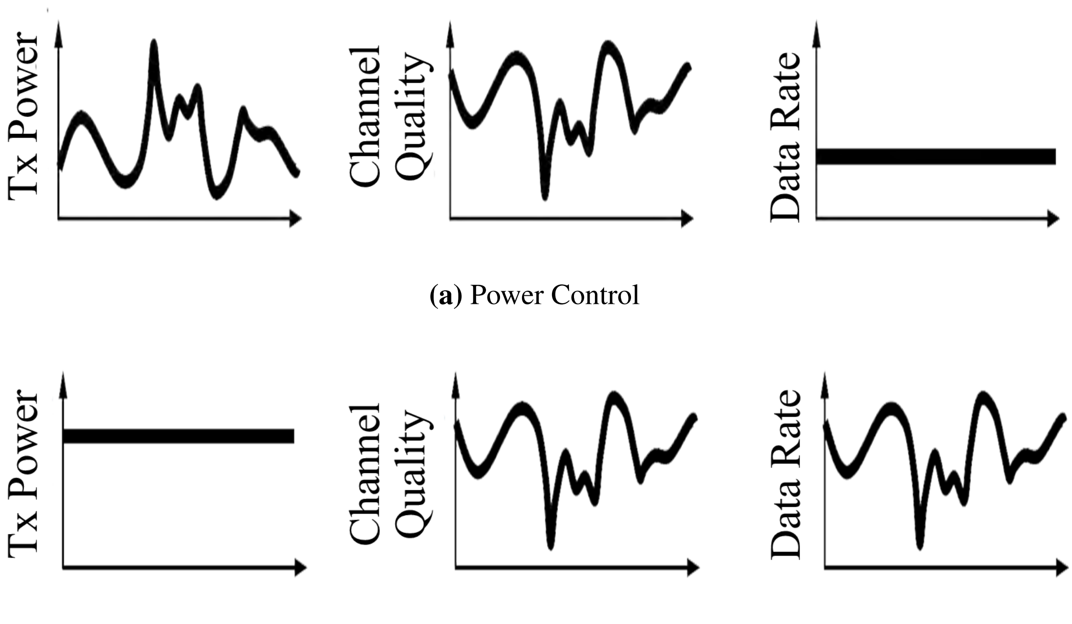
The LTE network provides two methods to control the UEs' power consumption: "Closed-Loop Power Control" and "Open-Loop Power Control." The Closed-Loop power control mechanism uses the feedback provided by the UEs to the eNB through sounding reference signals [106]. These SR signals provide information about the path gain and shadowing through the path between the UE and the eNB and are used to calculate the Signal-to-Inference plus Noise Ratio (SINR) employed to make decisions about the MCS and transmit power required for the transmission. MCS affects the amount of information transmitted per transmission and the power consumption [92, 85]. The power control for the PUCCH is defined as follows.

$$\begin{aligned} P_{PUCCH} =& min\{P^{Tx}_{max,c}, P_{0,PUCCH} + PL_{DL} \\ & + \triangle_{Format} + \delta\} \end{aligned} \tag{2.19}$$

The power control for PUSCH is defined as follows.

$$P_{PUSCH,c} = min\{P^{Tx}_{max,c} - P_{PUCCH}, P_{0,PUSCH} + \alpha.PL_{DL} + 10.\log(M) + \triangle_{MCS} + \delta\} \tag{2.20}$$

Where; $P^{Tx}_{max,c}$ is per-carrier maximum transmit power, $P_{PUSCH,c}$ is allocated power for PUSCH over carrier c, $P_{0,PUCCH}$ is the target received power, P_{PUCCH} is allocated power for PUCCH, $P_{0,PUSCH}$ is cell-specific parameter, PL_{DL} is downlink path loss, α is partial path-loss compensation, $\triangle_{Format}$ is P_{tx} power offset, and δ is explicit power-control commands. The $\triangle_{MCS}$ shows the requirements of different transmit power levels for the different MCS. The term $10.\log(M)$ reflects the power required per resource block. A larger resource block size requires more power to transmit data. Thus, (2.19) and (2.20) show that the transmission rate and resource block size affect the power consumption of the transmitter[92].



(a) Power Control

(b) Rate Control

Figure 2.7: Power and rate control in LTE [92]

The discontinuous Reception (DRx) scheme also improves the UE's energy efficiency. UE only consumes power when transmitting or receiving data to or from eNB; otherwise, it gets into sleep mode and consumes much less power. A UE can be configured to monitor PDCCH discontinuously. DRx can be set in two ways, Long DRx and Short DRx, as per the requirements of the UE's application. DRx scheme maintains various timers, such as the duration timer, DRx inactivity timer, and DRx re-transmission timer. DRx mechanism sets the time offset and sets or resets the timers as per the configurations [48, 92]. A typical DRx scheme with PDCCH reception is shown in Fig. 2.8 -

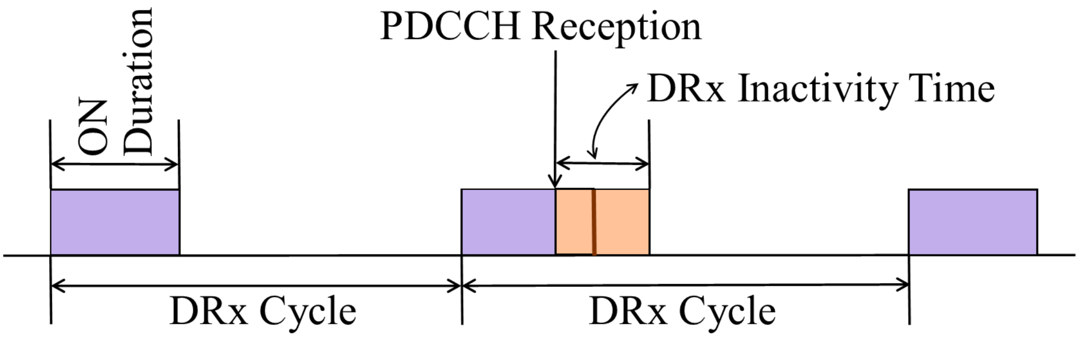


Figure 2.8: DRx scheme with PDCCH reception [48]

Rekhissa *et al.* [93] proposed two energy-efficient uplink allocation strategies in H2H / M2M co-existence scenario by modifying Carrier-By-Carrier (CBC) and Recursive Maximum Expansion (RME) algorithms for UEs as well as MTCDs. The authors define UE metrics for each RB and allocate RBs such that RB_i is allocated to the UE, which has the highest metric for the i_{th} RB. The CBC approach consists of choosing the best CQI and allocating the corresponding RB to the MTCD; this process is repeated until no one UE remains or all RBs are assigned. RME recursively expands allocation toward the left/right of an allocated RB for a device.

Azari *et al.* [107] modeled energy consumption and network lifetimes based on transmission and circuit energy consumption and proposed an algorithm that maximizes network lifetime by allocating devices that have the most effect on network lifetime first. The authors define network lifetime as the shortest, longest, and average individual lifetime and the expected lifetime metric.

$$L(t) \triangleq \frac{E(t)}{\varepsilon_s + \varepsilon_d} T \tag{2.21}$$

Where $E(t)$ is the remaining power, T is the reporting interval, ε_s is static power use, and ε_d is the average power of UE. The authors modeled the problem as Min-Max optimization and proposed variations of this algorithm that work with limited Channel State Information (CSI). Results demonstrate that the spectral and energy efficiency show an inverse trend, i.e., Increasing data sent in each resource block increases spectral efficiency and decreases energy efficiency.

In [108], the authors modeled machines' energy consumption as a constraint minimization problem and defined it as follows.

$$min \sum_{f=1}^{T} \sum_{n=1}^{N} E^n(f) \tag{2.22}$$

Where $E^n(f)$ is a function of the device's power consumption in the data and signal transmission rates, it goes into a sleep state when the device is not in any transmission state. The authors consider devices as sensory nodes and propose two energy-efficient scheduling algorithms. The first algorithm is used when the distance between eNB and the device is long; it schedules data with short deadlines first and tries all possible allocations of RBs to reduce power consumption. The second algorithm minimizes the number of active sub-frames to achieve efficiency and is used when the distance between the device and eNB is short. The algorithms were compared with EDF, WF^2Q, and Chen's algorithm and were found to consume less energy with satisfactory fairness and scheduling success ratios.

In [109], Azari *et al.* proposed an energy-efficient scheme to enhance the network's lifetime using an optimum size cluster. The expected lifetime of a cluster is defined as the ratio of remaining power to the average power consumption of nodes in each duty cycle and is expressed as follows.

$$L_c = \frac{E_0}{\frac{1}{g}E_h + (1-\frac{1}{g})E_m} T_c \tag{2.23}$$

Where E_0 is the remaining power, g is cluster size, E_h is the average power consumption of the cluster head, and E_m is the average power consumption of other devices in the cluster. The authors have reduced energy consumption by selecting an optimum value for g and proposed a distributed clustering scheme for massive M2M devices by modeling power consumption and creating clusters of optimal size. The authors also proposed a lifetime-aware scheduling technique that maximizes network lifetime. Results indicate that this technique consumes less energy than standard scheduling schemes.

In [110], the authors proposed an energy-efficient scheduling technique for small data transmissions in an LTE network. The proposed algorithm selects an optimal MCS according to the payload size to achieve energy efficiency. The authors defined the energy-efficiency of a UE as the ratio of the number of transmitted payloads bits L by UE to the energy consumed by the UE in the transmission E_T and are calculated as follows.

$$\eta = \frac{L}{E_T} \tag{2.24}$$

The simulation results show that their approach maximizes the battery lifetime of MTCDs. The authors have also suggested a simple PRB allocation in which all necessary PRBs are allocated to send the entire packet, which maximizes energy efficiency.

In [86], the problem is reduced to an NP-hard mixed-integer linear fractional programming problem consisting of MCS assignment, allocation of resources, power, and data scheduling. The authors achieved the global optimum using the Charnes-Cooper transformation and another technique called the Glover linearization scheme to obtain the global optimum. The authors compared the performance of their technique with Greedy and EES [108], and results indicate that their technique outperforms both in terms of packet-dropping ratio and provides optimal energy efficiency when the number of resource blocks is limited.

Throughput and spectral efficient techniques

The spectral efficiency of a radio link is defined as the achieved data rate over a fixed channel bandwidth. Spectral efficiency is also termed as normalized throughput and measured in bits/second/Hz [115, 116]. Spectral efficiency and throughput efficiency are proportionally related to each other. It can be enhanced through resource optimization, spectrum sharing among multiple users, optimum allocation of MCS, and optimum resource grid size. These strategies are also related to the device's utility. The parameters that broadly affect a radio link's spectral efficiency are MCS and SNR. Higher MCS and SNR give higher spectral efficiency [92, 117, 118]. 3GPP defines 15 MCS indexes ranging from 1 to 15. A sufficient SNIR is required at the receiver to maintain the BER acceptable for a selected MCS index.

Theoretical throughput for an LTE network can be defined as follows.

$$T_{the}(Bits/ms/TTI) = PRBs * REs * S_{TTI} * Bits_{MCS} \tag{2.25}$$

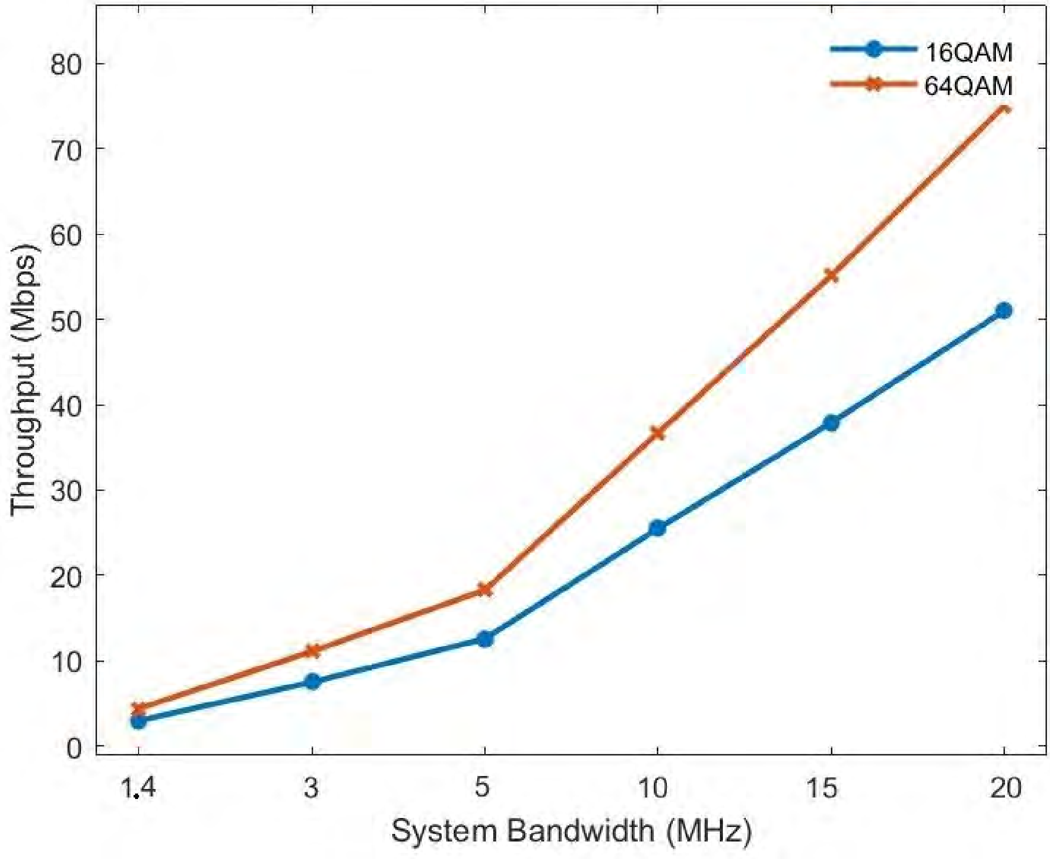
Where $PRBs$ is the number of resource blocks in a given bandwidth, REs is the number of resource elements in a resource block, and S_{TTI} is the number of slots per TTI and $Bits_{MCS}$ is the number of bits per symbol.

Figure 2.9 shows the SNR, MCS, and Throughput relationship. Multiple approaches have been proposed for resource scheduling to increase throughput and spectral efficiency. For example, Yaacoub *et al.* [111] proposed a resource scheduling scheme to increase spectral efficiency and define UE's utility as a function of

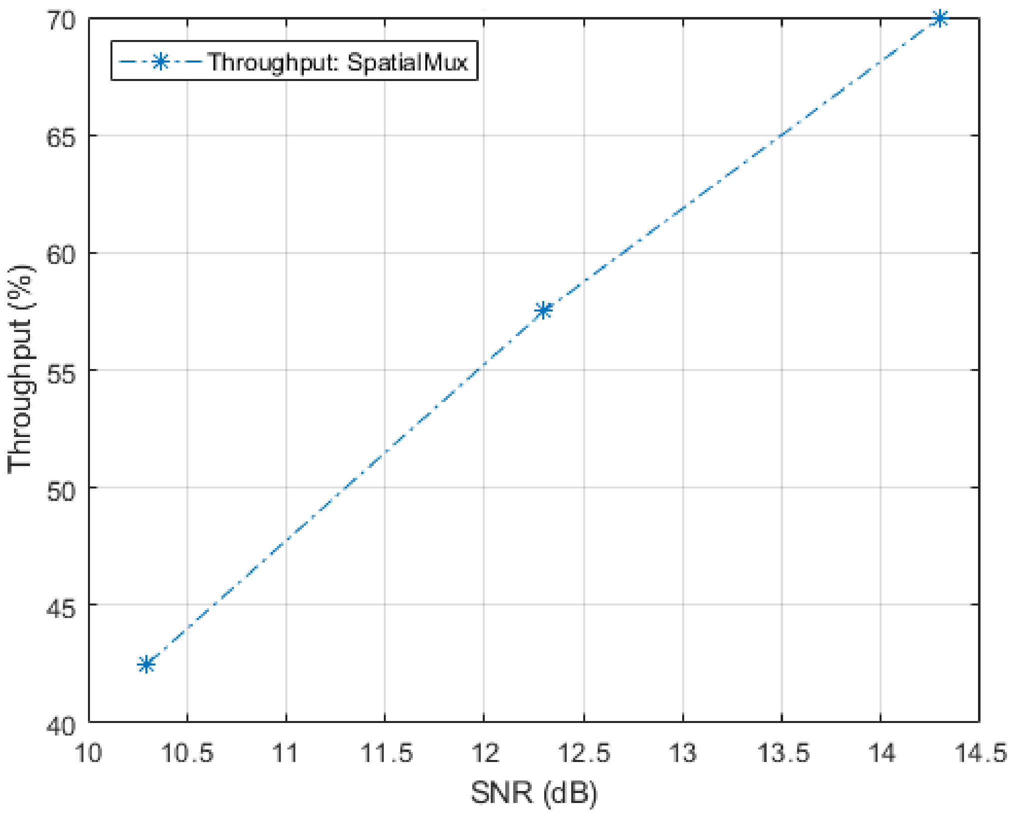
Table 2.2: Comparison of selected efficiency-focused scheduling techniques

Reference	**Model / Approach**	**Cell Type**	**EE**	**SE**	**TE**	**QoS**	**IA**	**FA**	**Priority**	**CLO**	**M2M?**
Rekhissa *et. al.* [93]	Resource Partitioning	S/MC	✓	N	✓	×	×	×	CQI	×	M2M/H2H
Azari *et. al.* [107]	Min-Max	S/MC	✓	✓	×	×	×	×	Energy	×	H2H
Shen *et. al.* [108]	Reduced PRB	S/MC	✓	×	×	✓	×	✓	Delay	✓	M2M
Azari *et. al.* [109]	Optimal Cluster Size	S/MC	✓	×	×	✓	×	×	Life Time	×	M2M
Wang *et. al.* [110]	Optimal MCS	S/MC	✓	×	✓	×	×	×	None	×	M2M
Li *et. al.* [86]	MILP	S/MC	✓	×	✓	×	×	×	None	✓	M2M
Yaacoub *et. al.* [111]	Gready	S/MC	×	✓	✓	×	×	×	None	×	H2H
Lin *et. al.* [87]	Gready	S/MC	×	×	✓	✓	×	✓	CQI / BSR	×	H2H
Alawi *et. al.* [40]	Game-Theoretic	S/MC	×	✓	✓	×	×	×	Delay/GBR	✓	H2H
Wang *et. al.* [112]	Game-Theoretic	S/MC	×	✓	✓	×	✓	×	None	×	M2M/H2H
Safdar *et. al.* [113]	Game-Theoretic	S/FC	×	✓	×	×	×	×	None	×	H2H
Tseng *et. al.* [114]	Genetic	S/MC	×	✓	✓	×	×	×	None	×	H2H

Note:- Macro Cell (MC) Femto Cell(FC), Single-Cell(S), Multi-Cell(Mu), Energy Efficient(EE), Spectrum Efficient (SE), Throughput Efficient (TE), QoS Support (QoS), Interference Avoidance (IA), Fairness Achievement (FA), Cross-Layer Optimization (CLO)



(a) Throughput vs MCS [84]



(b) Throughput vs SNR

Figure 2.9: Relationship between MCS, SNR, and throughput

achievable throughput and the number of PRBs as follows.

$$max \sum_{K=1}^{k} \bigcup (R_k | I_{RB,k}) \tag{2.26}$$

Where R_k and $I_{RB,k}$ are achievable throughput and allocation of *RBs* to the user k, the authors proposed an algorithm that greedily allocates each RB to a corresponding user, which causes a maximum increase in throughput. In [87], Lin *et al.* proposed a channel-aware and buffer-aware technique that sets priorities based on CQI and BSR values. The technique performs better than proportional fairness, purely opportunistic, and round robin against fairness, throughput, and packet loss probability.

Alawi *et al.* [40] proposed a scheme to meet the user's minimum rate and delay requirements by considering MAC and physical layer information. The authors used the game-theoretic approach by applying two cooperative games, nontransferable utility (NTU) and Transferable utility(TU). The authors use the Nash bargaining solution for NTU and a coalition-based method for solving TU.

Similarly, Wang *et al.* [112] proposed a Nash Bargaining-based game theoretic model for optimal resource allocation to maximize throughput as per the QoS of UEs and MTCDs. The authors divide the problem into two sub-problems. The first is channel allocation, and the second sub-problem is power allocation. The authors model the channel allocation problem as a matching problem between UEs and MTCDs, where MTCDs (max 2) share channels with a UE. The authors use Exhausted and KM algorithms to solve this matching problem to maximize unit and system earnings. According to the paper, maximizing unit and system earnings is equivalent to reducing interference in the common channel. The power allocation problem is solved by restricting the power of MTCDs to a threshold value. The UEs are allocated power to maximize throughput using the Lagrangian multiplier method. Safdar *et al.* [113] proposed another approach where the authors use both cooperative and non-cooperative games for the femtocell environment.

Tseng *et al.* [114] proposed a genetic algorithm-based technique that uses binary bit chromosome mutation based on fitness values. The fitness value is defined using resource block pairs. The selection procedure for next-generation parents is carried out using two methods: roulette wheel selection, which is similar to Russian roulette in which larger blocks are allocated to chromosomes with larger fitness values, and the tournament selection method in which random mating is performed, and the best-performing offspring becomes the subsequent parent. The procedure continues until either the desired convergent rate is obtained or the number of generations reaches a threshold. The authors compare their algorithm against the random allocation method, and their algorithm performed better in terms of throughput and packet service rates over the range of N_u (number of users).

2.2.2 Group-based techniques

M2M communication has many connected devices and diverse traffic patterns. Sometimes, data (i.e., sensory data) need not be sent immediately to the server due to a high correlation or delay-tolerant nature. So the data can be preprocessed at intermediary nodes (i.e., MTCGs) to reduce traffic and energy consumption of MTCDs through data aggregation and preprocessing [31, 32, 119]. For example, a temperature sensor sends temperature readings (T_S) every 30 seconds, but the temperature up to a threshold (T_{thr}) is acceptable. Then, sending readings if $T_s < T_{thr}$ is unnecessary. These intermediary nodes are known as MTCGs or Aggregators [120, 121]. The main objectives of group communication in the LTE network are as follows.

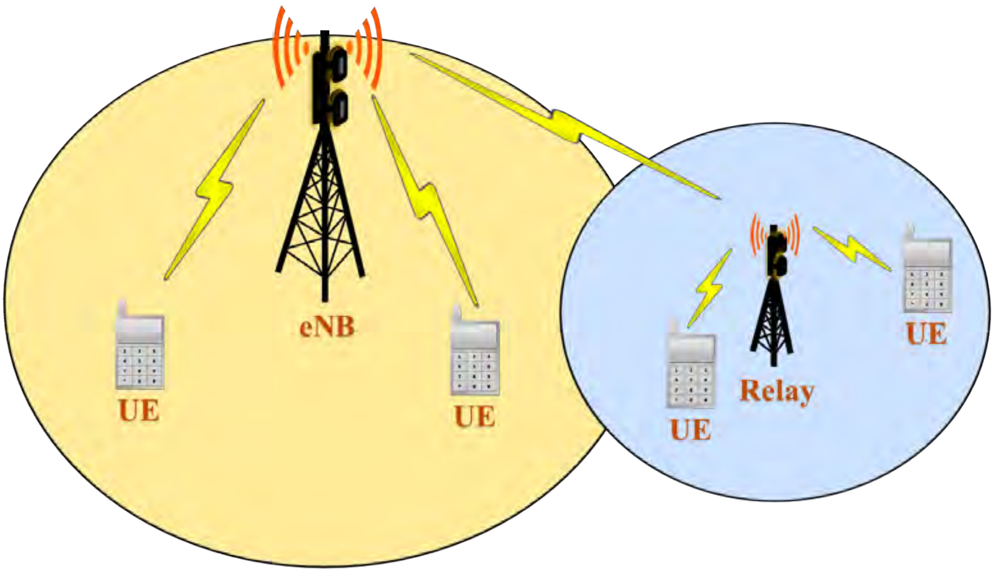
- *Data aggregation*:- When multiple MTCDs transmit small packets and delay-tolerant co-related data, then the data packets from different MTCDs can be aggregated at an intermediary node and sent together to save bandwidth and to reduce frequent scheduling [31].

- *Connectivity support*:- Gateways, which have the dual connectivity facility, can provide connectivity support to non-LTE devices [122].

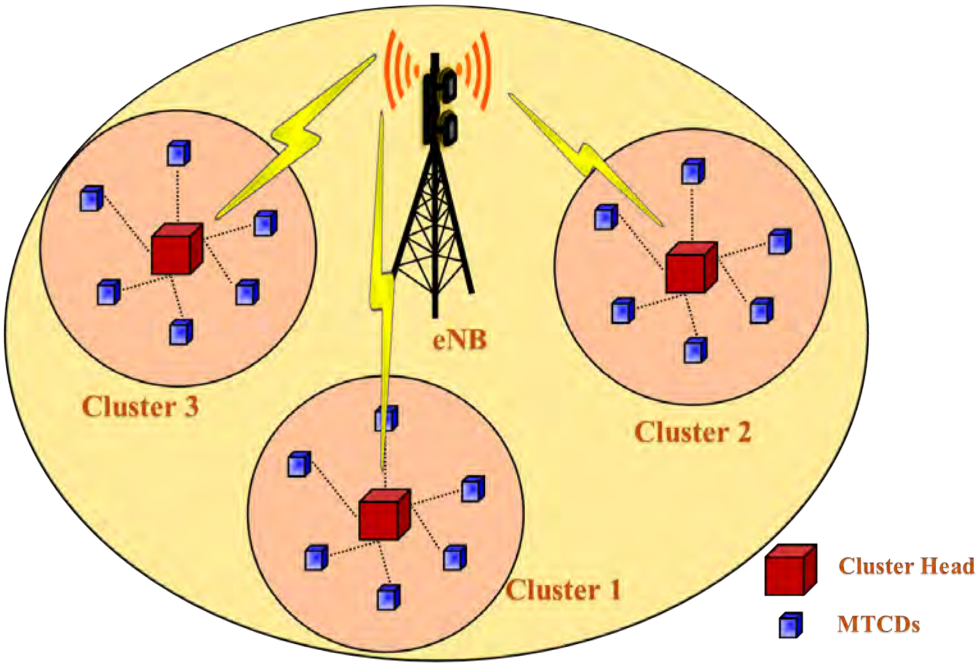
- *Reduced network traffic*:- Unnecessary data transmission can be avoided through preprocessing data at the gateway to reduce the network traffic, for example, sending only the temperature readings above the threshold value [65].
- *Reduced energy consumption*:- Energy consumption can be optimized by limiting packet transmission frequency and reducing transmission time [72].
- *QoS support for MTCDs*:- Gateway-based communication can provide QoS support to MTCDs through preprocessing and intelligent decision approach [32].

The grouping of devices is generally based on the characteristics and requirements of devices, which can be classified as the following criteria [123, 68].

- *QoS requirements of devices*:- the MTCDs having the same QoS requirements can be grouped to support a QoS-aware scheduling decision [124].
- *Communication protocol*:- MTCDs can be grouped, which have the same communication protocols, such as WiFi, BLE, ZigBee, etc., to support ease of connectivity with the gateway.
- *Data generation and traffic pattern*:- MTCDs that have the same data generation pattern (i.e., Time Trigger or Event Trigger) and same traffic characteristics (i.e., Periodic, Burst, Frequent) can be grouped to avoid the frequent scheduling [125].
- *Payload size*:- To support the data aggregation approach, MTCDs can be grouped based on the payload size (i.e., Small/Medium/Large).
- *Physical layer parameter*:- MTCDs can be grouped based on the physical layer properties (i.e., Channel Quality, Tx power) to support better resource utilization.



(a) Relay Based



(b) Cluster Based

Figure 2.10: Group communication paradigms in LTE [41]

- *Locality of device*:- MTCDs can be grouped based on the distance from different gateways. This approach can improve energy efficiency through short-range communication.

There are two types of group communication: relay-based and cluster-based. In relay-based communication, the relay serves as an eNB for the devices that cannot directly communicate with the eNB. A UE can, with higher capacity, work as a relay. In cluster-based communication, a UE acts as a cluster head, forwards data from UE to eNB, and vice-versa [123, 31]. Fig. 2.10 shows a typical group communication paradigm in an LTE network. The research community has extensive work in group-based resource scheduling for the LTE network, mostly focused on energy consumption and data traffic.

In [123], Songsong *et al.* proposed a proportional fairness algorithm using user grouping. The algorithm groups devices based on the number of carriers they can be assigned. The carriers with the same bandwidth are grouped into L aggregated carriers. Each aggregated carrier has V resource blocks. The power allocated to a carrier in a group carrier is given as $P_c = V/P_T$. Results indicate that this algorithm provides better fairness than proportional fairness (PF) with degradation in throughput.

In [72] Ho *et al.* proposed an energy-conserving 2-hop transmission-based allocation scheme. The author defines the maximum achievable bit rate for each subcarrier as follows.

$$f_{c_j} = \frac{r_{c_j}}{p_{c_j} + p_{cir}} = \frac{log_2(1 + p_{c_j}|h_{c_j}|^2/N_0 * B_c}{p_{c_j} + p_{cir}} \tag{2.27}$$

The author gets power for the group coordinator by maximizing Eq. (2.27) through an iterative process. The authors also propose an optimal number of coordinators in a group.

Gotsis *et al.* [68] proposed a queue-aware QoS-based scheduling technique. The UEs have been grouped into L clusters based on the individual QoS requirements of UEs. Physical resources are shared among clusters rather than directly to individual

Table 2.3: Comparison of selected group-based scheduling techniques

Reference	**Model / Approach**	**Cell Type**	**EE**	**SE**	**TE**	**QoS**	**IA**	**FA**	**Priority**	**CLO**	**M2M?**
Songsong *et. al.* [123]	Carrier Aggregation	S/MC	×	✓	×	×	×	✓	None	×	H2H
Ho *et. al.* [72]	Cooperative	S/MC	✓	✓	×	×	×	×	None	×	H2H
Gotsis *et. al.* [68]	Probabilistic	S/MC	×	×	×	✓	×	×	QCI	×	H2H
Xu *et. al.* [31]	QCI Clustering	S/MC	×	×	×	✓	×	×	Delay	×	H2H
Frank *et. al.* [65]	Cooperative	S/MC	×	✓	×	×	✓	✓	None	×	H2H
Hsu *et. al.* [32]	QCI Clustering	S/MC	×	✓	×	✓	×	×	QCI	×	M2M/H2H
Bayat *et. al.* [126]	Distributed Coalition	S/MC	✓	✓	×	✓	×	×	Delay	×	M2M

Note:- Macro Cell (MC) Femto Cell(FC), Single-Cell(S), Multi-Cell(Mu), Energy Efficient(EE), Spectrum Efficient (SE), Throughput Efficient (TE), QoS Support (QoS), Interference Avoidance (IA), Fairness Achievement (FA), Cross-Layer Optimization (CLO)

UE. The authors defined a probabilistic model such that the probability of violation δ of maximum delay threshold $\triangle$ is given as follows.

$$\delta = Prob\{W_q > \triangle\} \approx e^{-r.\theta(r).\triangle} \tag{2.28}$$

Where W_q is the experienced packet delay and θ is the QoS factor of the group.

Xu *et al.* [31] propose a group-based scheme for Random Access (RA) and uplink scheduling procedure. A UE (M) can be a member of the group (k) if it satisfies the following condition.

$$M_k = (n_{TB} * C_{TB})/C_{MTCD} \tag{2.29}$$

Where n_{TB} is the number of resource blocks allocated for transmission, C_{TB}, and C_{MTCD} are the capacity of resource block and UE, respectively. The authors consider a priority for a group of UE for uplink scheduling. The authors propose a group paging scheme for resource allocation that improves delay and access probability.

Frank *et al.* [65] defined a scheme to reduce the effects of cell-edge interference, which involves multiple adjacent base stations to communicate multi-cellular CSI reports through a fast back-haul network to a central scheduling unit. The authors proposed an interference-aware uplink scheduling algorithm based on a proportional fairness approach to avoid inter-cell interference.

Hsu *et al.* [32] proposed an enhanced cooperative access class barring and traffic adaptive radio resource management (ECACB + TARRM) for M2M devices. This technique builds upon enhanced cooperative access class barring (ECACB), to which the authors add support for UEs and TARRM. TARRM implies that UEs and MTCs use different PRBs, preambles, and MTC devices, which are clustered based on their data and random access rates. To determine better parameters for access class barring, several MTC devices are used as a factor over the factors used by CACB that connect to an eNB.

Bayat *et al.* [126] proposed a distributed coalition forming algorithm that involves the rules called *"merge-and-split."* The authors used data aggregation for machine-type devices with different delay requirements and proposed a game-theoretic approach using coalition games. The algorithm allows MTCDs to organize into groups independently, with each group head handling data to and from the machines in each group.

2.2.3 Quality of Service Based Techniques

The M2M communication differs from regular H2H communication regarding the number of connected devices, packet size, data transmission frequency, a broad range of delay budgets, throughput requirements, priority, etc. ETSI defines QoS class identifier for M2M communications in LTE [88, 127, 128], as shown in Table 1.1. As the M2M communication system has many devices with varying QoS requirements, designing scheduling schemes for M2M communication over the cellular network is a challenging task [67, 129]. Many MTC devices are infrequently sending varying data packets; the LTE bandwidth offers limited physical resources optimized for H2H communication [89]. Therefore, it is required to design a solution for M2M communication that optimizes the available physical resources while satisfying the unique QoS requirements of M2M communication [76, 130, 49].

AI-Rawi *et al.* [37] proposed an opportunistic channel adaptive radio resource scheduling algorithm for dynamic traffic patterns based on the buffer sizes of users. The scheduler estimates the expected rate $\mu_{n,i}$ of device i if the resource block n is allocated to the device with an estimated throughput of x_i.

$$y(t) = arg\,\underset{y}{max} \sum_{i=1}^{N} \sum_{n=1}^{C} u_i(x_i)\mu_{n,i}y_{i,n} \tag{2.30}$$

To find optimal allocation, the scheduler maximizes the Eq. (2.30). The authors evaluated pruning, a process of recovering weaker bands for use by other UEs. The authors evaluated the effect of delays in receiving buffer information and concluded

Table 2.4: Comparison of selected QoS-based scheduling techniques

Reference	Model / Approach	Cell Type	EE	SE	TE	QoS	IA	FA	Priority	CLO	M2M?
AI-Rawi *et. al.* [37]	Heuristic	S/MC	×	✓	×	✓	×	✓	Delay	×	H2H
Delgado *et. al.* [60]	Greedy	S/MC	×	×	✓	✓	×	×	GBR	×	H2H
Afrin *et. al.* [15]	Greedy	S/MC	×	×	×	✓	×	×	Delay	✓	H2H
Safa *et. al.* [131]	Greedy	S/MC	×	×	×	✓	×	✓	QCI	×	M2M
Mata *et. al.* [132]	Genetic	S/MC	×	×	✓	✓	×	×	GBR/Delay	×	H2H
Maia *et. al.* [133]	Deterministic	S/MC	×	✓	×	✓	×	✓	None	×	H2H/M2M
Afrin *et. al.* [134]	Deterministic	S/MC	×	×	×	✓	×	×	Delay	×	M2M
Giluka *et. al.* [66]	Utility Based	S/MC	×	×	✓	✓	×	×	GBR/Delay	×	H2H/M2M
Brown *et. al.* [135]	MLHE	S/MC	×	×	×	✓	×	×	Delay	×	M2M
Agdhmadi *et. al.* [136]	Flow Based	S/MC	×	×	✓	✓	×	×	GBR	×	M2M
Kumar *et. al.* [78]	Queue Awareness	S/MC	×	×	×	✓	×	✓	Delay	×	M2M
Abdelsadek *et. al.* [11]	Queuing Model	S/MC	×	✓	×	✓	×	×	Delay	×	H2H/M2M
Alaa *et. al.* [36]	Queue Awareness	S/MC	×	✓	×	✓	×	×	Delay	×	H2H/M2M
Kumar *et. al.* [79]	Queue Awareness	S/MC	×	✓	×	✓	×	×	Delay	×	M2M
Karadag *et. al.* [76]	Graph Theory	S/MC	×	×	×	✓	×	×	Delay	×	M2M
Ouaissa *et. al.* [100]	Heuristic	S/MC	×	✓	×	✓	×	×	QCI	×	M2M

Note:- Macro Cell (MC) Femto Cell(FC), Single-Cell(S), Multi-Cell(Mu), Energy Efficient(EE), Spectrum Efficient (SE), Throughput Efficient (TE), QoS Support (QoS), Interference Avoidance (IA), Fairness Achievement (FA), Cross-Layer Optimization (CLO)

that it would result in better fairness, whereas limited buffer information would lead to inefficient resource usage.

Delgado *et al.* [60] defined a utility function $\bigcup(R_k|S_k)$ as a function of throughput R_k and set of allotted resources S_k of device k and maximize the utility of the device for optimal allocation of resources as follows.

$$max \sum_{k \in K} \bigcup(R_k|S_k) \tag{2.31}$$

The authors proposed two algorithms that aim to reduce delay while having a minimum throughput constraint. The authors use two highly scalable greedy heuristics based on the problem.

Afrin *et al.* [15] defined an urgency metric U_i for the device i as a function of deadline d_i and BSR index B_i as follows.

$$U_i = \begin{cases} \frac{B_i}{max(B)} * \frac{T_{SF}}{d_i - t} & \text{if } d_i - t > 1 \\ 1 & \text{Otherwise} \end{cases} \tag{2.32}$$

The devices are selected based on the urgency metric U_i to improve the satisfaction of delay requirements. This approach allows the eNB to know the age of the oldest packet in their buffer using a new MAC control field in the MAC PDU. Afrin *et al.* [134] proposed a buffer-based adaptive semi-persistent scheduling (SPS) scheme, which does not have the same overheads as dynamic scheduling while offering the same flexibility. The authors examine the influence of semi-persistent scheduling on the QoS and compare it with fixed allocation SPS schemes.

Safa *et al.* [131] proposed a technique to satisfy the delay requirements of M2M devices. The authors defined a QoS aware allocation metric $\gamma_i^c(t)$ of a UE i for QoS introducer $\alpha_i(t)$ of UE as follows.

$$\gamma_i^c(t) = \frac{\lambda_i^c(t)}{\alpha_i(t)} \tag{2.33}$$

Their technique prioritizes the UEs having high-priority data while not starving others.

Mata *et al.* [132] proposed a genetic algorithm-based approach to optimize video streaming with a focus on video chat. The authors defined PF metric $\lambda_n^m(t)$ as a ratio of instantaneously achieved data rate $r_n^m(t)$ to the long-term rate $R_n(t)$ for a user n with assigned resources m over some time t as follows.

$$\lambda_n^m(t) = \frac{r_n^m(t)}{R_n(t)} \tag{2.34}$$

The authors also defined a metric based on the number of packets residing in the UE's buffer.

Maia *et al.* [133] proposed an extension to the QoS classes for M2M in two groups, event-based and time-based. The authors try to control the effect of M2M communication on H2H communication by calculating the current demand for H2H communication as a ratio of the average resource allocated to the average buffer size, expressed as follows.

$$\widehat{B_H}(u,t) = \frac{BS_{u,t} * RB_{u,t-1}^{avg}}{BS_{u,t-1}^{avg}} \tag{2.35}$$

Resources are shared among H2H and M2M devices based on the current demand. Maia *et al.* [137] proposed a genetic algorithm-based method and introduced a new scheme of initialization, crossover, mutation, and a QoS-aware fitness function.

Kumar *et al.* [78] proposed a multi-class scheduler for MTCDs with different delay requirements. The authors classify M2M data into periodic updates (PU) and event-driven (EU). The authors aimed to maximize utility by using heuristics and a sigmoid-based utility function for each device type. Their algorithm prioritizes ED data over PU data as long as PU deadlines are met and ensures that congestion due to failed updates is reduced. Kumar *et al.* [79] proposed a delay in optimal scheduling strategy in which multiple M2M devices communicate with an application server from multiple M2M aggregators.

Giluka *et al.* [66] proposed a classification and prioritization scheme of M2M and H2H service flows based on QoS requirements. In a given class, H2H devices are given higher priority, while a limit is set as the maximum limit for radio resource block assignment to MTCDs. The authors define the utility of a QoS class C_i as follows.

$$C_i = \sum S(H) + \beta_i * \sum S(M) \tag{2.36}$$

Where $S(H)$ and $S(M)$ are the satisfiability functions of H2H and M2M communication requests.

Agdhmadi *et al.* [136] proposed a scheme to provide QoS to Guaranteed Bit Rate (GBR) services based on QCI and using priorities for M2M. Erpek *et al.* [138] proposed a scheme that prioritizes delay-bound traffic over delay-insensitive traffic. The authors implement a utility proportional fairness policy based on the same.

Brown *et al.* [135] proposed a predictive resource allocation scheme using Maximum likelihood estimation (MLHE) and defined MLHE as follows.

$$\begin{aligned} \mathscr{L}(t_1, t_2, \dots t_n) &= Pr(t_1, t_2, \dots t_n | \tau) \\ &= \{(t_1, t_2, \dots t_n | \tau) \end{aligned} \tag{2.37}$$

$$\Rightarrow \mathscr{L} = \begin{cases} \frac{a-b+1}{\sigma}, & \text{if } a \geq b \\ 0 & \text{Otherwise} \end{cases} \tag{2.38}$$

Where,

$$a = min\{r_1 - 1, r_2 - 1 - \tau, \dots r_n - 1 - (n-1)\tau\}$$

$$b = max\{r_1 - \sigma, r_2 - \sigma - \tau, \dots r_n - \sigma - (n-1)\tau\}$$

The authors used inter-sensor propagation time to determine when it will reach downstream sensors. This approach allows sensors to send fewer scheduling requests, reducing traffic and delay.

Abdelsadek *et al.* [11] proposed a scheme considering the scheduler as an M/D/1 queues model. The achieved throughput for the UEs u is given as follows.

$$R_u = \begin{cases} \mu_u & \text{if } \mu_u \leq \lambda_u \\ \lambda_u & \text{if } \mu_u \geq \lambda_u \end{cases} \tag{2.39}$$

Where λ_u and μ_u are the average arrival rate and average service rate, respectively, the authors improved the computational efficiency of the optimization problem.

Alaa *et al.* [36] proposed a non-preemptive queuing model and investigated the scheduling performance for different QoS classes of M2M and H2H devices with dynamic access grant time interval scheduling. The authors use the M/G/c/c model to improve bandwidth utilization and QoS satisfaction.

Karadag *et al.* [76] proposed an optimization approach for MTCD transmissions considering the repetitive nature of these transmissions. The authors proposed semi-persistent scheduling and implementation using the Depth-First and minimum frequency-fit approaches to reduce the frequency bands used by MTCDs while maintaining delay requirements. The authors proposed a heuristic algorithm in polynomial time with fixed priority assignments to solve this problem. Ouaissa *et al.* [100] proposed a hybrid model of RR, First Maximum Expansion, and Maximum Throughput.

Abdalla *et al.* [139] proposed a technique that aims to retain the Quality-of-Experience (QoE) of UEs while processing the message requests of M2M devices. The authors proposed a new set of QCIs for M2M to ensure end-to-end QoS. Hassebo *et al.* [70] proposed a technique that aims to manage QoS requirements of M2M devices along with massive access while protecting H2H devices from a lapse in service quality. The authors used a semi-persistent approach for scheduling many MTCDs while using typical dynamic scheduling for H2H devices.

2.2.4 Hybrid / Multi-Objective techniques

Multiple applications running on a single device can require optimized connectivity on multiple parameters. For example, smartwatches have streaming applications and blood pressure sensors. The streaming application requires high bandwidth, while blood pressure sensors require urgent network access. This scenario requires a scheduling methodology optimized for multiple parameters like throughput, delay, and priority. For such scenarios, researchers proposed hybrid scheduling mechanisms based on multiple metrics. In this section, we consider the scheduling algorithms that focus on multiple objectives in combination with any scheduling objective, like priority, QoS, throughput, energy, and fairness.

For example, Elhamy *et al.* [63] proposed a technique called *"BAT"* that aims to balance throughput and delay requirements. This hybrid technique simultaneously allocates M2M resources and UE using an RME-like expansion method. Maia *et al.* [137] proposed a technique that aims to reduce congestion, satisfy QoS requirements, and ensure fairness of the allocation of M2M devices while minimizing the effect on H2H traffic. Their algorithm uses a state transition function with three states to evaluate the probability of allocating M2M devices while optimizing for the said factors. Kwan *et al.* [80] proposed the classic throughput and fairness balancing approach called PF scheduler, which aimed to improve the fairness of the Max-Rate scheduler while taking some loss in throughput. AlQahtani *et al.* [45] proposed a scheduling technique that borrows from RR and Best-CQI (B-CQI) to solve the fairness and throughput trade-off. RR provides ideal fairness, whereas B-CQI provides high data rates with weak fairness. Results by testing against RR and B-CQI indicated that the technique provides a balance between fairness and throughput. AlQahtani *et al.* [140] proposed a queuing-model-based access strategy for H2H and M2M coexistence. The authors evaluated the system performance using a continuous-time Markov-Chain model. Results indicate that this technique increases overall resource utilization while decreasing blocking probability.

Table 2.5: Comparison of selected hybrid/multi-objective scheduling techniques

Reference	Model / Approach	Cell Type	EE	SE	TE	QoS	IA	FA	Priority	CLO	M2M?
Elhamy *et. al.* [63]	Weighted Sum	S/MC	×	✓	✓	✓	×	×	Delay	×	M2M/H2H
Maia *et. al.* [137]	Weighted Sum	S/MC	×	✓	×	✓	×	✓	None	×	M2M/H2H
Kwan *et. al.* [80]	Weighted Sum	S/MC	×	×	✓	×	×	✓	Max Weight	×	H2H
AlQahtani *et. al.* [45]	Weighted Sum	S/MC	×	×	✓	×	×	✓	Max Weight	×	H2H
AlQahtani *et. al.* [140]	Markov Chain	S/MC	×	✓	✓	×	×	×	None	×	M2M/H2H
Mardani *et. al.* [98]	Fuzzy	S/MC	✓	×	×	×	×	×	None	×	M2M/H2H
Mardani *et. al.* [99]	MINLP	S/MC	✓	×	✓	✓	×	×	Delay	×	H2H
Aijaz *et. al.* [141]	MIP	S/MC	✓	×	✓	✓	×	×	Statistical QoS	×	M2M/H2H
Dawaliby *et. al.* [59]	0/1 Knapsack	S/MC	×	×	✓	✓	×	×	Delay	×	H2H
Dawaliby *et. al.* [48]	Memetic	S/MC	✓	×	×	✓	×	×	QCI	✓	M2M
Kalil *et. al.* [73]	Genetic	S/MC	×	×	✓	×	×	✓	None	×	H2H
Tagarian *et. al.* [94]	Genetic	S/MC	✓	×	×	✓	×	×	Delay	×	H2H
Fagan *et. al.* [64]	Deep-Learning	S/MC	×	✓	×	×	×	×	GBR	×	M2M/H2H
Comsa *et. al.* [142]	Neural Network	S/MC	×	×	✓	×	×	✓	None	×	H2H
Chen *et. al.* [91]	Heuristic	S/MC	✓	✓	×	✓	×	×	QCI	×	H2H
Abrignani *et. al.* [12]	Heuristic / MILP	S/MC	×	✓	✓	×	✓	×	None	×	H2H
Hamdoun *et. al.* [69]	Game-Theoretic	S/MC	✓	✓	×	✓	×	×	Group QoS	×	M2M/H2H
Lin *et. al.* [143]	Iterative	S/MC	✓	×	×	✓	×	×	Delay	×	M2M
Salam *et. al.* [144]	Queue Awareness	S/MC	✓	×	✓	✓	×	×	Delay	×	M2M
Edemacu *et. al.* [62]	Queuing Model	S/MC	×	✓	×	✓	×	×	QCI	×	M2M/H2H

Note:- Macro Cell (MC) Femto Cell(FC), Single-Cell(S), Multi-Cell(Mu), Energy Efficient(EE), Spectrum Efficient (SE), Throughput Efficient (TE), QoS Support (QoS), Interference Avoidance (IA), Fairness Achievement (FA), Cross-Layer Optimization (CLO)

Mardani *et al.* [98] proposed a technique that aims to minimize energy consumption while maintaining the QoS requirements of H2H devices. The authors used a fuzzy logic-based controller that anticipates and manages uncertainties and obtained an optimal bandwidth ratio for each type of service flow. Mardani *et al.* [99] proposed a technique to maximize throughput and satisfy power budget constraints and statistical QoS delay requirements. The authors defined the problem as a mixed-integer non-linear problem and proposed a solution using "Lagrange multipliers". Aijaz *et al.* [141] proposed a technique to minimize energy consumption and statistical QoS provisioning for M2M and H2H devices. The authors defined the problem as a Mixed Integer Programming problem to maximize effective energy efficiency in bits-per-joule capacity. The authors solved this using the Canonical Duality Theory. The authors also proposed another approach, proposing two low-complexity heuristic techniques.

Dawaliby *et al.* [59] proposed a technique that aims to maximize throughput and reduce delays in the case of LTE-M protocol. The authors model the problem to the 0/1 knapsack problem. Dawaliby *et al.* [48] proposed a technique to minimize energy consumption while maintaining the QoS requirements of M2M devices. The authors employ a cross-layer scheme using a memetic-based algorithm. Memetic algorithms (MAs) are evolutionary algorithms that use another local search rather than global search algorithms. The authors consider the QoS requirements while minimizing energy consumption using discontinuous reception switching. Kalil *et al.* [73] evaluated a genetic algorithm that considers multiple constraints for the uplink scheduling problem. This approach is evaluated against the optimal allocation binary-integer programming problem (BIP). It offers performance comparable to the optimal solution for low population levels (<300 UEs) while having comparatively lower time complexity. Tagarian *et al.* [94] proposed a technique that aims to minimize energy consumption while maintaining delayed QoS requirements of machines. The authors used a gateway-based approach where the use of clustering manages massive access. The optimization problem is solved using genetic algorithms for maximizing energy efficiency. To manage delay, the authors used an existing scheduling approach.

Fagan et al. [64] applied a deep learning approach for downlink scheduling. The data set is derived using a genetic algorithm over many simulated random UE data reports and used to train the deep learning network. This approach allowed for approximating the genetic algorithm schedule without the delay of a genetic algorithm. Comsa *et al.* [142] proposed a scheduling scheme using the q-learning method to adjust the fairness and system capacity trade-off dynamically during each transmission time interval. The proposed algorithm decides allocation using CQI for each class of users.

Chen *et al.* [91] proposed a heuristic technique to minimize the energy consumption of MTCDs while guaranteeing the QoS. The authors minimized MTCDs' energy consumption using lower modulation, coding, and spatial reuse. Abrignani *et al.* [12] considered the problem of improving throughput while reducing resource usage and minimizing Inter-Cell Interference (ICI) in the case of a densely populated heterogeneous network. The authors modeled the problem using Mixed Integer Linear Programming (MILP). The authors employed a heuristics approach to solving MILP. The algorithm was compared against RR and performed better in terms of throughput.

Hamdoun et al. [69] considered an evolutionary game approach to preserve UE's QoS while preserving the battery life of MTCD. Here, MTCDs are in the same group and share a spectrum with a UE, which is matched to it. The MTCDs switch dynamically from non-cooperative to cooperative strategies. Results indicate that this adaptive technique performs better than a fixed discrete power allocation strategy and a non-cooperative strategy regarding power consumption and QoS satisfaction.

Salam et al. [144] proposed a technique to improve outage probability, energy efficiency, and system capacity called the cooperative data aggregation scheme, which employs fixed data aggregators and mobile data aggregators. These aggregators serve M2M devices with varied QoS requirements. The authors also considered the parameters of queuing delay and the number of devices not served in a class. Edema *et al.*

[62] presented a study on existing Fixed access grant time interval (AGTI) and Time-controlled scheduling and proposed a dynamic AGTI scheme based on M2M and H2H traffic intensities focusing on resource utilization and QoS satisfaction.

Lin et al. [143] proposed a technique to minimize energy consumption while maintaining the QoS requirements of M2M devices. Their algorithm used the concept of Multi-access edge computing. The authors considered packet processing time and travel time in latency calculation.

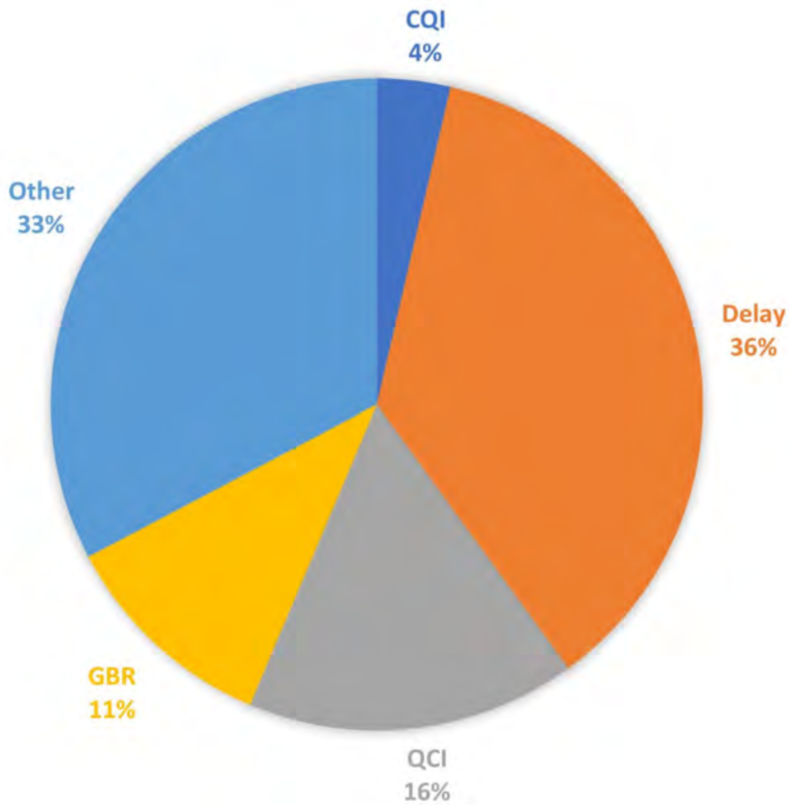
2.3 Findings from the Literature

Numerous works have been carried out on LTE radio resource scheduling in literature. As shown in figure 2.11a, based on the priority metrics, about 36% focused on delay budget satisfaction. Further, QoS class identifier and guaranteed bit rate are considered in 16% and 11% proposals in the literature, respectively. The authors worked on QoS, throughput, utilization, and energy efficiency in 30%, 20%, 21%, and 15% proposals as scheduling objectives, as shown in figure 2.11b. From figure 2.11b, we observe that 30% of work was done for QoS support. To provide QoS support, researchers focused on the three metrics: delay, priority, and GBR. The authors mostly used optimization and queuing theory about 27% and 15% in their proposals to meet scheduling objectives. Figure 2.12 shows the classification of literature based on the methodology used.

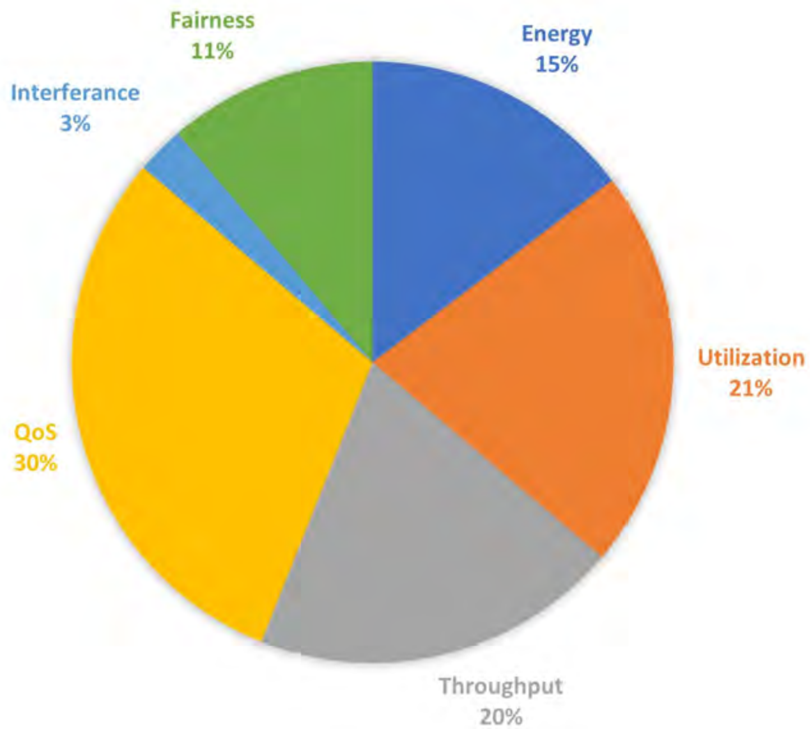
2.3.1 Research gaps

We find the following research gaps in LTE radio resource scheduling for M2M communication through the literature.

1. Scheduling schemes that use channel quality as prioritization metrics improve resource utilization and throughput. Because LTE can use higher modulation and coding schemes with higher channel quality, enabling a bigger transport



(a) Classification based on parameters used as priority metrics.



(b) Classification based on scheduling objectives.

Figure 2.11: Classification of radio resource scheduling in LTE.

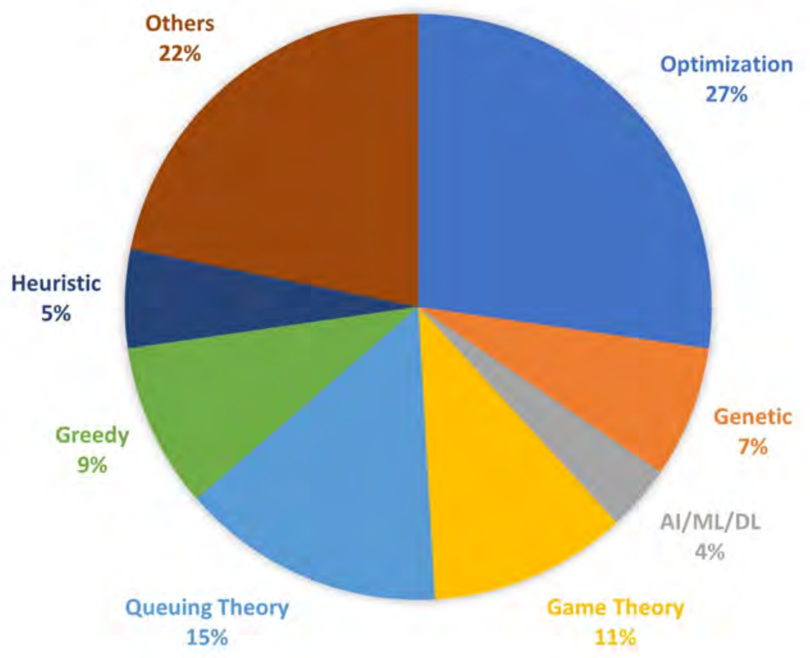


Figure 2.12: Classification based on the methodology used.

block size (TBS). These schemes lack priority support for mission-critical applications [93, 107, 108, 109, 110, 86].

2. If channel quality is preferred in allocation, it can cause starvation for the devices with poor channel quality. Generally, M2M devices are stationary and have a scarce chance of changes in channel conditions.

3. Mission critical applications, i.e., fire alarm required Non-Guaranteed bit rate (N-GBR) resource type (reference, table 1.1). Taking Guaranteed Bit Rate (GBR) as a prioritization metric can fail to support mission-critical applications [60, 66, 136, 64, 40].

4. Delay budget is aligned with the application priority (reference, table 1.1). Taking application priority as a resource allocation metric improve urgent channel access for mission-critical applications and delay budget violation. It reduces the chance for lower-priority devices [108, 40, 31, 126, 37, 15, 132, 66, 78, 11, 36, 79, 76, 63, 98, 94, 143, 144].

5. When the number of devices and the number of different types of applications is high, the lower-priority devices rarely get access to the network due to the presence of high-priority applications, even if high-priority applications have some delay budget [63, 137, 99, 141, 59, 48, 94, 91, 69, 143, 144, 62].

6. Researchers gave prioritization as a ratio of channel quality and application priority. This improves throughput, resource utilization, and priority support. However, in M2M communication, devices are static and have predetermined application priorities. There are rare chances of improvement in channel quality. Therefore, devices with poor channel quality and low application priority can be in starvation [48].

7. If more than one device has the same application priority and different channel conditions. A device with good channel conditions is always preferred by giving prioritization as a ratio of channel quality and application priority, even if the device is already given a chance to access the network [48].

8. In some proposals, a few traffic classes are considered, i.e., event-driven and periodic-update traffic types. Event-driven traffic is preferred over periodic-update traffic. This traffic prioritization does not align with priorities defined by standardization bodies [79].

9. Numerous works were presented in the literature to address the QoS requirement using delay budget, resource type, and application priority. Although, no work addresses the application priority decision problem in surveyed literature.

10. We observe from the surveyed literature that the scheduling mechanisms that focus on QoS parameters, i.e., priority and delay, lack in system performance parameters, i.e., throughput and resource utilization.

2.3.2 Test case scenario

Among the use cases for M2M communication are smart homes, smart agriculture, smart animal husbandry, smart cities, smart environments, smart water, smart metering, security and emergency, retail, logistics, and industrial control. M2M system has a heterogeneous environment where different types of applications work together. Numerous solutions fit best for the particular application domains. These solutions lack one performance parameter while focusing on others. Therefore, to support a heterogeneous environment for the M2M system, there is a need for a solution that provides a balanced performance of a set of parameters. Considering the above point, we choose the smart-building scenario as a test case for our work. Figure 2.13 illustrates the smart-building scenario, and table 2.6 provides a list of sensor-based applications in a smart building.

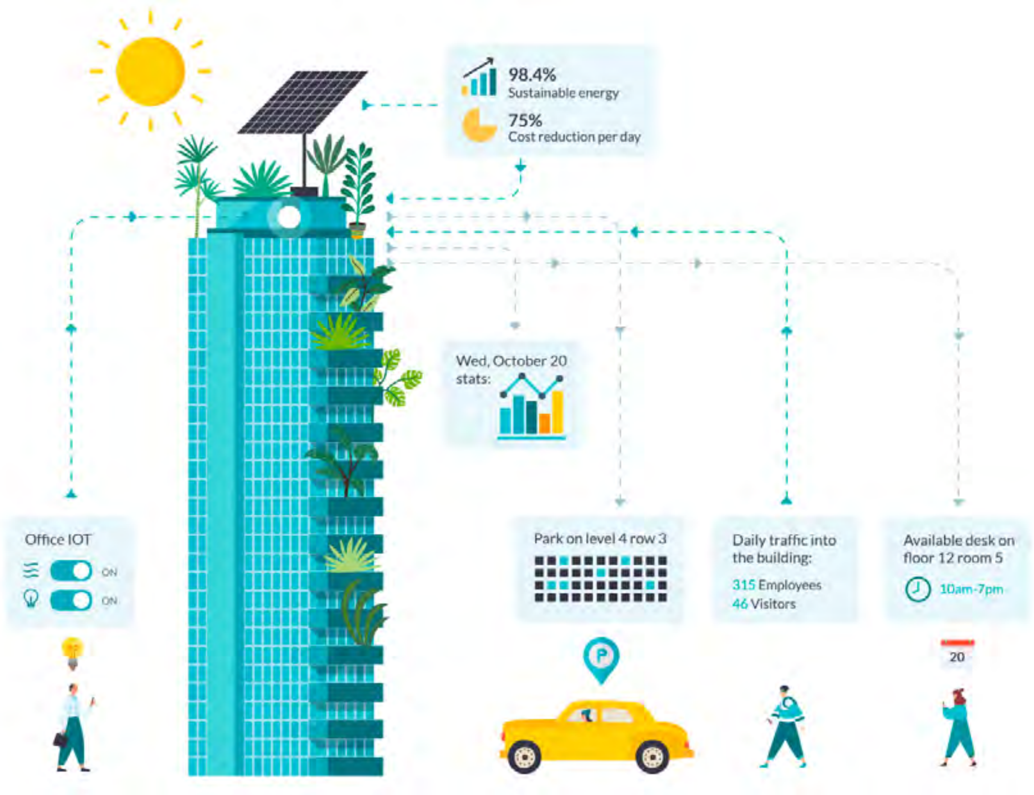


Figure 2.13: An illustration of smart-building scenario (taken from [145]).

Table 2.6: List of sensor-based applications in a smart building.

QCI	Priority	Application	Description
1	2	Perimeter access control	Access control to restricted areas and detection of people in non-authorized areas
1	2	Fall detection	Assistance for elderly or disabled people living independently
1	2	Intrusion detection	Detection of window and door openings
2	4	UV radiation	Measurement of UV sun rays to warn people
2	4	EMF levels	Energy radiated by cell stations and WiFi routers
2	4	Indoor air quality	Monitoring of toxic gas and oxygen levels.
3	3	Structural health	Monitoring of vibrations and material conditions in a building.
3	3	Patient surveillance	Monitoring of conditions of patients
4	5	Smart parking	Monitoring of parking spaces available.
4	5	Smartphone detection	WiFi or Bluetooth enabled device detection
4	5	Temperature monitoring	Controlling temperature inside the building.
4	5	Indoor location	Asset location using active and passive tags
4	5	Appliances control	Switching on and off appliances remotely.
5	1	Explosive detection	Detection of hazardous gas levels and leakages.
5	1	Fire alarm	Accidental warnings.
5	1	Earthquake alarm	Accidental warnings.
6	6	Smart lighting	Intelligent weather and adaptive lighting
6	6	Waste management	Detection of rubbish levels to optimize trash collection.
6	6	Water monitoring	Monitoring the quality of tap water.
6	6	Tank level	Monotring auto refilling water tank.
7	7	Swimming pool remote measurement	Controlling swimming pool conditions such as temperature and pH remotely
7	7	Water leakages	Detection of a liquid presence outside tanks and pressure variations along pipes
7	7	Smart Electricity and Water Meter	Energy and water consumption monitoring and management
7	7	Smart Gas Meter	Gas consumption monitoring and management

2.3.3 Possible scenarios for an incorrect priority claim

Each MTCD has a single application installed. Application has a priority corresponding to their service type. This priority is pre-decided according to the type of application. The priority can be falsely increased due to a bug in the application code, in the API, or in the host operating system being used. M2M devices can behave arbitrarily due to environmental conditions like overheating or moisture that can impact hardware/ chipset levels.

Due to a false priority (High priority) claim, a malfunctioning device can restrict the lower priority devices and limit high priority devices from getting a chance of communication channel access. Mission-critical applications have high priority and require urgent access. In a false priority claim scenario, mission-critical applications are highly affected due to limited channel availability. Lower-priority applications are rarely affected in a false-priority claim scenario due to the delay-tolerant nature of the application. Thus, a malfunctioning device's false priority (High Priority) claims can destabilize the M2M system.

2.3.4 Problem formulation

The M2M communication has many devices and diverse QoS requirements; for example, the delay budget ranges from 10 ms to several days, and the packet size ranges from 20 to 2000 bytes (ref. 3GPP TR 43.868 release 12). Numerous applications, like fire alerts in smart buildings, require urgent network access. As discussed in section 2.3.1, resource allocation mechanisms based on channel quality, application priority, resource types (GBR/N-GBR), traffic type, and the ratio of channel quality to application priority are not sufficient for the M2M communication. So, there is a requirement for a more feasible resource allocation mechanism to support the M2M requirement. So keeping given the above, the problem formulation for this thesis consists of the following steps:

1. A model is constructed to take decisions on application priority and integrate the model with a priority-based algorithm to validate the system performance and QoS trade-off problem in a heterogeneous environment. We assume the presence of malfunctioning devices in the environment.

2. A resource allocation mechanism is designed to minimize trade-offs between system performance and QoS support. The mechanism jointly optimizes the channel quality and application priority and introduces a scaling variable considering the total number of devices and QoS classes with delay-budget constraints.

2.3.5 Research objectives

Based on the above observation, our thesis objectives are defined as follows.

1. To develop a model to control the strategic behavior of MTCDs and provide a stable M2M system by revealing accurate application priority when malfunctioning devices are present in the network.

2. To investigate and validate the trade-off between QoS support and the system's performance and utilization using a priority-based resource allocation algorithm.

3. To provide support for application priority to enable urgent network access for mission-critical applications based on the application priority.

4. To improve resource utilization while providing QoS support by jointly optimizing channel quality and application priority.

5. To provide social fairness to support many connected devices and minimize starvation by adding a scaling mechanism.

We construct an auction-game model to control the strategic behavior of devices and implement an application priority-based resource allocation algorithm to address objectives 1 and 2 in chapter 3. We develop a scalable priority-based algorithm to address objectives 3, 4, and 5 in chapter 4.

2.4 Chapter Summery

- The chapter provides a comprehensive foundation of radio resource scheduling techniques in the LTE environment from the perspective of M2M communication.
- This chapter helps us identify the gaps in existing research for potential future research in M2M radio resource scheduling and highlights the primary methodologies employed by researchers.
- Through literature, we get an overview of the LTE network's fundamental architecture and physical layer concepts.
- It helps us to understand the basic architecture of M2M communication in the LTE network.
- This chapter helps to understand the following critical aspects of the LTE radio resource allocation process.
  - *Scheduling metrics* - The scheduling metrics are vital in preferring or selecting a particular UE over others while assigning resources to the UEs.
  - *Scheduling objectives* - The study provides an insight into the scheduling objectives like efficiency, QoS, etc., that were focused through previous research works.
  - *Scheduling methodologies* - The study helps to understand the pros and cons of previously employed approaches like game theory, queuing theory, etc. and provides direction toward the possibilities of implementing new methodologies in resource scheduling.
  - *Constraints and limitations* - In this literature review, we find out the constraints and limitations of LTE that draw a boundary for the resource scheduling process.
  - *Parameters that affect scheduling performance* - By comparing different scheduling work, we find out the parameters, i.e., MCS, number of PRBs, etc., that affect the performance of the scheduling methodology.
  - *Current state of LTE resource scheduling* - This study helps us get an integrated and synthesized overview of the current state of LTE scheduling.

Chapter 3

Decision on QoS Class Identifier (QCI) : An Auction Model

The content of this chapter is partially published in the following article.
U. Singh, A. Dua, N. Kumar and M. Guizani, "QoS Aware Uplink Scheduling for M2M Communication in LTE / LTE-A Network: A Game Theoretic Approach," in IEEE Transactions on Vehicular Technology, doi: 10.1109/TVT.2021.3132535.
[Q1, SCI, IF-5.987]
https://ieeexplore.ieee.org/document/9635654

3 | Decision on QoS Class Identifier (QCI): An Auction Model

"All outstanding work, in art as well as in science, results from immense zeal applied to a great idea."

Santiago Ramón y Cajal

Key Points:

- A combinatorial auction game model is proposed to make decisions on application priority in the presence of malfunctioning devices.
- We have worked in a scenario where some devices can malfunction for any reason (Ref. subsection 2.3.3) and can claim false priority.
- The constructed auction game model provides stability to the system with malfunctioning devices.
- We integrate the game-theoretic model with the application-priority-based resource allocation algorithm to investigate the system's performance and QoS support trade-offs.
- To investigate the trade-off between system performance and QoS support, the results of the proposed priority-based algorithm are compared to the state-of-the-art scheduler from the system's performance domain and proposals in the literature.

3.1 State-of-the-Art

Let's study the previous works on the quality-of-service-based scheduling approaches proposed by the research community. As discussed in chapter 1, M2M traffic is uplink-dominated, mostly data flow from the device to the application server. The M2M device generates small-size packets, and packet generation frequency depends on the application installed on the device. The nature of M2M applications is very diverse compared to H2H communication regarding the QoS requirements, as shown in table 2.1. Therefore scheduling approaches developed for H2H communication are not directly adaptable for M2M communication. Many researchers are contributing to developing solutions for scheduling in the LTE network to support M2M communication. Some researchers have worked on the coexistence of H2H and M2M communication. The following pages show that the work on QoS-aware LTE uplink scheduling is broadly grouped into four categories based on the QoS metric and summarized in Table 3.1.

3.1.1 Delay aware scheduling

In delay-aware scheduling, the scheduler prioritizes devices based on the delay budget of the application installed on the device. In [150], the authors presented a group-based delay-aware heuristic approach for the multi-cell environment to support the M2M communication in the LTE network. The authors in [151] [76] [79] proposed a delay-sensitive approach for a single-cell scenario while considering channel awareness (channel quality between UE and eNB) as the allocation metric. In [76] [79], the uplink resource allocation problem is modeled as an optimization problem. In [79], the author proposed a queuing model to give an optimal allocation of resources by using the application-specific parameters, i.e., traffic rate, as an allocation metric. The delay budget is used as a constraint in optimizing other physical layer parameters, and the lowest delay budget as an allocation metric with heuristic approaches.

Table 3.1: Comparison of Selected work on Uplink Scheduling in LTE

Ref.	Year	Authors	Network	Allocation Metric	QoS Metric	Algorithm	3GPP Compliant	M2M Support
[146]	2008	Calabrese, et al.	M/Mu	Fairness	Max Profit	Optimal	N	N
[147]	2008	De Temino, et al.	M/S	Channel Aware	Many	Heuristic	N	N
[102]	2009	Lee, S-B., et al.	S	Max Throughput	Max Throughput	Heuristic	N	N
[65]	2010	Frank, et al.	M/Mu	Channel Aware	Fairness	Heuristic	N	N
[148]	2011	Ruby, et al.	M/S/Mu	Channel Aware	QoS Classes	Heuristic	N	N
[149]	2011	Yaacoub, et al.	M/S	Max Throughput	Max Throughput	Polynomial	N	N
[150]	2011	Lien, et al.	Mu	Group Based	Delay	Heuristic	N	Y
[151]	2011	Lioumpas, et al.	S	Channel Aware	Delay	Heuristic	N	Y
[139]	2011	Abdalla, et al.	S	Bit Rate Aware	QoS Classes	Heuristic	N	Y
[152]	2012	Ren, et al.	M/Mu	Channel Aware	Max Profit	Heuristic	Y	N
[153]	2013	Afifi, et al.	M/S	Channel Aware	QoS Classes	Heuristic	N	N
[74]	2014	Yang, Kai, et al.	M/Mu	Channel Aware	Fairness	Heuristic	N	N
[133]	2014	Maia, et al.	S	Channel Aware	QoS Classes	Heuristic	Y	Y
[141]	2014	Aijaz, et al.	S	Channel Aware	QoS Classes	Heuristic	N	Y
[154]	2014	Si, Peng, et al	S	Channel Aware	QoS Classes	Greedy	Y	Y
[155]	2014	Hussain, et al.	S	NS	Max Throughput	Exponential	N	Y
[156]	2015	Xiang, et al.	M/F	Max Throughput	Fairness	Heuristic	N	N
[157]	2015	Liao, et al.	S	App Specific	Quality of Video	Heuristic	N	Y
[109]	2015	Azari, et al.	S	Channel Aware	QoS Classes	Heuristic	N	Y
[158]	2016	Hsieh, et al.	S	Group Based	QoS Classes	Heuristic	N	Y
[49]	2016	Ghavimi, et al.	S	Channel Aware	QoS Classes	Greedy	Y	Y
[130]	2017	Mostafa, et al.	S	Channel Aware	App Specific Priority	Greedy	N	Y
[48]	2018	Dawaliby, et al.	S	Channel Aware	QoS Classes	Genetic	Y	Y
[76]	2019	Karadag, et al.	S	Channel Aware	Delay	Heuristic	Y	Y
[79]	2019	A Kumar et al.	S	App Specific	Delay	Optimal	N	Y

Note:- macro-cell(M), femtocell(F), single-cell(S), multi-cell(Mu)

3.1.2 Fairness based scheduling

The authors proposed a fair distribution approach of radio resources among the devices in this category. This scheduling approach does not directly depend on the device's physical resource status and priority, which can lead to some devices' starvation due to the poor condition of physical resources and lower priority level of the device. In [65], the author proposed a cooperative interference-aware heuristic approach for fair distribution of resources among the devices, and multi-cell channel awareness is used as an allocation metric. In [74], the resource allocation problem was modelled as a mixed integer programming (MIP) problem and gave a channel-aware approach for multi-cell for fair distribution of radio resources. In [156], the authors proposed a Markov Chain-based model to provide the maximum possible throughput in macrocell and femtocell networks. A Markov chain is a stochastic model depicting a series of potential occurrences where each event's probability is solely determined by the state obtained in the preceding event. The authors focused on the fairness of LTE radio resource distribution among UEs for H2H communication only. The authors addressed fairness that depends on the previous stats i.e. previous throughput requirements.

3.1.3 QCI priority aware scheduling

A QoS Class Identifier (QCI) refers to a group of communication attributes for a particular scenario, such as application priority, required bit rate, delay budgets, etc., jointly defining a QoS class. The authors proposed various models like queuing model [154], MIP [141, 109], BIP [7], and stochastic modeling [158] for QoS. Some of proposed scheduling works provide support for both H2H and M2M communication [133, 141, 154], whereas others are only for H2H communication [148, 153], or for M2M communication [139, 158, 49, 48]. The application priority is an allocation metric, and other parameters, i.e., delay budget and throughput are used as a constraint in the optimization.

3.1.4 Throughput aware scheduling

Throughput-aware scheduling methodology focuses on the throughput achieved by the UEs in the unit time interval and tries to maximize it. The throughput depends on the number of resource blocks assigned and the MCS index of UE related to the communication channel condition. All the scheduling approaches in this category use channel awareness (channel quality between UE and eNB) as the allocation metric. The authors formulated the scheduling problem as a search tree [146], game-theoretic [149], MIP [152], and mixed-integer linear programming problem [155] for different network scenarios.

3.1.5 Other scheduling work

The authors have proposed a channel-aware resource allocation methodology for application-specific requirements in [147, 130]. In [157], the author considers Quality-of-Video (QoV) as a QoS metric and has used application-specific parameters as an allocation metric to propose a scheduling approach for a single-cell environment.
The scheduling approaches presented in the literature are best-fit for a particular environment and objective and may lack performance in other scenarios.

3.2 Research Contribution of this Chapter

Following are the research contributions of this chapter concerning surveyed literature.

1. Priority decision problem is formulated as a function of the device's valuation and modeled as a combinatorial auction game, which is a computationally efficient solution that can be computed in polynomial time.

2. The proposed auction game model controls the strategic behavior of the MTCDs and provides stability to the M2M system in the presence of malfunctioning devices.

3. This chapter highlights the effect of priority-based resource allocation on other key performance indicators (KPIs), i.e., resource utilization, throughput, and fairness.

4. The proposed solution provides QoS support for MTCDs and stabilizes the M2M system when malfunctioning devices are present in a heterogeneous environment.

3.3 Combinatorial Auction Game

Game theory helps control the players' strategic behavior to get an efficient outcome for the game. The application field of game theory is vast, for example, networking, resource management, economics, and politics. Games can be classified as cooperative or non-cooperative, zero-sum or non-zero-sum, simultaneous move or sequential move, and one-shot or repeated in the context of game theory. The auction game is a non-cooperative game where players (i.e., bidders) make their strategy to win. Auction is a combinatorial game where an item or group of items are distributed among a set of players. Players bid for an item or a bundle of items. The bidding amount of an item or bundle depends on the player's private valuation. The player with the highest bidding amount wins the auction, and the auctioneer assigns the item or a bundle of items to the winning player. In our work, the bidder is the device, the auctioneer is the eNB, and the item is the resource [159, 160].

The resource allocation problem is modeled as an auction. Figure 3.1 depicts the auction model of resource allocation. Each device bids for a resource(s) with a valuation (Need of the communication). The eNB acts as an auctioneer for the resource allocation. The resource allocation consists of three steps.

1. *Resource valuation:* In this step, devices calculate the valuation for the resources and submit their bids.

2. *Resource allocation:* The eNB performs the resource allocation task based on device bids.

3. *Payment:* The eNB assigns a cost to each winner of the auction process.

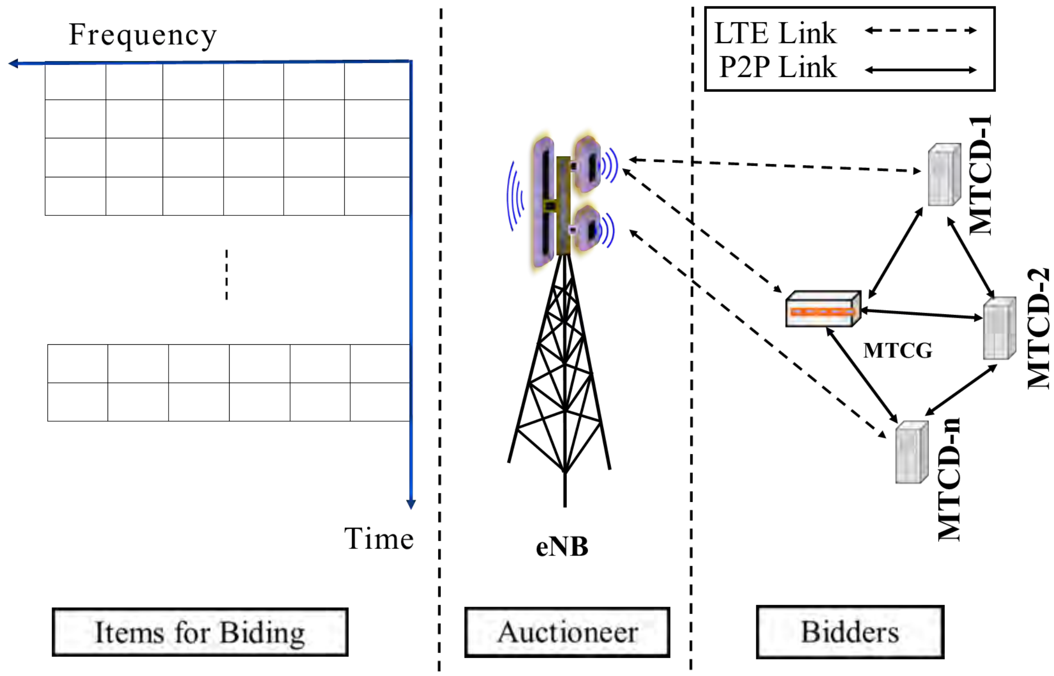


Figure 3.1: Auction Model of Resource Allocation Process.

3.3.1 Resource valuation

A valuation is a real-value function $v : (V_R)^n \rightarrow \mathbb{R}^+$ for each subset R of resources, indicating the value that bidder(device) gets if the resource(s) is/are allocated to the device. if device does not get any resource than $v(R) = \phi$. The valuation function of device i is a private-value function, $s.t.$ no other device or eNB knows. In this work, we assume that devices are a kind of single-minded bidders and that each device is only interested in a single set of resources R. If the device gets the requested set of resource R, it gains a scalar value of welfare and zeroes otherwise [161].

A valuation v is called single-minded valuation if there is a set of resources R^* and a value $v^* \in \mathbb{R}^+$ s.t. $v(R) = v^*; \forall\, R \supseteq R^*$, and $v(R) = 0$ otherwise. A single-minded bid is a pair of (R^*, v^*) [161].

3.3.2 Resource allocation

We need to define a socially efficient (imposed cost of communication in a TTI should be the same for all devices in that TTI) and computationally efficient allocation mechanism to allocate resources among devices. Given an allocation A_r of resources $R = r_1, r_2, r_3, ..., r_m$ among devices such that $R_i^* \cap R_j^* = \phi; \forall\, i \neq j$, the social welfare achieved by the allocation is given as follows [162].

$$Welfare(A_r) = \sum_i v_i(R_i^*) \tag{3.1}$$

Welfare is the sum of the total valuation of devices allocated to a set of resources. We aim to maximize the total welfare, i.e., maximizing the total QoS satisfied by the allocation. An allocation A_r is socially efficient if it maximizes the total welfare among all possible allocations. A socially efficient allocation increases the utility of the device. The device's utility d_i is given as follows [162].

$$Utility(d_i) = Welfare(d_i) - Payment(p_i) \tag{3.2}$$

Resource allocation gives disjoint sets of resources R_i^* for all the devices $i \in 1...n$, and maximize the total welfare $\sum_i v_i(R_i^*)$. An optimal resource allocation can allocate exactly the requested set of resources to each device $R_i = R_i^*$ or nothing $R_i = \phi$. Let V be the set of valuations of m resources by n devices and A_r is the set of resource allocation for n devices. Then, the resource allocation mechanism is defined as a function $a : V^n \to A_r$. This allocation problem is similar to the "$Weighted - Packing$" problem, and it is an NP-complete problem that can be proven by the reduction from the "$INDEPENDENT - SET$" problem. A mechanism is computationally efficient only if it can be completed in polynomial time [162, 163].

3.3.3 Payment function

Let e_i be a payment function and assume that resources can be allocated continuously and each device has a single valuation for all identical resources. Then the payment function can be defined as $e_i : V \rightarrow \mathbb{R}^+; \quad \forall\, i = 1,2,...,n$. We need to define a payment function so that no device can report a false valuation for a set of resources instead of reporting true information to the eNB. This type of payment mechanism is called dominant strategy incentive compatible (DSIC). A mechanism is universally truthful and dominant strategy incentive compatible if the device reports its true valuation with probability 1. Let, v_i is the true valuation of device i, v_{-i} is the true valuation of other devices and $v_i^{'}$ is reported valuation of device i [162][164]. A device i can only report a potential lie if the device gets a higher utility with a lie than the utility with a true valuation. It can be defined as follows.

$$v_i(A_i) - e_i(v_i, v_{-i)} \geq v_i(A_i^{'}) - e_i(v_i^{'}, v_i) \tag{3.3}$$

$$\forall \quad v_1, ...v_n, v_i^{'} \in V$$

where, $A_i = v(v_i, v_{-i})$, $A_i^{'} = v(v_i^{'}, v_{-i})$ if device i wins in allocation A_r and zero if device lost in auction.

We use the Vickery-Clark-Groves (VCG) mechanism to define the payment function, as it is a DSIC mechanism with a socially efficient outcome. In the VCG mechanism, a bidder, in this case, the device is charged by its externalities. A device must pay a cost equivalent to the loss to other devices by participating in the auction, i.e., the amount of bidder lost due to the device i. A_r is an allocation of resources for k number of devices those wins and device i report its private value p_i then the payment of device i can be defined as follows [162].

$$e_i = \sum_{i \neq j}^{M} v_j(p_j, A^*_{r-i}) - \sum_{j}^{M} v_j(p_j, A^*_r) \tag{3.4}$$

where $\sum_{i \neq j}^{M} v_j(p_j, A^*_{r-i})$ is the welfare of other devices when device i is not participating in the auction and $\sum_{i \neq j}^{M} p_j(v_j, A^*_r)$ is the welfare to other devices with the participation of device i. Thus if K devices win in the auction process, device i needs to pay the valuation of $(K+1)^{th}$ device, which is lost due to the participation of device i. All devices pay the same amount of $v(K+1)$, which is socially feasible because all winning devices communicate in the same time slot in time domain scheduling. By reporting the valuation v_i of device i, the chosen allocation is A^*_r, and the utility for device i is defined as follows.

$$U_i = v_i(p_i, A^*_r) - \left\{ \sum_{i \neq j}^{M} v_j(p_j, A^*_{r-i}) - \sum_{j}^{M} v_j(p_j, A^*_r) \right\} \tag{3.5}$$

$$= v_i(p_i, A^*_r) + \sum_{j}^{M} v_j(p_j, A^*_r) - \left\{ \sum_{i \neq j}^{M} v_j(p_j, A^*_{r-i}) \right\}$$

$$= \sum_{j=1}^{M} v_j(p_j, A^*_r) - \left\{ \sum_{i \neq j}^{M} v_j(p_j, A^*_{r-i}) \right\}$$

$$U_i = Welfare(A^*_r) - \left\{ \sum_{i \neq j}^{M} v_j(p_j, A^*_{r-i}) \right\} \tag{3.6}$$

If the device reports a fake valuation to eNB and can change the allocation A^*_r. If the new allocation is X^*_r, then the changed allocation is good if and only if

$$Welfare(X^*_r) - \left\{ \sum_{i \neq j}^{M} v_j(p_j, A^*_{r-i}) \right\}$$

$$> Welfare(A^*_r) - \left\{ \sum_{i \neq j}^{M} v_j(p_j, A^*_{r-i}) \right\} \tag{3.7}$$

Constraint equations (3.3) and (3.7) show that a device can only lie if the welfare received by the device is more than the welfare received with true valuation. The VCG payment function defined in equation (3.4) maximizes the device's welfare with true valuation. Since the device's utility depends on the total efficiency, the device has to choose the efficient outcome of the resource allocation process [162].

Example of VCG mechanism:

Let there be five bidders A, B, C, D, and E. All bidders submit the bids as <A-3$>, <B-6$>, <C-4$>, <D-7$>, and <E-5$>. There are three items to be auctioned. So based on the bids submitted by the bidders, bidders D, B, and E win in the auction. Bidder C has the bid 4$ that just lost the auction. All the winner has to pay 4$ as payment. Thus the welfare for bidders D, B, and E is 3$, 2$, and 1$ and zero for others. Now, assume that bidder C makes a false bid of 6$ (actual valuation for C is 4$) to win in the auction. Bidders D, B, and C win the auction with valuations of 7$, 6$, and 6$, and bidder E with a valuation of 5$, lost the auction. Now all the winners will pay 5$, then the welfare for bidder C is 4$ - 5$ = -1$. This way, the VCG auction controls the strategic behavior of the bidders.

3.3.4 Feasibility of VCG mechanism

Theoretically, the VCG mechanism provides perfect efficiency with DSIC truth-revealing strategies. However, VCG has some serious limitations regarding the implementation [165]. In this subsection of this chapter, we address the feasibility of these issues concerning our scenario.

1. *Weak equilibrium and bidding cost:* As losing the bid sets the price to be paid by the winner. The auctioneer may be worried about revenue generation. Furthermore, submitting the bids for n items will require $2^n - 1$ combinations of bidding information. Nevertheless, in this scenario, revenue is not an objective of the whole process, and all the resources are identical; therefore, only one individual bidding information is sufficient for the auction process. Thus $2^n - 1$ bidding price combinations turn into a linear relation of problem size n.

2. *Winner determination effort:* To determine the winner from n bidders requires $2^n - 1$ information from each bidder. This is an NP-complete problem [162, 163]. If the bidders are single-minded and desire continuous allocation of identical resources (SC-FDMA constraint) $s.t.$, $\forall\, R^* = \{j^i, j^i + 1, ...k^i\}$ for some $1 \leq j^i \leq k^i \leq m$ then this can be performed in the polynomial time [162].

3. *Budgets limitations and information revelation:* Budget constraints of a bidder can destroy the truthful auction process. There is no way to ascertain which will be the highest bidding [166]. In our scenario, every device must submit the bid, which is a sealed bid process; therefore, it does not affect the process. Furthermore, the complete process of auction is carried out by eNodeB. Thus, there is no risk of information revelation.

4. *Possibility of cheating:* In the VCG process, there is a possibility of cheating by the bidder and bid taker [165]. In this scenario, all the devices can submit only a single bid and are unaware of each other. So conspiracies by bidders are not possible. Furthermore, the bid taker (eNodeB) is assumed to be a trusted device.

3.4 Priority-Based Resource Allocation with Auction Game

We consider an LTE single-cell network scenario, where M2M devices can communicate directly to the LTE base station or indirectly through a gateway. In this model, we focus only on LTE communication. Let's consider n number of devices (MTCDs) randomly distributed in an LTE cell. Each device tries to get uplink access for communication in a time slot for m number of shared resources. Devices have a priority and a fixed valuation for uplink communication to happen. The objective of the work is to devise a scheduling policy for access grant to a device for uplink data transmission in the time domain and allocate resources in the frequency domain as per the priority and valuation such that communication per time slot maximizes the total system valuation and also prevent device's strategic behavior in a socially efficient manner. System valuation is the sum of all valuations of communicating devices in a particular time slot; an efficient system maximizes the total system valuation per time slot. The term "socially efficient" refers to the cost of getting access in a time slot for a device, which is the lowest cost with which the system can control the device's strategic behavior from claiming false valuation.

3.4.1 Problem formulation

Let $D = \{d_1, d_2, \ldots d_n\}$ be a set of devices, randomly deployed in an LTE cell and ready for communication, and $R = \{r_1, r_2, \ldots r_m\}$ is a set of resources in the frequency domain for uplink scheduling. Device i sends its preference value η_i to eNB at the time of bearer establishment. The preference value of a device d_i depends on all QoS parameters (i.e., Delay Budget, Throughput Required, Application Priority). The packet scheduler uses this preference value. Then, the scheduler allocates m resources among devices such that the allocation A^* maximizes the total valuation from k number of alternatives.

$$v(A^*) = \text{argmax}\{\bigcup_{i=1}^{k} v(A_i)\} \tag{3.8}$$

where $v(A_i)$ is an allocation. The selected valuation in equation (3.8) maximizes the system utility.

As the UE sends its preference vector to eNB, which is used to calculate the UE's valuation, UE may send false information in the preference vector to eNB to gain priority over other communicating UEs. Therefore the allocation process should be such that it can not make a false claim and can be enforced to release true information by applying some constraints. The constraint should be strong enough to restrict a UE from making a false claim and soft so that it should not burden UE. Suppose k number of UEs get scheduled for communication in current TTI, c_i is the cost of communication for the UE_i and c_{k+1} is the cost of communication for the UE_{k+1}. Then the cost of communication of UE_i can be expressed as follows.

$$c_i = c_{k+1} + \varepsilon; \quad \forall 1 \leq i \leq k \tag{3.9}$$

Where ε stands for small positive quantity. That provides c_i slightly greater than c_{k+1}. Equation (3.8) chooses an efficient QoS aware allocation, and equation (3.9) applies a soft constraint to restrict the strategic behavior of UEs.

3.4.2 Methodology

Consider a scenario where n number of devices wait for their turn to communicate over M shared resources in the LTE network. If the number of available resources m is more than or equal to the number of devices n in the network, all the devices can be satisfied in the same TTI. Nevertheless, with the available number of resources $m < n$ number of devices, there is a need for an efficient allocation mechanism to allocate the resources between the devices. In this case, the problem is to satisfy the QoS (i.e., Delay, Throughput, HARQ) requirements. An Auction game-theoretic approach is used to allocate resources in a computationally efficient and socially efficient manner (the cost of communication for all devices communicating in the same TTI should be the same) along with taking care of the QoS requirements of devices. Figure 3.2 shows the proposed scheduling scheme.

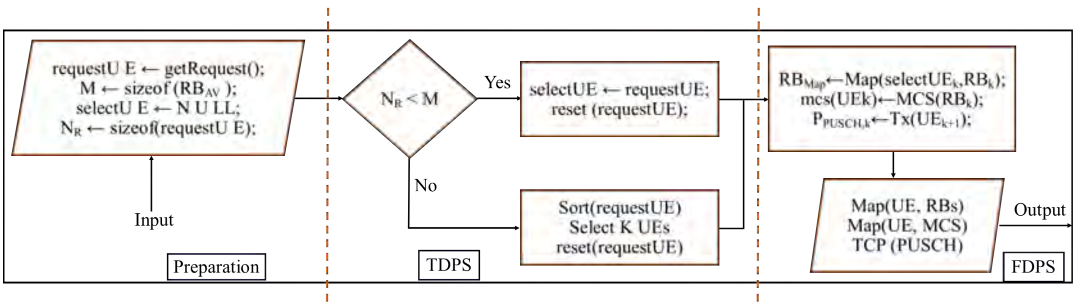


Figure 3.2: Proposed Scheduling Scheme

In this work, the priority of QCI, defined in the table 1.1, is used as the valuation for resource allocation purposes. The higher the valuation, the higher the preference for that device over other devices. UE's preference (valuation) is a function of the QoS attributes of UE such that $v : \text{QoS} \rightarrow \mathbb{Z}^{+}$. In the radio resource allocation in the LTE network, valuation is a random variable that is uniformly distributed in $[0, p] = [0, 1]$. The probability distribution function (PDF) of the uniform random variable in $[0, 1]$ is defined as follows [167].

$$f(p) = \begin{cases} 1 & if\, 0 \leq p \leq 1 \\ 0 & Otherwise \end{cases} \tag{3.10}$$

The equation 3.10 shows that any device in the network can have a value for preference from QCI_1 to QCI_{max}. Moreover, devices do not have different valuations for different sets of resources. At eNB, TDPS sorts the UEs in non-increasing order and selects the k number of UEs out of n UEs with higher preference. The number of selected UEs k depends on available resources. The eNB allocates a set of the required number of physical resources (PRBs) for the UEs so that each UE gets precisely the number of resources required; otherwise, it gets zero.

The FDPS allocates a set of resources to an UE depending on the requirement and assigns MCS index and power limitations for the data transmission. The UE i gets subset R_i^* (consecutive resource blocks) of the available resource set R and is defined as follows.

$$R_i^* \subseteq R \tag{3.11}$$

The equation (3.11) is subject to following constraints -

$$R_{i,j}^* \cap R_{i',j}^* = \phi; \quad \forall\, i \in I,\, j \in J$$

$$R_{i,j}^* = \phi \quad \forall\, j \geq m+1\, if\, R_{i,m} = \phi$$

Where $R_{i,j}$ are resources allocated to device i and $R_{i',j}$ are resources allocated to the device i'. The eNB informs the UE by sending a scheduling grant using DCI-0 over PDCCH. The UE inspects the received DCI-0 for scheduling information. Then, UE starts the actual data transmission over PUCCH in response to the uplink (UL) grant. The timing among scheduling requests (SR), scheduling grants, and PUSCH depends on the configured transportation mode.

The number of high-priority applications is less compared to lower-priority applications, as observed from table 2.6. There is a significant difference in the number of packets generated by higher-priority applications to lower-priority applications. High-priority applications generate fewer packets in a specific time interval. On the other hand, the low-priority application generates many packets and therefore demands more transmission chances [168, 169, 170].

Let a device i with application priority P_i and the total required transmission chances C_i^T in a time interval T claim their priority as a bid when they connect to eNB. We consider application priority as valuation and scheduling chances as a cost in the auction. We can penalize the device i with the number of scheduling chances if it changes the application priority. We reduce scheduling chances for every scheduling grant of device i. Reduction in the scheduling chances depends on the priority of the winner and loser device as the winner pays externalities in VCG auction and defined in equation 3.4. A device i with priority P_i^w wins in the auction, and a device k with priority P_k^l loses in the auction, then the reduction in transmission chances C_i^T is defined as follows.

$$C_i^T = C_i^T - \left(1 + \left\lfloor \frac{1}{2} + \frac{P_i^w}{P_k^l} \right\rfloor\right) \tag{3.12}$$

For example, if a device has a priority of 7 and claims its priority as 1 the device wins in the auction. If the loser device's priority is 2 or 3, we reduce 2 scheduling chances, and if the loser device's priority is 5 or 6, we reduce 1 scheduling chance from total transmission chances. If the application generates packets as per its real priority and claims a high priority, it will get fewer chances and exit from the scheduling opportunity. This approach provides a stable M2M system with malfunctioning devices.

Note: There is a probability that fair devices can suffer with the scheduling opportunities as the reduction in transmission chances is the same for all winning devices in this approach. Therefore the time interval T and the total required transmission chances C_i^T need to be defined precisely. Thus the total required transmission chances should be sufficient enough in a longer time interval.

3.4.3 Feasibility of proposed methodology

To evaluate the feasibility of the proposed model, we use a three-step approach as follows.

- Whenever we define an efficient allocation mechanism, the first question arises; Is the allocation maximizing the system's utility and computationally efficient? It should give optimal and computationally efficient output. As the constraint imposed by SC-FDMA, resources can only be allocated in a consecutive resource blocks subset of resource blocks $s.t.$, $\forall\, R^* = \{j^i, j^i+1, ...k^i\}$ for some $1 \leq j^i \leq k^i \leq m$ and all the resource blocks are identical, so no one device has a different valuation for two different resources. Due to this, the allocation functions $a : V^n \to A_r$ turn into a linear allocation function such that $a : V \to A_r$. it leads to the allocation problem as a linear mapping of sorted elements. The resource allocation problem can be solved efficiently in polynomial time. The scheduler selects the highest valuation for each TTI. Thus the allocation scheme maximizes the total system utility in context to the total QoS satisfied [161].

- The second question is; Is the allocation mechanism strategy-proof? To control UEs' strategic behavior so that it reveals accurate information to eNB about their valuation, the utility of the UE should be independent of its self-valuation [162]. In the VCG mechanism, the device pays its externalities (as shown in equation 3.6) and enforces the UE to reveal accurate information equation (3.7). As in LTE, UEs have a single valuation; thus, the payment function can be computed in polynomial time.

- Is the cost of communication socially efficient? Socially efficient means the cost of getting a chance to communicate should not differentiate between the UEs. It should be sufficient to control strategic behavior and not impose a huge burden on the UEs. The equation (3.12) provides such a mechanism.

3.5 Scheduler Algorithm

The main steps of the proposed scheduling scheme are listed in the algorithm 1. The UE sends its buffer status via BSR to eNB. If the UE has data in its buffer for the uplink transmission, it is considered for the current scheduling process. In the initialization phase, the scheduler collects the information on available data in UE's buffer, available resources, and the number of active UEs. In the algorithm 1, steps from $1-5$ perform the initialization of the scheduler.

If a device $i \in X$ needs RB_i resources from $|M|$ available resources, there are two possibilities for the scheduling based on the initial information.

1. **if** $\sum_{i=1}^{|X|} RB_i \leq |M|; \forall i \in X$**:** In this case, the number of UEs with data in their buffer is less than the number of available resources, TDPS selects all UEs for data transmission in steps 8 and 9. In the steps from 10 to 14 in algorithm 1, FDPS performs the resource allocation with MCS and power assignment.

2. **if** $\sum_{i=1}^{|X|} RB_i > |M|; \forall i \in X$**:** In this case, if the number of UEs is more than the available resources, then TDPS selects the appropriate number of UEs from the request queue (sorted in non-increasing order of preference) and stores the radio network temporary identifiers (RNTI) of these UEs in the select queue and remove the selected UEs from the request queue in steps from 16 to 20 of

Algorithm 1: QCI_Priority_Scheduler

Data: QoS Parameter: Application Priority of UEs.
Other Parameter: CQI, SINR, MCS Matrix, UEs Request Queue
Result: Set of selected UEs with assigned RBs

```
1  Initialization:
2  requestUE ← getRequest();
3  M ← sizeof(RB_AV);
4  selectUE ← NULL;
5  N_R ← sizeof(requestUE);
6  begin
7      if N_r < M then
8          selectUE ← requestUE;
9          reset(requestUE);
10         for j ← 0 to N_R do
11             allocateRB ← Map(selectUE_i, RB_k);
12             mcsIndex(selectUE_j) ← MCS(RB_j);
13             P_PUSCH,j ← Tx(UE_j);
14         end
15     else
16         requestUE ← Sort(requestUE, Priority);
17         for i ← 0 to M do
18             selectUE_i ← requestUE_i;
19             remove(requestUE_i);
20         end
21         for j ← 0 to M and k ← M to 0 do
22             allocateRB ← Map(selectUE_i, RB_k);
23             mcsIndex(selectUE_j‘) ← MCS(RB_k);
24             P_PUSCH,j ← Tx(UE_j)
25             updateTChances(UE_j)
26         end
27     end
28 end
```

algorithm 1. TDPS passes the select queue to the FDPS to allocate resource blocks. From 22 to 30, the FDPS allocates the physical resources to the UEs. MCS index is assigned as per the assigned resource block. If k number of UEs are selected for the physical resource allocation, then the transmission chances for each UE, C_i^T are updated as defined in the equation (3.12). Then eNB sends scheduling grants to the UEs in DCI-0 format. The UE inspects the received DCI-0 and starts actual data transmission. If the UE receives a scheduling grant in n^{th} TTI, then it can transfer the data in $(n+3)^{th}$ TTI.

In both above cases, the scheduling task is performed in two phases; 1) Node selection in TDPS and 2) Resource block allocation in FDPS. The proposed algorithm considers the QCI as the allocation metric. UEs are sorted in non-increasing order of QCI before being selected by TDPS. TDPS selects the appropriate UE number in $O(n \log n)$ time. TDPS selects the m number of UEs, and then FDPS allocates the physical resources to the selected devices in $O(m)$. Thus the running complexity of the proposed algorithm is $O(m+n \log n)$. If the number of available resources is lower than the active UEs, then the proposed algorithm's complexity is $O(n \log n)$. The proposed algorithm can be implemented in polynomial time complexity with diverse QoS support for M2M communication.

3.6 Scheduler Performance Evaluation

The simulation is performed with the 5G NR Toolbox in MATLAB R2020b. Figure 3.3 shows the layered architecture of MATLAB simulation modules. MATLAB 5G NR toolbox supports RLC, MAC, and physical layer simulation. This toolbox uses a traffic generation module to generate data packets per the application requirement. At the MAC layer, the 5G NR toolbox supports the simulation of custom schedulers. We have run a test simulation to decide on parameters. After four hundred devices, the average system buffer level increases abruptly with 5 MHz bandwidth. Therefore, we have chosen five hundred as the maximum number of devices. We run simulations for 50, 100, ..., 450, and 500 frames. We have observed a pattern in the results of up to 500 frames. Therefore, we have chosen five hundred frames for a single iteration as simulation time. The results are stable after 25 iterations for a single scenario. Consequently, we have simulated for forty iterations.

We create a single-cell LTE network of 1 KM radius. The base station is placed at the center of the cell. The devices are randomly placed in the cell without mobility for a single simulation. We use the hNRUEPassThroughPhy module for devices and

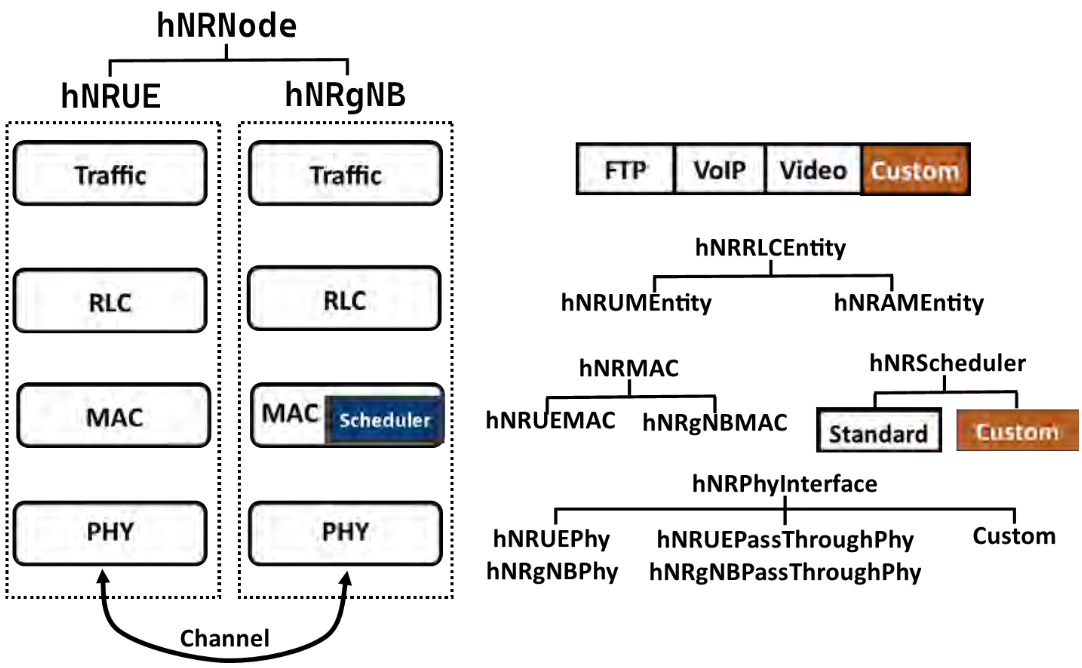


Figure 3.3: MATLAB LTE simulation environment.

hNRgNBPassThroughPhy for the base station to simplify the physical layer. We assume the physical layer works correctly, and the parameters are listed in table 3.2. At the RLC layer, the hNRUMEntity module is used to reduce the packet transfer complexity. The hNRUMEntity is an unacknowledged packet transfer mode at the RLC layer. The hNRScheduler module is only on the base station and attached to the hNRgNBMAC module. The hNRScheduler module supports the attachment of a custom packet scheduler.

3GPP defines 13 QCIs for M2M communication as listed in table 1.1. Initially, we took 7 QCIs out of 13 for simulation to test resource utilization, throughput, fairness, delay budget violation, and priority violation. Further, we increased up to 10 to test the effect of an increase in heterogeneity on priority and delay budget violations. Applications, number of devices, and traffic models are chosen as defined in the 3GPP TR 43.868 release 12. QCI for the applications is adopted from *Dawaliby et al.* [48]. A single application is installed on each device. Devices with the exact application requirements are grouped and connected to a single MTCG. As specified in 3GPP

Table 3.2: Physical layer parameters

Parameter	Value
Network Type	Single cell (1 KM Radius)
Number of Frame	500
Frame Duration	10ms
Number of Sub-frame/frame	10
Scheduler Periodicity	4 slots
Sub-frame Configuration	1
Number of PUSCH RBs	25 (5MHz)
UL Carrier Freq.	2.515e9 Hz
Sub-carrier spacing	15 KHz
Number of Sub-carrier/PRB	12
Number of Symbols/sub-carrier	7(Normal CP)
RBG Allocation Limit UL	1 RBG
RBG Size Configuration	1
No. Logical Channels	3
Channel Model	AWGN
Transportation Mode	TDD / Frame Type-2/ Config-0
Antenna Mode	Single
UE Distribution	Random
BSR Periodicity	5ms
Channel Update Periodicity	0.1 sec

TR 43.868 release 12, we generate custom traffic in MATLAB simulation. The application characteristics like traffic type, number of devices, and traffic intensity are listed in table 3.3. The values in the table 3.3 are normalized according to the simulation time. We update channel quality with a random operation (± 2) every 100ms to model the channel quality changes.

3GPP defines 13 QCIs, as listed in the table 1.1. We include 7 QCIs for a smart-building scenario in simulation. The core objective of this thesis is to provide a balanced system performance and QoS support. Therefore, we evaluate the scheduler's performance on the metrics explained in section 3.6.1.

3.6.1 Performance evaluation metrics

We consider two metrics, system performance, and QoS support, to evaluate the balanced performance of the priority-based algorithm.

Table 3.3: Application traffic patterns

Application Type	Traffic Type	QCI	Priority	App Count	UE Count (%)	Pkt /minute
Accidental Warning	Event-Driven	5	1	3	12	5
Security Breaches	Event-Driven	1	2	3	7	10
Delay Sensitive Monitoring	Periodic Update	3	3	2	12	30
Environment Monitoring	Event-Driven / Periodic Update	2	4	3	12	30
Home Automation	Event-Driven / Periodic Update	4	5	5	19	40
Utility Application	Periodic Update	6	6	4	19	60
Smart Meters	Periodic Update	7	7	5	19	60

- We have selected fairness, throughput, and resource utilization parameters as system performance metrics, and plots include the aggregate results for all devices. The proposed algorithm's simulation results are compared to state-of-the-art schedulers like Proportional Fair (PF) and BestCQI schedulers for system performance evaluation.

- QCI priority and delay budget are selected as QoS metrics, and plots include the sum of results for individual devices. The proposed algorithm's simulation results are compared to schedulers from literature (RCQ [48], and Queue [79]) for QoS support evaluation.

Following metrics are used to evaluate the performance of the proposed scheme.

1. *Resource sharing fairness*: It shows how equally resources are shared among the communicating devices. In this thesis, resource-sharing fairness is based on Jain's fairness, and we measure the deviation of resources shared among

the devices. Value one in the fairness index shows that resources are equally shared among the devices.

2. *Average cell throughput*: It defines how many bits of information (excluding control information) are transmitted per unit of time. In this thesis, throughput is used as a system performance parameter. Therefore, we measured for cell throughput rather than individual device throughput. The average cell throughput is measured as an average throughput for the entire simulation time of each scenario.

3. *Average resource utilization*: It shows the percentage of resources used from the total available resources in a time interval. The resource utilization is measured as an average resource utilization per transmission time interval for each scenario.

4. *QCI priority support* : It shows whether data transmission preference is given to high-priority applications over lower-priority ones. The QCI violation index shows how often a lower-priority device is served; even a high-priority device is ready to send a data packet. The stats are measured as per the TTI average.

5. *Delay budget violation*: It shows the average number of events when one of the communicating devices cannot transmit data before its delay budget expires. The stats for delay-budget violation are measured as an average per TTI for the number of incidents when the devices do not get a chance to send a packet before the delay budget expires. This metric is used to show delay budget violations in the view of the M2M system rather than to count delay budget violations for individual devices or groups of devices. To highlight system performance, we count delay budget violations per TTI and average it for the simulation time.

The selection of the schedulers for each performance metric is as listed in the table 3.4.

Table 3.4: Performance evaluation metrics

Algorithm	Objective	Evaluation Metric	Domain
Proportional Fair	Fairness	Resource sharing fairness	System Performance
BestCQI	Throughput	Average cell throughput	
	System Utilization	Resource Utilization	
RCQ [48]	QoS	Delay budget	QoS Support
		QCI Priority	
Queue [79]	QoS	Delay budget	QoS Support
		QCI Priority	

M2M communication has diverse characteristics as stated in section 1.1. Therefore, we can evaluate the metrics mentioned above against the following events.

- Number of nodes.
- Traffic types.
- Number of applications.
- Traffic generation frequencies.
- Number of priorities or QCI classes.
- Packet size.

- The traffic generation frequency, packet size, and traffic type are characteristics of an application (refer to table 3.3). Each application has a unique priority (refer to table 1.1). We have selected a heterogeneous environment with different types of applications. Therefore, there are only two different events, the number of nodes and applications / QCIs, to compare the performance evaluation metrics.

- We have evaluated all five parameters (listed in table 3.4) against the variable number of nodes.

- We have evaluated QoS metrics delay-budget violation and priority support against a variable number of QoS class identifiers (QCIs) to show the effect of environment heterogeneity on QoS support.

3.6.2 Resource sharing fairness

Resource sharing fairness is crucial for the performance of the resource allocation algorithm. It shows how likely the individual devices have been treated. The resource fairness index I_{RB} shows whether the resources are fairly shared among the devices. The resource fairness index varies in the interval [0 1], 0 represents entirely unfair, and 1 represents fairness entirely [96]. The fairness index is inspired by Jain's fairness and is defined as follows.

$$I_{RB} = \left(\frac{1}{1+\hat{x}}\right) \tag{3.13}$$

$$\hat{x} = \sqrt{\frac{\sum(RB_i - RB_{Avg})^2}{N_{UEs}}}$$

Where RB_i is the resource allocated to device i and $\hat{x}$ is the standard deviation in re-

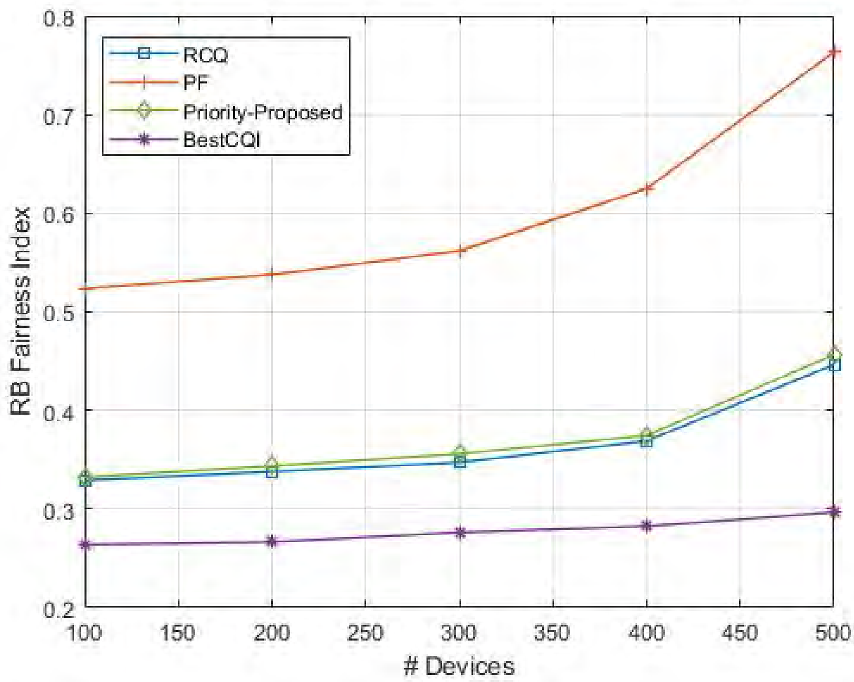


Figure 3.4: Resource sharing fairness.

source share among devices. The resource-sharing fairness of different algorithms is shown in figure 3.4. The highest value for the fairness index of the proportional fair algorithm is 0.8, which is 0.5 (50%) more than the fairness index of BestCQI. The fairness index values for both priority and RCQ algorithms are below the average for the highest and lowest values of the fairness index of all algorithms. This phenomenon is because the BestCQI and RCQ algorithms consider the channel quality as an allocation metric. The wireless channel is much more unpredictable; therefore, the BestCQI and RCQ algorithms do not share the devices' resources fairly. The QCI priority-based (proposed algorithm) considers traffic priority. This information is not sufficient for fairness. Therefore, the proposed algorithm performs below average with a margin of approximately 0.33. Although fairness increases with the number of devices, it is highly unlikely to distribute resources equally among devices due to unpredicted network conditions.

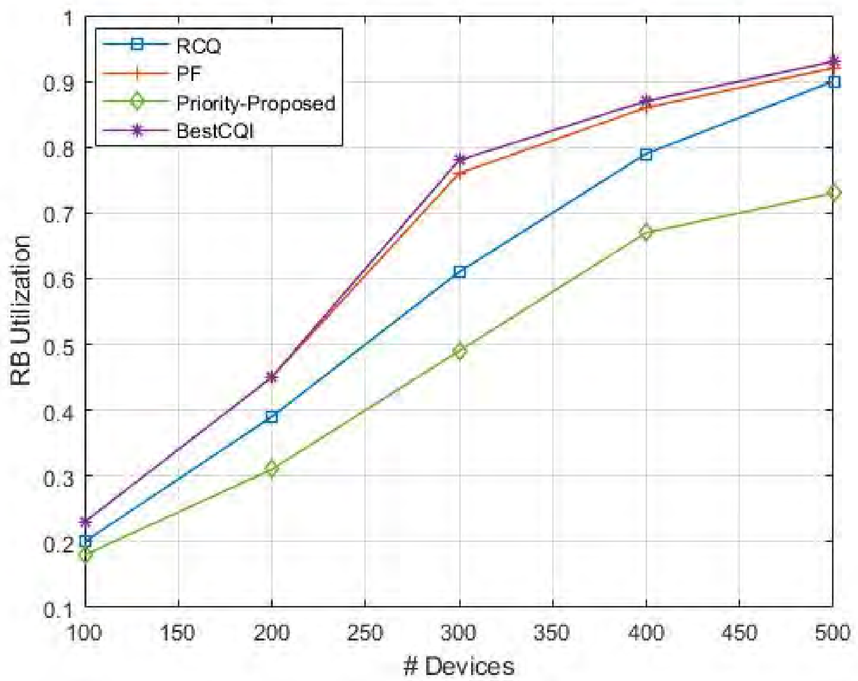


Figure 3.5: Resource utilization.

3.6.3 Resource utilization

Resource utilization is the percentage of radio resources utilized from the total available resources. Resource utilization R_u is defined as follows.

$$R_u = \sum_{i=1}^{N_{UE}} \left(\frac{RBG_{avg}^{i}}{RBG_{Total}} \right) \tag{3.14}$$

Where RBG_{avg}^{i} is the average number of RBGs per transmission the device i uses, figure 3.5 shows the resource utilization of implemented algorithms. For a lower number of devices (100 devices), resource utilization is around 0.2 (20%) with a limited difference of 0.06 (6%) among implemented algorithms. As the number of devices increases (500), the resource utilization for PF, BestCQI, and RCQ increases to 0.91 (91%). The resource utilization is limited to 0.74 (74%) for the priority algorithm, with a difference of approximately 20%. BestCQI algorithms allocate resources irrespective of the device's QoS requirements. The RCQ algorithm also considers channel quality as the primary metric in the allocation. Therefore these algorithms can highly utilize the resources. PF algorithm does not directly consider QoS; however, it considers the device's requirement based on past data transmission activities. The resource utilization incurs a saturation after 300 devices for all algorithms. This implies that resource utilization increases with the number of devices but moves towards saturation after an average number of devices (300 devices). It will not be possible to use 100% of resources as it shows a logarithmic pattern.

3.6.4 Average cell throughput

The throughput is the rate of successful data transmission over some time. The average cell throughput for implemented scheduling algorithms is shown in figure 3.6. The average cell throughput is likely to equal (3.8 Mbps with 200 devices) for all the algorithms for a small number of devices. As the number of devices increases, the difference in average cell throughput for different algorithms increases. The BestCQI

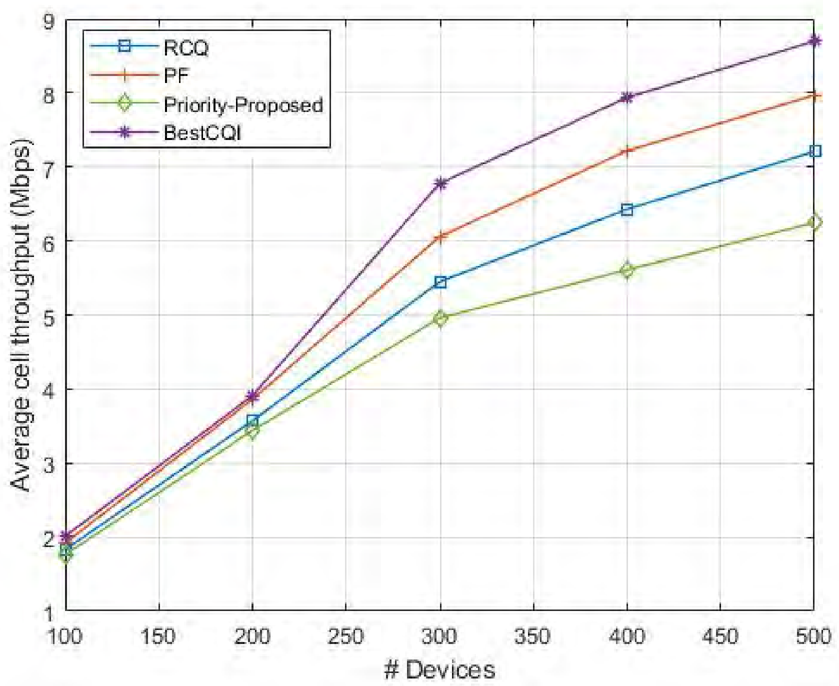
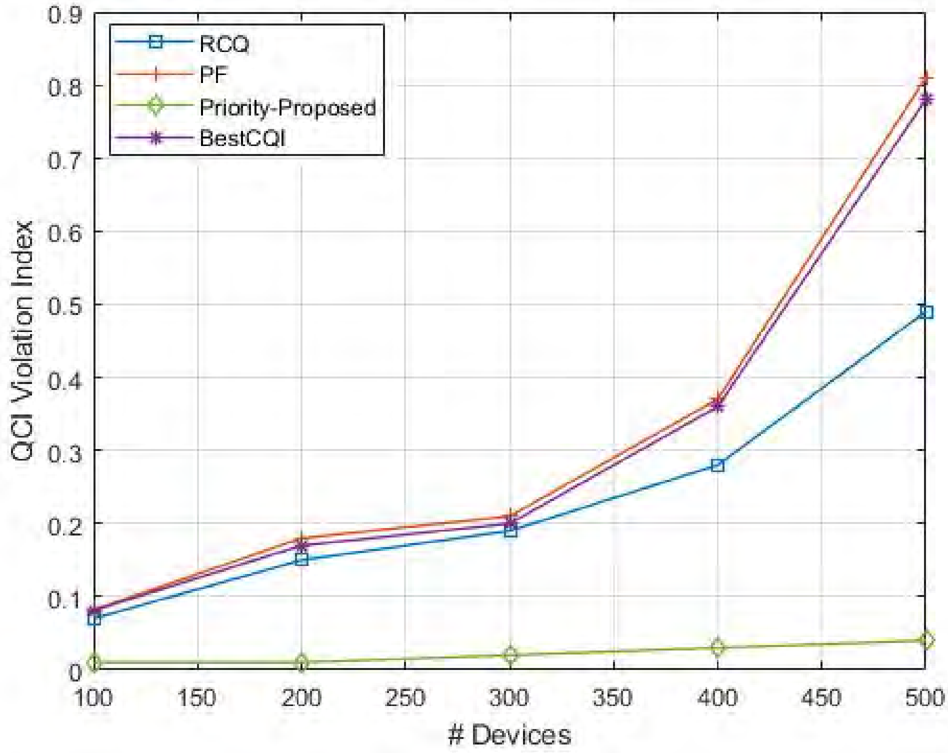


Figure 3.6: Average cell throughput.

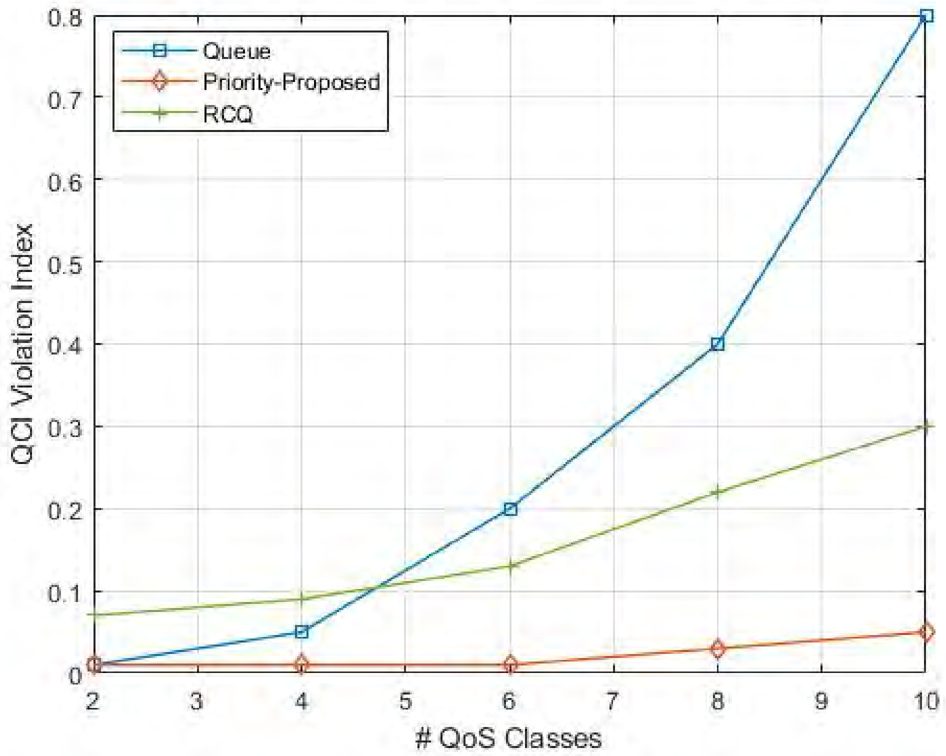
provides the highest throughput of 8.7 Mbps, and the priority algorithm provides the lowest throughput of 6.2 Mbps for 500 devices. The proportional fair and RCQ algorithms provide average results. The reasoning behind this phenomenon is that the throughput is closely related to the channel quality, which is crucial in deciding the MCS and TBS. The channels with high CQI can use higher MCS and TBS to provide a high data transmission rate. The algorithms that consider channel quality as an allocation metric provide high throughput. The average cell throughput improves with the number of devices but is bounded by Shannon capacity (2.2), as shown in figure 3.6.

3.6.5 QCI violation index

QCI violation index shows how often the algorithm fails to satisfy the QCI priority requirement. We have measured QCI satisfaction as the average number of QCI violations per frame. The number of frames in a single iteration of the simulation is 500. The high QCI violation index value shows that the algorithm frequently fails



(a) QCI violation index (For 7 QCIs).



(b) QCI violation index (For 250 devices).

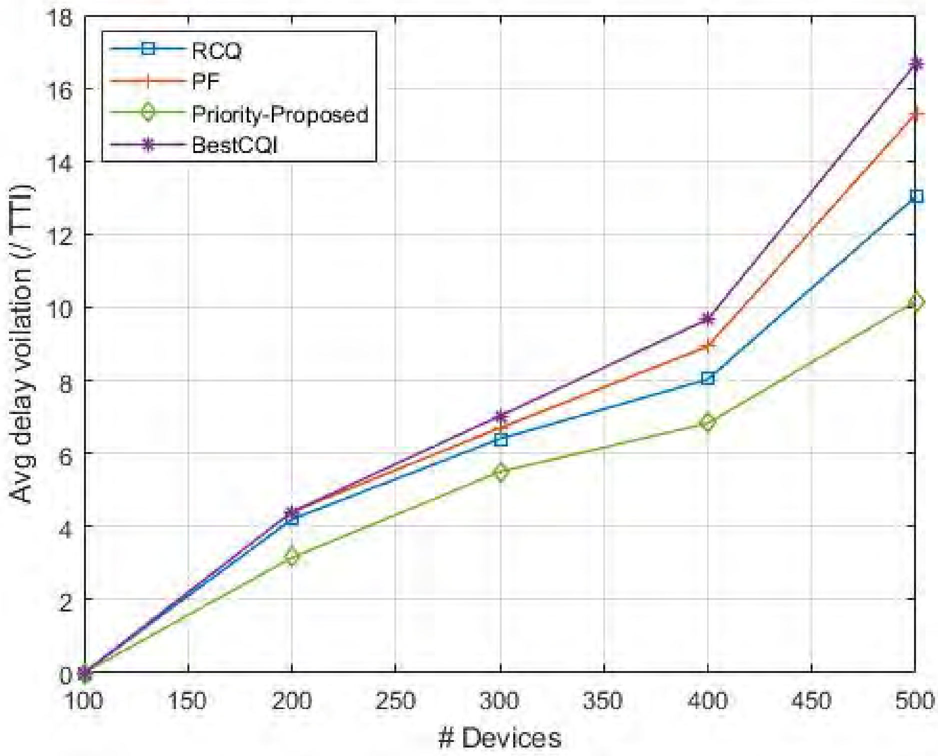
Figure 3.7: QCI violation index (Avg. violation per frame).

to support the QCI priority. The proportional fair, BestCQI, and RCQ algorithms perform up to 300 devices equally. The QCI violation index increases exponentially for BestCQI and PF scheduling schemes after 300 devices and reaches 0.8 for five hundred devices. The RCQ algorithm manages the growth to some extent, using the CQI and QCI priority ratio and limiting the QCI violation index to 0.5. The proposed priority-based algorithm maintains the QCI violation index below 0.05. As shown in figure 3.7a, the priority-based algorithm outperforms all other algorithms in QCI priority satisfaction. The QCI violation increases abruptly after 300 devices for the algorithms other than the priority-based algorithm.

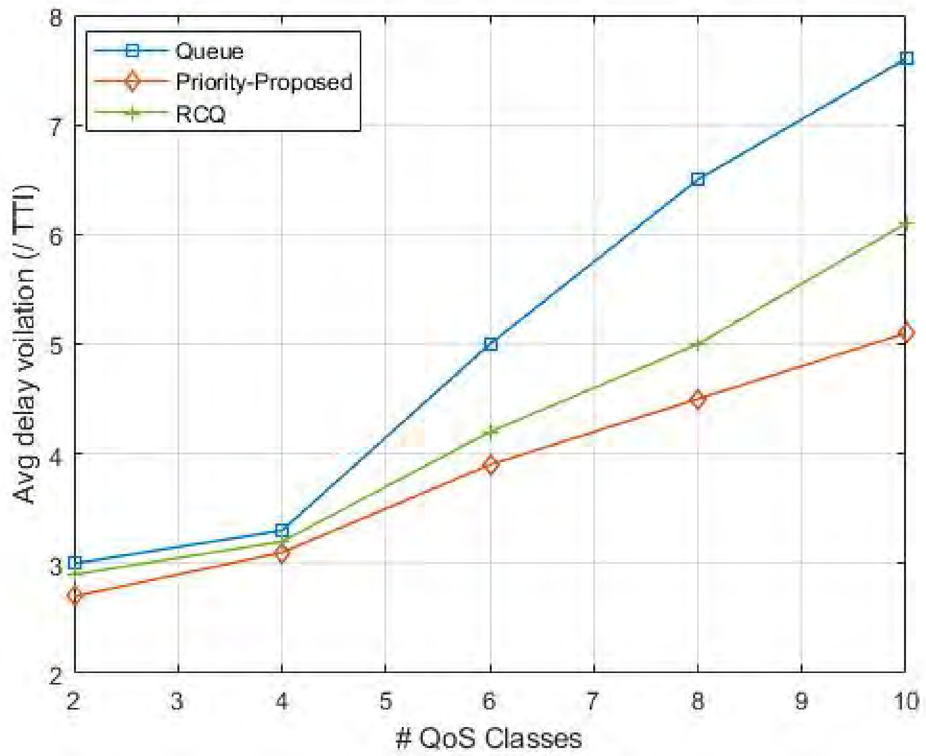
We compare the results of priority algorithms with two schedulers from the literature, RCQ, and Queue, for a different number of QoS classes. QCI violation for the priority algorithm increases gradually, while for other algorithms, it increases rapidly with the increasing number of QoS classes. Figure 3.7b shows the increment of QCI violation for the number of QCIs. The Queue algorithm uses two classes of traffic. Therefore at two QCIs, it performs equally to the QCI priority-based algorithm. After that, QCI-priority violations increase exponentially. The RCQ algorithm uses QCI priority with a combination of channel quality. Therefore, it has restricted growth in QCI priority violation.

3.6.6 Delay budget violation

The delay violation is the average number of incidents per TTI when the devices miss the delay budget. The average delay budget violation for different scheduling algorithms is shown in fig 3.8a. As shown in fig 3.8a, there is no incident when a device misses the delay budget for 100 devices with 5 MHz(25 PRBs) bandwidth. The delay budget violation is gradually increasing as the number of devices increases. The delay violation increases abruptly after 400 devices. The proposed priority algorithm better results in delay violation with ten average incidents. Delay budget violation increases rapidly as the number of devices increases. For an increasing number of QoS



(a) Delay budget violation (For 7 QCIs).



(b) Delay budget violation (For 250 devices).

Figure 3.8: Average delay budget violation per frame.

classes, delay budget violation increases promptly for RCQ and Queue algorithms after four QCIs with 250 devices, while it has sub-linear growth for the priority algorithm. Figure 3.8b shows that all algorithms have similar growth patterns after 16 QCIs.

Observations from results

1. A device should have data and an appropriate channel quality to get resources. The higher volume of devices increases the probability of having data and appropriate channel conditions for many devices. This increases the chances of fair distribution of resources among devices (equation 3.13). Thus, The fairness index value increases for an increasing number of devices. Although fairness increases with the number of devices, it cannot reach up to 1 because it is practically impossible to distribute resources equally due to unpredicted network conditions.

2. The resource utilization for all implemented algorithms observes a slack saturation after 300 devices and shows a logarithmic behavior. This implies that resource utilization increases with the number of devices, and it will not be possible to use 100% of resources.

3. The algorithms that consider channel quality as an allocation metric provide high throughput. The average cell throughput improves with the number of devices but is bounded by Shannon's capacity (equation 2.2), as shown in figure 3.6.

4. QCI priority violation increases rapidly for all the algorithms other than the priority-based algorithm for both varying numbers of devices and QCIs.

5. Delay budget violation moderately increases for priority-based algorithm for an increasing number of devices with seven QCI classes. But delay-budget

increases rapidly for all algorithms, including priority-based algorithms with more QCIs.

6. The simulation results show that the proposed priority algorithm performs poorly in throughput and resource utilization. It gives average results in resource-sharing fairness.

7. The priority-based algorithms perform better than others in priority support and delay budget violations.

8. For the higher number of QoS classes, the priority algorithm gives good results for priority support and average results for delay-budget violation.

9. The results conclude that the QoS satisfaction of the end-users degrades if the resource utilization and throughput increase and vice versa.

Preface to chapter 4

- The priority-based algorithm performs best in application priority support. Therefore, we continue using application priority in resource allocation in the proposed scheme in chapter 4.

- The priority-based algorithm performs the poorest in resource utilization and throughput. We jointly optimize the channel quality and application priority (equation 4.3) to improve resource utilization and throughput in chapter 4.

- The priority-based algorithm performs below average in the fair distribution of resources among the devices. We include a scaling factor regarding the number of scheduling grants, devices, and QCIs (equation 4.5) to improve fairness in resource distribution in chapter 4.

- The priority-based algorithm performs well for delay budget violation for seven QCIs but does not perform well for a larger number of QCIs. Therefore, to improve delay budget violation, we include delay constraint (equation 4.6) in the resource allocation process in chapter 4.

3.7 Chapter Summary

- The proposed game-theoretic approach provides a stable M2M system with malfunctioning devices.

- The results show that the throughput is directly proportional to the CQI between the eNB and UE. CQI is used to decide the transport block size (TBS). UEs having a significant TBS with better CQI can send enormous amounts of data using a higher modulation and coding scheme (MCS).

- From Fig.3.6, it can be observed that the allocation algorithms that use CQI as an allocation metric, directly or indirectly, provide better results regarding cell throughput. Therefore, for the system (M2M system) with a diverse range of QoS requirements, if the allocation is more focused on QoS satisfaction, it compromises resource utilization.

- There is a trade-off between QoS satisfaction and resource utilization. The degradation in resource utilization affects the overall throughput.

- Defining the correct time interval and the number of scheduling chances is a very challenging task to minimize the probability of getting fewer scheduling chances for fair devices.

Chapter 4

Balancing Performance and QCI support for M2M

The content of this chapter is submitted for publication in the following article.
U. Singh, A. Dua, N. Kumar, S. Tanwar, R Iqbal, M. Hijji, R. Sharma, "Scalable Priority-based Resource Allocation Scheme for M2M Communication in LTE/LTE-A Network," in Computers and Electrical Engineering,2022 Oct 1;103:108321 **[Q1, SCIe, IF-4.15]**
https://www.sciencedirect.com/science/article/pii/S0045790622005432

4 | Balancing Performance and QCI support for M2M

"Research is to see what everybody else has seen and to think what nobody else has thought."

Albert Szent-Györgyi

Key Points:

- The solution in this chapter is intended to balance the M2M-LTE system's performance and QoS support for M2M applications.
- We proposed a scheduling scheme that provides social fairness regarding the number of devices rather than considering past data rates or channel quality.
- We give preferences to the devices for scheduling grants on social bases rather than on the physical and application layer parameters.
- In this chapter, we have used the term "socially fair or social fairness," which means we are concerned with the number of devices and type of applications installed on the devices.
- We introduce a virtual QCI concept to achieve social fairness, system performance, and QoS support and minimize the starvation for lower CQI and lower-priority devices. The virtual QCI is based on the application priority, channel quality, delay budget, number of QCIs, and the number of devices.

4.1 State-of-the-art

Dawaliby *et al.* [48] addressed the problem of allocating LTE radio resources in a collective energy-efficient and QoS-aware manner for both the time and frequency domains. This paper used a memetic-based optimization strategy to consider a cross-layer resource allocation for M2M devices over LTE-M. It optimizes resource allocation in the time and frequency domain to reduce the energy consumption of LTE-M devices while considering, at the same time, their delay requirements. Karadag *et al.* [76] used a heuristic approach to propose an aware semi-persistence radio resource allocation scheme for M2M communication in an LTE network. Using the heuristic approach, they efficiently used the frequency bands to increase the number of serving devices per schedule.

Ghavimi *et al.* [49] used a group-based radio resource allocation with identical transmission protocols and QoS requirements to ensure QoS guarantees for M2M devices and efficiently tackle the overload problems for M2M communications in 3GPP LTE-A networks. The authors presented a framework as a sum throughput maximization problem while respecting all the constraints associated with RB and power allocation in the LTE-A uplink networks. Abrignani *et al.* [81] formulated the radio resource allocation problem as mixed-integer linear programming (MILP) to balance the trade-off between throughput versus delay and throughput versus fairness. Girici *et al.* [171] proposed a proportional fair scheduling with QoS constraints. The authors optimized the required bandwidth and power to improve energy efficiency throughput.

Wei Fu *et al.* [128], used priority queue model to improve E2E QoS. The proposed mechanism prefers guaranteed-bit-rate (GBR) traffic over Non-GBR traffic. The solution fails to perform in terms of system throughput. Zaki *et al.* [172] also used the priority queue model to balance the system throughput and QoS support trade-offs. J. Yin *et al.* [173] used the queue model to support Quality-of-Experience (QoE) in their work. They focused on queue overflow and system stability. W.K. Lai *et al.* [174] used the queue model and focused on the access delay. The proposed solution imposes a high packet loss.

Hajer *et al.* [93] used a recursive expansion approach in frequency domain resource allocation to improve the energy efficiency of MTCDs. MTCD's delay budget is used as a QoS parameter, and the proposed scheduler provides average performance for QoS support. Shafinaz *et al.* [175] used a recursive expansion approach in the FDPS stage to satisfy the delay budget and increase the throughput of the individual device in vehicular communication. The authors proposed traffic classes for vehicular communication to prioritize the devices. The proposed solution lacks fairness compared to other solutions. Mostafa *et al.* [130] also implemented a recursive expansion mechanism for resource allocation in the frequency domain to ensure the End-to-End (E2E) QoS for machine type communication in the LTE network. The authors proposed a statistical priority based on traffic statistics. Maia *et al.* [52] proposed a dynamic uplink scheduler for M2M communication using a resource reservation approach for a different type of traffic. The authors divided traffic among event-driven, time-triggered, and H2H. They prioritize H2H traffic and reserve some frequency resources for that. The proposed solution lacks system throughput and fairness compared to other solutions.

4.2 Motivation

The M2M devices either have low or no mobility, so changes in the channel conditions are limited. Focusing only on channel quality indicator (CQI) as an allocation metric can cause starvation for devices with poor channel quality. M2M has a vast application domain from a mission-critical application (requires urgent transfer) to smart meters (very long delay budget). Therefore, application priority should be considered as an allocation metric. Moreover, considering application priority as the only allocation metric degrades resource utilization as the traffic pattern of M2M is very diverse. if high-priority devices are continuously present in the network. It may lead to starvation for lower-priority devices. Therefore, designing a solution that jointly considers channel quality and application priority as allocation metrics is desirable to minimize the trade-off between channel utilization and QoS satisfaction. However, joint optimization of CQI and application priority minimizes the trade-off between resource utilization and QoS support but shortfalls in minimizing the starvation for MTCDs with poor CQI and lower application priority. The application priority needs to be scalable to minimize the probability of starvation for some M2M devices.

Motivated by the discussion as mentioned earlier, this chapter presents a scalable priority-based resource allocation scheme for the M2M communication under the LTE/LTE-Advance network.

4.3 Contribution of This Chapter

This chapter provides LTE physical resource allocation schemes based on the scalable priority of the application installed on the MTCDs. Scalable priority minimizes the trade-off between resource utilization and priority handling. The proposed solution also minimizes the starvation of the device having poor channel quality and lower application priority.

Following are the research contributions of this chapter.

- This chapter provides LTE physical resource allocation schemes based on the scalable priority of the application installed on the MTCDs.
- The scalable priority minimizes the trade-off between resource utilization and priority handling.
- The proposed scheme also minimizes the starvation for the device having poor channel quality and lower application priority.

4.4 Weighted Priority Proportional Fair Scheduling

Let's assume a network scenario where M2M devices are randomly placed in LTE cells and communicating to eNB. M2M devices are stationary. Devices have a single application installed, and the application priority and the traffic pattern are as per the 3GPP QCI classification. This section discusses the proposed scheduling mechanism, focusing on resource utilization, fairness, and QoS support for M2M communication.

4.4.1 Problem formulation

Let D be a set of active devices randomly placed in a macro cell with one eNB, and M denote a set of available physical resource blocks. Each device has a single application installed on it with static priority. Priority is an integer number in $\mathbb{Z}^+$. The smallest priority number denotes the highest application priority. Let $X \subseteq D$ denote a set of eligible devices with data in their buffer and waiting for the transmission with available resources $|M|$. The scheduling of physical resources occurs at an integral time, numbered from zero. The scheduling time slot $[t, t+1); t \in T$ (including t, excluding $t+1$) is referred to as transmission time interval (TTI). If a device $i \in X$ need RB_i resources, then scheduling is required with $|M|$ resource if and only if $\sum_{i=1}^{|X|} RB_i > |M|; \forall\, i \in X$. Based on the buffer status report (BSR) received by eNB from UE, the LTE scheduler selects k devices for the transmission where $\sum_{i=1}^{k} RB_i \leq |M|;\ i \in X$ based on the allocation metrics. The proposed resource allocation scheme combines weighted priority with proportional fairness.

To improve channel utilization and priority support, the priority of devices p_i is weighted to the channel quality c^i. The weighted priority metric is defined as follows.

$$f(x_i^t) = \omega_i . p_i \quad \forall\, t \in T \tag{4.1}$$

Where $\omega_i = 1/c^i$ is priority weight such that $0 < \omega_i < 1; \forall i \in X$. Moreover, we combined the weighted priority of the previous scheduling grant time and the weight of the total number of devices and QCIs to improve fairness. A M2M device i has a priority weight $\omega_i = 1/c^i$ such that $0 < \omega_i < 1; \forall i \in X$ with priority p_i and c^i is channel quality. The fitness function $f(x_i^t)$ at time t for device i is defined as follows.

$$f(x_i^t) = \omega_i . p_i + \alpha . (f(x_i^{t-1}) + \xi) \quad \forall\, t \in T \tag{4.2}$$

Where ξ is a scaling factor in integrating the proportional fairness for a total number of priority classes and devices. The α is a scaling constraint and $\alpha \in \{0, 1\}$. The time $t-1$ represents the time of the last scheduling grant for device i.

Equation 4.2 is a recursive function bounded by the device i's delay budget D_b^i. Moreover, it is a convex function in the interval $[t, t']$ where $(t' - t) \leq D_b^i$. To match the LTE scheduling parameters, we define a virtual QoS class identifier based on the problem formulated in equation 4.2 and discuss it in the next section.

4.4.2 Virtual QoS class identifier

We define a virtual QoS class identifier (vQCI) $V_{qci}^i \, \forall i \in D$ for each device in set D to select devices in a proportional fair manner, considering channel quality to improve resource utilization, and QCIs to improve the priority handling for urgent services. The vQCI, V_{qci}^i, is used to decide the device's fitness for the scheduling in the current TTI. The metric $V_{qci}^{i,t}$ for a device i at time t is defined as a virtual QCI.

$$V_{qci}^{i,t} = \omega_i . p_i + \alpha \times \left(V_{qci}^{i,t-1} + \xi \right) \tag{4.3}$$

The scheduler per iteration selects a device that has a minimum for V_{qci}^i from the eligible devices. Further, the value of vQCI defined in equation (4.3) is increased per scheduling grant concerning the total number of QCIs and devices.

Priorities assigned to applications are static, and M2M devices have minimal or no mobility. In this case, changes in the value of V_{qci}^{i} for device i are rare. This scenario leads to starvation for devices with poor channel quality and lower-priority applications. Therefore, we are concerned about the total number of devices and QoS classes in scaling the priority. Thus, the scaling factor ξ is defined as follows.

$$\xi = \frac{qci^{i,t}}{|qci|} \times |D| \tag{4.4}$$

Thus, with priority scaling, the V_{qci}^{t} at time t is defined as follows.

$$V_{qci}^{i,t} = \frac{qci^{i,t}}{cqi^{i,t}} + \alpha \times \left(V_{qci}^{i,t-1} + \left(\frac{qci^{i,t}}{|qci|} \times |D| \right) \right) \tag{4.5}$$

Where $|qci|$ represents the total number of QoS classes that scale the $qci^{i,t}$ to the total number of QCIs, and $|D|$ represents the total number of devices and socially scales the allocation metric. The increment in the value of $V_{qci}^{i,t}$ for device i at time t is constrained by equation (2.5). Although the value of $V_{qci}^{i,t}$ is increased for every grant, we reset it to zero if the difference between the last grant and the current scheduling time violates the delay budget of the device. The term α in equation (4.5) is defined as follows.

$$\alpha = \begin{cases} o & if\, t_{cg}^{i} - t_{lg}^{i} \geq D_{b}^{i} \\ 1 & Otherwise \end{cases} \tag{4.6}$$

Where t_{cg}^{i} is the current scheduling grant time, and t_{lg}^{i} is the last scheduling grant time for the device i. Our proposed algorithm uses the device selection metric defined in the equation 4.5 as the primary allocation metric.

4.4.3 Resource allocation algorithm

The proposed scheduling algorithm uses virtual QCI as a primary allocation metric. The proposed scheduling scheme minimizes the value of fitness function $f(x_i^t)$ for

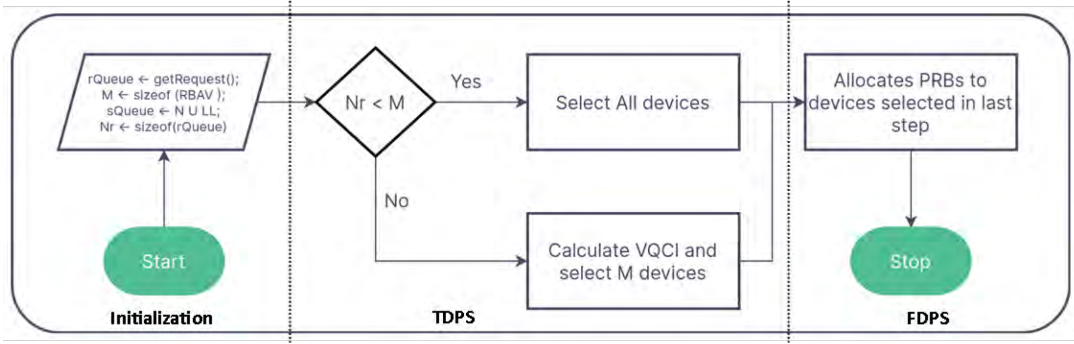


Figure 4.1: Proposed scheduling scheme.

the current allocation for k number of selected devices for the current TTI.

$$argMin \sum_{i=1}^{k} f(x_i^t) \tag{4.7}$$

The device with a minimum value of virtual QCI is given priority over other devices. The proposed scheduling scheme is divided into three parts: initialization, TDPS, and FDPS, as shown in figure 4.1. The virtual QoS (vQCI) class identifier defined in section 4.4.2 selects devices in the time domain. We use the red-black tree to store vQCI and other information related to devices to reduce the time complexity of the resource allocation process.

Red-Black tree building in TDPS

A red-black tree is a self-balancing binary search tree with one extra bit at each node, generally referred to as the color (red or black). As insertions and deletions are made, these colors balance the tree. The balance of the tree is not perfect, but it is good enough to reduce searching time to about O(log n), where n is the total number of elements in the tree. These trees have the same memory footprint as a regular (uncolored) binary search tree because each node only needs 1 bit of memory to hold the color information. Figure 4.2 illustrates the structure of a red-black tree. Every Red-Black tree maintains the following properties.

- Every node has a color, either red or black.

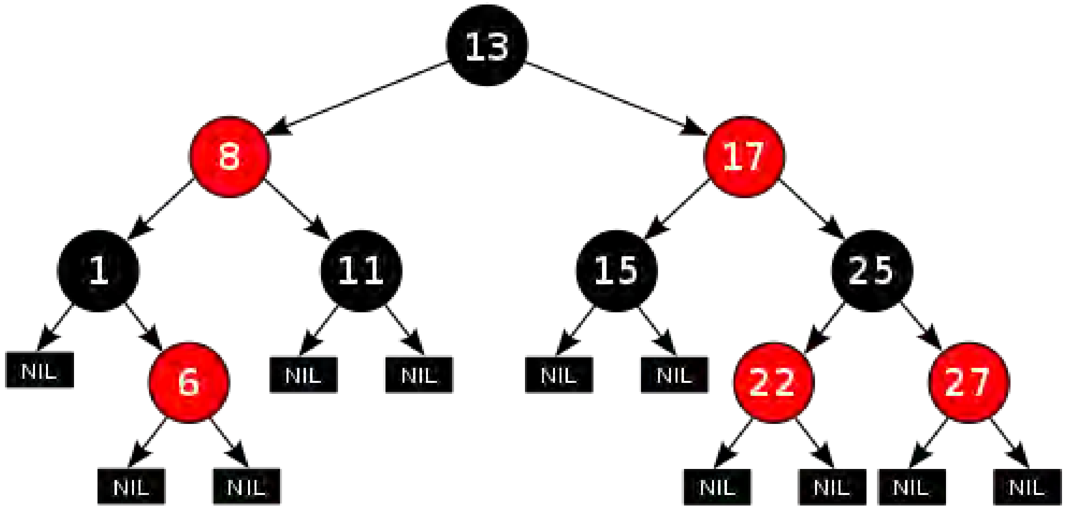


Figure 4.2: Structure of a red-black tree.

- The root of the tree is always black.
- There are no two adjacent red nodes.
- Every path from a node (including root) to any of its descendant's NULL nodes has the same number of black nodes.
- All leaf nodes are black nodes.

Compared to the Red-Black tree, the AVL tree is more balanced, although the AVL tree may produce more rotations during insertion and deletion. If the application requires a lot of insertions and deletions (As in our case), the Red-Black tree is the way to go. If insertions and deletions are few and searching is the most frequent operation, the AVL tree should be used over the Red-Black Tree.

Proposed scheduling algorithm builds a red-black (RB) tree of eligible devices. An eligible device has data and is ready for transmission. The balancing of the RB tree is performed using the value of virtual QCI ($V_{qci}^{i,t}$). RB tree is an approximately balanced tree and has height at most $2lg(n+1)$ for n nodes [176].

Lemma 4.4.1. *A red-black tree with n internal nodes has a height at most* $2lg(n+1)$ *[176].*

Insertion and deletion in a binary search tree (BST) of height h takes $O(h)$ time. Thus, the height of a red-black tree is $O(log\ n)$, which ensures that an insertion or deletion operation on a red-black tree can be performed in $O(log\ n)$. Thus, building a red-black tree of eligible devices takes $O(log\ n)$ time [176]. Due to the ease of insertion and deletion operation in the red-black tree, the proposed algorithm uses a red-black tree as a primary data structure for storing information regarding the scheduling process. The algorithm 2 presents the scheduling scheme.

Proposed algorithm

The main steps of the proposed scheduling scheme are listed in the algorithm 2. The UE sends its buffer status via BSR to eNB. If the UE has data in its buffer for the uplink transmission, it is considered for the current scheduling process. In the initialization phase, the scheduler collects the information on available data in UE's buffer, available resources, and the number of active UEs. In the algorithm 2, steps from $1-4$ perform the initialization of the scheduler.

If a device $i \in X$ needs RB_i resources from $|M|$ available resources. There are two possibilities for the scheduling based on the initial information.

1. **if** $\sum_{i=1}^{|X|} RB_i \leq |M|; \forall i \in X$**:** In this case, The number of UEs with data in their buffer is less than the number of available resources, TDPS selects all UEs for data transmission in steps 7. In steps 8 to 12 in algorithm 2, FDPS performs the resource allocation with MCS and TBS assignment.

2. **if** $\sum_{i=1}^{|X|} RB_i > |M|; \forall i \in X$**:** In this case, the number of eligible UEs is more than the available resources. Here, the scheduler selects the devices for resource allocation using the value of the virtual QCI of each UE. Steps from 14 to 22 in algorithm 2 update the value for virtual QCI depending on the value of α. In step 23, the proposed algorithm builds the red-black tree for the eligible devices. In steps 24 to 29, the FDPS assigns the radio resource for the selected devices with appropriate MCS and TBS and updates the last scheduled grant

Algorithm 2: WPPF_Scheduler

Data: QoS Parameter: Application Priority p, QCI
Other Parameter: CQI, BSR, MCS Matrix, rQueue, sQueue, $last\_G^t$, $curr\_G^t$
Result: Set of selected MTCDs with assigned RBs

```
Initialization:
rQueue ← getRequest();
M ← sizeof (RB_AV);
sQueue ← NULL; & N_r ← sizeof(rQueue);
begin
    if N_r < M then
        sQueue ← rQueue;
        for j ← 0 to N_r do
            RB_Map(sQueue_j) ← Map(sQueue_j, RB_AV);
            mcsIndex(sQueue_j) ← MCS (RB_Map(sQueue_j));
            TBS(sQueue_j) ← tbSize (mcsIndex(sQueue_j));
        end
    else
        for j ← 0 to N_r do
            if (curr_G^t - last_G^t) ≥ DB then
                α = 0;
                computeVQCI (V_qci, rQueue_j, C, p_j, Q);
            else
                α = 1;
                computeVQCI (V_qci, rQueue_j, C, p_j, Q);
            end
        end
        buildRBTree (T_RB, rQueue);
        for j ← 0 to M do
            RB_Map(sQueue_j) ← Map(removeMTCD (T_RB, RB_AV));
            mcsIndex(sQueue_j) ← MCS (RB_Map(sQueue_j));
            TBS(sQueue_j) ← tbSize (mcsIndex(sQueue_j));
            updateLastGrantTime (sQueue_j, curr_G^t);
            updatevQCI (sQueue_j, vQCI);
        end
    end
end
```

time to the current grant time. Then, eNB sends scheduling grants to the UEs in DCI-0 format. The UE inspects the received DCI-0 and starts actual data transmission. If the UE receives a scheduling grant in n^{th} TTI, then it can transfer the data in $(n+4)^{th}$ TTI.

The proposed scheduling scheme uses virtual QCI values for selecting devices, which are updated after every scheduling grant of the devices. The proposed algorithm uses two for-loops that iterated up to m time and took $O(m)$ time to complete. Another for-loop iterates n times and takes $O(n)$ time to complete. For dense M2M networks, $n >> m$ and the loops can be completed in $O(n)$ time. The proposed algorithm uses a red-black tree that takes $O(n\ lg\ n)$ with n active devices. Thus, the proposed algorithm takes time to complete $O(n\ lg\ n)$.

Numerical illustration of proposed algorithm

The proposed weighted priority proportional fair (WPPF) algorithm considers the following metrics in selecting devices for radio resource allocation as specified in equation 4.5.

- QCI priority of M2M application.
- QCI delay budget to control scaling.
- Channel quality (CQI) as a weight for QCI priority.
- Total number of devices, QCIs, and vQCI of M2M device at last grant as a scaling factor.

The critical steps of the proposed algorithm are first to select devices with weighted priority and then lower the priority with a scaling factor. If the delay budget of devices is close, then the proposed algorithm resets to the original priority of the device. Figure 4.3 and 4.4 show the calculation of all steps. The cells shaded in light green are selected RNTIs in that step and shaded in light red to show the delay budget time-out for that device.

Let's consider two physical resource blocks (PRBs) available. The number of devices is 6, and the total number of QCIs is 3. There are two devices from each QCI category. We assume that the delay budget for QCIs 1, 2, and 3 are 2, 3, and 4 steps. The initial vQCI for each device is zero, as shown in Table 1 of figure 4.3.
In step 1, the scheduler calculates CQI and weighted priority for each device. The vQCI is calculated as per the equation 4.5. In step 1, vQCI' is received from the initial stage. So, based on the value of vQCI, devices 1 and 2 are selected for allocation grants. The value of vQCI and the last grant step are updated for devices 1 and 2.
In step 2, the value of vQCI' for devices 1 and 2 is 2.11; for the remaining devices, it is still 0 from the initial step. Now, the scheduler selects devices 1 and 2 based on the vQCI values. The value of vQCI and the last grant step are updated for devices 1 and 2 to 4.18 and 4.22.

In step 3, values of vQCI' for devices 1 and 2 are 4.18 and 4.22; for others, it is 0. Based on the vQCI calculation scheduler, select devices 3 and 4 for the scheduling grant and update the vQCI and last grant step value for devices 3 and 4. This process continues for all steps.
In step 5, the delay budget of two steps for devices 1 and 2 is expired. Thus, the value of vQCI' for devices 1 and 2 becomes zero. It significantly reduces the value of vQCI for devices 1 and 2. Therefore, devices 1 and 2 are selected for scheduling.
Step 7's delay budgets for devices 3 and 4 are timed out. Thus, the value of vQCI' becomes zero, and devices 3 and 4 are selected for scheduling.
Step 9's delay budgets are exceeded for devices 2, 5, and 6. However, we can serve only two devices. Therefore, devices 2 and 6 are selected for scheduling.

Initial

RNTI	1	2	3	4	5	6
QCI Priority	1	1	2	2	3	3
vQCI	0	0	0	0	0	0

Step - 1

RNTI	1	2	3	4	5	6
QCI Priority	1	1	2	2	3	3
CQI	9	9	5	13	13	15
X = QCI/CQI	0.11	0.11	0.40	0.15	0.23	0.20
vQCI = X + vQCI'	2.11	2.11	4.40	4.15	6.23	6.20
Updated vQCI'	2.11	2.11	0.00	0.00	0.00	0.00
Last Grant Step	1	1	0	0	0	0
Selected RNTI	1	2	4	3	6	5

Step - 2

RNTI	1	2	3	4	5	6
QCI Priority	1	1	2	2	3	3
CQI	14	9	4	5	14	4
X = QCI/CQI	0.07	0.11	0.50	0.40	0.21	0.75
vQCI = X + vQCI'	4.18	4.22	4.50	4.40	6.21	6.75
Updated vQCI'	4.18	4.22	0.00	0.00	0.00	0.00
Last Grant Step	2	2	0	0	0	0
Selected RNTI	1	2	4	3	5	6

Step - 3

RNTI	1	2	3	4	5	6
QCI Priority	1	1	2	2	3	3
CQI	12	3	15	3	13	10
X = QCI/CQI	0.08	0.33	0.13	0.67	0.23	0.30
vQCI = X + vQCI'	6.27	6.56	4.13	4.67	6.23	6.30
Updated vQCI'	4.18	4.22	4.13	4.67	0.00	0.00
Last Grant Step	2	2	3	3	0	0
Selected RNTI	3	4	5	1	6	2

Step - 4

RNTI	1	2	3	4	5	6
QCI Priority	1	1	2	2	3	3
CQI	3	6	1	13	12	15
X = QCI/CQI	0.33	0.17	2.00	0.15	0.25	0.20
vQCI = X + vQCI'	6.52	6.39	10.13	8.82	6.25	6.20
Updated vQCI'	4.18	4.22	4.13	4.67	6.25	6.20
Last Grant Step	2	2	3	3	4	4
Selected RNTI	6	5	2	1	4	3

Step - 5

RNTI	1	2	3	4	5	6
QCI Priority	1	1	2	2	3	3
CQI	11	3	13	4	5	14
X = QCI/CQI	0.09	0.33	0.15	0.50	0.60	0.21
vQCI = X + vQCI'	0.09	0.33	8.29	9.17	12.85	12.41
Updated vQCI'	0.09	0.33	4.13	4.67	6.25	6.20
Last Grant Step	5	5	3	3	4	4
Selected RNTI	1	2	3	4	6	5

Figure 4.3: Numerical illustration of the proposed algorithm -1

4.5 Performance Evaluation of Proposed Scheduling Scheme

In chapter 3, we have analyzed the effect of priority support on the system performance. The proposed algorithm minimizes the trade-offs between system performance and QCI priority support. Therefore, we consider the same network and sim-

Step - 6

RNTI	1	2	3	4	5	6
QCI Priority	1	1	2	2	3	3
CQI	12	11	9	10	13	8
X = QCI/CQI	0.08	0.09	0.22	0.20	0.23	0.38
vQCI = X + vQCI'	2.17	2.42	8.35	8.87	12.48	12.58
Updated vQCI'	2.17	2.42	4.13	4.67	6.25	6.20
Last Grant Step	6	6	3	3	4	4
Selected RNTI	1	2	3	4	5	6

Step - 7

RNTI	1	2	3	4	5	6
QCI Priority	1	1	2	2	3	3
CQI	4	8	9	7	1	6
X = QCI/CQI	0.25	0.13	0.22	0.29	3.00	0.50
vQCI = X + vQCI'	4.42	4.55	0.22	0.29	15.25	12.70
Updated vQCI'	2.17	2.42	0.22	0.29	6.25	6.20
Last Grant Step	6	6	7	7	4	4
Selected RNTI	3	4	1	2	6	5

Step - 8

RNTI	1	2	3	4	5	6
QCI Priority	1	1	2	2	3	3
CQI	9	9	5	13	13	15
X = QCI/CQI	0.11	0.11	0.40	0.15	0.23	0.20
vQCI = X + vQCI'	4.28	4.53	4.62	4.44	12.48	12.40
Updated vQCI'	4.28	2.42	0.22	4.44	6.25	6.20
Last Grant Step	8	6	7	8	4	4
Selected RNTI	1	4	2	3	6	5

Step - 9

RNTI	1	2	3	4	5	6
QCI Priority	1	1	2	2	3	3
CQI	15	12	1	6	10	15
X = QCI/CQI	0.07	0.08	2.00	0.33	0.30	0.20
vQCI = X + vQCI'	6.35	0.08	6.22	8.77	0.30	0.20
Updated vQCI'	4.28	0.08	0.22	4.44	0.30	0.20
Last Grant Step	8	9	7	8	4	9
Selected RNTI	2	6	5	3	1	4

Step - 10

RNTI	1	2	3	4	5	6
QCI Priority	1	1	2	2	3	3
CQI	4	7	1	8	4	15
X = QCI/CQI	0.25	0.14	2.00	0.25	0.75	0.20
vQCI = X + vQCI'	6.53	2.22	6.22	8.69	0.75	6.40
Updated vQCI'	4.28	2.22	0.22	0.29	0.75	0.20
Last Grant Step	8	10	7	8	10	9
Selected RNTI	5	2	3	6	1	4

Step - 11

RNTI	1	2	3	4	5	6
QCI Priority	1	1	2	2	3	3
CQI	9	10	2	4	2	12
X = QCI/CQI	0.11	0.10	1.00	0.50	1.50	0.25
vQCI = X + vQCI'	0.11	4.32	1.00	0.50	8.25	6.45
Updated vQCI'	0.11	2.22	0.22	0.50	0.75	0.20
Last Grant Step	8	10	7	8	10	9
Selected RNTI	1	4	3	2	6	5

Figure 4.4: Numerical illustration of the proposed algorithm-2

of the proposed weighted priority proportional fair (WPPF) scheduler. We implement the application traffic and QCI priorities in table 3.3 and physical layer parameters in table 3.2. We compare the result of the proposed WPPF algorithm with standard schedulers' proportional fair and BestCQI scheduling algorithms. We also compare with state-of-art scheduler RCQ from Dawaliby *et al.* [48] and the QCI priority algorithm from chapter 3. The updated performance evaluation metrics are shown in table 4.1

Table 4.1: Performance evaluation metrics

Algorithm	**Objective**	**Evaluation Metric**	**Domain**
Proportional Fair	Fairness	Resource sharing fairness	System Performance
BestCQI	Throughput	Average cell throughput	
	System Utilization	Resource Utilization	
RCQ [48]	QoS	Delay budget	QoS Support
		QCI Priority	
Queue [79]	QoS	Delay budget	QoS Support
		QCI Priority	
Priority Chapter 3	QoS	Delay budget	QoS Support
		QCI Priority	

4.5.1 Average cell throughput

The average cell throughput for the proposed scheduling algorithm and other algorithms is shown in figure 4.5. The priority-based algorithm proposed in chapter 3 provides minimum average cell throughput ranging from 2 to 6.3 Mbps. The BestCQI algorithm provides the highest average cell throughput ranges from 2 to 8.5 Mbps. The difference between the highest and lowest achieved throughput is 3.2 Mbps. The proposed WPPF algorithm and PF and RCQ algorithm provide average throughput. We can conclude that the achievable throughput over a specific bandwidth depends on the channel quality between the device and eNB. Higher channel quality permits a larger transport block size (TBS) with a higher modulation and coding scheme(MCS). More bits can be transmitted in a single resource element

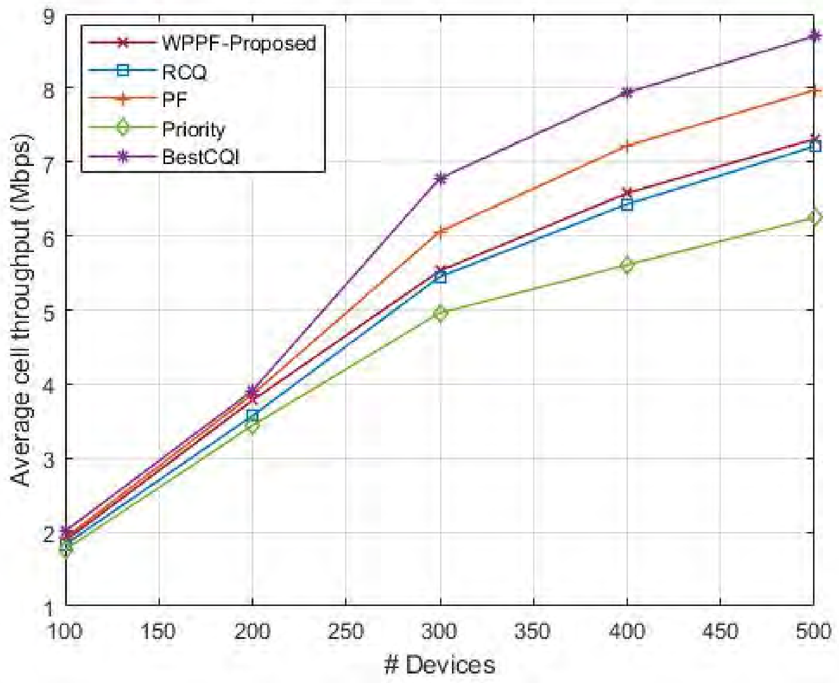


Figure 4.5: Average cell throughput.

with high MCS. The scheduling algorithm that uses CQI as an allocation metric, like BestCQI and PF, provides a higher cell throughput. The proposed algorithm (WPPF) and RCQ use the ratio of CQI to priority to provide better average cell throughput.

4.5.2 Resource sharing fairness

The resource sharing fairness is defined in equation 2.8. The value one for the fairness index shows that the allocation algorithm is entirely fair, and zero indicates that the algorithm is entirely unfair. The simulation results for the resource sharing fairness are shown in figure 4.6. The BestCQI algorithm is the lowest in resource-sharing fairness, as the value of the fairness index is below 0.3 for all cases. The priority algorithm and RCQ are slightly fairer than the BestCQI algorithm, and the fairness index value ranges from 0.35 to 0.45. The proposed WPPF algorithm provides better fairness than all other algorithms, and the value of the fairness index varies from 0.55 to 0.85 for 100 devices to 500 devices. The difference in the highest fairness index value for proposed WPPF and BestCQI is 0.55 for 500 devices,

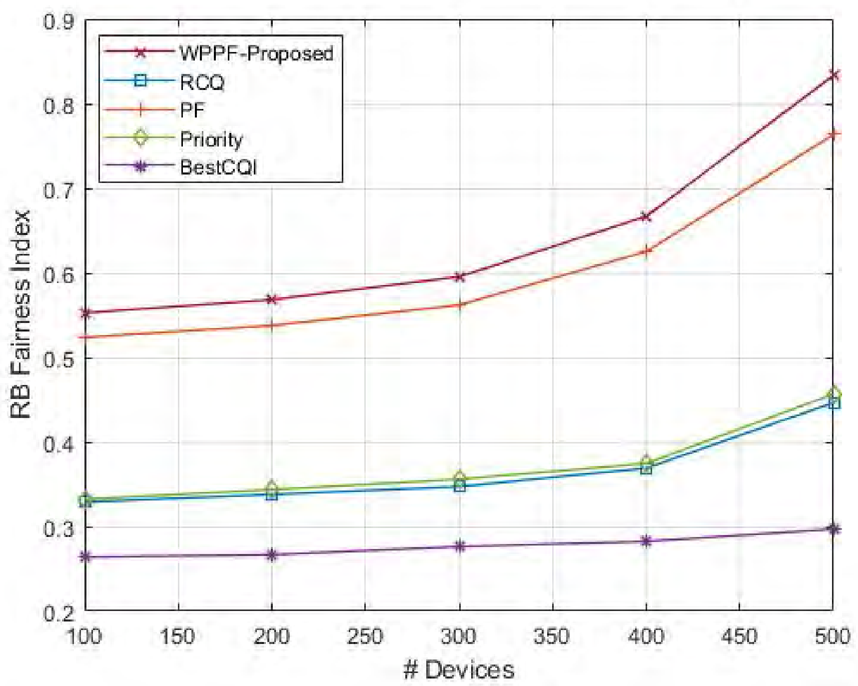


Figure 4.6: Resource sharing fairness.

which is more than 50% of the fairness range. The simulation results show that the scheduling schemes independent of device attributes like BestCQI perform poorly in fairness. The proposed algorithm performed better in fairness, considering total devices and total CQIs in the allocation metric.

4.5.3 Resource utilization

Resource utilization is defined as the percentage of resources used from available resources. The simulation results for average resource utilization per TTI are shown in figure 4.7. The priority algorithm from chapter 3 provides the lowest resource utilization in 20% to 72% for the 100 to 500 devices. BestCQI provides the best resource utilization with 92% for 500 devices. The proposed algorithm provides 90% resource utilization, which is 18% more than the priority algorithm. We can conclude that resource utilization directly depends on the channel quality between the

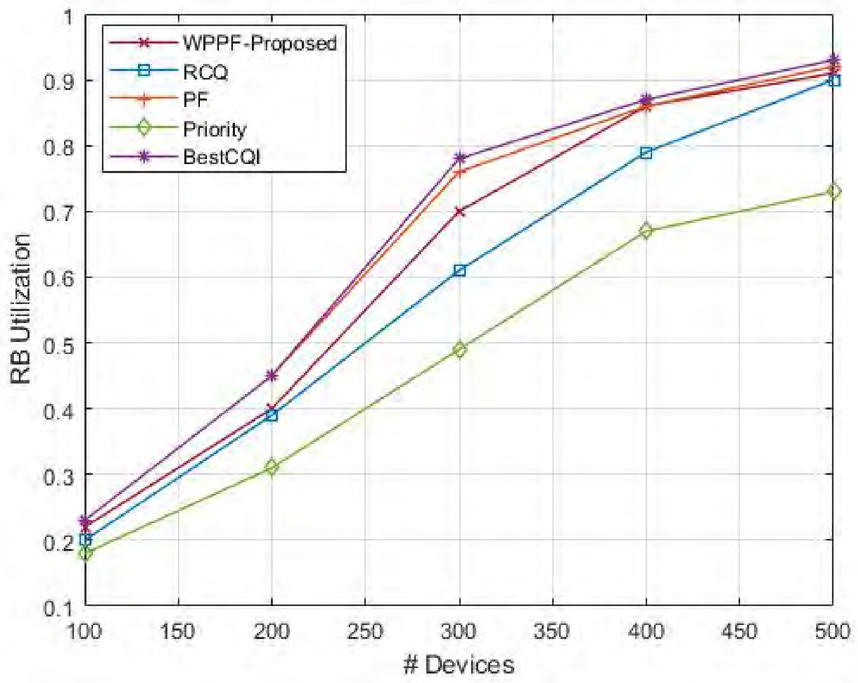
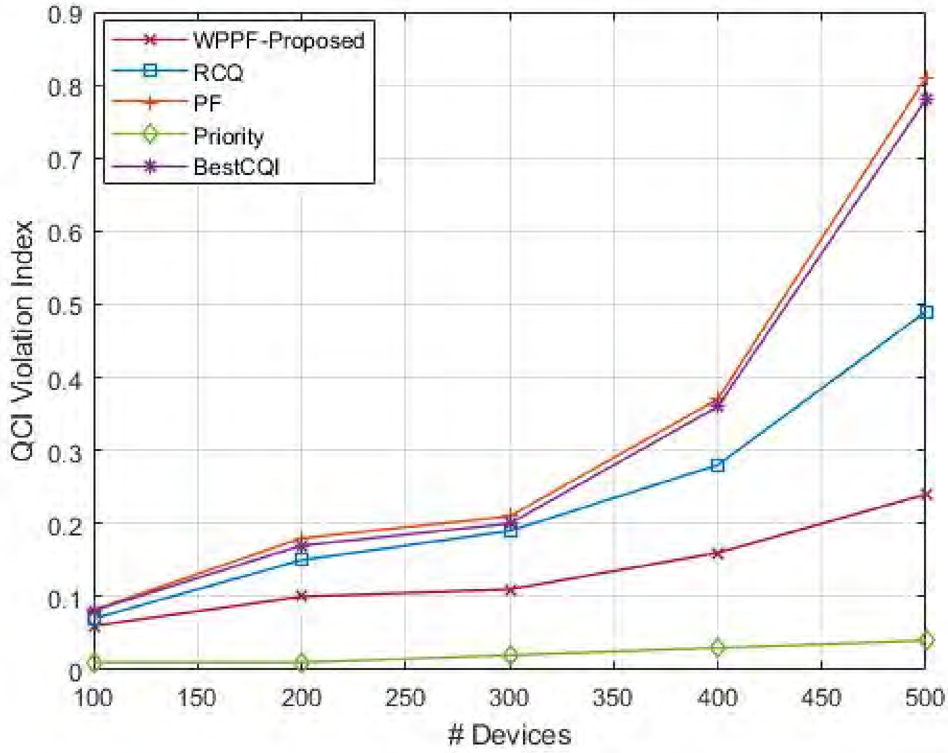


Figure 4.7: Average resource utilization per TTI.

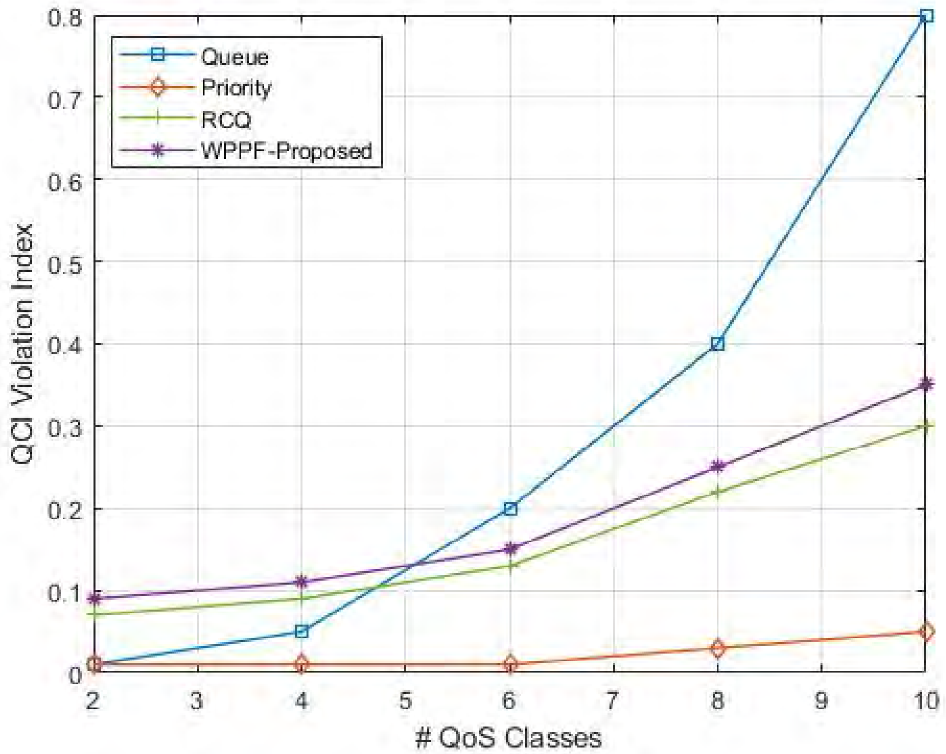
device and eNB. If optimization of the scheduling scheme focused on channel quality, the scheduling scheme could provide higher resource utilization. The scheduling scheme, BestCQI, provides the highest resource utilization. In comparison, the priority-based algorithm provides the lowest resource utilization. However, the proposed algorithm provides satisfactory results in resource utilization.

4.5.4 QCI priority support

M2M communication systems sometimes require urgent transmission for critical services, e.g., road accident alerts. To evaluate the performance of urgent services, we use the QCI violation index. The QCI violation index is defined as the average number of times per frame when a lower-priority device is granted access, even if the higher-priority device has data for transmission. The QCI priority-based algorithm performs best in this category, and the QCI violation index value remains below 0.05. The proposed WPPF scheduling algorithm performs satisfactorily. The QCI violation index value ranges from 0.05 to 0.25 and has a difference of 0.55 from the

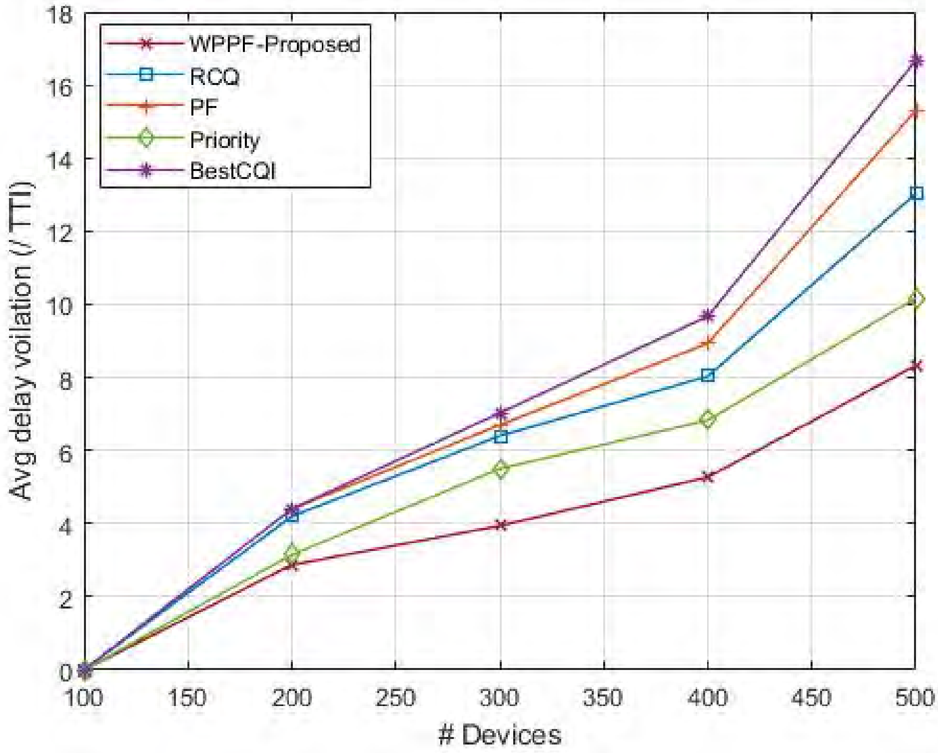


(a) QCI violation index (For 7 QCIs).

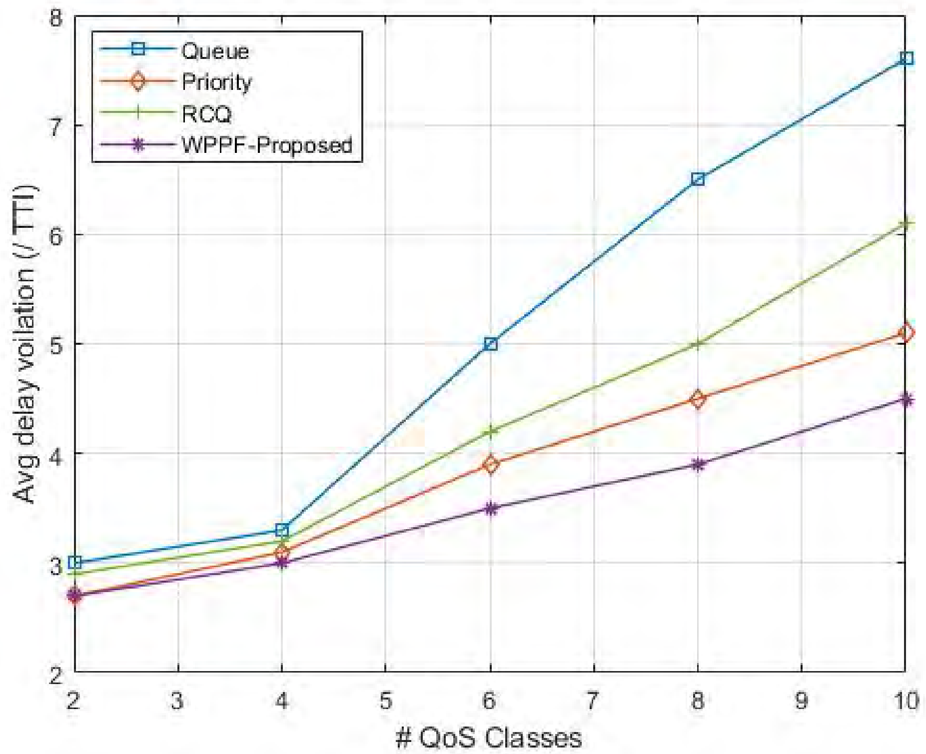


(b) QCI violation index (For 250 devices).

Figure 4.8: QCI violation index (Avg. violation per frame).



(a) Delay budget violation (For 7 QCIs).



(b) Delay budget violation (For 250 devices).

Figure 4.9: Average delay budget violation per frame.

worst case of the proportional fair algorithm and 0.2 from the best case. The RCQ algorithm performs average in the QCI priority support. The results for QCI priority violation for the different number of devices are shown in figure 4.8a, and for the different number of QoS classes are shown in figure 4.9b. QCI violation index for the algorithm queue increases quickly after four QCIs, and the RCQ and WPPF increase moderately for the increasing number of QoS classes. The queue algorithm reaches up to 0.8 average QCI violations per TTI while priority remains below 0.1 average QCI violation per frame.

4.5.5 Delay budget violation

As shown in table 1.1, an MTC application has a specific delay budget. After the violation of the delayed budget, information lost its significance. The proposed algorithm has a scaling factor and the delay constraint as shown in equation (4.5). The average delay budget violation of up to 200 devices does not differ much. After 200 devices, the average delay budget violation increases drastically. BestQCI algorithm has 17 average delay budget violations. The proposed WPPF algorithm has eight average delay budget violation incidents, which is approximately 50% of the BestCQI algorithm. The algorithms PF, RCQ, and Priority provide average results. Due to the scaling and delay constraints, the proposed algorithm provides better results than others. The simulation results are shown in figure 4.9a. The delay-budget violations for RCQ and Queue algorithms increase exponentially. In contrast, for the proposed WPPF algorithm, growth in delay-budget violations is very low for the increasing number of QoS classes. Figure 4.9 shows results for delay violation with variable QoS classes.

Observations from results

- The results show that the proposed algorithm performs best in delay budget satisfaction and fairness and better in QCI priority support due to fast-forwarding in the priority line using scaling and the delay constraint.

- Using a combination of channel quality and QCI priority improves the proposed algorithm's resource utilization and average cell throughput performance.
- The proposed WPPF algorithm provides better results for many QoS classes regarding QCI priority and the delay-budget violation.
- The proposed WPPF algorithm performs more than average in throughput, resource utilization, and priority support. The WPPF algorithm performs best in fairness and delay-budget violation.
- Thus, the proposed WPPF algorithm perfectly balanced the system performance and QoS support for M2M communication in a heterogeneous environment.
- The improved resource-sharing fairness also minimizes the cases of starvation.

4.5.6 Comparison to recent proposals in the literature

We have compared our proposed algorithm with the recent proposals of LTE scheduling strategies in the literature concerning the parameters, i.e., energy efficiency, throughput efficiency, resource sharing fairness, resource utilization, delay budget, and balance between system performance and QoS. We do not implement all these proposed schedulers as those schedulers do not completely match our implementation scenario. Table 4.2 compares different algorithms.

4.6 Chapter Conclusion

- This chapter presents an uplink resource allocation algorithm for M2M communication in the SC-FDMA-based LTE/LTE-Advance network.
- The proposed algorithm combines channel quality and application priority as allocation metrics. It scales the allocation metric concerning the total QCI classes and the number of devices for every scheduling grant.

Table 4.2: Comparison to recent proposals in the literature.

Year	Author	Methodology	EE	TE	Fairness	RU	DB	PR	BPQ	Traffic Classes	3GPP	M2M?
2019	Huei-Wen *et al.* [177]	Weighted average	N	Y	Y	N	Y	Y	N	Multi	Y	N
2020	Moustafa *et al.* [178]	Traffic Classification	N	Y	Y	Y	Y	N	Y	Multi	N	N
2020	Mohammed *et al.* [179]	Partition and Matching	N	Y	N	N	N	N	N	H2H and M2M	N	Y
2021	Jeremy *et al.* [180]	Statistical	N	Y	Y	N	Y	N	N	Video, VoIP	N	N
2021	Gunasekaran *et al.* [181]	Statistical	N	Y	N	Y	Y	N	N	Multi	N	N
2021	Mukhopadhyay *et al.* [182]	ILP	N	Y	Y	N	Y	Y	Y	Video, VoIP, Data	N	N
2021	Heba *et al.* [183]	BestCQI with delay	N	Y	N	N	Y	N	Y	NA	NA	NA
2022	Leeban *et al.* [184]	Memetic-Based	Y	Y	Y	N	N	N	N	NA	N	Y
2022	Abdulhakeem *et al.* [185]	Resource Partition	N	Y	N	N	N	Y	N	GBR vs NGBR	Y	N
2022	This Thesis -	Scalable Priority	N	Y	Y	Y	Y	Y	Y	Multi	Y	Y

Note:- Energy Efficient(EE), Throughput Efficient (TE), Resource Utilization (RU), Delay Budget (DB), Priority (PR), Balance in Performance and QoS (BPQ).

- The delay-budget constraint restricts scaling. The proposed scheme implements a virtual QoS class identifier (vQCI) concept to scale the application priority.
- The proposed scheme balances resource utilization and application priority support in resource allocation.
- The results show that the proposed scheduling scheme performs better than state-of-the-art schemes concerning resource sharing fairness, QCI priority support, and delay budget violation.
- Moreover, the proposed scheduling algorithm performs satisfactorily in resource utilization and average cell throughput.
- The proposed scheduling scheme performs satisfactorily in our scenario. However, the proposed scheduling scheme is not suitable for a small number of QoS classes and devices.

5 | Thesis Summary

"Science is not only a disciple of reason but also one of romance and passion."

Stephen Hawking

In this chapter, the primary outcomes of the thesis are encapsulated. Further, a few research directions are provided that may be studied in future research work.

5.1 Conclusion

This thesis provides a foundation for the M2M communication in chapter 1 and LTE radio resource scheduling mechanism in chapter 2. We found some research gaps for M2M communication in LTE / LTE-A cellular networks based on the literature review in section 2.2 as described in section 2.3.1 and addressed these research gaps through chapter 3 to chapter 4.

Chapter 3

Issue If malfunction or intruder devices are present in the network, it can claim the wrong application priority, and the M2M system can be unstable. Is it feasible to schedule resources based on the priority concerning system performance? Because the focus on application priority degrades the system's performance.

Solution We have constructed a game-theoretic model to stabilise the M2M system with the malfunctioning/intruder devices. We have used the VCG combinatorial auction game to provide stability, and we penalize the devices by the number of transmission opportunities. Further, we integrate the constructed game theoretic model with a priority-based resource allocation mechanism to investigate the effect of priority-based resource allocation on the system/s performance parameters like throughput and resource utilization.

Result The implemented priority-based algorithm performs better in application priority and delay-budget support and stability of the M2M system with intruder/malfunctioning devices. But it lacks in the system's performance. There is a trade-off between QoS support and system performance.

Lacks The priority-based resource allocation mechanism degrade system performance. Deciding the exact number of transmission chances and time intervals is challenging.

Chapter 4

Issue As discussed in chapter 1, M2M communication has many devices and a vast range of QoS classes. Therefore, static priority allocation leads to starvation for devices with poor channel conditions and lower priority.

We observe from the results in chapter 3 that channel quality must be used as a device selection metric in resource scheduling to improve resource utilization and throughput.

The priority-based algorithm performs below average in the fair distribution of resources among the devices.

The priority-based algorithm performs well for delay budget violations for seven QCIs but does not perform well for many QCIs.

Solution A weighted-priority proportional fair (WPPF) scheduling scheme based on the virtual priority is designed in chapter 4 to improve resource utilization, resource utilization, and proportional fairness to support many devices with mentioned challenges. The proposed mechanism uses application priority as a weight to channel quality and scales down the weighted priority for each scheduling grant. Scaling of the priority is done proportionally fair concerning the total number of MTCDs and QCIs. The delay budget of the device restricts the scaling of priority if the delay budget of the device is about to time out.

Result The results show that the proposed algorithm performs best in delay budget satisfaction and better in QCI priority support and fairness and minimizes starvation.

Using a combination of channel quality and QCI priority improves the proposed algorithm's resource utilization and average cell throughput performance with application priority support.

The proposed WPPF algorithm provides better results for many QoS classes regarding QCI priority and the delay-budget violation.

The proposed WPPF algorithm perfectly balanced the system performance and QoS support for M2M communication in a heterogeneous environment.

Lacks The proposed scheduling scheme performs satisfactorily in our scenario. However, the proposed scheduling scheme may provide similar results to other implemented scheduling schemes for a small number of QoS classes and devices.

5.2 Application of Proposed Work

The M2M system is a heterogeneous network with many applications running on many devices with diverse QoS requirements. When the heterogeneity applications increase, a scheduling scheme may fail on one parameter while focusing on another parameter. For example, while focusing on application priority, the scheduling scheme may degrade resource utilization or vice versa. The proposed scheduling scheme in this thesis best fits in a heterogeneous environment, i.e., Smart-City scenario. The smart city scenario has many applications related to home automation, smart buildings, smart grid, intelligent transportation, traffic management, retail, logistics, etc.

5.3 Pertinence of the Proposed Solution towards 5G

We can directly apply the proposed solution with higher frequencies if the available resources are shared using orthogonal multiple access (OMA). A single resource of orthogonal multiple access (OMA) can be shared among more than one device in the power domain in NOMA [186, 187]. If the distribution of resources uses the power domain, the proposed solution needs significant improvement before implementation. The proposed solution is not compatible with slot aggregation. However, it supports mini slots and heavy DL and UL transmissions with OMA distribution. Consequently, the proposed solution can be applied in 5G NR if the OMA is used.

5.4 Future Research Directions

However in the last decade, extensive research has been carried out by industry and academia in the field of radio resource scheduling in LTE networks. Researchers have proposed various techniques to improve the efficiency of the scheduling process and optimize the radio resources.

There is a scope for improvement in the proposed WPPF resource allocation schemes for Non-Orthogonal Multiple Access (NOMA) technologies, which provide different power labels in resource allocation. NOMA is suitable for devices having various power labels and the co-existence of H2H and M2M communication.

Massive-multiple Input Multiple Output (massive-MIMO) is a futuristic technology; the MIMO communication model must optimize resource allocation.

Although the QoS classes are well-defined in the literature, they all are statically defined. So, the resource allocation methodologies can be improved to support the dynamic QoS classes.

Bibliography

[1] S. Asif, "5g mobile communications- concepts and technologies," in *5G Mobile Communications- Concepts and Technologies*, S. Asif, Ed., Boca Raton: CRC Press, 2018, Saad Asif, ISBN: 978-0-12-385489-6. DOI: https://doi.org/10.1201/9780429466342. [Online]. Available: https://www.taylorfrancis.com/books/9780429466342.

[2] J. Kim, J. Lee, J. Kim, and J. Yun, "M2m service platforms: Survey, issues, and enabling technologies," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 1, pp. 61–76, 2013.

[3] C. Anton-Haro and M. Dohler, "1 - introduction to machine-to-machine (m2m) communications," in *Machine-to-machine (M2M) Communications*, C. Anton-Haro and M. Dohler, Eds., Oxford: Woodhead Publishing, 2015, pp. 1–23, ISBN: 978-1-78242-102-3. DOI: https://doi.org/10.1016/B978-1-78242-102-3.00001-0. [Online]. Available: http://www.sciencedirect.com/science/article/pii/B97-81782421-023000010.

[4] *Machine-to-machine (m2m) communications*, accessed on 9 Dec, 2022. [Online]. Available: https://www.gartner.com/en/information-technology/glossary/machine-to-machine-m2m-communications.

[5] *Machine to machine communications*, accessed on 9 Dec, 2022. [Online]. Available: https://dot.gov.in/machine-machine-communications.

[6] *Machine to machine communications (m2m); use cases of automotive applications in m2m capable networks*, accessed on 10 August 2020, 2013. [Online]. Available: https://www.etsi.org/deliver/etsi_tr/102800_102899/102898/01.01.01_60/tr_102898v010101p.pdf.

[7] F. Ghavimi and H.-H. Chen, "M2m communications in 3gpp lte/lte-a networks: Architectures, service requirements, challenges, and applications," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 2, pp. 525–549, 2014.

[8] M. Chen, J. Wan, S. Gonzalez, X. Liao, and V. C. Leung, "A survey of recent developments in home m2m networks," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 1, pp. 98–114, 2013.

[9] H. Wang and A. O. Fapojuwo, "A survey of enabling technologies of low power and long range machine-to-machine communications," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 4, pp. 2621–2639, 2017.

[10] M. Chen, J. Wan, and F. Li, "Machine-to-machine communications: Architectures, standards and applications.," *Ksii transactions on internet & information systems*, vol. 6, no. 2, 2012.

[11] M. Y. Abdelsadek, Y. Gadallah, and M. H. Ahmed, "An lte-based optimal resource allocation scheme for delay-sensitive m2m deployments coexistent with h2h users," in *2017 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, IEEE, 2017, pp. 139–144.

[12] M. D. Abrignani, L. Giupponi, A. Lodi, and R. Verdone, "Scheduling m2m traffic over lte uplink of a dense small cells network," in *2015 International*

Symposium on Wireless Communication Systems (ISWCS), IEEE, 2015, pp. 760–764.

[13] H. S. Jang, S. M. Kim, H.-S. Park, and D. K. Sung, "Message-embedded random access for cellular m2m communications," *IEEE Communications Letters*, vol. 20, no. 5, pp. 902–905, 2016.

[14] E. Soltanmohammadi, K. Ghavami, and M. Naraghi-Pour, "A survey of traffic issues in machine-to-machine communications over lte," *IEEE Internet of Things Journal*, vol. 3, no. 6, pp. 865–884, 2016.

[15] N. Afrin, J. Brown, and J. Y. Khan, "A delay sensitive lte uplink packet scheduler for m2m traffic," in *2013 IEEE Globecom Workshops (GC Wkshps)*, IEEE, 2013, pp. 941–946.

[16] T. P. de Andrade, C. A. Astudillo, and N. L. da Fonseca, "Allocation of control resources for machine-to-machine and human-to-human communications over lte/lte-a networks," *IEEE Internet of Things Journal*, vol. 3, no. 3, pp. 366–377, 2016.

[17] N. Kumar, *The evolution of m2m technology — from m2m to iot*, accessed on 10 August 2020. [Online]. Available: https://medium.com/plasma-business-intelligence/the-evolution-of-m2m-techno-logy-from-m2m-to-iot.

[18] *M2m history begins in 1885*, accessed on 10 August 2020. [Online]. Available: https://nimbelink.com/blog/m2m-history-begins-in-1885/.

[19] *What is m2m – a history*, accessed on 10 August 2020. [Online]. Available: https://nimbelink.com/blog/m2m-history-begins-in-1885/.

[20] S. Chen, R. Ma, H.-H. Chen, H. Zhang, W. Meng, and J. Liu, "Machine-to-machine communications in ultra-dense networks - a survey," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 3, pp. 1478–1503, 2017.

[21] A. Nordrum, *Popular internet of things forecast of 50 billion devices by 2020 is outdated*, accessed on 10 August 2020. [Online]. Available: https://spectrum.ieee.org/tech-talk/telecom/internet/popular-internet-of-things-forecast-of-50-billion-devices-by-2020-is-outdated.

[22] *M2m connections market - growth, trends, and forecast (2020 - 2025)*, accessed on 10 August 2020. [Online]. Available: https://www.mordorintelligence.com/industry-reports/m2m-connections-market.

[23] D. Evans, "The internet of things: How the next evolution of the internet is changing everything," *CISCO white paper*, vol. 1, no. 2011, pp. 1–11, 2011.

[24] *Internet of things (iot) connected devices installed base worldwide from 2015 to 2025*, accessed on 10 August 2020. [Online]. Available: https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide/.

[25] A. Laya, K. Wang, L. Alonso, J. Alonso-Zarate, and J. Markendahl, "7 - supporting machine-to-machine communications in long-term evolution networks," in *Machine-to-machine (M2M) Communications*, C. Anton-Haro and M. Dohler, Eds., Oxford: Woodhead Publishing, 2015, pp. 109–129, ISBN: 978-1-78242-102-3. DOI: https://doi.org/10.1016/B978-1-78242-102-3.00007-1. [Online]. Available: http://www.sciencedirect.com/science/article/pii/B9781782421023000071.

[26] P. Martigne, "2 - overview of etsi machine-to-machine and onem2m architectures," in *Machine-to-machine (M2M) Communications*, C. Anton-Haro

and M. Dohler, Eds., Oxford: Woodhead Publishing, 2015, pp. 27–46, ISBN: 978-1-78242-102-3. DOI: https://doi.org/10.1016/B978-1-78242-102-3.00002-2. [Online]. Available: http://www.sciencedirect.com/science/article/pii/B9781782421023000022.

[27] P. Bhat and M. Dohler, "3 - overview of 3gpp machine-type communication standardization," in *Machine-to-machine (M2M) Communications*, C. Anton-Haro and M. Dohler, Eds., Oxford: Woodhead Publishing, 2015, pp. 47–62, ISBN: 978-1-78242-102-3. DOI: https://doi.org/10.1016/B978-1-78242-102-3.00003-4. [Online]. Available: http://www.sciencedirect.com/science/article/pii/B9781782421023000034.

[28] J. H. Huh, D. H. Kim, and J. Deok Kim, "Onem2m : Extension of protocol binding: Reuse of binding protocol's legacy services," in *2016 International Conference on Information Networking (ICOIN)*, 2016, pp. 363–365. DOI: 10.1109/ICOIN.2016.7427133.

[29] J. Brown, N. Afrin, and J. Y. Khan, "Delay models for static and adaptive persistent resource allocations in wireless systems," *IEEE Transactions on Mobile Computing*, vol. 15, no. 9, pp. 2193–2205, 2015.

[30] D. T. Wiriaatmadja and K. W. Choi, "Hybrid random access and data transmission protocol for machine-to-machine communications in cellular networks," *IEEE Transactions on Wireless Communications*, vol. 14, no. 1, pp. 33–46, 2014.

[31] S. Xu, J. Li, and X. Qin, "Grouping based uplink resource allocation for massive m2m communications over lte-a," in *2017 IEEE 86th Vehicular Technology Conference (VTC-Fall)*, IEEE, 2017, pp. 1–5.

[32] Y.-H. Hsu, K. Wang, and Y.-C. Tseng, "Enhanced cooperative access class barring and traffic adaptive radio resource management for m2m communications over lte-a," in *2013 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, IEEE, 2013, pp. 1–6.

[33] H. Zhang, B. Di, K. Bian, and L. Song, "Iot-u: Cellular internet-of-things networks over unlicensed spectrum," *IEEE Transactions on Wireless Communications*, vol. 18, no. 5, pp. 2477–2492, 2019.

[34] *Etsi m2m solution introduction : Etsi standard is enabling global m2m communications*, accessed on 10 August 2020. [Online]. Available: https://www.etsi.org/images/files/Events/2014/201405_DGCONNECT_SmartM2MAppliances/ETSI_M2M_introduction_main.pdf.

[35] *Machine-to-machine communications (m2m) : Functional architecture*, accessed on 10 August 2020, 2013-16. [Online]. Available: https://www.etsi.org/deliver/etsi_ts/102600_102699/102690/01.02.01_60/ts_102690v010201p.pdf.

[36] Y. Alaa, H. M. ElAttar, F. Digham, L. H. Afify, and H. Elbadawy, "Lte dynamic scheduling scheme for massive m2m and h2h communication," in *2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON)*, IEEE, 2017, pp. 478–482.

[37] M. Al-Rawi, R. Jantti, J. Torsner, and M. Sagfors, "On the performance of heuristic opportunistic scheduling in the uplink of 3g lte networks," in *2008 IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications*, IEEE, 2008, pp. 1–6.

[38] *Digital cellular telecommunications system (phase 2+)*, accessed on 10 August 2020, 2018. [Online]. Available: https://www.etsi.org/deliver/etsi_ts/123000_123099/123032/15.01.00_60/ts_123032v150100p.pdf.

[39] C. Troncoso, G. Danezis, E. Kosta, J. Balasch, and B. Preneel, "Pripayd: Privacy-friendly pay-as-you-drive insurance," *IEEE Transactions on Dependable and Secure Computing*, vol. 8, no. 5, pp. 742–755, 2010.

[40] S. M. Alavi and C. Zhou, "Resource allocation scheme for orthogonal frequency division multiple access networks based on cooperative game theory," *International Journal of Communication Systems*, vol. 27, no. 8, pp. 1105–1125, 2014.

[41] M. Dohler and T. Nakamura, *5G Mobile and Wireless Communications Technology*. Cambridge University Press, 2016. DOI: 10.1017/CBO9781316417744.

[42] A. Barki, A. Bouabdallah, S. Gharout, and J. Traore, "M2m security: Challenges and solutions," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 1241–1254, 2016.

[43] S. Han, S. Han, and K. Sezaki, "Estimation of achievable power capacity from plug-in electric vehicles for v2g frequency regulation: Case studies for market participation," *IEEE Transactions on Smart Grid*, vol. 2, no. 4, pp. 632–641, 2011.

[44] Z. M. Fadlullah, M. M. Fouda, N. Kato, A. Takeuchi, N. Iwasaki, and Y. Nozaki, "Toward intelligent machine-to-machine communications in smart grid," *IEEE Communications Magazine*, vol. 49, no. 4, pp. 60–65, 2011.

[45] S. A. AlQahtani and M. AlHassany, "Performance modeling and evaluation of novel scheduling algorithm for lte networks," in *2013 IEEE 12th International Symposium on Network Computing and Applications*, IEEE, 2013, pp. 101–105.

[46] I. M. Ozcelik, I. Korpeoglu, and A. Agrawala, *Energy efficient ip-connectivity with ieee 802.11 for home m2m networks*, 2017.

[47] R. Liu, W. Wu, H. Zhu, and D. Yang, “M2m-oriented qos categorization in cellular network,” in *2011 7th International Conference on Wireless Communications, Networking and Mobile Computing*, IEEE, 2011, pp. 1–5.

[48] S. Dawaliby, A. Bradai, Y. Pousset, and C. Chatellier, “Joint energy and qos-aware memetic-based scheduling for m2m communications in lte-m,” *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 3, no. 3, pp. 217–229, 2018.

[49] F. Ghavimi, Y.-W. Lu, and H.-H. Chen, “Uplink scheduling and power allocation for m2m communications in sc-fdma-based lte-a networks with qos guarantees,” *IEEE Transactions on Vehicular Technology*, vol. 66, no. 7, pp. 6160–6170, 2016.

[50] Q.-V. Pham *et al.*, “A survey of multi-access edge computing in 5g and beyond: Fundamentals, technology integration, and state-of-the-art,” *IEEE Access*, vol. 8, pp. 116 974–117 017, 2020.

[51] N. Hassan, K.-L. A. Yau, and C. Wu, “Edge computing in 5g: A review,” *IEEE Access*, vol. 7, pp. 127 276–127 289, 2019.

[52] A. M. Maia, M. F. de Castro, and D. Vieira, “A dynamic lte uplink packet scheduler for machine-to-machine communication,” in *2014 IEEE 25th annual international symposium on personal, indoor, and mobile radio communication (PIMRC)*, IEEE, 2014, pp. 1609–1614.

[53] M. A. Mehaseb, Y. Gadallah, A. Elhamy, and H. Elhennawy, “Classification of lte uplink scheduling techniques: An m2m perspective,” *IEEE Communications surveys & tutorials*, vol. 18, no. 2, pp. 1310–1335, 2015.

[54] M. Shafi *et al.*, “5g: A tutorial overview of standards, trials, challenges, deployment, and practice,” *IEEE journal on selected areas in communications*, vol. 35, no. 6, pp. 1201–1221, 2017.

[55] K. Shafique, B. A. Khawaja, F. Sabir, S. Qazi, and M. Mustaqim, “Internet of things (iot) for next-generation smart systems: A review of current challenges, future trends and prospects for emerging 5g-iot scenarios,” *Ieee Access*, vol. 8, pp. 23 022–23 040, 2020.

[56] J. Brown and J. Y. Khan, “A predictive resource allocation algorithm in the lte uplink for event based m2m applications,” *IEEE Transactions on Mobile Computing*, vol. 14, no. 12, pp. 2433–2446, 2015.

[57] F. Morvari and A. Ghasemi, “Two-stage resource allocation for random access m2m communications in lte network,” *IEEE Communications Letters*, vol. 20, no. 5, pp. 982–985, 2016.

[58] M. Rumney *et al.*, *LTE and the evolution to 4G wireless: Design and measurement challenges*. John Wiley & Sons, 2013.

[59] S. Dawaliby, A. Bradai, and Y. Pousset, “Scheduling optimization for m2m communications in lte-m,” in *2017 IEEE International Conference on Consumer Electronics (ICCE)*, IEEE, 2017, pp. 126–128.

[60] O. Delgado and B. Jaumard, “Scheduling and resource allocation in lte uplink with a delay requirement,” in *2010 8th Annual Communication Networks and Services Research Conference*, IEEE, 2010, pp. 268–275.

[61] L. F. Ibrahim, H. A. Salman, Z. F. Taha, N. Akkari, G. Aldabbagh, and O. Bamasak, “A survey on heterogeneous mobile networks planning in indoor dense areas,” *Personal and Ubiquitous Computing*, vol. 24, no. 4, pp. 487–498, 2020.

[62] K. Edemacu and T. Bulega, “Resource sharing between m2m and h2h traffic under time-controlled scheduling scheme in lte networks,” in *2014 8th International Conference on Telecommunication Systems Services and Applications (TSSA)*, IEEE, 2014, pp. 1–6.

[63] A. Elhamy and Y. Gadallah, “Bat: A balanced alternating technique for m2m uplink scheduling over lte,” in *2015 IEEE 81st Vehicular Technology Conference (VTC Spring)*, IEEE, 2015, pp. 1–6.

[64] D. Fagan, M. Fenton, D. Lynch, S. Kucera, H. Claussen, and M. O’Neill, “Deep learning through evolution: A hybrid approach to scheduling in a dynamic environment,” in *2017 International Joint Conference on Neural Networks (IJCNN)*, IEEE, 2017, pp. 775–782.

[65] P. Frank, A. Müller, H. Droste, and J. Speidel, “Cooperative interference-aware joint scheduling for the 3gpp lte uplink,” in *21st Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, IEEE, 2010, pp. 2216–2221.

[66] M. K. Giluka, N. S. Kumar, N. Rajoria, and B. R. Tamma, “Class based priority scheduling to support machine to machine communications in lte systems,” in *2014 Twentieth National Conference on Communications (NCC)*, IEEE, 2014, pp. 1–6.

[67] A. Ahmad, S. Ahmad, M. H. Rehmani, and N. U. Hassan, “A survey on radio resource allocation in cognitive radio sensor networks,” *IEEE Communications Surveys & Tutorials*, vol. 17, no. 2, pp. 888–917, 2015.

[68] A. G. Gotsis, A. S. Lioumpas, and A. Alexiou, “Evolution of packet scheduling for machine-type communications over lte: Algorithmic design and performance analysis,” in *2012 IEEE Globecom Workshops*, IEEE, 2012, pp. 1620–1625.

[69] S. Hamdoun, A. Rachedi, H. Tembine, and Y. Ghamri-Doudane, “Efficient transmission strategy selection algorithm for m2m communications: An evolutionary game approach,” in *2016 IEEE 15th International Symposium on Network Computing and Applications (NCA)*, IEEE, 2016, pp. 286–293.

[70] A. Hassebo, A. Rezk, and M. A. Ali, “A hybrid uplink scheduling approach for supporting mission-critical smart grid applications in commercial 4g cellular networks,” in *2018 9th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, IEEE, 2018, pp. 303–308.

[71] W. Qiu, P. Tarasak, and H. Minn, “Orthogonal multicarrier division multiple access for multipoint-to-multipoint networks,” *IEEE transactions on communications*, vol. 61, no. 9, pp. 3841–3853, 2013.

[72] C. Y. Ho and C.-Y. Huang, “Energy-saving massive access control and resource allocation schemes for m2m communications in ofdma cellular networks,” *IEEE Wireless Communications Letters*, vol. 1, no. 3, pp. 209–212, 2012.

[73] M. Kalil, J Samarabandu, A. Shami, and A. Al-Dweik, “Performance evaluation of genetic algorithms for resource scheduling in tle uplink,” in *2014 International Wireless Communications and Mobile Computing Conference (IWCMC)*, IEEE, 2014, pp. 948–952.

[74] K. Yang, S. Martin, and T. A. Yahiya, “Interference aware resource allocation for lte uplink transmission,” in *2014 IEEE Symposium on Computers and Communications (ISCC)*, IEEE, 2014, pp. 1–6.

[75] K. K. S. Bhardwaj and H. Venkataraman, “Different network coding techniques for multihop machine-to-machine (m2m) communication in next generation smart environment,” in *2017 IEEE International Conference on Advanced Intelligent Mechatronics (AIM)*, IEEE, 2017, pp. 142–147.

[76] G. Karadag, R. Gul, Y. Sadi, and S. C. Ergen, “Qos-constrained semi-persistent scheduling of machine-type communications in cellular networks,” *IEEE Transactions on Wireless Communications*, vol. 18, no. 5, pp. 2737–2750, 2019.

[77] Q. Cui *et al.*, “A unified protocol stack solution for lte and wlan in future mobile converged networks,” *IEEE wireless communications*, vol. 21, no. 6, pp. 24–33, 2014.

[78] A. Kumar, A. Abdelhadi, and C. Clancy, “A delay efficient multiclass packet scheduler for heterogeneous m2m uplink,” in *MILCOM 2016-2016 IEEE Military Communications Conference*, IEEE, 2016, pp. 313–318.

[79] A. Kumar, A. Abdelhadi, and C. Clancy, “A delay optimal multiclass packet scheduler for general m2m uplink,” *IEEE Systems Journal*, vol. 13, no. 4, pp. 3815–3826, 2019.

[80] R. Kwan, C. Leung, and J. Zhang, “Proportional fair multiuser scheduling in lte,” *IEEE Signal Processing Letters*, vol. 16, no. 6, pp. 461–464, 2009.

[81] M. D. Abrignani, L. Giupponi, A. Lodi, and R. Verdone, “Scheduling m2m traffic over lte uplink of a dense small cell network,” *EURASIP Journal on Wireless Communications and Networking*, vol. 2018, no. 1, p. 193, 2018.

[82] S. A. AlQahtani, “Analysis and modelling of power consumption-aware priority-based scheduling for m2m data aggregation over long-term-evolution networks,” *IET Communications*, vol. 11, no. 2, pp. 177–184, 2017.

[83] M.-R. Hojeij, C. Abdel Nour, J. Farah, and C. Douillard, “Weighted proportional fair scheduling for downlink nonorthogonal multiple access,” *Wireless Communications and Mobile Computing*, vol. 2018, 2018.

[84] A. Ghosh, J. Zhang, J. G. Andrews, and R. Muhamed, *Fundamentals of LTE*. Pearson Education, 2010.

[85] C. Han and S. Armour, “Computational complexity and energy consumption analysis of dynamic resource scheduling algorithms for lte,” in *2012 IEEE 75th Vehicular Technology Conference (VTC Spring)*, IEEE, 2012, pp. 1–5.

[86] Q. Li, Y. Ge, Y. Yang, Y. Zhu, W. Sun, and J. Li, "An energy efficient uplink scheduling and resource allocation for m2m communications in sc-fdma based lte-a networks," *Mobile Networks and Applications*, pp. 1–12, 2019.

[87] Y. Lin and G. Yue, "Channel-adapted and buffer-aware packet scheduling in lte wireless communication system," in *2008 4th International Conference on Wireless Communications, Networking and Mobile Computing*, IEEE, 2008, pp. 1–4.

[88] N. Li, C. Cao, and C. Wang, "Dynamic resource allocation and access class barring scheme for delay-sensitive devices in machine to machine (m2m) communications," *Sensors*, vol. 17, no. 6, p. 1407, 2017.

[89] Z. Dawy, W. Saad, A. Ghosh, J. G. Andrews, and E. Yaacoub, "Toward massive machine type cellular communications," *IEEE Wireless Communications*, vol. 24, no. 1, pp. 120–128, 2016.

[90] P. Rawat, K. D. Singh, and J. M. Bonnin, "Cognitive radio for m2m and internet of things: A survey," *Computer Communications*, vol. 94, pp. 1–29, 2016.

[91] J.-J. Chen, J.-M. Liang, and Z.-Y. Chen, "Energy-efficient uplink radio resource management in lte-advanced relay networks for internet of things," in *2014 international wireless communications and mobile computing conference (IWCMC)*, IEEE, 2014, pp. 745–750.

[92] E. Dahlman, S. Parkvall, and J. Skold, "Chapter 13 - power control, scheduling, and interference handling," in *4G LTE/LTE-Advanced for Mobile Broadband*, E. Dahlman, S. Parkvall, and J. Skold, Eds., Oxford: Academic Press, 2011, pp. 265–299, ISBN: 978-0-12-385489-6. DOI: https://doi.org/10.1016/B978-0-12-385489-6.00013-8. [Online]. Available: http://www.sciencedirect.com/science/article/pii/B9780123854896000138.

[93] H. B. Rekhissa, C. Belleudy, and P. Bessaguet, "Energy efficient resource allocation for m2m devices in lte/lte-a," *Sensors*, vol. 19, no. 24, p. 5337, 2019.

[94] Z. Tagarian and B. Shahgholi Ghahfarokhi, "Energy and delay aware massive access management in machine-to-machine communications," *Transactions on Emerging Telecommunications Technologies*, vol. 30, no. 10, e3618, 2019.

[95] J. Wu, E. W. M. Wong, Y.-C. Chan, and M. Zukerman, "Energy efficiency-qos tradeoff in cellular networks with base-station sleeping," in *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, 2017, pp. 1–7. DOI: 10.1109/GLOCOM.2017.8254042.

[96] U. U. Khan, N. Dilshad, M. H. Rehmani, and T. Umer, "Fairness in cognitive radio networks: Models, measurement methods, applications, and future research directions," *Journal of Network and Computer Applications*, vol. 73, pp. 12–26, 2016.

[97] S. Hwang and S. Park, "On the effects of resource usage ratio on data rate in lte systems," in *2017 19th International Conference on Advanced Communication Technology (ICACT)*, IEEE, 2017, pp. 78–80.

[98] M. R. Mardani, S. Mohebi, and H. Bobarshad, "Robust uplink resource allocation in lte networks with m2m devices as an infrastructure of internet of things," in *2016 IEEE 4th International Conference on Future Internet of Things and Cloud (FiCloud)*, IEEE, 2016, pp. 186–193.

[99] M. R. Mardani, S. Mohebi, B. Maham, and M. Bennis, "Delay-sensitive resource allocation for relay-aided m2m communication over lte-advanced networks," in *2017 IEEE symposium on computers and communications (ISCC)*, IEEE, 2017, pp. 1033–1038.

[100] M. Ouaissa and A. Rhattoy, "Qos hybrid uplink scheduler based on service type for m2m communications in lte networks," *Indones. J. Electr. Eng. Comput. Sci*, vol. 14, pp. 1460–1470, 2019.

[101] J. W. Raymond, T. O. Olwal, and A. M. Kurien, "Cooperative communications in machine to machine (m2m): Solutions, challenges and future work," *IEEE Access*, vol. 6, pp. 9750–9766, 2018.

[102] S.-B. Lee, I. Pefkianakis, A. Meyerson, S. Xu, and S. Lu, "Proportional fair frequency-domain packet scheduling for 3gpp lte uplink," in *IEEE INFOCOM 2009*, IEEE, 2009, pp. 2611–2615.

[103] M. Lauridsen, L. Noël, T. B. Sørensen, and P. Mogensen, "An empirical lte smartphone power model with a view to energy efficiency evolution.," *Intel Technology Journal*, vol. 18, no. 1, 2014.

[104] E. Dahlman, S. Parkvall, and J. Skold, "Chapter 6 - scheduling, link adaptation, and hybrid arq," in *4G LTE/LTE-Advanced for Mobile Broadband*, E. Dahlman, S. Parkvall, and J. Skold, Eds., Oxford: Academic Press, 2011, pp. 79–93, ISBN: 978-0-12-385489-6. DOI: https://doi.org/10.1016/B978-0-12-385489-6.00006-0. [Online]. Available: http://www.sciencedirect.com/science/article/pii/B9780123854896000060.

[105] M.-J. Shih, G.-Y. Lin, and H.-Y. Wei, "Two paradigms in cellular internet-of-things access for energy-harvesting machine-to-machine devices: Push-based versus pull-based," *IET Wireless Sensor Systems*, vol. 6, no. 4, pp. 121–129, 2016.

[106] Z. Yang, W. Xu, H. Xu, J. Shi, and M. Chen, "Energy efficient non-orthogonal multiple access for machine-to-machine communications," *IEEE communications letters*, vol. 21, no. 4, pp. 817–820, 2016.

[107] A. Azari and G. Miao, "Network lifetime maximization for cellular-based m2m networks," *IEEE Access*, vol. 5, pp. 18 927–18 940, 2017.

[108] P.-C. Shen, S.-R. Yang, and P. Lin, "Energy-efficient scheduling algorithms for real-time data reporting in 4g lte machine-to-machine communication

networks," in *2016 International Wireless Communications and Mobile Computing Conference (IWCMC)*, IEEE, 2016, pp. 1057–1062.

[109] A. Azari and G. Miao, "Lifetime-aware scheduling and power control for m2m communications in lte networks," in *2015 IEEE Wireless Communications and Networking Conference (WCNC)*, IEEE, 2015, pp. 1171–1176.

[110] K. Wang, J. Alonso-Zarate, and M. Dohler, "Energy-efficiency of lte for small data machine-to-machine communications," in *2013 IEEE International Conference on Communications (ICC)*, IEEE, 2013, pp. 4120–4124.

[111] E. Yaacoub, H. Al-Asadi, and Z. Dawy, "Low complexity scheduling algorithms for the lte uplink," in *2009 IEEE Symposium on Computers and Communications*, IEEE, 2009, pp. 266–270.

[112] G. Wang and T. Liu, "Resource allocation for m2m-enabled cellular network using nash bargaining game theory," *Peer-to-Peer Networking and Applications*, vol. 11, no. 1, pp. 110–123, 2018.

[113] H. Safdar *et al.*, "Resource allocation for uplink m2m communication: A game theory approach," in *2013 IEEE Symposium on Wireless Technology & Applications (ISWTA)*, IEEE, 2013, pp. 48–52.

[114] H.-W. Tseng, Y.-H. Lee, C.-Y. Lo, L.-Y. Yen, and Y.-G. Jan, "Applying genetic algorithms to the data traffic scheduling and performance analysis of a long-term evolution system," *Wireless Personal Communications*, vol. 81, no. 1, pp. 387–403, 2015.

[115] N. Zhang, G. Kang, J. Wang, Y. Guo, and F. Labeau, "Resource allocation in a new random access for m2m communications," *IEEE Communications Letters*, vol. 19, no. 5, pp. 843–846, 2015.

[116] H. S. Dhillon, H. Huang, H. Viswanathan, and R. A. Valenzuela, "Fundamentals of throughput maximization with random arrivals for m2m communications," *IEEE Transactions on Communications*, vol. 62, no. 11, pp. 4094–4109, 2014.

[117] P. Rysavy, "Challenges and considerations in defining spectrum efficiency," *Proceedings of the IEEE*, vol. 102, no. 3, pp. 386–392, 2014.

[118] A Merwaday and I Güvenç, "Optimisation of feicic for energy efficiency and spectrum efficiency in lte-advanced hetnets," *Electronics Letters*, vol. 52, no. 11, pp. 982–984, 2016.

[119] H. S. Jang, B. C. Jung, and D. K. Sung, "Dynamic access control with resource limitation for group paging-based cellular iot systems," *IEEE Internet of Things Journal*, vol. 5, no. 6, pp. 5065–5075, 2018.

[120] S. Duan, V. Shah-Mansouri, Z. Wang, and V. W. Wong, "D-acb: Adaptive congestion control algorithm for bursty m2m traffic in lte networks," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 12, pp. 9847–9861, 2016.

[121] F. Hussain, R. Hussain, A. Anpalagan, and A. Benslimane, "A new block-based reinforcement learning approach for distributed resource allocation in clustered iot networks," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 3, pp. 2891–2904, 2020.

[122] A. Aijaz and A. H. Aghvami, "Cognitive machine-to-machine communications for internet-of-things: A protocol stack perspective," *IEEE Internet of Things Journal*, vol. 2, no. 2, pp. 103–112, 2015.

[123] S. Shi, C. Feng, and C. Guo, "A resource scheduling algorithm based on user grouping for lte-advanced system with carrier aggregation," in *2009 International Symposium on Computer Network and Multimedia Technology*, IEEE, 2009, pp. 1–4.

[124] V. Rodoplu, M. Nakıp, D. T. Eliiyi, and C. Güzelis, "A multi-scale algorithm for joint forecasting-scheduling to solve the massive access problem of iot," *IEEE Internet of Things Journal*, 2020.

[125] J. Liu, L. Song, *et al.*, "A novel congestion reduction scheme for massive machine-to-machine communication," *IEEE Access*, vol. 5, pp. 18 765–18 777, 2017.

[126] S. Bayat, Y. Li, Z. Han, M. Dohler, and B. Vucetic, "Distributed data aggregation in machine-to-machine communication networks based on coalitional game," in *2014 IEEE Wireless Communications and Networking Conference (WCNC)*, IEEE, 2014, pp. 2026–2031.

[127] A. A. Dowhuszko, G. Corral-Briones, J. Hämäläinen, and R. Wichman, "Performance of quantized random beamforming in delay-tolerant machine-type communication," *IEEE Transactions on Wireless Communications*, vol. 15, no. 8, pp. 5664–5680, 2016.

[128] W. H. Fu, Q. L. Kong, W. X. Tian, C. Wang, and L. L. Ma, "A qos-aware scheduling algorithm based on service type for lte downlink," in *Applied Mechanics and Materials*, Trans Tech Publ, vol. 347, 2013, pp. 2468–2473.

[129] W. U. Rehman, T. Salam, A. Almogren, K. Haseeb, I. U. Din, and S. H. Bouk, "Improved resource allocation in 5g mtc networks," *IEEE Access*, vol. 8, pp. 49 187–49 197, 2020.

[130] A. E. Mostafa and Y. Gadallah, "A statistical priority-based scheduling metric for m2m communications in lte networks," *IEEE Access*, vol. 5, pp. 8106–8117, 2017.

[131] H. Safa, W. El-Hajj, and K. Tohme, "A qos-aware uplink scheduling paradigm for lte networks," in *2013 IEEE 27th International Conference on Advanced Information Networking and Applications (AINA)*, IEEE, 2013, pp. 1097–1104.

[132] S. H. da Mata and P. R. Guardieiro, “A genetic algorithm based approach for resource allocation in lte uplink,” in *2014 International Telecommunications Symposium (ITS)*, IEEE, 2014, pp. 1–5.

[133] A. M. Maia, D. Vieira, M. F. de Castro, and Y. Ghamri-Doudane, “A mechanism for uplink packet scheduler in lte network in the context of machine-to-machine communication,” in *2014 IEEE Global Communications Conference*, IEEE, 2014, pp. 2776–2782.

[134] N. Afrin, J. Brown, and J. Y. Khan, “An adaptive buffer based semi-persistent scheduling scheme for machine-to-machine communications over lte,” in *2014 Eighth International Conference on Next Generation Mobile Apps, Services and Technologies*, IEEE, 2014, pp. 260–265.

[135] J. Brown and J. Y. Khan, “Predictive allocation of resources in the lte uplink based on maximum likelihood estimation of event propagation characteristics for m2m applications,” in *2014 IEEE Global Communications Conference*, IEEE, 2014, pp. 4965–4970.

[136] A. Aghmadi, I. Bouksim, A. Kobbane, and T. Taleb, “A mtc traffic generation and qci priority-first scheduling algorithm over lte,” in *2015 International Conference on Wireless Networks and Mobile Communications (WINCOM)*, IEEE, 2015, pp. 1–6.

[137] A. M. Maia, D. Vieira, M. F. de Castro, and Y. Ghamri-Doudane, “A fair qos-aware dynamic lte scheduler for machine-to-machine communication,” *Computer Communications*, vol. 89, pp. 75–86, 2016.

[138] T. Erpek, A. Abdelhadi, and T. C. Clancy, “An optimal application-aware resource block scheduling in lte,” in *2015 International Conference on Computing, Networking and Communications (ICNC)*, IEEE, 2015, pp. 275–279.

[139] I. Abdalla and S. Venkatesan, “A qoe preserving m2m-aware hybrid scheduler for lte uplink,” in *2013 international conference on selected topics in mobile and wireless networking (MoWNeT)*, IEEE, 2013, pp. 127–132.

[140] S. A. AlQahtani, "Performance analysis of cognitive-based radio resource allocation in multi-channel lte-a networks with m2m/h2h coexistence," *IET Communications*, vol. 11, no. 5, pp. 655–663, 2016.

[141] A. Aijaz, M. Tshangini, M. R. Nakhai, X. Chu, and A.-H. Aghvami, "Energy-efficient uplink resource allocation in lte networks with m2m/h2h co-existence under statistical qos guarantees," *IEEE Transactions on Communications*, vol. 62, no. 7, pp. 2353–2365, 2014.

[142] I. S. Comsa, S. Zhang, M. Aydin, P. Kuonen, and J.-F. Wagen, "A novel dynamic q-learning-based scheduler technique for lte-advanced technologies using neural networks," in *37th Annual IEEE Conference on Local Computer Networks*, IEEE, 2012, pp. 332–335.

[143] Z.-N. Lin, S.-R. Yang, and P. Lin, "Edge computing-enhanced uplink scheduling for energy-constrained cellular internet of things," in *2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC)*, IEEE, 2019, pp. 1391–1396.

[144] T. Salam, W. U. Rehman, and X. Tao, "Cooperative data aggregation and dynamic resource allocation for massive machine type communication," *IEEE Access*, vol. 6, pp. 4145–4158, 2018.

[145] *What is a smart building?* accessed on 9 Dec, 2022. [Online]. Available: https://www.resonai.com/blog/what-is-a-smart-building.

[146] F. D. Calabrese *et al.*, "Search-tree based uplink channel aware packet scheduling for utran lte," in *VTC Spring 2008-IEEE Vehicular Technology Conference*, IEEE, 2008, pp. 1949–1953.

[147] L. A. M. R. de Temino, G. Berardinelli, S. Frattasi, and P. Mogensen, "Channel-aware scheduling algorithms for sc-fdma in lte uplink," in *2008 IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications*, IEEE, 2008, pp. 1–6.

[148] R. Ruby and V. Leung, "Towards qos assurance with revenue maximization of lte uplink scheduling," in *2011 Ninth Annual Communication Networks and Services Research Conference*, IEEE, 2011, pp. 202–209.

[149] E. Yaacoub and Z. Dawy, "Achieving the nash bargaining solution in ofdma uplink using distributed scheduling with limited feedback," *AEU-International Journal of Electronics and Communications*, vol. 65, no. 4, pp. 320–330, 2011.

[150] S.-Y. Lien, K.-C. Chen, and Y. Lin, "Toward ubiquitous massive accesses in 3gpp machine-to-machine communications," *IEEE Communications Magazine*, vol. 49, no. 4, pp. 66–74, 2011.

[151] A. S. Lioumpas and A. Alexiou, "Uplink scheduling for machine-to-machine communications in lte-based cellular systems," in *2011 IEEE GLOBECOM Workshops (GC Wkshps)*, IEEE, 2011, pp. 353–357.

[152] F. Ren, Y. Xu, H. Yang, J. Zhang, and C. Lin, "Frequency domain packet scheduling with stability analysis for 3gpp lte uplink," *IEEE Transactions on Mobile Computing*, vol. 12, no. 12, pp. 2412–2426, 2012.

[153] A. Afifi, K. M. Elsayed, and A. Khattab, "Interference-aware radio resource management framework for the 3gpp lte uplink with qos constraints," in *2013 IEEE Symposium on Computers and Communications (ISCC)*, IEEE, 2013, pp. 000 693–000 698.

[154] P. Si, J. Yang, S. Chen, and H. Xi, "Adaptive massive access management for qos guarantees in m2m communications," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 7, pp. 3152–3166, 2014.

[155] F. Hussain, A. Anpalagan, and M. Naeem, "Multi-objective mtc device controller resource optimization in m2m communication," in *2014 27th biennial symposium on communications (QBSC)*, IEEE, 2014, pp. 184–188.

[156] X. Xiang, C. Lin, X. Chen, and X. Shen, "Toward optimal admission control and resource allocation for lte-a femtocell uplink," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 7, pp. 3247–3261, 2014.

[157] Y.-K. Liao, C.-H. Wang, D.-N. Yang, and W.-T. Chen, "Uplink scheduling for lte 4g video surveillance system," in *2015 IEEE Wireless Communications and Networking Conference (WCNC)*, IEEE, 2015, pp. 1380–1385.

[158] H.-Y. Hsieh, T.-C. Juan, Y.-D. Tsai, and H.-C. Huang, "Minimizing radio resource usage for machine-to-machine communications through data-centric clustering," *IEEE Transactions on Mobile Computing*, vol. 15, no. 12, pp. 3072–3086, 2016.

[159] N. Kumar, R. Iqbal, S. Misra, and J. J. Rodrigues, "Bayesian coalition game for contention-aware reliable data forwarding in vehicular mobile cloud," *Future Generation Computer Systems*, vol. 48, pp. 60–72, 2015.

[160] N. Kumar, S. Misra, J. J. Rodrigues, and M. S. Obaidat, "Coalition games for spatio-temporal big data in internet of vehicles environment: A comparative analysis," *IEEE Internet of Things Journal*, vol. 2, no. 4, pp. 310–320, 2015.

[161] L. Blumrosen and N. Nisan, "Combinatorial auctions," in *Algorithmic Game Theory*, N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani, Eds. Cambridge University Press, 2007, pp. 267–300. DOI: 10.1017/CBO9780511800481.013.

[162] N. Nisan, "Introduction to mechanism design (for computer scientists)," in *Algorithmic Game Theory*, N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani, Eds. Cambridge University Press, 2007, pp. 209–242. DOI: 10.1017/CBO9780511800481.011.

[163] M. Jofre-Bonet and M. Pesendorfer, "Estimation of a dynamic auction game," *Econometrica*, vol. 71, no. 5, pp. 1443–1489, 2003.

[164] P. Milgrom and P. R. Milgrom, *Putting auction theory to work*. Cambridge University Press, 2004.

[165] M. H. Rothkopf, "Thirteen reasons why the vickrey-clarke-groves process is not practical," *Operations Research*, vol. 55, no. 2, pp. 191–197, 2007.

[166] C. Borgs, J. Chayes, N. Immorlica, M. Mahdian, and A. Saberi, "Multi-unit auctions with budget-constrained bidders," in *Proceedings of the 6th ACM Conference on Electronic Commerce*, 2005, pp. 44–51.

[167] "N-person nonzero sum games with a continuum of strategies," in *Game Theory*. John Wiley and Sons, Ltd, 2011, ch. 4, pp. 163–218, ISBN: 9781118032398. DOI: 10.1002/9781118032398.ch4. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118032398.ch4.

[168] Y. Mehmood, C. Görg, M. Muehleisen, and A. Timm-Giel, "Mobile m2m communication architectures, upcoming challenges, applications, and future directions," *EURASIP Journal on Wireless Communications and Networking*, vol. 2015, no. 1, pp. 1–37, 2015.

[169] M. Laner, N. Nikaein, D Drajic, P. Svoboda, M Popovic, and S Krco, "M2m traffic and models," in *Machine-to-Machine Communications: Architectures, Technology, Standards, and Applications*, Taylor & Francis, 2014.

[170] *Geran improvements for machine-type communications (mtc)*, accessed on 9 Dec, 2022. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=500.

[171] T. Girici, C. Zhu, J. R. Agre, and A. Ephremides, "Proportional fair scheduling algorithm in ofdma-based wireless systems with qos constraints," *Journal of communications and networks*, vol. 12, no. 1, pp. 30–42, 2010.

[172] Y. Zaki, T. Weerawardane, C. Gorg, and A. Timm-Giel, "Multi-qos-aware fair scheduling for lte," in *2011 IEEE 73rd Vehicular Technology Conference (VTC Spring)*, IEEE, 2011, pp. 1–5.

[173] J. Yin, Y. Chen, G. Sang, B. Liao, and X. Wang, "Qoe-oriented rate control and resource allocation for cognitive m2m communication in spectrum-sharing ofdm networks," *IEEE Access*, vol. 7, pp. 43 318–43 330, 2019.

[174] W. K. Lai and C.-L. Tang, "Qos-aware downlink packet scheduling for lte networks," *Computer Networks*, vol. 57, no. 7, pp. 1689–1698, 2013.

[175] S. Ismail, D. M. Ali, and A. L. Yusof, "Mecc scheduling algorithm in vehicular environment for uplink transmission in lte networks.," *International Journal of Electrical & Computer Engineering (2088-8708)*, vol. 9, no. 2, 2019.

[176] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to algorithms*. MIT press, 2009.

[177] H.-W. Ferng, C.-Y. Lee, J.-J. Huang, and Y.-J. Liang, "Urgency-based fair scheduling for lte to improve packet loss and fairness: Design and evaluation," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 3, pp. 2825–2836, 2019.

[178] M. M. Nasralla, "A hybrid downlink scheduling approach for multi-traffic classes in lte wireless systems," *IEEE Access*, vol. 8, pp. 82 173–82 186, 2020.

[179] M. Y. Abdelsadek, M. H. Ahmed, and Y. Gadallah, "An lte matching-based scheduling scheme for critical-mtc with shortened transmission time intervals," in *2020 IEEE 92nd Vehicular Technology Conference (VTC2020-Fall)*, IEEE, 2020, pp. 1–6.

[180] J. Van den Eynde and C. Blondia, “A minimal delay violation downlink lte scheduler,” in *2021 IEEE 46th Conference on Local Computer Networks (LCN)*, IEEE, 2021, pp. 387–390.

[181] G. Manogaran and B. S. Rawal, “An efficient enb selection and traffic scheduling method for lte overlay iot communication networks,” in *2021 IEEE Global Communications Conference (GLOBECOM)*, IEEE, 2021, pp. 1–6.

[182] A. Mukhopadhyay and G. Das, “Low complexity fair scheduling in lte/lte-a uplink involving multiple traffic classes,” *IEEE Systems Journal*, vol. 15, no. 2, pp. 1616–1627, 2020.

[183] H. M. Zegallai and S. A. Awad, “Packet scheduling with carrier aggregation for balancing qos in lte advanced networks,” in *2021 IEEE 1st International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering MI-STA*, IEEE, 2021, pp. 882–886.

[184] M. L. Moses, G. Chandar, *et al.*, “Joint energy and qos-aware cross-layer uplink resource allocation for m2m data aggregation over lte-a networks,” in *2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS)*, IEEE, 2022, pp. 1485–1490.

[185] A. Abdulazeez, M. M. Yahaya, I. B. Yabo, A. Bello, M. M. Umar, and A. Mohammed, “Prioritized quality of service-aware downlink scheduling algorithm for lte network,” in *2022 IEEE Nigeria 4th International Conference on Disruptive Technologies for Sustainable Development (NIGERCON)*, IEEE, 2022, pp. 1–5.

[186] A. Osseiran, J. F. Monserrat, and P. Marsch, *5G mobile and wireless communications technology*. Cambridge University Press, 2016.

[187] P. Marsch, Ö. Bulakci, O. Queseth, and M. Boldi, *5G system design: architectural and functional considerations and long term research*. John Wiley & Sons, 2018.

Publications

List of research article(s) in journal

1. **U. Singh**, A. Dua, S. Tanwar, N. Kumar and M. Alazab, "A Survey on LTE/LTE-A Radio Resource Allocation Techniques for Machine-to-Machine Communication for B5G Networks," in IEEE Access, vol. 9, pp. 107976-107997, 2021, doi: 10.1109/ACCESS.2021.3100541. **[Q1, SCIe, IF 3.367]** https://ieeexplore.ieee.org/document/9497084

2. **U. Singh**, A. Dua, N. Kumar and M. Guizani, "QoS Aware Uplink Scheduling for M2M Communication in LTE / LTE-A Network: A Game Theoretic Approach," in IEEE Transactions on Vehicular Technology, doi: 10.1109/TVT.2021.3132535. **[Q1, SCI, IF-5.987]** https://ieeexplore.ieee.org/document/9635654

3. **U. Singh**, A. Ramaswamy, A. Dua, N. Kumar, S. Tanwar, G. Sharma, I. E. Davidson, R. Sharma, "Coalition Games for Performance Evaluation in 5G and Beyond Networks: A Survey," in IEEE Access, doi: 10.1109/ACCESS.2022.3146158. **[Q1, SCIe, IF 3.367]** https://ieeexplore.ieee.org/document/9691369

4. **U. Singh**, A. Dua, N. Kumar, S. Tanwar, R Iqbal, M. Hijji, R. Sharma, "Scalable Priority-based Resource Allocation Scheme for M2M Communication in LTE/LTE-A Network," in Computers and Electrical Engineering,2022 Oct 1;103:108321 **[Q1, SCIe, IF-4.15]** https://www.sciencedirect.com/science/article/pii/S0045790622005432

Brief Biography of the Candidate

Upendra Singh received his B.E. degree in Computer Engineering from the University of Rajasthan, Jaipur, India, in 2009. He received his M. Tech. in Computer Science from Jagannath University, Jaipur, India in 2013. He worked as an Assistant Professor for four years in the Department of Computer Science and Engineering at Engineering College Bharatpur, Bharatpur, Rajasthan, India affiliated to RTU Kota, Rajasthan, India. Upendra Singh has completed his Ph.D. as a full-time scholar at BITS, Pilani under the guidance of Dr. Amit Dua, Associate Professor, Department of Computer Science & Information Systems, BITS, Pilani. His primary area of interest is Computer Networks, Information Security, M2M Communication, IoT, and 4G-5G Cellular Network.

Brief Biography of the Supervisor

Dr. Amit Dua is an Associate Professor in the Computer Science and Information System Department at BITS Pilani. He did a BE (Hons) in computer science and an ME in software systems both from BITS Pilani. He has completed his Ph.D. thesis from Thapar University in Ad hoc networks. He has seven years of experience in teaching, industry, and research. He has worked as a lecturer at NIT Kurukshetra and was part of the software team at UPSINVERTER.COM.

His areas of interest include ad hoc networks, network security, game theory, fuzzy logic, and smart grid. He is a reviewer of the International Journal of Communication Systems, KSII Transactions on Internet and Information Systems, and a TPC member of several international Conferences. He is a member of IEEE and ACM.

Outside the work, he is fond of various outdoor sports activities. He has won prizes for table tennis and lawn tennis at the state level and at annual institutional sports meets.

In addition to being a member of the Board of Studies and Doctoral Research Committee of many universities, he has delivered invited talks at universities and has visited many research organizations, government organizations, and industry. He has an association with CBSE-AIPMT, National Service Scheme (NSS), BITSAT, UKPSC, and many more.