DESIGN AND ANALYSIS OF LTE AND WI-FI SCHEMES FOR COMMUNICATIONS OF MASSIVE MACHINE DEVICES

AYOADE OLUWAFEMI ILORI

Doctor of Philosophy

January 2017

ASTON UNIVERSITY

Summary

Existing communication technologies are designed with specific use cases in mind, however, extending these use cases usually throw up interesting challenges. For example, extending the use of existing cellular networks to emerging applications such as Internet of Things (IoT) devices throws up the challenge of handling massive number of devices. In this thesis, we are motivated to investigate existing schemes used in LTE and Wi-Fi for supporting massive machine devices and improve on observed performance gaps by designing new ones that outperform the former.

This thesis investigates the existing random access protocol in LTE and proposes three schemes to combat massive device access challenge. The first is a root index reuse and allocation scheme which uses link budget calculations in extracting a safe distance for preamble reuse under variable cell size and also proposes an index allocation algorithm. Secondly, a dynamic subframe optimization scheme that combats the challenge from an optimisation solution perspective. Thirdly, the use of small cells for random access. Simulation and numerical analysis shows performance improvements against existing schemes in terms of throughput, access delay and probability of collision. In some cases, over 20% increase in performance was observed. The proposed schemes provide quicker and more guaranteed opportunities for machine devices to communicate.

Also, in Wi-Fi networks, adaptation of the transmission rates to the dynamic channel conditions is a major challenge. Two algorithms were proposed to combat this. The first makes use of contextual information to determine the network state and respond appropriately whilst the second samples candidate transmission modes and uses the effective throughput to make a decision. The proposed algorithms were compared to several existing rate adaptation algorithms by simulations and under various system and channel configurations. They show significant performance improvements, in terms of throughput, thus, confirming their suitability for dynamic channel conditions.

Keywords: *RACH, IEEE 802.11, LTE, Rate Adaptation, M2M, MTC, Wi-Fi*

To the Almighty God,
for reassurance at every step of the way
and
my beloved parents,
Chief Oluwadare & Mrs Adenike Ilori,
without whom I wouldn't even have embarked on this journey.

Acknowledgements

I would like to extend my most sincere gratitude to my supervisors, Dr. Zuoyin Tang and Dr. Jianhua He. They have been a constant source of support and encouragement in every aspect. Their unending insight and ideas made the journey easier.

I would also like to appreciate my adorable wife, Mrs Feyisayo Ilori, for her prayers, patience, inexhaustible support and encouragement through the highs and lows of this journey. You complete me.

Special appreciation to my parents, Chief & Mrs Ilori and Prince & Dr. (Mrs) Amuwa for their elderly words of wisdom, spiritual and moral support and giving the most needed pushes when the going got tougher. I could not have asked for better parents.

My siblings, Adewole, Tinuade and Adedapo Ilori also deserve enormous credits for their unparalleled understanding through the entire period. You guys are the best.

Finally, a befitting thank you to colleagues and friends who have been there for me, encouraged me, laughed with me, and supported me in some way throughout the journey. Such invaluable persons include: Adegoke Falaiye, Anayo Chukwukere, Akinbiyi Akindoyin, Michael Allo, Timothy Ayoade, 'Bose Ayoade, Bolaji Lawal, Adewale Adetayo, Pst. Olusola Ojo, Pst. 'Ranti Oyewale, Dr. Francis Elewuwa, Afolabi Olayemi, Fidelis Odinaka Offor and Dr. Muktar Bashir.

Contents

Acronyms

3GPP Third Generation Partnership Project.

4G Fourth Generation.

AARF Adaptive Auto Rate Fallback.

AC Access Class.

ACB Access Class Barring.

ARF Auto Rate Fallback.

AS Application Server.

ATIS Alliance for Telecommunications Industries Solution.

BPSK Binary Phase Shift Keying.

CARS Context Aware Rate Selection for Vehicle networks.

CCK Complementary Clock Keying.

CCSA China Communications Standards Association.

CDMA Code Division Multiple Access.

CFP Contention Free Period.

CSMA/CA Carrier Sense Multiple Access with Collision Avoidance.

CTS Clear to Send.

DL Down Link.

DL-SCH Down Link Shared Channel.

DRX Discontinuous Reception.

DSSS Direct Sequence Spread Spectrum.

EAB Extended Access Class Barring.

EC-GSM-IoT Extended Coverage - GSM - Internet of Things.

EDGE Enhanced Data rate for GSM Evolution.

eMTC enhanced Machine Type Communications.

eNodeB A base station, also called Evolved NodeB.

EPC Evolved Packet Core.

ETSI European Telecommunications Standards Institute.

FCS Frame Check Sequence.

FDD Frequency division duplex.

FDMA Frequency Division Multiple Access.

FHSS Frequency Hopping Spread Spectrum.

GPRS General Packet Radio Service.

GSM Global System for Mobile telecommunications.

HARQ Hybrid Automatic Repeat reQuest.

HSPA High Speed Packet Access.

HSS Home Subscription Server.

IEEE Institute of Electrical and Electronics Engineers.

IoT Internet of Things.

ISM Industrial Scientific and Medical.

ITS Intelligent Transport Systems.

ITU International telecommunications union.

LAN Local Area Network.

LLC Logical Link control.

LTE Long Term Evolution.

LTE-A Long term Evolution Advanced.

M2M Machine to Machine.

MAC Media Access Control.

MME Mobility Management Entity.

MO Mobile Originating.

MSISDN Mobile Subscriber Integrated Services Digital Network Number.

MTC Machine Type Communication.

MTC-IWF MTC Inter Working Function.

NAS Non Access Stratum.

NB-IoT Narrowband Internet of Things.

OFDM Orthogonal Frequency Division Multiplexing.

OFDMA Orthogonal Frequency Division Multiple Access.

OSI Open Systems Interconnection.

P-GW Packet Gateway.

PCRF Policy and Charging Resource Function.

PDCP Packet Data Convergence Protocol.

PHY Physical layer.

PIFS PCF Inter Frame Spacing.

PRACH Physical Random Access Channel.

PS Packet Switched.

QAM Quadrature Amplitude Modulation.

QoS Quality of Service.

QPSK Quadrature Phase Shift Keying.

RA Random Access.

RACH Random Access Channel.

RACH Random Access Channel.

RCA Random Channel Access.

RLC Radio Link Control.

RRC Radio Resource Control.

RTS Request To Send.

S-GW Serving Gateway.

SC-FDMA Single Carrier - Frequency Division Multiple Access.

SCS Services Capability Server.

SIFS Short Inter-frame Spacing.

SNR Signal to Noise Ratio.

TDD Time division duplex.

TIA Telecommunications Industry Association.

TTA Telecommunications Technology Association.

TTC Telecommunications Committee.

UDP User Equipment.

UE User Datagram Protocol.

UL-SCH Uplink Shared Access Channel.

UMTS Universal Mobile Telecommunications System.

UP User Plane.

UTRAN Universal Terrestrial Radio Access Network.

V2V Vehicle to Vehicle.

VANET Vehicular Ad Hoc Network.

WAVE Wireless Access for Vehicular Environment.

WLAN Wireless LAN.

List of Figures

List of Tables

1 Introduction

The Internet of Things (IoT) is a constantly growing network of Internet enabled devices for the purpose of data collection and exchange, remote diagnostics, device configuration and control amongst other things. The IoT has changed the face of industrial, commercial and domestic sectors of our society with interesting applications being developed in these areas.

For instance, Smart meters have been introduced in the energy sector making it possible for remote meter reading collection, vehicle to vehicle (V2V) and vehicle to infrastructure (V2I) communications have gained further prominence in the transportation industry leading to the birth of safety applications and other intelligent transport systems. Other interesting application areas of the IoT includes security - the operations of surveillance systems and access control to buildings, payments - point of sales , gaming and vending machines, smart grids, smart home, personal health monitoring, controlling vending machines remotely. On a larger scale, governments are also embracing the IoT with the introduction of smart cities, the MKSmart project in the United Kingdom is a very good example.

Based on the massive application areas of the IoT, billions of devices would need to be interconnected for this technology to be a success. Short range radio technologies such as Wireless Fidelity (Wi-Fi), Bluetooth, Zigbee and wide area connectivity solutions such as cellular media would be used to provide connectivity solutions to devices in this emerging ecosystem. The IoT has also generated interests amongst standard bodies such as the Institute of Electrical and Electronics Engineers (IEEE), third Generation Partnership Project (3GPP) and the European Telecommunications Standards Institute (ETSI). IEEE, 3GPP and ETSI are standard bodies composed of academics, technocrats, engineers and companies that contribute to the advancement of wireless communication technology. These bodies define rules and protocols that govern the software and hardware systems that operate under them. The IEEE plays major role in Wi-Fi, while cellular communications is the forte of 3GPP and ETSI.

In academia, a large amount of research has been carried out in various aspects of IoT communication technology including protocol design, performance analysis and modelling, compatibility with existing technologies and safety. It is easy to predict that the IoT would have a massive impact on our everyday lives in the nearest future just as it is now.

1.1 Application of IoT

The IoT is being applied in every segment of industry and the society at large. These application areas are further discussed in the following subsections.

Agriculture

Potential application of IoT in agriculture can be found in tracking livestock including plant and soil monitoring. These include sensing for soil moisture and nutrients, control of water usage for optimal plant growth and determination of what fertilizers to be used in a farm based on soil profile. Also, IoT applications can be deployed to gather data about the location, health and well-being of animals. A typical example of the application of IoT in agriculture is stated in [1].

Utilities

Smart grid management and smart metering are typical application areas of IoT in utilities. Monitoring of energy consumption and generation patterns can help load balancing within the grid and reduce cost for consumers. Water distribution, quality and consumption patterns can also be monitored with the help of IoT and this data can be used for futuristic planning by the relevant bodies [2] [3].

Transportation and Logistics

Through IoT, goods can be fitted with sensors making them easy to track from production points to warehouses or other retail points. Also, by installing sensors on vehicles, the location and condition of vehicles within a fleet can easily be determined making them easier to manage. IoT can also be used in traffic management. Intelligent transport applications can help predict congestion points and patterns and suggest alternate routes to the destination. IoT technology can also help in managing incidences, such as accidents, and responding to emergencies [4] [5].

Manufacturing and Retail

IoT can be used to monitor and control a manufacturing process. Readings of the vital signs of manufacturing machines can be sent to a central device which would help in scheduling maintenance [6] [7].

Smart Cities

Smart cities aim to make proficient use of public resources, increase the quality of service offered to the populace and reduce the governments' operational costs. All these can be made possible using IoT technology [8] [9]. This is an interesting application area especially for government and there are a couple of trials springing up in order to expedite the implementation of the smart city vision. A typical example is the UK's MKSmart project [10].

Smart homes

The IoT can help in making homes more comfortable through smart homes. A smart home is an interconnection of household appliances, such as washing machines, televisions and heaters, for easy maintenance and control [11]. For example, during winter, an individual can remotely switch on heating in the home making it cosier.

Healthcare

IoT technology can assist in the provision of healthcare including patient monitoring. A patients vital signs can be sent to a doctor remotely who can then advise on the best course of action for the patient. Thus, except it is absolutely necessary, a patient would not need to visit the hospital. This would in turn save the patient the stress and time of visiting hospitals [12].

1.2 Communication Technologies in IoT

Different connectivity alternatives exists for IoT devices, some proprietary, others standard. However, the choice of connectivity for any IoT application depends majorly on how it will be used (use case). As discussed earlier, IoT applications can be found in a variety of environment, some indoors others outdoors, and also some in difficult to reach terrains like in desert. We can classify the various connectivity options, as short and long range, based on their coverage range.

1.2.1 Short Range

IoT applications that are designed for home or indoor use typically make use of short range connectivity technologies. Prevalent in this category are the Wi-Fi, Bluetooth, and Zigbee. It is also interesting to note that these technology operate in the unlicensed spectrum and the quality of service (QoS) and security they offer are limited compared to their long range counterparts.

Wi-Fi

IoT connectivity based on Wi-Fi takes advantage of existing technology. Already, a large number of devices are able to connect to the Internet using Wi-Fi. This technology is standardized by the IEEE and developments are ongoing on how to better adapt it for IoT applications. The de-facto standard of Wi-Fi is the IEEE 802.11 and its implementation started with theoretical maximum rates of 1 and 2 Mbps. This rate is capable of supporting many indoor IoT applications. Wi-Fi also offers a good quality of service and is relatively secure.

Bluetooth

Bluetooth is a short range wireless technology that operates in the Ultra High frequency (UHF) range of 2.4 to 2.485 GHz. More details on the technical specification of this technology can be found in [13]. Recently, Bluetooth is being improved to make it more suitable for the IoT by providing significant improvements in terms of speed, range and broadcast capacity and this has also been studied in literature [14].

Zigbee

Zigbee is specifically designed for device to device communications which is a critical part of the IoT [15]. It is also a short range technology that was developed by the Zigbee Alliance. Zigbee operates in the unlicensed 2.4 GHz frequency band and has found wide use in smart homes and utilities industry amongst others. Thus, this technology finds its application majorly in short range IoT applications.

1.2.2 Long Range

IoT applications can also require connectivity between devices that are hundreds of kilometres apart and this has led to the rise in use of wide area network technologies for implementing IoT connectivity. Cellular technologies, which were originally designed for mobile voice and data, are now being adapted to suit applications in the IoT ecosystem. The quality of service, security and ubiquity amongst others, make this technology an attractive option for implementing IoT. Cellular technologies hold sway in catering for the needs of long range IoT applications. Particularly, 3GPP technologies such as Global System for Mobile Communications (GSM), Wideband Code Division Multiple Access (WCDMA), Long Term Evolution (LTE) and fifth generation mobile networks (5G) which operate in the licensed spectrum are now evolving with functionalities to suit the IoT landscape.

1.3 Motivation and Research Problem

The world is gradually embracing the IoT and this emerging technology has permeated almost every sector of human existence, from transportation to healthcare and energy, just to mention a few. There has also been a steady growth in the number of IoT capable devices within the IoT stratosphere. In fact, Ericsson predicted an over 100% rise in the number of devices from 14 billion in the year 2014 to over 28 billion in 2021 and that the IoT industry would be worth about $1.9 trillion [16]. All these points to the fact that the relevance of this technology, now and in future, cannot be overemphasized.

However, for all of the potential benefits of this budding technology to be realised, a major challenge would be in the choice of communication media to be used in the IoT network. Several mediums exists that can be used for this but the challenge is to ascertain whether the existing mediums can cope with such massive load and demands of the IoT and if not to propose solutions on how communication can be effectively achieved. The consensus, both in academia and industry, seem to point to the fact that a large proportion of communication devices in the IoT would be done through wireless means. But this also comes with its own issues.

Two popular candidates for IoT device communication are the IEEE 802.11 technology and cellular technologies. The ease of access and installation of Wi-Fi coupled with the ubiquity of cellular medium have endeared these technologies in the heart of players in the IoT space. But there are several major obstacles to contend with in order to provide efficient and robust communications for practical IoT applications to be developed and deployed, a few of which are listed as follows:

- In both cellular and Wi-Fi networks, the sheer number of devices that might seek to use the network has the potential of making both mediums suffer degradation in terms of

Quality of Service (QoS) and throughput.

- Being contention based technologies, the initial access schemes in both cellular and Wi-Fi would potentially be congested from the amount of load projected to be put on them.
- A natural consequence of the above point is that devices might have to wait a long time to access the network. This is particularly undesirable scenario especially for time critical applications.

To the best of our knowledge, there are very few works out there that seek to address these challenges from the perspective of combining a number of communication medium. However, the 3GPP recently came up with the idea of a cellular and wireless combination called LTE-Unlicensed (LTE-U) and there are a number of ways in which this technology is being proposed to be implemented including:

- As a standalone that would involve extending the reach of LTE communication to the unlicensed band. However, LTE is designed specifically for the licensed band and making it work in the unlicensed band would involve optimizations, not to mention potential disruption to other wireless technology who operate solely in the unlicensed band, for example, Wi-Fi. This is not so much convergence but a way of obtaining additional spectrum for LTE.
- As LTE- license assisted access (LAA-LTE) which would involve running all control channels in the licensed band and using unlicensed bands for additional data plane capacity. This is now infused by the 3GPP for standardization in Release 13.
- An appealing option is that of LTE + Wi-Fi Link Aggregation (LWA). Traditionally, LTE operates in the licensed frequency spectrum and Wi-Fi in the unlicensed, however, both technologies are combined to provide an appealing user experience. This approach is also being promoted by Qualcomm. In this way, both technologies can focus on what they are best at whilst users can enjoy a good QoS.

Of the three options above, the LWA is most appealing especially for implementation in the IoT. The main motivations of this thesis are to develop analytical and simulation tools to investigate the performance of using LTE and Wi-Fi technology to support massive device access in the IoT and also propose effective solutions in mitigating potential challenges.

1.4 Research Objectives

The implications of this research are expected to contribute directly to investigating the effects of an ever increasing network load on communication mediums and proposing solutions to identified challenges. The aims of this research are to design scalable and reliable communication protocols and control algorithms to support the communication of a large number of IoT enabled devices over wireless networks, this would, in effect, provide Quality of Services (QoS) support for the applications built for the IoT.

The main objectives of this thesis are summarized as follows:

- To design system level simulators that can be used to validate models and reinforce system performance evaluations in IoT using LTE and Wi-Fi as communication mediums.
- To investigate and address issues arising in support of massive device communications in IoT including medium access control (MAC), link adaptation, and resource management support for devices over wireless networks.
- To propose new and/or improved algorithms once issues are identified.
- To devise performance evaluation tools for analysing the performance of LTE and Wi-Fi networks under different scenarios.
- To evaluate the performance of proposed algorithms through the designed tools.

1.5 Thesis Outline and Contributions

This section outlines the contributions of this thesis and further elaborates on its contents. In chapter 2, we present a detailed overview of the LTE technology including its system architecture, interfaces, channels and protocol. We also give detailed insight into how the 3GPP is adapting its present technologies and creating new ones for IoT. Furthermore, we delve into key issues that affect IoT communications in 3GPP before proceeding to discuss the random access channel (RACH) of the LTE and its challenges when used for IoT. The state of the art with respect to proposed solutions to the RACH congestion challenge is also presented in this chapter. From existing proposals, it is observed that there are performance gaps that need to be filled in order to effectively tackle the challenge of RACH congestion in LTE networks.

In chapter 3, an insight into link adaptation in IEEE 802.11 Wireless Local Area Networks (WLAN) is given. We also discuss the challenges confronting the MAC layer of this technology. In particular, how the IEEE standard body implements IoT communications and the challenges facing this implementation is also presented. Rate adaptation in IEEE 802.11 WLANs is studied including various algorithms that aim to proffer solution to the rate adaptation challenge in these networks. From our studies, it is evident that majority of the rate adaptation algorithms proposed are specifically for static networks whilst a few exists for dynamic networks. The main contributions are then presented in the subsequent three chapters.

In chapter 4, a root index allocation scheme for massive device access in LTE networks is proposed to tackle the challenge of congestion in the RACH. We propose an effective preamble reuse scheme, based on safe distance, to provide more random channel access opportunities to support multiple machine devices accessing LTE cellular networks. It is observed that this solution significantly increase random access throughput, reduce access delays and preamble collision probability whilst offering network operators more flexibility in planning their preamble configurations and deployment. A novel preamble allocation algorithm is also proposed to allocate the root sequences to cells based on saturation colouring, which can be applied to cellular networks with irregular layouts. The sequence allocation algorithm is flexible and can be applied to meet different RCA traffic loads and operational requirements. The works in this chapter has been submitted to the IEEE transaction on Industrial Informatics for review and possible publication.

In chapter 5, congestion in RACH is tackled from two different perspectives. The first is from the perspective of resource optimization. In this way, the congestion problem is viewed as one of optimisation to be solved with a utility function.Based on this and a method of estimating the number of machine devices, an adaptive subframe allocation scheme is proposed. Numerical and simulation results verify the effectiveness of the proposed frame adaptation scheme in combating RACH congestion.
The second is from the view of using small cells within LTE networks to alleviate network loads on macro cells in case of congestion. At present, there seems to be a steady implementation of small cells with the network by operators. Thus, it is logical to investigate the performance of LTE network, with small cell support, in terms of combating the RACH overload challenge. It is observed that the capacity of the networks in terms of the number of supported machine devices with small cell support can be increased significantly. This second part is based on the following publication:

- A. Ilori, Z. Tang, J. He, K. Blow and H. Chen, "A random channel access scheme for massive machine devices in LTE cellular networks." 2015 IEEE International Conference on Communications (ICC). IEEE, 2015, pp. 2985 - 2990.

In chapter 6, rate adaptation algorithms in static and dynamic network scenarios are studied. The idea is that some of the direct application areas of IoT, such as, vehicle to vehicle (V2V) or vehicle to infrastructure (V2I) communications would gain ground, especially in the transportation industry and as a result, the use of IEEE 802.11 wireless technology in this scenario would be quite important. This chapter is also divided broadly into two sections. The first section proposes a vehicular rate adaptation algorithm that is based on the best performer out of a number of existing algorithms used in static networks. It is observed that the proposed algorithm performs much better than existing ones. This section is based on the following publication:

- A.Ilori, Z. Tang, J. He and Y.Li, "Context-Aware Rate Adaptation Algorithm for Vehicular Networks", 2013 HET-NETs conference, UK, pp 221 - 225.

The second section also proposes a Throughput-based rate adaptation algorithm for IEEE 802.11 vehicle networks and an ideal signal to noise ratio (SNR) based algorithm. It is observed that rate adaptation algorithms tend to benefit from using contextual information in order to determine the state of the network before or during a transmission. Thus, the throughput based algorithm, exploits throughput of various transmission modes to determine channel conditions and make appropriate changes to data rates. The ideal rate adaptation algorithm assumes accurate channel SNR is available and is implemented for benchmark performance comparison. This section is based on the following journal publication:

- A. Ilori, Z. Tang, J. He, and Y. Li, "Throughput-based rate adaptation algorithm for IEEE 802.11 vehicle networks." IET networks, vol. 4, no. 2, pp.111-118, 2015.

This thesis is concluded in chapter 7. The chapter contains a summary of the works presented in the previous chapters and also discusses ways in which these works can be extended in future.

1.6 Related Publications

- A. Ilori, Z. Tang, J. He and K. Blow, "Root index allocation scheme for massive device access in LTE networks." submitted to IEEE transaction on Industrial Informatics (Under major revision), 2016.
- A. Ilori, Z. Tang, J. He, K. Blow and H. Chen, "A random channel access scheme for massive machine devices in LTE cellular networks." 2015 IEEE International Conference on Communications (ICC). IEEE, 2015, pp. 2985 - 2990.
- A. Ilori, Z. Tang, J. He, and Y. Li, "Throughput-based rate adaptation algorithm for IEEE 802.11 vehicle networks." IET networks, vol. 4, no. 2, pp.111-118, 2015.
- A.Ilori, Z. Tang, J. He and Y.Li, "Context-aware Rate Adaptation Algorithm for Vehicular Networks", 2013 HET-NETs conference, Uk, pp 221 - 225.

2 IoT Technologies in 3GPP Networks

Cellular technology has grown over the years from the GSM to the present LTE systems. This advancement comes with a lot of advantages including increased system efficiency, increasing upload and download speed and reduction in system complexity just to name a few. The 3GPP has been actively involved in the development of cellular telecommunications networks and consist of a broad variety of members most of whom are major players in the telecommunications industry.

Since IoT involves communication between various entities that make up the IoT ecosystem, devices, also referred to as machines, are key participants within the IoT. A machine in this context refers to any device that is capable of receiving, sending and acting on information.
Machine to machine (M2M), also known as Machine Type communication (MTC) could be best described as machines communicating with each other without or with minimal human intervention. Typical examples of machines include smart sensors, smart meters and embedded processors.

Technical standardization for machine type communications (MTC) is ongoing in various standards developing organizations such as Third Generation Partnership Project (3GPP), European Telecommunications Standards Institute (ETSI), Institute of Electrical and Electronics Engineering (IEEE) and the telecommunications industry association (TIA) [17]. Different network architecture and functions that are specific to each organization to support M2M communications are clearly stated by individual standard bodies. However, by joint discussions between seven regional standards development organizations (SDOs) an agreement has been reached to form a global alliance in order to synergize the development of a common global standard for M2M communications. The seven SDOs are Association of Radio Industries Business (ARIB) based in Japan, Telecommunications Committee (TTC) based in Japan, Alliance for Telecommunications Industries Solution (ATIS) based in USA, China Communications Standards Association (CCSA), European Telecommunications Standards Institute (ETSI) based in Europe, Telecommunications Technology Association (TTA) based in Korea and Telecommunications Industry Association (TIA). This union seeks to come up with end-to-end specifications for M2M communications with a starting point on the services layer using use cases and architecture principles across the many divides of MTC applications. The initiative, launched in July 2012 has been christened "oneM2M".

This chapter focuses specifically on technical specifications defined for MTC communications by the 3GPP standards body. The 3GPP standard was chosen due to the fact that it cuts across a wide range of telecommunication bodies from all over the world. It consists of six organizational partners and these include ARIB, ATIS, CCSA, ETSI, TTC and TTA.

The 3GPP has been working on optimizations of the network to accommodate Machine type communications (MTC) and specifically focusing on the fundamental connections between M2M application servers and devices through the mobile network. With the advancement in network technology within the 3GPP, specific attention is now being paid to machine communications within each new release.

In this chapter, details of the LTE implementation of MTC technology is given. First of all, the system architecture of LTE is presented in section 2.1, including discussions on the interfaces and protocols in LTE. The network architecture of LTE implementation of MTC is presented in section 2.2 whilst in section 2.3 we consider the various technologies being proposed by the 3GPP for IoT. In section 2.4, key issues prevalent in MTC are discussed while we delve into the random access procedure including its potential challenges in section 2.5. We conclude this chapter in section 2.6.

2.1 System Architecture of LTE

LTE allows network operators to use new and wider spectrum and complements third generation (3G) networks with higher user data rates, lower latency, and a flat Internet Protocol (IP)-based network architecture [18]. The system architecture of LTE technology showing majority of the functional nodes present in the LTE architecture and the interfaces through which these nodes communicate can be represented as shown in Figure 2.1. The dotted line indicates control signalling information route while the thick lines show user plane signalling route.

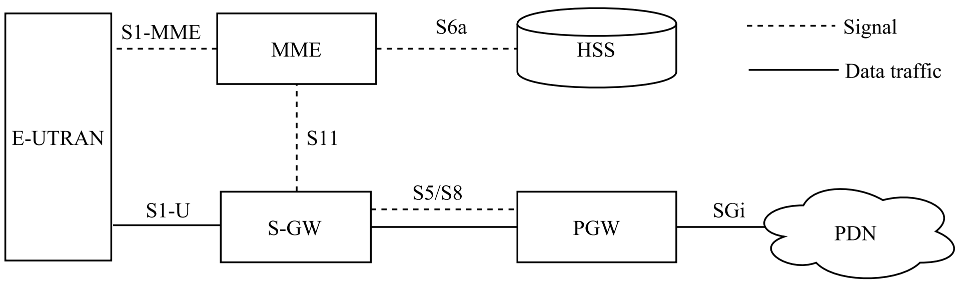

Figure 2.1: Functional nodes in LTE architecture

A description of each of the nodes including its functions is given as follows:

Evolved Universal Terrestrial Radio Access Network (E-UTRAN)

The E-UTRAN is made up of a combination of eNodeBs. An eNodeB covers a specific geographical area of the network and the entire coverage area of a telecommunications network operator is covered by groups of eNodeBs. eNodeBs communicate with each other via the X2 protocol. Functionally, an eNodeB acts as a bridge between the UE and the evolved packet core (EPC) and is the termination point of all radio protocols towards the UE. The eNodeB is also responsible for control plane functions and radio resource management functions such as the control of the usage of the radio interface i.e allocation of resources based on requests, prioritization, scheduling and constant monitoring of traffic according to required QoS. Mobility management functions are also performed by the eNodeB in conjuction with the mobility management entity (MME). This is done through handovers and the eNodeB analyses radio signal level measurements done by the UE and compares this to its own self measured value and then makes the decision whether a handover should occur or not. Signalling information is exchanged between the eNodeB and the MME via the S1-MME interface. The eNodeB also performs user plane data tunnelling for uplink data and data delivery for downlink data by working closely with the serving gateway (S-GW) [19].

Mobility Management Entity (MME)

This is the main control element of the EPC and is usually a server located in operator premises. Being a control element, the MME is not involved in user plane data [18]. The main functions of the MME include:

- Authentication and Security: A UE entering the network for the first time is authenticated by the MME which performs series of steps that involves the UE responding to a challenge. An incorrect response means the UE is denied access to the network. Authentication is performed periodically by the MME.
- Mobility Management: MME keeps track of all the UEs in its service area. When a UE is in active state, it is tracked at the eNodeB level and when it is in idle mode, its location is gotten via tracking area, which is essentially a group of eNodeBs.
- Managing Subscriber profile and service connectivity: when a UE registers with the network, the MME retrieves its subscription profile from the home network and stores the information for the duration it is serving the UE.

Serving Gateway (S-GW)

The function of the S-GW is user plane tunnel management and switching. When the S5/S8 interface is based on GPRS tunnelling proxy (GTP), the S-GW will have GTP tunnels on all its user plane interfaces but if S5/S8 is based on proxy mobile internet protocol (PMIP) the IP service flows hold sway. Mapping between the IP service flows and GTP tunnels is done in the packet data networks gateway (P-GW). The S-GW does very little when it comes to control functions. It is responsible for its own resources and allocates them based on requests from MME or P-GW. The S-GW also carries out bearer binding and event reporting function (BBERF) by mapping IP flows in S5/S8 interface to bearers in the S1 interface [20].

Packet Data Network Gateway (P-GW)

The P-GW is usually an edge router between the evolved packet system (EPS) and external packet data networks. It is the highest mobility anchor of the system and acts as the IP point of attachment for the UE. It supplies IP addresses to the UE based on requests and also performs dynamic host configuration protocol (DHCP) by querying a DHCP server. Thus, functionally, the P-GW carries out IP flows of user data, does policy and charging control requests and controls user plane tunnels for uplink and downlink data delivery.

Home Subscription Server (HSS)

This is the subscription data repository for all permanent user data. The HSS is a database maintained centrally by the network operator and stores a master copy of the subscriber profile and this usually contain services applicable to the user, roaming and allowed data network connections. Due to its importance in authenticating users, the HSS needs to be able to connect to all the MMEs in the network [21]..

Packet Data Network (PDN)

The PDN could be external servers that are outside the LTE network. They are basically used to interface with external networks such as the internet [21].

2.1.1 Interfaces and Protocols in LTE

The interfaces and protocols used by the various nodes in the LTE is described in this section. These could be subdivided into control and user plane protocol. Some protocols are common to both control and user planes whilst others are specific. A brief description of each protocol is also given.

- Non-Access Stratum (NAS): This is the topmost layer of the control plane and it consists of two separate protocols that are carried on direct signalling transport between the UE and MME. The content of NAS is not visible to eNodeB. The NAS layer protocols are: EPS Mobility Management: This protocol is responsible for handling UE mobility within the system. It includes functions for attaching to and detaching from the network and performing location updating in between.

  EPS Session Management: this protocol is used to handle bearer management between the UE and MME.

- Radio Resource Control (RRC): This protocol controls radio resource usage. It also manages UE's signalling and data connections and includes functions for hand-overs.

- Packet Data Convergence Protocol (PDCP): helps in IP header compression in the user plane. It also handles encryption and integrity protection in the control plane.

- Radio Link Control (RLC): This protocol is responsible for segmentation and concatenation of PDCP-PDUs for radio interface transmission. Error correction using the automatic repeat request (ARQ) is also part of the functions of this protocol.

- Physical Layer: this is the first layer of the radio interface and it takes care of OFDMA and SC-FDMA. The S1 interface connects the E-UTRAN to EPC.

- S1 Application Protocol (S1AP): this handles the UE's control plane and user plane connections between the EPC and E-UTRAN. It also participates in handovers when EPC is involved.
- SCTP/IP Signalling Transport: the stream control transmission protocol (SCTP) and IP signalling transport aids in the transportation of signaling messages. SCTP provides reliable transport and sequenced delivery functions.
- GPRS Tunnelling Protocol (GTP): this could either be used in the control plane or in the user plane. In the control plane it is called GTP-C and its function is to manage control plane data including QoS and signalling It also performs mobility management functions within the EPC. In the user plane, GTP-U holds sway and it used when S5/S8 is GTP based. It forms the GTP-U tunnel that is used to send the end users IP packets belonging to an EPS bearer.
- UDP/IP Transport: The User Datagram protocol (UDP) and Internet Protocol (IP) are basic standards defined by the OSI model. UDP is used instead of TCP because higher layers already provide reliable transport with the necessary error correction and recovery mechanisms.

2.1.2 Channels in LTE

Channels carry information between the different layers of the LTE protocol stack and can be grouped into either physical, transport and logical channels.

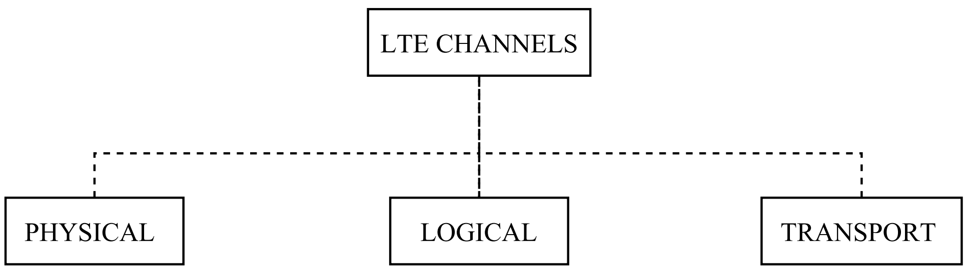

Figure 2.2: Channels in LTE architecture

2.1.2.1 Physical Channel

Physical channels carry user data and control message. A brief description of physical channels existing in LTE is given as follows:

- Physical broadcast channel (PBCH): this operates in the downlink and it broadcasts a limited number of parameters that are essential for the initial access of the cell e.g DL system bandwidth.
- Physical Control Format Indicator Channel (PCFICH) : Information about the format of the signal to be received is sent on the PCFICH. It also indicates the number of OFDM symbols used for the physical downlink control channels (PDCCH).

- PDCCH informs the user equipment (UE) about resource allocation of paging channel (PCH) and downlink shared channel (DL-SCH) and hybrid ARQ information related to DL-SCH. It also gives uplink power control instructions.
- Physical Hybrid ARQ Indicator Channel (PHICH): The Hybrid ARQ status is reported using this channel. The correct receipt of a transport block is indicated using the HARQ ACK/NACK signal and this is carried in this channel. It operates in the uplink.
- Physical Uplink Control Channel (PUCCH) : This channel carries HARQ ACK/NACKs in response to downlink (DL) transmission, Scheduling Requests (SR) and CQI reports
- Physical Uplink Shared Channel (PUSCH): This channel carries the uplink shared channel (UL-SCH).
- Physical Random Access Channel (PRACH): This channel carries the preamble that is used for random access procedure. It is functional in the uplink.

2.1.2.2 Transport Channel

Transport channels provide information transfer to MAC layers. There are a number of channels that are under this category including:

- Broadcast Channel (BCH) : This channel operates in the downlink and is mapped to the Broadcast Control Channel (BCCH).
- Downlink Shared Channel (DL-SCH): This is also a downlink channel that is mainly used to for downlink data transfers.
- Paging Channel (PCH) : This downlink transport channel is used to convey the PCCH
- Multicast Channel (MCH): Also operational in the downlink, this channel helps in the setup of multicast transmissions.
- Uplink Shared Channel (UL-SCH): This uplink transport channel is used for uplink data transfer.
- Random Access Channel (RACH): This is another uplink channel that is used in the random access process.

2.1.2.3 Logical Channel

Logical channels are provided by the Service Access Points (SAP) and they serve as a shield over the data being transmitted. The following channels are classified as logical channels:

- Broadcast Control Channel (BCCH): All UEs connected to an eNodeB are provided system information by this channel.

- Paging Control Channel (PCCH): This channel carries paging information.
- Common Control Channel (CCCH): This channel contains information used for connection setup.
- Multicast Control Channel (MCCH): Multicast reception is a forte of the MCCH. This channel carries required data for receiving multicast transmissions.
- Dedicated Control Channel (DCCH): This channel provides user specific control information.
- Dedicated Traffic Channel (DTCH): This traffic channel is used for the transmission of user data.
- Multicast Traffic Channel (MTCH): Multicast data are transmitted on this channel.

2.2 Network Architecture for IoT Communications

As earlier stated, M2M communication is seen as the bedrock of IoT. Therefore, the 3GPP defines two M2M communication scenarios namely:

1. MTC devices communicating with one or more MTC Server
2. MTC devices communicating with each other

MTC devices communicating with one or more MTC Server

In this scenario, MTC server(s) controls the operations of MTC devices. An MTC user, usually a person or control center, would be in charge of operating the MTC server(s) via an application Programming Interface (API) provided by a network operator.The MTC server could be located in the same operator domain with other MTC communication elements or in a different operator domain. This scenario is depicted in Figure 2.3.

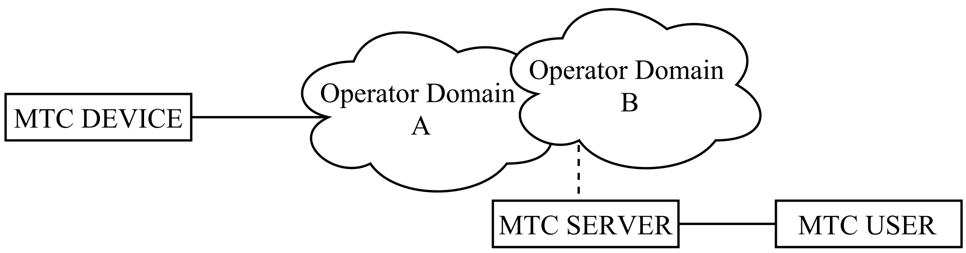

Figure 2.3: MTC communication scenario in LTE

Inter-MTC device communication

This scenario involves communication between MTC devices without any intermediate MTC server. The devices can be located within the same operator domain or in a different domain as depicted in Figure 2.4

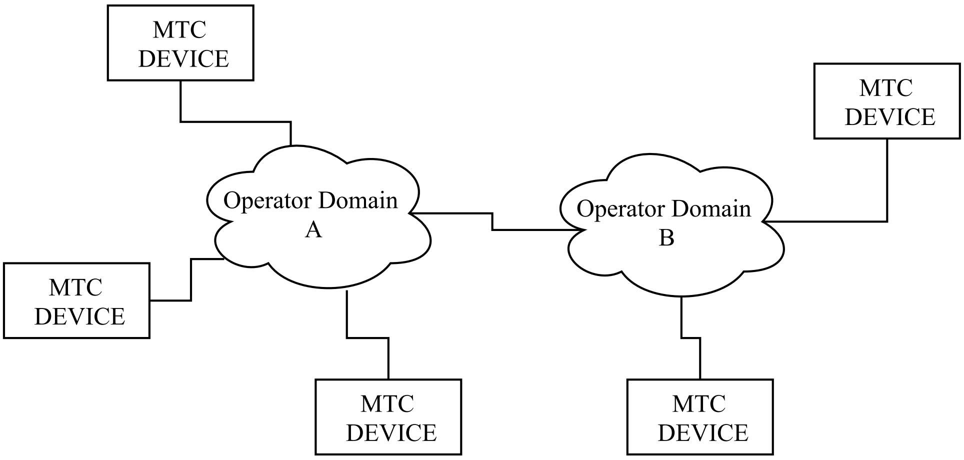

Figure 2.4: MTC devices communicating with each other without MTC server

2.3 3GPP Standards for IoT

Due to the rapid advancements of the IoT, a lot of interest in academia and industry has been generated [22]. However, in terms of standardization efforts, the 3GPP is providing solutions for the potential challenges and adaptability of this budding technology into her existing and future communication technology standards. It is well known that standards provide a common platform for the development and synergy across many industrial players, thus, making the adoption of a technology and solution to many challenges quicker [23]. In release 13 of her standards [24], the 3GPP has made a major attempt to address the IoT requirements using different portfolios of technologies. Although, the works in release 13 has been frozen in June 2016, additional works is still going on for release 14 which will address more requirements for IoT applications using 3GPP technologies. Release 13 also included a number of projects which have direct bearing with the implementation of IoT in 3GPP. The works in this release have generated huge interests in academia but of particular interest to this thesis are the following:

- LTE in unlicensed spectrum (Licensed-Assisted Access) Enhancement for LTE operation in unlicensed spectrum are being studied. Whilst it is preferable to use the licensed spectrum for delivery of advanced services and user experiences, the unlicensed spectrum could also be used to complement the ever increasing traffic demands. Licensed-Assisted Access will provide operators with an opportunity of saving costs, improving spectral efficiency and providing the users with better services by taking advantage of the unlicensed spectrum.

  The works in Release 13 aims to have a primary cell, which will operate in the licensed spectrum and offer guaranteed QoS and a secondary cell which will operate in the unlicensed band to provide higher data rate. However, in order to implement this technology, the coexistence between LTE LAA and Wifi should be as fair as possible.

- LTE enhancements for MTC

  The works to specify key physical layer and RF enablers to enhance the use of LTE for

IoT began in Release 12 [25] of the 3GPP specifications. However, in Release 13, the focus is on defining a new low complexity UE category that can support reduced bandwidth, reduced transmit power, extended coverage and reduced power consumption amongst other things. In terms of coverage, the aim is to improve the coverage performance of MTC devices that can tolerate delays by 15db.

Continuing the normative work started in Release 12 to specify key physical layer and RF enablers to enhance LTE's suitability for the promising IoT market, the key focus for Release 13 is to define a new low complexity UE category type that supports reduced bandwidth, reduced transmit power, reduced support for downlink transmission modes, ultra-long battery life via power consumption reduction techniques and extended coverage operation [26]. The advantage of this is that the network will be able to provide adequate services even for devices that are placed in hard to reach areas - such as meters in basements.

The 3GPP considers three technologies that can be used to address various market requirements and they are the enhanced machine type communications (eMTC), narrowband internet of things (NB-IoT) and Extended Coverage - GSM - Internet of Things (EC-GSM-IoT). An overview of technologies is given in the following subsections.

2.3.1 eMTC

The aim of eMTC is to create a set of physical layer features that can support low cost devices with long battery lives over large distances, variable rate transmission and extended coverage whilst reusing majority of the LTE physical layer procedures [27]. eMTC should be compatible with existing LTE technologies and be capable of being deployed in any existing LTE spectrum, support transmission modes such as, Time division duplexing (TDD), Frequency division duplexing (FDD) and half duplex, and also coexist with other LTE services within the same bandwidth without disrupting their services. From the eNodeB point of view, implementing eMTC is relatively easy as a simple software update will enable existing eNodeBs to be reused for eMTC.

Specifically, in terms of cost, eMTC devices should be about the same price as GSM and GPRS devices, should be able to last about ten years operating with 5 watts of battery power, and transmit at between $10kbit/s$ and $1Mbit/s$ [28]. A new power class of $20db$ is being proposed and it is envisaged that eMTC's narrowband operation will only require 1.08Mhz of bandwidth.

Summarily, 3GPP for eMTC has the following objectives [29]:

- Considering a reduced bandwidth of $1.4MHz$ and a reduced maximum transmit power of $20dBm$, define a new device category that can cater for LTE duplex mode operation of M2M communications based on the Release 12.
- Improve LTE coverage performance by $15dB$ in FDD specifically for delay tolerant applications that are in hard to reach network coverage areas including the device category

defined above.

- Optimise battery life by enhancing discontinuous reception (DRX) cycle to allow for longer periods of inactivity.

2.3.2 NarrowBand - IoT (NB-IoT)

An interesting technology proposed by the 3GPP specifically for the IoT is the NB-IoT. The standardization process for this new narrowband radio technology was recently concluded by the 3GPP in June, 2016 with the changes being implemented into Release 13 of the 3GPP standards. This technology is becoming quite popular in industry and has also generated a lot of comments form major players [30] [16].

Similar to eMTC, the idea behind NB-IoT was to support key requirements in supporting IoT such as flexibility in deployment, low device complexity, extended battery life, massive device support in a cell, and significant coverage extension beyond existing cellular technologies. However, NB-IoT was specifically designed for ultra-low-end IoT applications.

The minimum bandwidth requirement for NB-IoT is $180kHz$ for both uplink and downlink allowing for three distinct deployment options. For co-existence with GSM, a GSM carrier ($200kHz$ bandwidth) can simply be replaced with NB-IoT while an LTE operator can simply allocate one of the Physical Resource Blocks (PRB) of $180kHz$ to it. These are scenarios of Inband deployment of NB-IoT, there is also the scenario in which NB-IoT is deployed in the guard band of the LTE carrier and a third option where NB-IoT is deployed as standalone where the technology is allocated in its own frequency spectrum.

As with many of the evolving technologies of the 3GPP, there is extensive reuse of existing technologies. Thus, LTE designs such as downlink OFDMA, uplink SC-FDMA, interleaving and channel coding amongst others are used in NB-IoT. This is one of the reasons why the core specifications of NB-IoT in 3GPP was completed in good time. However, commercial deployment of NB-IoT products and services is expected to be around the end of 2016 and the beginning of 2017.

Although, NB-IoT is not fully backward compatible with existing 3GPP devices , it is still able to achieve great co-existence with LTE, General Packet Radio Services (GPRS) and legacy GSM technologies [31]. An interesting area of distinction between NB-IoT and LTE is in their initial network access protocols. Whilst they both use contention based random access protocols, LTE is based on selection of preambles generated using Zadoff-Chu root sequences, NB-IoT uses single or multi-tone transmission of preambles which are based on frequency hopping [32].

2.3.3 Extended Coverage - GSM - Internet of Things (EC-GSM-IoT)

The EC-GSM-IoT aims to support longer battery life of around 10 years of operation and low device cost when compared to GPRS/GSM devices. Extended coverage and variable rates of up to 240Kbps are also included, as well as the ability to support massive numbers of devices (50,000 per cell). This technology is to operate at a of 200KHz per channel with system bandwidth of

2.4MHz [33].

EC-GSM-IoT proposes new logical channels for coverage extension and overlaid Code Division Multiple Access (CDMA) to increase cell capacity. Extended DRX, optimized system information, storage and use of coverage levels in Serving GPRS Support Node (SGSN) to avoid unnecessary over the air repetitions are also features included in this technology. Also, there is consideration for much greater security with stronger ciphering algorithms, mutual authentication and integrity protection being looked into [34].

However, works for this technology is to be included in Release 14 with the objective of specifying radio interface enhancement that allows the use of alternative mappings of blind physical layer transmissions, positioning supports and on all uplink channels, a minumum of 3db Minimum Coupling Loss (MCL) improvements should be achieved.

The three technologies discussed are compared in table 2.1.

Table 2.1: Comparison of 3GPP IoT technologies

	eMTC	NB-IoT	EC-GSM-IoT
Deployment	In-band LTE	standalone, guard band and In-band	In-band GSM
Downlink	OFDMA	OFDMA	TDMA/FDMA, GMSK and 8PSK (optional)
Uplink	SC-FDMA	Single tone, Turbo code and SC-FDMA	TDMA/FDMA, GMSK and 8PSK (optional)
Bandwidth	1.08 MHz	180 KHz	200 KHz per channel. system bandwidth 2.4MHz
Peak Data rate (DL/UL)	1 Mbps DL and UL	DL of 50 kbps and UL - 50 kbps multitone and 20 kbps single tone	DL and UL - 70 kbps using GMSK, 240kbps using 8PSK
Duplexing	FDD and TDD	FDD	FDD
Power class	23 dbm, 20 dbm	23 dbm	23 dbm, 33 dbm

2.4 Key Issues in 3GPP Standards for IoT

There are a number of challenges that plague cellular media implementation of M2M communications [35], however, in this section we focus on the those identified by the 3GPP. Based on the service requirements treated by 3GPP SA WG1, the system architecture working group (3GPP SA WG2) came up with key functional issues that were documented in [25]. Figure 2.5 shows some of the key architectural issues. A brief description of each issues is described below.

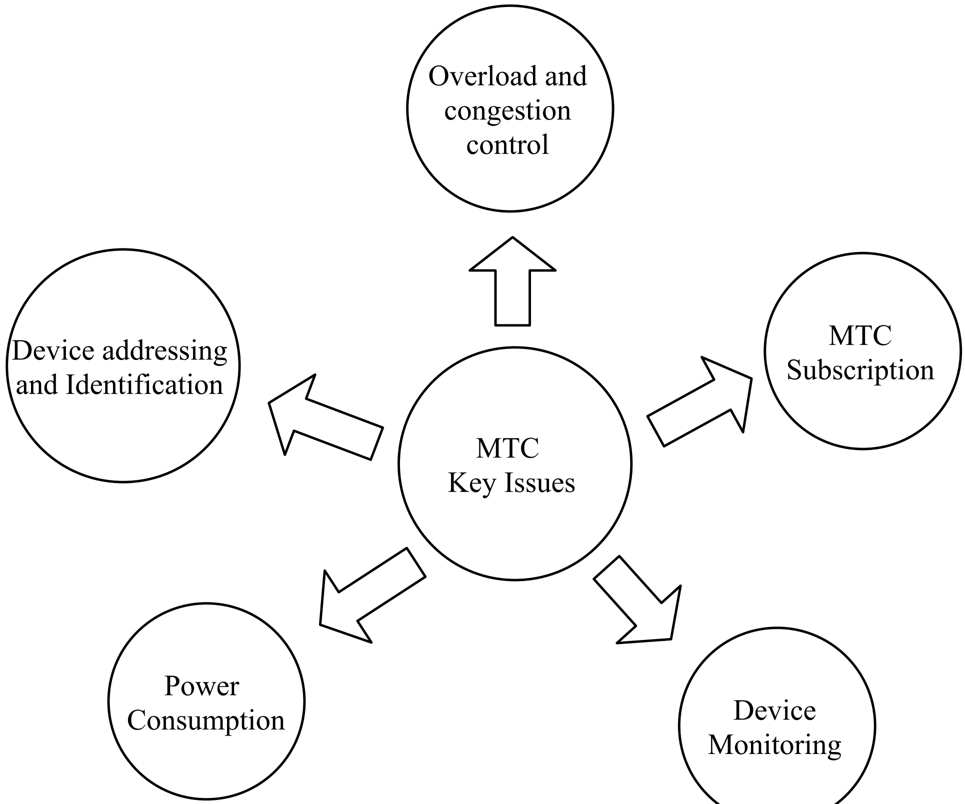

Figure 2.5: Challenges confronting MTC in LTE

Overload and congestion control

Signalling congestion and overload issues could arise from massive MTC devices in the network. This can spiral into the entire network leading to an overload. Thus, ways must be sought to curtail congestion and overload in the network.

Device Addressing and identification

Some MTC devices may use private IPv4 addresses which are non-routable, thus, making such devices unreachable. How will MTC devices be addressed? Even as Mobile Subscriber Integrated Services Digital Network Number (MSISDNs) are running short.

Power consumption

MTC devices should consume as low power as possible. This challenge is still an open area of research in industry and academia.

Device monitoring

It should be possible to monitor MTC devices placed in area of high risk. Ways in which this is to be possible is still a challenge.

MTC Subscription

Activation and deactivation of MTC features should be possible. This could be controlled either from the MTC server or device end. However, an MTC device should not be able to activate a feature which it is not subscribed to at the network end.

2.5 Random Access in LTE

The random access procedure is used by devices seeking initial network access and also during handovers. The RACH acts as an interface between non-synchronized devices and the orthogonal transmission scheme of the LTE uplink radio access. Only devices that have their uplink transmission timing synchronized with the eNodeB are allocated resources. Apart from initial network access, handovers, device positioning (timing alignment) and device transitioning (from idle to connected states) also require the completion of the RACH process. Therefore, RACH is a critical gateway to the efficient operation of the network. The RACH process can either be initiated by the device (contention based) or the eNodeB (contention free).

A summary of the steps involved in the contention based RACH procedure is shown in Figure 2.6 below.

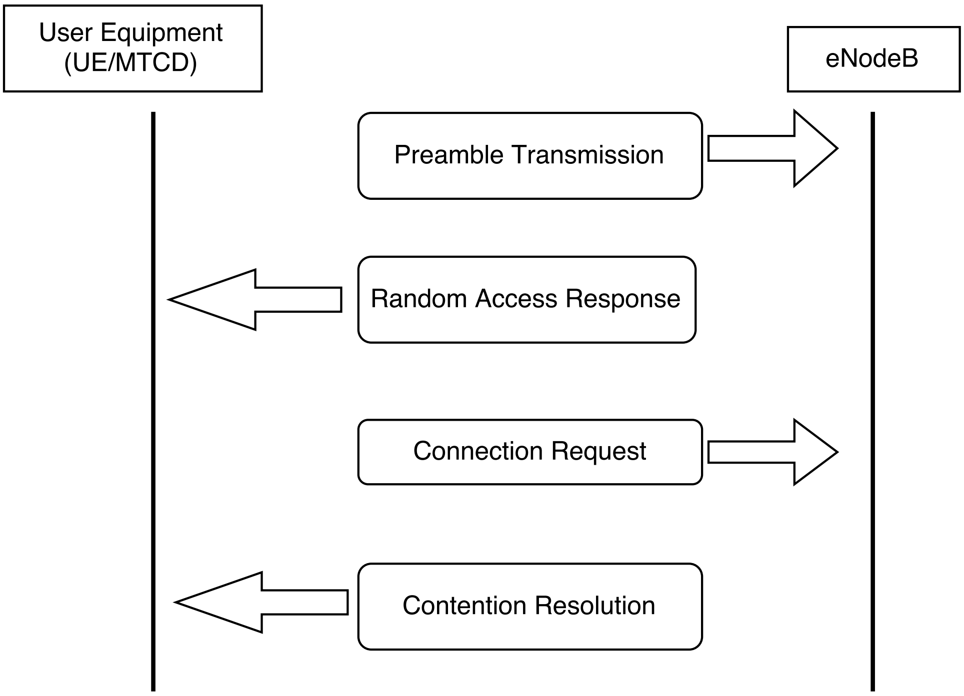

Figure 2.6: Random Access procedure in LTE

The steps are further explained as follows:

- Selection of a random access preamble from a pool of available (equal selection probability) preambles for onward transmission over the PRACH channel to the eNodeB.
- Upon receipt of a preamble, eNodeB sends RA response (RAR) containing uplink synchronization parameters including an initial resource grant for transmission of the step 3 message. If multiple devices select the same preamble, they are sent the same RAR.

- If the RAR is successfully decoded, a L2/L3 message containing the actual random access message, such as a radio resource control (RRC) message or scheduling request is sent by the device on the uplink time-frequency resource allocated to it. Otherwise, the device restarts the RA process.

- A contention resolution message is sent by the eNodeB. This is simply an echo of the UE identity contained in the L2/L3 message. Devices able to detect their own identity send feedback to the eNodeB, other devices understand there is a collision and they restart the RA process again.

The receipt of similar RAR by multiple devices in step 2 is the first indication of a preamble collision which can go undetected until the fourth step of the RA process. Early detection and avoidance of selection of the same preamble can be seen as a major step towards efficient prevention of preamble collision.

2.5.1 Random Access Improvements

Due to the strategic importance of the random access channel (RACH) of the LTE when devices connect to the network, there has been a surge of interest in academia in resolving the potential bottlenecks that could likely occur through massive device access attempt. Various solutions have been proposed and there are a number of similarities between certain groups of solutions being proposed. Thus, the solutions could be classified broadly under five categories as shown in Figure 2.7.

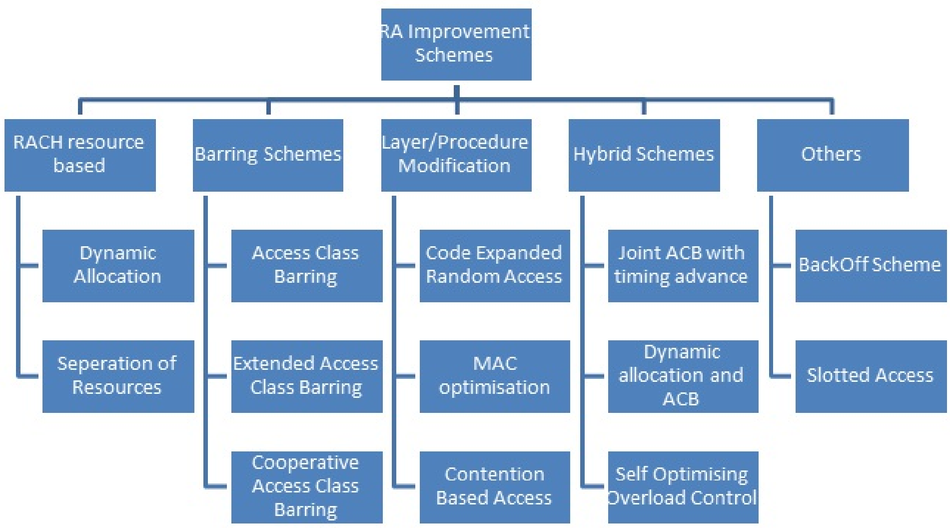

Figure 2.7: Classification of Random Access overload schemes

A detailed description of proposals in each groups is given below:

2.5.2 RACH resource based schemes

Protocols that are based on a certain distribution or allocation of RACH resources are grouped under this heading. Allocation could be dynamic and resources might be for the exclusive use of a certain class of devices accessing the network. A typical example of resources as used here is the RA preambles and the random access slots.

- Dynamic allocation: if the network is able to predict a surge in access traffic due to the presence of MTC devices (for example, thousands of smart meters sending their readings at a particular time), then the eNB can dynamically allocate resources to cater for the escalation in access attempt [37]. Simulation results in [38] shows that this scheme would be beneficial in resolving the RACH congestion with minimal impact on human to human (H2H) communications. However, a major issue to be considered is the tradeoff between the number of random access slots (resource) available in a frame and the number of slots that are available to carry data. Based on current frame structure, an inverse relationship would exist between slots reserved for random access contention and those reserved for data transmission. The more the number of slots for random access, the lesser the number of slots available for data transmission. This would significantly affect the performance of this scheme when applied in high data traffic load cases.

- Separation of Resources: Resources could be split in such a way that a particular resource or group is only for the exclusive use of a class of device accessing the network. Preambles could be split such that human to human communication devices and machine to machine communication devices use different group of preambles in accessing the network [37].[39] considered two methods that preambles could be split in, the aforementioned method and a method that involves giving priority to H2H devices by having a group of preambles for them whilst the other group of preambles would be available for use by both H2H and MTC devices. The key issue that arises with this kind of method is that there would be a reduction in resources available for any of the classes of devices that are accessing the network and this will impact on the network performance. Thus, under high loads of MTC devices, the system would suffer from severe degradation.

2.5.3 Barring Schemes

Barring devices that do not meet a certain criteria from accessing the network, at a time, is also a subset of proposed solutions. MTC devices can be grouped into classes based on quality of service (QoS) requirements or the urgency at which they need to access the network [25]. The 3GPP specifies access class barring as a means of mitigating network overload due to the surge in the number of devices contending for network access. Access class barring schemes follow a simple protocol in which the eNB broadcasts an access barring parameter which devices have to overcome before gaining access to the network. The access barring parameter is usually a value between 0 and 1 and devices generate a random number which is compared to the access parameter. The access barring parameter could be flexibly controlled by the eNB to cope with the prevailing network load at any particular instance. However, there have been a number of researches on using access class barring schemes to mitigate RACH overload and this has led to a number of variants which could be grouped as either one of the following:

- Individual Access Class Barring (IACB): This follows the aforementioned process without any modification and there is no differentiation between devices. This scheme improves the probability of successful access but delay sensitive applications would suffer under this scheme as access delay becomes a major concern because devices have to generate a suitable number before being granted contention access.
- Extended Access Class Barring (EACB): The need for a fine grain control of devices led to an extension of the ACB scheme in which delay intolerant classes of devices are given priority to contend for network access. The EACB follows the same method as the ACB but with the addition of classes to which devices may belong to. It is considered to be the most feasible baseline solution adoptable for network overload control [37]. Whilst there is a higher probability of delay intolerant devices gaining access to the network, this still comes at the price of a delay in network access with delay tolerant devices having to wait more.
- Cooperative access class barring: This scheme aims to take advantage of the various base stations that may be present in an LTE-Advanced system, for example, femtocells and picocells. [40] argues that with cooperation amongst base stations, stabilization and access load sharing would overcome the defects inherent in the legacy ACB scheme, particularly, improving access delay. However, in the absence of smaller base stations (picocells), the scheme becomes largely similar to the legacy barring schemes.

2.5.4 Layer/Procedure modification schemes

Some proposals argue that changes should be made to the RACH procedure. Whilst some solutions consider a partial modification of the procedure, others opt for a whole new approach to the contention based random access process. Representative of these proposals are:

- Code Expanded Random Access: This scheme seeks to introduce the concept of virtual frames that consist of random access slots only and an expansion of the contention space through the creation of codewords. Codewords are transmitted at each virtual subframe by MTC devices [41]. The major advantage of this scheme is that no extra resource is needed to increase the amount of contention resources. However, the decision as to whether an MTC device experiences collision is usually taken at the completion of a virtual frame and this could introduce some delays in network access. Also, the continuous transmission in every subframe raises the question of the energy efficiency of the proposed system.

- MAC Optimization: This focuses on changing the way machines with small amount of data to transmit data by changing the MAC layer protocol [42]. The idea is to transmit data either in message 1 of the random access procedure or in message three thus reducing the amount of resources that would actually be needed to transmit data. However, there is no mathematical analysis or simulation to prove the effectiveness of this method.
- Contention Based Access: proposes a RA procedure that is completely different from the existing one. In the scheme, MTC devices transmit packets without having to go through a contention phase as packets are transmitted on resources (slots) that are randomly selected by the device. Collision is taken care of by the use of MU-MIMO to detect the RNTI of collided MTC devices and special resources are scheduled for retransmission of collided packets [43]. A major plus of this scheme is that it cuts down signalling overhead required by devices to access the network and there is no need for additional resources in its implementation. There is also reduction in access time as devices no longer undergo a contention phase. However, the effectiveness of the scheme in combating collision is dependent on the number of receiving antennas present in the MTC device. More antennas could lead to an increase in energy consumption. Also, changes required in the eNB have not been fully discussed.

2.5.5 Hybrid Schemes

These are solutions that combine a number of other solutions either from the ones listed above or otherwise, into one singular solution.

- Joint ACB with timing advance: makes use of the ACB scheme described above and timing advance in order to combat the RA overload. A dynamic optimal barring parameter, that, when used in conjunction with the timing advance reduces the number of RA slots required to serve MTC devices in a particular coverage area [44]. This scheme has the advantage of not requiring extra resource for implementation. Also, a dynamic barring parameter would imply that the scheme would be able to adjust to load variations. However, the fact that this scheme is still based on the ACB means that the inherent problems of the ACB such as access delay would also be present here. Also, there is no consideration of different classes of MTC devices that may need to access the network as obtainable in the real world.
- Prioritized Random Access: This could be thought of as a combination of resource allocation and access class barring scheme. The scheme classifies MTC devices based on the urgency in which they need to access the network and allocates a set of resources for the different classes. Five different classes are specified and a dynamic ACB is used to

cater for a situation of heavy access load [45]. Simulation results show an improvement in access success probability and access delay.

- Self-Optimizing overload control: This scheme is a combination of RA separation scheme, the ACB, slotted access and the p-persistent scheme and it operates by continuously adapting network resources and parameters based on the prevailing load conditions. It allows the network to reduce or increase PRACH resources when it detects an increase or decrease in PRACH load [46]. The scheme is quite flexible has it seems to cater for a variety of likely scenarios that could occur in the network, however, there have been no numerical or simulation proofs as to the effectiveness of the theoretical claims of the proposed scheme.

2.5.6 Others

Another idea mooted in literature is the allocation of identical preambles to groups of machine devices that may be some away from each other by some distance. [47] and [48] proposed quite similar ideas that involves the spatial partitioning of a cell into multiple groups and evaluated only the collision probability performance. Their proposal may reduce the number of cyclic shifts required during preamble generation. But it is noted that the research works presented in the above papers are focused on the user centred grouping and preamble reuse, which is in contrast to our eNodeB centred grouping and preamble reuse. We believe it is more natural and scalable to use eNodeB centred grouping, as the IoT devices are associated to the eNodeBs and access to the networks through the eNodeBs. With our eNodeB centred grouping and preamble reuse, the network management overhead and preamble detection performance can be controlled as expected. For the user centred group, there are several major technical problems, including at least:
a) the user grouping requires the existence information and the positions of the IoT devices in the networks, which is very hard to be obtained for massive IoT device. Therefore, their proposal is not scalable, and it is not able to keep preamble interference under control to ensure preamble detection performance;
b) the user centered preamble reuse method does not address the network wide interference coordination problem, which leads to uncontrolled inter-cell preamble interference and reliable preamble detection becomes a big problem;
c) the implementation of their proposed scheme only considers a single cell scenario. A network wide implementation of their proposal will pose a number of challenges.

Other solutions that do not fit into any of the aforementioned come under this category. They include:

- BackOff Adjustment: this schemes aim to alleviate RACH overload by adjusting the back-off timer. Simulation results in [49] show improvement in rach congestion but only for low congestion cases. These schemes would be unable to cope with cases of heavy congestion in the network and would result in an increase in collision probability and access delay times.

- Slotted Access: This scheme involves dedicating RA slots for each MTC device to access the network. A RA cycle and the device identities help MTC devices to calculate their RA slot which they can transmit on [37]. Dedicating RA slots to devices reduces drastically the probability of collision, however, allocating dedicated RA slots per device would mean that in there has to be enough RA slots for every device trying to gain network access else insufficient RA slot would lead to delays in devices accessing the network.

2.6 Conclusion

Cellular communication technologies are preferred by a large number of business operators for implementing long range communications due to their ubiquity and popularity amongst others factors, although, legacy implementations such as LoRa also exists.

An in-depth overview of the existing LTE implementation has been given and how it is being adapted for use with MTC (which is a backbone of the IoT). Challenges being faced by the adaptation of existing LTE implementation for MTC are also discussed. These challenges including massive device access, energy consumption, and signalling congestion were examined.

Due to the sheer numbers of devices that would require communication in IoT, massive device access is an area of research that is generating a lot of interest in industry and academia. Thus, the random access channel plays an important role in the conquest of this challenge. We discuss the random access channel in details and also review the state of the art in literature.

In order to proffer solutions to this challenge, we propose a number of ideas which are presented in different chapters of this work. In chapter 5, we investigate how small cells are being used to combat massive device access and also propose a solution based on subframe resource optimization. Chapter 4 also presents a root index allocation scheme to help mitigate congestion that could arise due to massive device access in LTE.

3 IEEE 802.11 Technologies in IoT

With tremendous growth, widespread acceptance and use of various wireless technologies increasing over the years, their development have followed similar trend. Of particular interest is the IEEE implementation of wireless technologies.

The IEEE, through her standards committee, provides specifications that govern communication in various media. The Local/Metropolitan Area Network (LAN)/(MAN) standards committee of the IEEE created and maintains the 802.11 WLAN standard. In particular, IEEE 802.11 is a group of media access control (MAC) and physical layer (PHY) rules that govern wireless device communication in the 900 MHz and 2.4, 3.6, 5, and 60 GHz frequency bands. With its first version released in 1997, subsequent amendments have been proposed and implemented [50].

In particular, IEEE 802.11 is becoming quite a popular implementation technology option for consideration in IoT communications. This is because, Wi-Fi is ubiquitous and already well understood both in industry and academia. It also scales well in terms of data rate which can be crucial in determining power consumption.

In this chapter, we provide an in-depth explanation of the background theory of the IEEE 802.11 wireless technology, a state of art review, including current developments, challenges, and research directions. Of particular interest is the rate adaptation challenge in wireless networks which is also discussed in detail. The remainder of this chapter is organised as follows, we discuss the system architecture of IEEE 802.11 technology in section 3.1 and then move on to discuss the IEEE 802.11 standard for IoT in section 3.2. The key challenges confronting IoT implementation using IEEE 802.11 communication technology are presented in section 3.3 whilst rate adaptation in IEEE 802.11 wireless technology is discussed in section 3.4. This chapter is then concluded in section 3.5.

3.1 System Architecture of IEEE 802.11

3.1.1 Physical Architecture

In terms of physical architecture, the 802.11 supports two topologies namely:

- Basic service set (BSS)
- Extended service Set (ESS).

Basic service set (BSS)

The Basic Service Set (BSS) is the fundamental building block of 802.11 physical architecture. A BSS could be thought of as a group of devices that can communicate with each other directly. A BSS also covers a geographical area called the Basic Service Area (BSA) and connection to the wireless BSS is impossible outside the BSA. BSS are classified as either independent or infrastructure.

Independent BSS also known as an Ad-hoc network, does not require an access point for communication between nodes and traffic is channelled directly between nodes themselves while an Infrastructure BSS comprises of both nodes and Access Point (AP). The AP acts as an intermediary and all communications in goes through it. Information destined for a node is sent to the AP which then forwards it to the intended recipient(s) [51].

Extended service set (ESS)

An AP is also involved in this architecture and range extension is provided by connecting two or more BSSes, thus forming an extended service set. A backbone network, called a distribution system, whose primary responsibility is MAC level transportation of MAC service data units connects BSSes together. This distributed system can be a wired IEEE 802.3 Ethernet LAN, a MAN or another 802.11 wireless medium. An ESS can also serve as a gateway to a wired network, even the Internet. This can be achieved via a device known as a portal. The portal is a logical entity that specifies the integration point on the distribution system where the IEEE 802.11 network integrates with the non IEEE 802.11 network [51].

Figure 3.1: Topology of 802.11 Wireless Network

Figure 3.1 gives a general overview of the topology of a wireless network. An ESS and BSS each consisting of three stations and an access point is shown. The access points are connected together by means of a distribution system. The BSS shown is an infrastructure BSS whilst the ESS consists of the whole network including the DS as earlier described.

Layers	*Function*
Application	Resource sharing and network management
Presentation	Data conversion, encryption and compression
Session	Session establishment, maintenance and termination
Transport	Message segmentation, acknowledgement and flow control
Network	Routing and frame fragmentation
Data Link	Media access control
Physical	Physical medium attachment and transmission techniques

Table 3.1: The OSI model layers including their functions

3.1.2 Logical Architecture

A logical architecture describes network operation specifically on a layer level [52]. The Open Systems Interconnection (OSI) model describes network communications on a layer level. It consists of seven layers with each performing specific functions, a few of which are as shown in Table 3.1. The PHY and MAC layers is at the core of IEEE 802.11 networks and is discussed in more details in this section.

3.1.3 The Medium Access Control Layer

Apart from providing a means of accessing the wireless medium, the MAC layer also carries out important functions which includes frame check sequence generation, addressing, carrier sensing, and inter-frame spacing. During transmission in wireless medium, the carrier sense multiple access with collision avoidance (CSMA/CA) is used and this is because it is impossible to detect a collision in wireless networks. A collision occurs when two nodes transmits data at exactly the same time, through the same medium. Typically before any node transmits data on a wireless medium, it listens to the medium to see if there is any transmission going on (carrier sensing). If the medium is idle, the node waits then waits for a predetermined period of time before it actually begins its own transmission. This wait period help minimize collision on the medium [52]. The MAC layer also defines special functional behaviour to reserve the medium using request to send (RTS) / clear to send (CTS) polling interaction and point coordination. Packet fragmentation is also considered at this layer . The Link Layer Control (LLC) is a sub layer of the MAC layer which is responsible for channel allocation procedures, error checking, frame formatting and a host of other functions. A standard frame format for the MAC layer consists of the frame control field, duration Identifier, destination and receiving address, sequence control, the data and a frame check sequence [50]. This, including the size of each field (in bytes), is shown in Table 3.2 below.

Frame control	Duration ID	AD 1	AD 2	AD 3	Sequence Control	AD 4	Data	FCS
2	2	6	6	6	2	6	0-2312	4

Table 3.2: MAC Frame Format

The frame control is a two byte long field that gives information about the protocol version.

The duration ID is also 2 bytes long and it gives an estimate of the duration of the expected frame. There are four address block, labelled AD 1, 2 3 and 4 in Table 3.2, each with a length of 6 bytes and they represent the MAC addresses of the initial sending device, final destination of frame, next immediate station to receive the frame and the address of the station that transmitted the frame respectively. The Data field carries data between 0 and 2312 bytes long whilst the frame check sequence (FCS) is provided in a four byte format.

Generally, transmission in IEEE 802.11 makes use of carrier sense multiple access with collision avoidance (CSMA/CA) with positive acknowledgement and retransmission. A random back off timer is used to guard against simultaneous sending of frames by nodes in a wireless network.

3.1.3.1 MAC Layer Functions

The functions of the MAC layer include the following:

1. **Inter-frame spacing:** During transmission of frames in wireless medium, there is a time period between the sending of successive frames. The minimum idle period between transmissions of Ethernet frames is referred to as Inter frame Spacing (IFS). Depending on the events that occur during the transmission, there are different types of IFS namely:

   Short Inter Frame Spacing (SIFS): this is the smallest duration between the transmission of a frame and the acknowledgement of delivery of the frame.
   Point Coordination Function Inter frame space (PIFS): is the sum of the SIFS and the slot time.
   Distributed Coordination Function Inter frame space (DIFS): this is the sum of PIFS and the slot time. The slot time is the basic unit of timing in the wireless protocol and it varies based on the 802.11 standard being considered.

2. **Collision Avoidance:** Whenever the channel is busy, nodes that intend to transmit wait a DIFS period, after which they calculate a random back off time and wait till it gets to zero. The node whose back off time first gets to zero has the privilege of transmitting. Other nodes will have to wait till the transmitting station finishes its transmission. Usually, the random back off time is an integer value that corresponds to a number of time slots usually between 0 and 7 [52]. That idle period after the DIFS is referred to as contention window.

3. **Distributed coordinated function (DCF):** The DCF is the fundamental access method that is used to support asynchronous data transfers in 802.11 networks. In networks that do not have an access point (ad-hoc networks), the DCF is the choice access method. Since there is no coordinator in ad hoc networks, control is distributed between nodes depending on which of the nodes is transmitting. The transmission of a frame still follows

the process aforementioned. Priority access to the medium is determined by the value of SIFS, PIFS or DIFS in that order with the DIFS having the least priority.

4. **Point Coordinated Function:** In networks that have an access point (Infrastructure networks), control of data transmission can be carried out by the access point. Thus, there is a coexistence of the DCF and PCF in this kind of network. The PCF is, however, an optional feature and it provides connection oriented and contention free frame transfer. A point coordinator performs polling and there is no channel access contention during this period which is known as the contention free period (CFP). The frequency of occurrence of the PCF period is determined by a rate called the CFP repetition rate.

   Prior to the commencement of a contention free period (CFP), the point coordinator waits a DIFS period after when the medium is idle and then transmits a beacon frame to start the CFP. The CFP is started after waiting an SIFS interval after the beacon frame is sent out by sending out a CF-Poll (no data), DATA or DATA + CF-Poll frame. The point coordinator can terminate the contention free period by sending a CF-END frame.

Carrier Sense Multiple Access with Collision Avoidance

Collisions cannot be detected in wireless networks, so the CSMA/CA access method is used. It consist of nodes transmitting only when the medium is sensed to be idle. The protocol can simply be broken down into two steps namely:

- Channel Sensing: Every node in the network has to sense the channel to ascertain whether there are other nodes transmitting or not.
- Collision Avoidance: In event of the channel being busy, nodes that have data to transmit waits for a time period, random back off time. At the expiration of the wait period, the node attempts senses the medium again to ascertain if it is free.

3.1.3.2 MAC Layer Challenges

The physical medium of wireless network makes it more challenging than traditional wired ones whose range is easily defined. There are three important issues that 802.11 wireless networks face namely:

- Half duplex operation
- Time varying channel
- Bursty channel errors

Half duplex operation

During transmission, a large proportion of signal energy leaks into the path of the receiver. This leakage signal typically has much higher power than the recieved signal thus making simultaneous reception and transmission of signals impossible. Thus, 802.11 wireless nodes operate in half duplex mode, where transmission of signals is in one direction at a time. Signal collision is not detectable by the sending node in wireless medium, thus, proposed protocols attempt to avoid collision by reducing the likelihood of a collision occurring. Specifically, 802.11 networks operate using the carrier sense multiple access with collision avoidance (CSMA/CA).

Time varying channel

As signals propagate through multiple paths to arrive at their destination, they are shifted in time and attenuated. The receiving node gets a superposition of these changed signals. However, in order to ensure that the link quality is capable of communication, handshaking is widely used. when a nodes received signal strength falls below a threshold, it is said to have faded. The challenge is that as the signal travels in time, its strength varies and this could introduce errors into the signal making it impossible to be decoded at the receiving end.

Bursty channel errors

This is a natural consequence of time varying channel. Errors can be introduced into the signal during transmission. Although, error correction schemes exists, they may not be able to correct all errors. Therefore, packets may be lost due to a significant number of errors in the bits. Several mitigation techniques such as smaller packets, forward error correction codes and packet retransmissions are popular techniques used to mitigate loss of packets due to bursty errors.

3.1.4 The Physical Layer

The physical layer is the first point of contact for any node attempting communication in a network. Thus, it only makes sense that technologies associated with the way wireless nodes initially access the transmission media are first implemented here. Such technologies may include the direct sequence spread spectrum (DSSS), orthogonal frequency division multiplexing (OFDM) and the frequency hopping spread spectrum (FHSS). The initial 802.11 standard operated at the 2.4 GHz of the Industrial Scientific and Medical (ISM) frequency and only supported transmission at both 1 and 2 Mbps.

Devices using FHSS hop through frequencies at predetermined pattern and set rate. FHSS further divides an available spectrum into non-overlapping channels. For example, the 2.402-2.480 GHz frequency range is divided into 79 non-overlapping channels with each channel being 1 MHz wide. In contrast, DSSS performs modulation by using a spreading code resulting in the transmitted signal being spread over a large bandwidth, thus, allowing multiple WLANs to operate within the same coverage area. On the other hand, OFDM uses a large number of closely spaced orthogonal sub-carriers to transport data on several parallel channels which are usually at different frequencies in order to help reduce interference and crosstalk [50]. Over the years, there has been advancement both in terms of the development of 802.11 standards and the data transmission rates up to about 54 Mbps.

A summary of popular 802.11 standards is given in Table 3.3.

Standard	Data Rate (Mbps)	Frequency (GHz)	Modulation	Bandwidth (MHz)
a	6, 9, 12, 18, 24, 36, 54	5	OFDM	20
b	1, 2, 5.5, 11	2.4	DSSS	20
g	6, 9, 12, 18, 24, 36, 54	2.4	OFDM , DSSS	20
n	7.2, 14.4, 21.7, 28.9, 43.3, 57.8, 65, 72.2, 15, 30, 45, 60, 90, 120, 135, 150	2.4	MIMO-OFDM	20,40
p	3, 4.5, 6, 9, 12, 18, 27	5.9	OFDM	5, 10, 20

Table 3.3: Summary of popular IEEE standards

Modulation and Coding Schemes

Environmental factors such as noise, interference and multipath fading affect the channel condition of any wireless network. Thus, data transmission needs to be done in such a way that the information being sent is capable of being received and decoded correctly by the recipient. MCS helps determine the data rate to be used in a wireless communication. It takes into account the channel conditions among other things. Modulation is the process of applying a stream of bits to a carrier at the operating frequency band and can be carried out at either the amplitude or phase level. Different modulation techniques including Binary phase shift keying (BPSK), Quadrature Amplitude modulation (QAM) and Quadrature phase shift keying (QPSK) are widely used in communication systems today. Since each 802.11 standard have different data rate options, their MCS also varies. Table 3.4 shows some of the MCS used in 802.11a today. It is obvious that the different data rates use varying constellation schemes.

Data rate (Mbps)	Constellation Scheme	Coded Bits per Sub Carrier	Data Bit per OFDM Symbol	Coded bits per Symbol	Convolution Rate Coding
6	BPSK	1	24	48	1/2
18	QPSK	2	72	96	3/4
36	16-QAM	4	144	192	3/4
48	64-QAM	6	192	288	2/3

Table 3.4: Summary of MCS used in 802.11a

Channel Errors

The amount of data that can be received over a link is based on the number of bits that can decoded correctly. Bit Error rate (BER) is a ratio of the number of bits received in error to the total number of bits transmitted over the link. Thus, the number of bits received correctly over a link is the product of the symbol rate, bit per symbol and the probability that a bit is correctly received which is given by (1-BER). In wireless medium, bits are usually grouped into packets and this allows for sharing of the transmission media by the nodes in the wireless system.

Similar to the bit error rate, the packet error rate (PER) is the ratio of incorrectly received packet by a receiver to the total number of packet that was sent by a transmitter. A packet is deemed to contain errors if at least one bit in the packet is an error. Thus, we can mathematically represent the packet error probability for a packet containing M bits by equation 3.1:

$$P_{\mathrm{p}} = 1 - (1 - P_{\mathrm{e}})^M \tag{3.1}$$

Where P_{p} is the packet error probability, P_{e} is the bit error rate and M the number of bits in the packet.

Channel errors in a system can affect the overall throughput. By definition, the rate of successful message delivery over a communication channel is referred to as throughput. In wireless systems, the bit error rate and data rate are crucial determinants of this metric. Mathematically, we can represent the throughput, T, of any wireless system as shown in 3.2

$$T = (1 - BER)^M \times D \tag{3.2}$$

where D is the data rate. Both T and D are measured in frames per second.

The BER is used due to the fact that data rate is measured in bits transferred per second. These metrics, throughput and error rate, helps determine the health of any wireless network.

802.11 Standards

A few of the existing 802.11 standards are briefly discussed here.

1. **802.11b:** This was introduced in 1999 and it made use of high rate direct sequence spread spectrum (HR-DSSS) which provides higher data rates of 5.5Mbps and 11Mbps in the 2.4GHz band. It is based on complementary code keying (CCK) as its modulation scheme. The HR-DSSS is similar to the DSSS and it uses the same channelization scheme with 22MHz bandwidth and 11 channels, with 3 of these channels non-overlapping [52]. The 802.11b is compatible with earlier 802.11 standards.

2. **802.11a:** The 802.11a was introduced at about the same time 802.11b was also introduced. It is based on the orthogonal frequency division multiplexing (OFDM) and it operates in the 5GHz frequency band (5.725 - 5.825GHz). The 802.11a utilizes 20MHz channels and provides a mandatory data rate of 24Mbps and optional rates of 54Mbps. The basic principle of the OFDM is to split high data stream into a number of lower data streams that are transmitted simultaneously over a number of sub-carriers. Inter symbol interference (ISI) is taken care of by the introduction of guard time for every OFDM symbol.

   The fact that 802.11a operates in the 5GHz frequency gives it an edge over the 802.11b. This is because the 5GHz has little interference when compared to the 2.4GHz which has a number of devices operating at its frequency. Although, this high frequency comes at the cost of distance, as the 802.11a signals cannot travel as far as the 802.11b signals and they are easily absorbed by obstructions such as buildings.

3. **802.11g:** This standard was approved in 2003 and it provides for data rates up to 54Mbps in the 2.4GHz frequency band. It is backward compatible with 802.11b and uses orthogonal frequency division multiplexing (OFDM) for data rates of 6, 9, 12, 18 and 54Mbps. It also uses the complementary code keying (CCK) for data rates of 5.5 and 11Mbps.

4. **802.11p:** This is an amendment aimed at adding wireless access to vehicular environments (WAVE). It defines enhancements to 802.11 required to support Intelligent Transportation Systems (ITS) applications [53]. WAVE is a mode of operation for use in environment where the physical layer properties are rapidly changing and where very short duration communication exchanges are required. Dedicated applications such as toll collection, red light duration broadcast at traffic lights or hot spots for transferring maps all make use of the WAVE standard. Presently, the approved 802.11p amendment was published in 2010 and it was allocated to work in the 5.9GHz band for inter-vehicle to infrastructure communications and priority road safety applications [53]. The 802.11p standard is meant to:

   - Describe the functions and services of WAVE compliant stations to operate in a rapidly varying environmental condition and exchange messages without having to join a Basic Service set (BSS), as in the traditional wireless cases.
   - Define the wave signalling technique and interface functions that are controlled by the IEEE 802.11 MAC.

   The 802.11p and 802.11a standards are very similar to each other, they use the same modulation (orthogonal frequency division multiplexing), coding scheme, as well as training sequence. Although, the difference between both standards is that the 802.11p exploits OFDM transmission with 10MHz spacing in the 5.9GHz band, the reason for this reduced channel spacing is to deal with increased delay spread in vehicular network environments [54].

5. **802.11n:** This standard aims to achieve higher data rates of at least 100Mbps [55]. This would require modifications to both the physical and MAC layers. Smart antennas and Multiple Input Multiple Output (MIMO-OFDM) are being considered by this group [56].

6. **802.11h:** This group provides solution for dynamic frequency selection and transmit power control [55]. It solves issues such as interference with satellite and radars using the same 5GHz frequency band. This standard was finalized in September 2003.

7. **802.11k:** This group is saddled with the responsibility of radio resource management enhancements for 802.11. Radio resource management involves the ability of each radio resource to be able to monitor and respond to changes in the environment.

3.2 IEEE 802.11 Standard for IoT

A new standard that is specifically designed for the IoT and particularly for low power, long range devices was conceived by the IEEE. This standard is an amendment of the earlier IEEE 802.11 one which was released in 2007 and is named as IEEE 802.11ah or WiFi HaLow.

802.11ah operates in the $900MHz$ spectrum and this gives it the advantage of providing low power connectivity for IoT applications such as sensors. It also allows HaLow to extend its operating range by almost twice of what is obtainable in previous 802.11 standards and further provide robust and reliable transmission signals even in challenging environments due to its ability to penetrate walls and other barriers quite easily [57].

HaLow differs from other 802.11 technologies in terms of its medium contention. Its protocol allows for stations to be grouped in such a way that contention on the air interface is minimized. It also allows relay APs to be used to extend its reach and with predefined wake periods, power consumption is reduced. Similar to 802.11a/g protocols, HaLow provides up to 100 kbps throughput using 26 channels that have each been down sampled [58].

In terms of data rate, about $347Mbps$ is acheiveable using four spatial streams in a channel with a width of $16MHz$. A $2MHz$ channel uses 64 fast fourier transform (FFT) comprising of 56 OFDM subcarriers - 52 are reserved for data while 4 are pilot tones with $31.25kHz$ carrier separation. Any one of BPSK, QPSK, 16-QAM, 64-QAM or 256-QAM can be employed by each of these subcarriers. The total symbol duration, including guard intervals (4-8 microseconds) is $36 - 40$ microseconds and total bandwidth is $2MHz$ [59].

There are three adopted use cases of IEEE 802.11ah and they include the following:

- Sensor Networks: This use case involves applications that would require short burst data transmissions such as smart meters and power distribution systems. With the advantage of increased penetration at lower frequencies, a massive number of devices can be accommodated in a single hop manner [60].

- Backhaul Aggregation: IEEE 802.11ah can also be used as a backhaul connector between machine devices and remote servers. Its large transmission range allows for routers to collate data from devices connected to it and forward them to servers using 802.11ah links [61].

- Extended Wi-Fi Range for Cellular Traffic Offloading: In an era of increasing coexistence between various wireless technologies, one of the use cases being considered by the IEEE is the ability to offload cellular traffic onto Wi-Fi using 802.11ah. Technical considerations such as spectral efficiency, system load and throughput are all being considered in developing the performance requirements for 802.11ah .

3.3 Key Challenges for IEEE 802.11 Standard in IoT

As common with any developing technology, there are a number of challenges that are to be overcome before the technology can be fully operational and effective. A few of these challenges as it affects the 802.11 standards designed for IoT is discussed in this section.

Requirements on IEEE 802.11ah PHY design

Whilst future IEEE 802.11ah standardized devices will operate using Multi Input Multi Output Orthogonal Frequency Division Multiplexing (MIMO-OFDM), the various ISM bands in which they will operate in different countries is an issue that does need to be considered. For example, the United States (US) proposes the operation of IEEE 802.11ah with a maximum bandwidth of 16MHz between 902MHz and 928MHz bandwidth, although, it can also operate using $1, 2, 4$ and 8 MHz. However, the 1 and 2 MHz are the least adopted and there is a need for the standardization group to support these bandwidths in the US and other countries [62].

Path Loss Models

It is envisaged that a separate path loss models for device to AP communications in indoors and outdoor environments would be considered [63]. In fact, [64] has investigated IEEE 802.11ah path loss models for indoor environment. However, device to device path loss models and Link budget investigations are also being carried out in academia in order to determine the suitability of the IEEE 802.11ah for various scenarios [58] [65].

Requirements on IEEE 802.11ah MAC design

The massive number of devices that would require communications in the IoT would mean that considerations should be given to contention free MAC protocol in addition to designing scalable and effective contention based algorithms. Alternatives to the DCF, might need to be investigated due to the potential network load of IoT. Already, IEEE 802.11ah is designed to support repetition and physical layer channelization [57].

Rate Adaptation

Another critical challenge of the IEEE 802.11ah is rate adaptation. Whilst there are a number of data rates which this technology can communicate, the determination of optimum transmission rates is still a subject for investigation. However, rate adaptation is not standardized by the IEEE standardization group.

3.4 Rate Adaptation in IEEE 802.11

The 802.11 rate adaptation challenge has been studied widely in literature. Several factors including the environmental conditions, degree of node mobility amongst others affect the performance of rate adaptation algorithms in wireless networks. Thus, some adaptation protocols perform very well in static environments with stationary or little mobility nodes and a relatively stable channel condition whilst others are designed for highly mobile nodes in dynamic networks with fast changing channel conditions. This section discusses a number of rate adaptation algorithms. However, particular attention is paid to the Auto rate fall back (AARF), ONOE and Samplerate algorithms as performance evaluation of these are carried out in later chapters of this thesis.

3.4.1 Auto Rate fall back (AARF)

The earliest rate adaptation algorithm was the Auto Rate Fallback (ARF) which was later modified to the Adaptive Auto Rate Fall Back (AARF) [66]. Rate change decisions were based on the outcome of frame transmissions which could either be success or failure.

Nodes operating using the AARF usually start transmission at the lowest data rate of the standard in use. Upon successfully sending a number of consecutive packets (threshold), N, or the expiration of a timer, the transmission rate is incremented to the next consecutive rate available in the standard and then the threshold is reset. However, if at the newly incremented data rate, the first transmission fails, the rate decrease follows equation 3.3 whilst the rate after two consecutive transmission failures is determined using equation 3.4

$$N = Min(N_{\max}, 2N) \tag{3.3}$$

$$N = (N_{\min}) \tag{3.4}$$

where $N_{\min}$ is the lowest number of consecutive successful transmission and it has a value of 10, $N_{\max}$, is the highest number 50.

The number of consecutive successful transmission makes this algorithm susceptible to spurious rate fluctuations. However, it is not likely to react quick to environments in which channel conditions change quickly.

A flowchart of AARF is shown in Figure 3.2.

3.4.2 ONOE

ONOE was developed by the MadWifi organization for wireless adapters with Atheros chips [67]. The algorithm is open source and is a credit based algorithm that uses the loss ratio as its main metric of operation. It aims to find the best data rate with a loss ratio of not up to 50% [68].

Starting at the lowest data rate in any standard, this algorithm does not make any rate changing decision until after one second. Its decision is based on the foregoing:

- If a transmission is successful, the number of transmitted frame being equal to or greater than ten (10) and the number of retries per frame greater than one, rate is decreased to the next consecutive rate else the number of retries per frame is checked, if this is less than

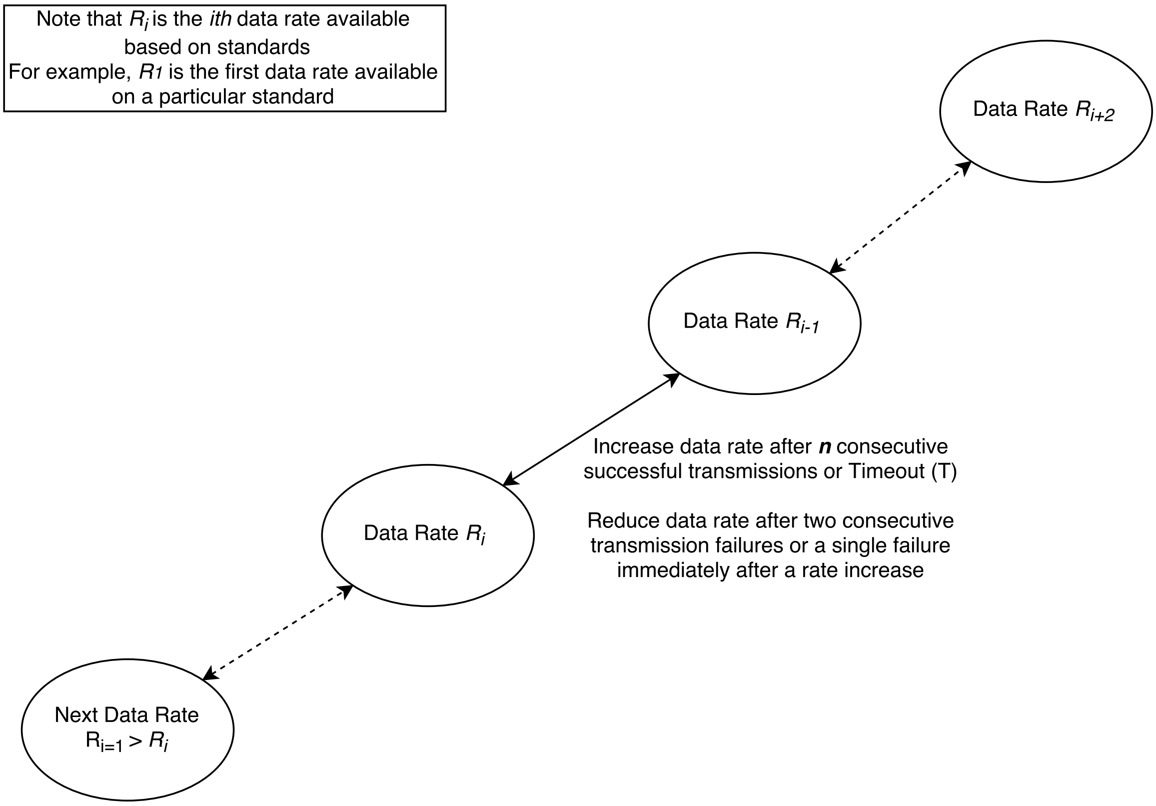

Figure 3.2: Schematics of operation of AARF

10% of transmitted frames, the number of credits is increased, otherwise it is reduced.

- Finally, a check is made on the total number of credits in the one second duration and if this is greater than or equal to ten, the data transmission rate is increased and all variables are reinitialized.

Due to wait period involved and the probability that a node would only be in the range of an access point for a particular period of time (in a V2I network), Onoe is slow to make rate changes and hence not suitable for vehicular environment as the communication between nodes take place very quickly. A flowchart of ONOE as described in [68] is shown in Figure 3.3.

3.4.3 Samplerate

Samplerate is the default rate adaptation algorithm in Linux open drivers. It takes advantage of wireless nodes ability to send packets at different data rates and maximizes throughput by sending packets at the bit rate that has the smallest average packet transmission time as measured by recent samples [69]. The samplerate algorithm could be broken down into the following steps:

- Initially samplerate starts by sending data at the highest possible rate available in the standard being used.

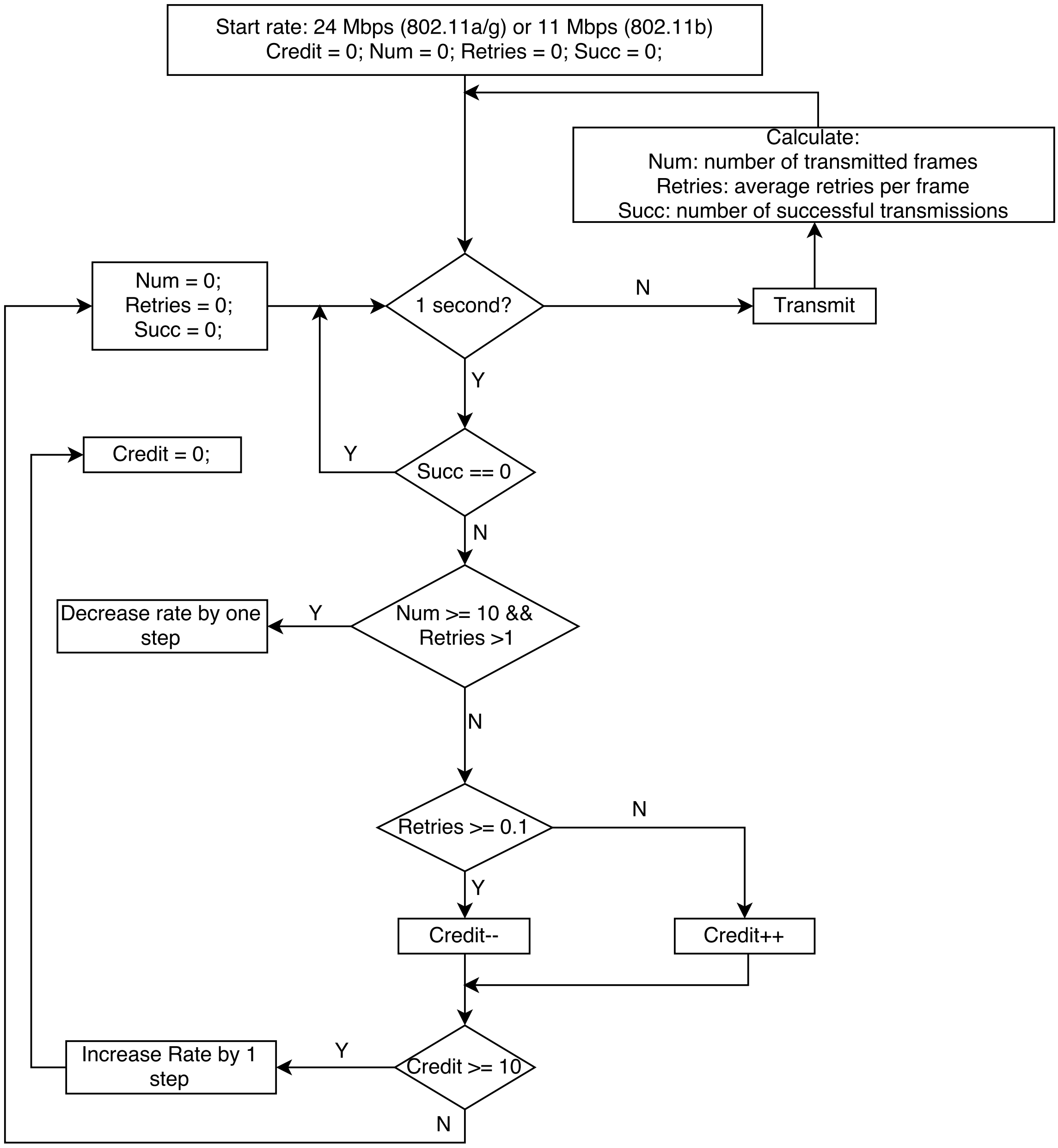

Figure 3.3: Schematics of operation of ONOE

- If that data rate experiences four consecutive transmission failures, it select the next lower data rate until one that is capable of sending four consecutive packets successfully is found.
- At every tenth packet, samplerate randomly select a rate from the set of bit rate that outperforms the current one and transmit with that bit rate. This process is known as sampling.
- a data rate is not eligible for sampling if four recent successive packets at that rate has not been acknowledged or if its transmission time without retries is greater than the average

transmission time of the current data rate.

- if the sampled rate is able to successfully transmit four consecutive packets, it becomes the current data rate.

Wireless cards provide feedback that include the number of retries a packet required and the outcome of the transmission, whether it was successful or not. These feedback is also used to calculate the average transmission time considering the duration of each packet transmission. The packet length, data rate and number of retries (including backoff time) are metrics used to determine the packet transmission time. Samplerate avoids using outdated information by basing its calculations on the last averaging window which is set to 10 $secs$.

SampleRate sends data at the data rate that has the smallest predicted average packet transmission time, including time required for retransmission. Whilst this is good, it is not the best method of determining the best performing rate even as transmission time could be affected by rapidly changing channel conditions as is often the case in a vehicular network.

However, samplerate is quite effective in static networks and analytic models have been proposed to study the steady state behaviour of the algorithm [70].

3.4.4 Throughput Enabled Rate Adaptation Algorithm (TERA)

TERA considers the effects of interference and congestion on rate changes but rate changing decisions are solely based on throughput measurements of the different data rates.

In TERA, the data rate is adjusted periodically after a review of transmission statistics. This period is referred to as a time window and has a value of 100 ms. The throughput attained by a node is obtained by taking into account the probability of success and the packet transmission time. In adjusting rates, a reference throughput, obtained using an exponential weighted moving average filter, is compared to the current one. Transmission data rate is increased only when the node is in the probing state. In this state, the node is able to search available rates in order to find the highest rate with the best throughput using a combination of two probing frequencies. When the current throughput is better than the reference throughput, rate increase is attempted either additively or multiplicatively. If the increased rate provides a better throughput, it is selected as the new current transmission rate and the algorithm exits the probing state. Data rate is decreased by selecting the next lower rate when compared with the current one that provides a better throughput [71].

TERA is likely to have a low responsiveness if used in mobile environments and with a maximum client speed of 1 m/s considered, its performance might suffer degradation when used in heavily mobile clients.

3.4.5 Full Auto Rate algorithm (FAR)

An algorithm capable of adapting transmission rate of both control and data frames is described in [72]. By using the extended interframe spacing (EIFS) duration, deferments the authors demonstrated that control frames can be transmitted at a higher rate with the help of EIFS deferment. FAR incorporates the rate adaptation algorithm for control frames and a modified virtual carrier sensing (MVCS) proposal.

At the sending node, in order to choose the transmission rate of the request to send (RTS), every node maintains a cache variable whose initial value is set to the lowest basic rate. A successful transmission of an RTS or data frame leads to an update whilst a failure would lead to the variable being reset to the lowest basic rate. If the time of transmission of the data frame is greater than the EIFS, the clear to send (CTS) control frame is sent at the lowest basic rate

At the receiving node, based on the signal to noise ratio (SNR), a data rate is proposed for the next data frame and this rate is piggybacked in the CTS as well. The determination of the rate at which the acknowledgement frame (ACK) is to be sent follows a similar decision as previously described.

Summarily, in FAR, the sending node does adaptation based on RTS/CTS frames whilst the receiving node uses ACK and DATA frames in order to adapt to environmental conditions. It also takes virtual carrier sensing into consideration and proposed a modified virtual sensing. The authors verified that FAR can adapt rate with the combination of both sender and receiver based protocols.

3.4.6 History Aware Robust Rate Adaptation Algorithm (HA - RRAA)

HA-RRAA aims to maximize aggregate throughput in the presence of various channel dynamics by being robust and responding quickly to dynamic environmental conditions. It uses short-term loss ratio to opportunistically guide its rate change decisions, a cost- effective adaptive RTS filter to prevent collision losses from triggering rate decrease and an adaptive time window to limit transmissions at high loss rates. The operation of HA - RRAA is based on three principles which are [73]:

- Loss estimation: By assessing channel conditions using the frame loss ratio with a specified time window (5-40 frames in our implementation), the algorithm is able to estimate losses.
- Rate change: Decision on whether to retain, increase or decrease transmission rate is based on the estimated loss ratio.
- Adaptive RTS filtering: Through selective activation of the RTS/CTS mechanism, the algorithm aims to suppress collision losses, and adapts the number of RTS-protected frames to the collision level.

3.4.7 Threshold Optimization Algorithm

One of the drawbacks of ARF was the fact that its rate changing threshold was fixed. However, in order to address this challenge, [74] proposed a threshold optimization algorithm.

In their proposal, an implicit objective function is derived by reverse engineering the ARF. The solution of the objective function converges to a stochastic optimum in an arbitrary stationary random channel environment. This solution is then used as threshold value for rate increase or decrease.

However, the performance of this algorithm was benchmarked using ARF and AARF but there are numerous algorithms that have outperformed both. It is fairly certain that this algorithm would have been outperformed.

3.4.8 Enhanced Adaptation of link Rate and Contention window Rate Adaptation Algorithm (EARC)

An enhanced adaptation of link rate and contention window (EARC) that jointly adapts data rate and backoff parameters is proposed in [75].

The algorithm operates by using the reserved bits in physical layer convergence protocol (PLCP) header and one additional byte in the data packet, the authors claims the algorithm incurs little communication overhead despite the fact that it is a receiver assisted algorithm. The optimal transmission rate is determined according to an empirically derived rate selection reference (RSR) table at the receiver. This table is derived from constant monitoring of reception behaviour and environmental energy levels at the receiver node.

How this algorithm would perform in fast changing environmental conditions is yet to be determined.

3.4.9 Minstrel

Another algorithm that supports multiple rate transmission and retries is Minstrel. This algorithm is also used in the Madwifi project.

It works on the basis of transmitting at different rates whenever possible and comparing throughput to determine the optimal rate. The Minstrel algorithm is based on the exponential weighted moving average (EWMA). This allows more emphasis to be placed on recent channel condition changes rather than past ones, thereby discarding stale results. The EWMA, for each rate, is calculated ten times per second. A formula is then used to calculate the successfulness of packet delivery and based on this, the optimal rate of transmission is chosen.

By dedicating 10% of data packets for transmission at rates other than the current rate, a list is populated that gives the history of successful transmissions and the time taken for such transmission.

Using the throughput as metric, data rates are ranked and the highest ranked data rate replaces the current one and is used for transmission.

3.4.10 Context-aware Rate Selection for Vehicular Networks (CARS)

CARS makes use of contextual information in making decisions on transmission rates. Metrics such as vehicle speed and distance from neighbours, and past history are used to evaluate channel conditions and maximize link throughput [76]. CARS start at the lowest rate available in the standard which it is being implemented and based on a variable, it determines whether to pay more attention to contextual information or the exponentially weighted moving average (EWMA) of past transmission history of frames per bit rate. The assignment of this variable is based on the vehicular speed. Packet error is estimated by a weighted decision function that is composed of two functions. One of which uses context information, transmission rate and packet length as input and generates packet error rate as output while the other uses an exponentially weighted moving average (EWMA) of past frame transmission statistics for each data rate.

CARS then calculates estimated throughput for each bit-rate and selects the bit rate it predicts will provide the most throughput [76].

However, CARS requires to know the signal propagation path loss exponent to estimate the channel condition, which can lead to significant performance deteriorates with large mismatch between the actual and estimated path loss exponents.

3.4.11 Generic Rate Adaptation (GERA)

GERA for vehicular networks makes use of contextual information and received signal strength to estimate channel conditions and make optimum rate changing decisions based on these information. It dynamically and adaptively switch rate selection resources between the context information empirical model and SNR prediction model according to prevailing environmental conditions [77]. GERA seeks to strike a balance between using contextual information and RSSI in making rate changing decisions. However, RSSI is not the best indicator of channel quality and this may have an adverse effect on the overall system throughput.

3.4.12 Random Forests Rate Adaptation for Vehicular Networks (RFRA)

RFRA is based on the random forest machine learning algorithm. The random forest is a learning based heuristics that recognizes and exploits statistical dependencies in multidimensional decision problems.

The RFRA algorithm undergoes a training phase where past statistics about the environmental condition in which it is to be used are collected, a prediction phase whereby during and before transmission of a new packet input variables are also collected and a rate selection phase [78]. The design output of RFRA is not a rate but a packet success rate (PSR), which is calculated by data rate and rate change decision is made based on either of three criteria namely threshold, raw goodput and MAC goodput [78]. The rate adaptation in mobile environments makes use of a receiver based approach to handle asymmetric channel conditions and an SNR prediction algorithm to handle channel fluctuations. It is implemented in the madwifi device driver and also takes SNR into consideration in its rate making decision [79].

3.5 Conclusion

Short range communication technologies such as Zigbee, Wi - Fi, and Bluetooth are quite popular today. In this chapter, we focus on the IEEE 802.11 wireless technology. We have given an in-depth insight into how this is implemented, and have taken a look at its physical and logical architectures,as well as, the physical and MAC layers.

Also, some challenges facing the IEEE 802.11 technology are also examined. For instance, the data rate selection and security are still areas where improvements can be made in IEEE 802.11 networks.

As this is one of the core areas of this thesis, ways have been discussed in which the rate selection challenge is being tackled in literature and considered a number of rate adaptation algorithm which exists in literature. It is interesting to note that this problem, albeit important, is not yet standardized by the IEEE. It is still an open area of research that is generating massive attracting interest both in academia and industry.

Contributions made in tackling this challenge are discussed extensively in chapter 6. This includes proposal of two rate adaptation algorithms for use in vehicular networks and evaluating their performances using simulations.

4 Root Index Allocation Scheme for Massive Device Access in LTE Networks

4.1 Introduction

As already highlighted in previous chapters, one of major challenges of implementing IoT communications using cellular media, particularly LTE, is the potential for congestion in the RACH due to exponential increase in the number of devices seeking network access. The demands for connection and traffic generated by the massive number of IoT devices will pose big challenges to wireless networks.

Therefore, the first mile of attachment to a cellular network, the radio access network (RAN), is expected to be overloaded with requests from millions of devices seeking network access. These demands would lead to an exponential increase in signalling, data traffic and ultimately lead to congestion in the RAN and core networks. The effects of such congestion on revenue and quality of service of the network cannot be overemphasized.

Although there are several recent enhancements (such as the access class barring (ACB) scheme) proposed to the 3GPP baseline RCA algorithm [80] and which have been discussed in chapter 2, none of them can meet the RCA demands from the rapidly increasing number of IoT devices. In this chapter, we propose to solve the RCA problem in LTE networks by more effective reuse of random access preambles. Our scheme provides more than the default 64 RCA opportunities per frame slot, thus enabling more devices to access the network. The major contributions in this chapter can be summarized as follows.

- We propose an effective preamble reuse scheme, based on safe distance, to provide more RCA opportunities to support multiple machine devices accessing LTE cellular networks. These solutions, as demonstrated, significantly increase random access throughput, reduce access delays and preamble collision probability whilst offering network operators more flexibility in planning their preamble configurations and deployment.
- In order to maximize the reuse factor of the preambles available to the operator of a cellular network, while not introducing detrimental interference due to the aggressive reuse of the preambles, we analyse the 3GPP preamble formats and derive a safe distance for preamble reuse under the conditions of variable cell radii and layouts.
- A novel preamble allocation algorithm is proposed to allocate the preambles to cells based on saturation colouring, which can be applied to cellular networks with irregular layouts.

The preamble allocation algorithm is flexible and can be applied to meet different RCA traffic loads and operational requirements.

The proposed preamble reuse scheme is complementary to the existing RCA algorithms and is implemented on top of the default RCA algorithm. Extensive simulation results show that the RCA performance is significantly improved with preamble reuse. The proposal could be broken down into three major steps each of which is explained in details in the relevant sections. Firstly, the concept of safe distance is discussed. By mathematical analysis, we are able to determine the distance within which preamble reuse can be carried out by each eNodeB. Once the safe distances are determined for the eNodeBs, the second step involves a series of calculations to determine how to implement preamble reuse in a network wide scenario, considering both homogeneous and heterogeneous cases. Finally, we developed a saturation colouring based algorithm for preamble reuse allocation. With the allocation algorithm, preambles can be dynamically allocated to all the enodeBs within a network.

The remainder of this chapter is organized as follows. In Section 4.2, we lay the foundation of this work by describing the reuse of RCA preambles and the concept of safe distance. we also discuss details the derivation of safe reuse distance and potential implications of preamble reuse on existing networks. Section 4.3 presents the preamble allocation algorithm. Section 4.4 discusses the system model and presents the simulation results with the proposed preamble reuse scheme. Section 4.5 concludes the chapter.

4.2 Reuse of RCA Preamble and Safe Distance

Under normal cellular network operation, the default number of preambles used for RCA purpose in each subframe is set to 64. However, as we analyse in this section, there can be many more RCA preambles generated and used for IoT devices, which can have significant positive impact on the success of the RCA procedure. We believe that effective reuse of RCA preambles can be a vital approach to tackle the challenge of massive industrial device access. In this section, we present a simple introduction on Zadoff-Chu sequences and the generation of RCA preambles. We then present the calculation of safe distance for preamble reuse.

4.2.1 Zadoff-Chu Sequence and RCA Preambles

With careful configurations, preambles used in LTE have specific properties that allows them to be detected by eNBs. The Zadoff-Chu sequence is the choice in preamble generation in LTE mainly due to its constant amplitude zero autocorrelation property and cyclic shifting of a single Zadoff-Chu sequence leads to the generation of several preambles [81] [82]. There are 838 root Zadoff-Chu sequences that are available for preamble generation and each root has a root sequence length of 839.

In the frequency domain, a preamble sequence is generated using equation 4.1 below:

$$x_{\mathrm{u}}(n) = \exp\left(-j\frac{\pi un(n+1)}{N_{\mathrm{zc}}}\right), 0 \leq n \leq N_{\mathrm{ZC}} - 1 \tag{4.1}$$

where u is the physical root index and N_{ZC} is the number of Zadoff Chu sequence. In dynamic preamble generation, an MTC device can select a logical root index from which the physical root sequence index, u, is obtained. By cyclically shifting generated sequences, a number of preambles can then be generated. However, in order to maintain orthogonality of preamble sequences, the number of cyclic shifts per sequence is dependent on some factors including the radius of the cell and the maximum delay spread amongst others. This relationship is given by equation 4.2:

$$N_{\mathrm{cs}} \geq \left[\left(\frac{20}{3}r - T_{\mathrm{ds}}\right)\frac{N_{\mathrm{zc}}}{T_{\mathrm{seq}}}\right] + n_{\mathrm{g}} \tag{4.2}$$

where r is the cell radius (km), T_{ds} is the maximum delay spread, $N_{\mathrm{zc}} = 839$, T_{seq} is the preamble sequence length (in μs) and n_{g} is the number of additional guard samples due to the receiver pulse shaping filter [19].

The value with which the generated sequence is to be shifted is defined below:

$$C_{\mathrm{v}} = \begin{cases} vN_{\mathrm{cs}}, & v = 0, 1, ..., [N_{\mathrm{zc}}/N_{\mathrm{cs}}] - 1, N_{\mathrm{cs}} \neq 0 \\ 0, & N_{\mathrm{cs}} = 0 \end{cases}$$

where C_{v} is the cyclic shift value.

The cyclic shift value is usually signalled as part of the system information and the number of roots from which the preambles are generated depends on the size of the cell and cyclic shift used. Cyclically shifted Zadoff-Chu sequences possess constant amplitude which ensures efficient power amplifier utilization, maintaining the low power to average (PAR) ratio.

Also, the number of preambles, N_{p}, that a single ZC root sequence can generate could be obtained as below:

$$N_{\mathrm{p}} = floor(839/N_{\mathrm{cs}}) \tag{4.3}$$

Therefore, to generate any number of preambles for a cell, the required amount of root sequence can easily be calculated from equation 4.2 and 4.3.

Also, cross correlation between sequences generated from the same root is zero at the receiver provided that the cyclic shift used when generating the preambles is larger than the maximum round trip time in the cell plus the maximum delay spread of the channel [19]. In addition, Zadoff-Chu sequences have a good periodic and simple detection mechanism.

Thus, no intra-cell interference on the preambles is expected from multiple random access attempts using different preambles derived from the same root sequence due to the ideal cross-correlation property of these sequences.

Assuming a constant delay spread from equation 4.2, it is evident a direct relationship exists between the cell radius and the number of cyclic shifts required to generate any amount of preambles.

4.2.2 Safe Distance for Preamble Reuse

The primary objective of preamble reuse is to generate more than the default 64 preambles for RCA in the cellular networks, in order to create more RCA opportunities in each frame to accommodate more IoT devices. In this subsection, we introduce the concept of safe distance for preamble reuse to help make the right trade-off to generate more preambles to meet the demands of IoT RCA traffics. The safe distance for preamble reuse is defined as the distance beyond which it is impossible for users located in neighbour regions to transmit a preamble capable of causing a collision or significant interference.

Interference in the RACH can either be from other preambles (intra-preamble interference) or from data which is carried in the PUSCH (physical uplink shared channel). However, to cater for interference caused by the PUSCH, guard sub-carriers are positioned on each side of the PRACH sub-carrier to aid the detection of the preamble transmitted by the eNodeB. In practice, the preamble is positioned centrally in a block of 864 available PRACH sub-carriers with 12.5 null sub-carriers on each side acting as a guard band.

In deriving the safe distance, a number of parameters need to be considered. Firstly, each frame in LTE is segmented into subframes, each of length 1 ms and since not all subframes in a frame are allocated for RCA, it will be effective to minimise the number of subframes for RCA whilst giving more room for data transmission using the other subframes. Thus, for efficiency, any sequence that is to be used for RCA should have a duration that fits into the LTE subframe structure. The duration of the Zadoff-Chu sequence is affected by its length and a single sequence must be long enough to maximise the potential number of orthogonal preambles it can generate.

Another major factor to be considered is the maximum round trip time. The sequence duration should be able to accommodate the round trip time of an edge user located in the largest expected cell radius. For preamble format 0, with a maximum cell radius of 14.53 Km, based on 3GPP estimates, the maximum delay spread, T_{ds}, expected in such cell is 6.25 μs. The sequence duration, T_{seq}, can be determined as follows,

$$T_{\text{seq}} \geq \frac{R_{\text{max}}}{c} + T_{\text{ds}} \tag{4.4}$$

where R_{max} is the cell radius (km), T_{ds} is the maximum delay spread and T_{seq} is the preamble sequence length (in μs).

Of utmost importance in the derivation of safe distance is the preamble coverage performance. A longer sequence gives better coverage but at the expense of a longer cyclic prefix and guard time in order to cushion the corresponding round trip delay. Performance coverage can be estimated by a link budget calculation and is also dependent on the environment in which the cell operates. Typically, for a low density, rural or medium to large suburban cell, we can use the Okumura-Hata model to evaluate the path loss which is given by,

$$P_{\text{loss}} = A + B\log(D_{\text{r}}) + C \tag{4.5}$$

$$A = 69.55 + 26.16\log(f) - 13.82\log(h_{\text{b}}) - a(h_{\text{m}}) \tag{4.6}$$

$$B = 44.9 - 6.55\log(h_{\text{b}}) \tag{4.7}$$

where D_{r} is the distance from the receiver in km, f is operating frequency given in MHz, h_{b} and h_{m} are eNodeB and device height respectively. The function $a(h_{\text{m}})$ and the factor C depend on the environment. In a metropolitan environment, these are calculated using,

$$a(h_{\text{m}}) = 3.2(\log(11.75h_{\text{m}})^2 - 4.97) \tag{4.8}$$

for $f > 400\ MHz$

$$C = 0 \tag{4.9}$$

The path loss model applied here describes an envisaged scenario and does not invalidate our results in any way. Other path loss models can also be applied with similar results expected.

The PRACH signal power, P_{RA}, received at the enodeB can then be estimated using equation 4.10

$$P_{\text{RA}} = P_{\text{max}} + G_{\text{a}} - P_{\text{loss}} - LF - P_{\text{L}}(dB) \tag{4.10}$$

The parameters for the link budget performance calculations are taken from [19] and are stated in Table 4.1.

Table 4.1: Estimation Parameters

Parameter	Value
Carrier Frequency, f	2000 MHz
eNodeB antenna Height	60 m
eNodeB receiver antenna Gain, G_{a}	14 dBi
Penetration loss, P_{L}	0 dB
Log-normal fade margin, LF	0 dB
Receiver noise figure, N_{f}	5 dB
Thermal noise density, N_0	-174 dBm/Hz
UE transmitter power, P_{max}	24 dBm (250 mW)

In 3GPP standards, a preamble sequence energy to thermal noise ratio (E_{p}/N_0) of 18 dB is required to meet the missed detection and false alarm probability of 10^{-2} and 10^{-3} [36]. Based on this,we can then estimate the required preamble sequence duration using,

$$T_{\text{seq}} = \frac{N_0 N_{\text{f}}}{P_{\text{RA}}}(E_{\text{p}}/N_0) \tag{4.11}$$

where N_0 is the thermal noise power density (in mW/Hz) and N_{f} is the receiver noise figure (in linear scale) and their values are stated in Table 4.1.

The performance coverage can then be obtained as a function of sequence length as shown in Fig. 4.1. We can see that for a 1 ms subframe, the potential coverage radius is about 14 km. This is consistent with preamble formats 0 and 1 stated in 3GPP standards, longer preamble sequences can be implemented by repeating the baseline preamble sequence but this would require different cyclic prefixes and guard times. It should be noted here that both cyclic prefixes and guard times are included within the 1 ms subframe.

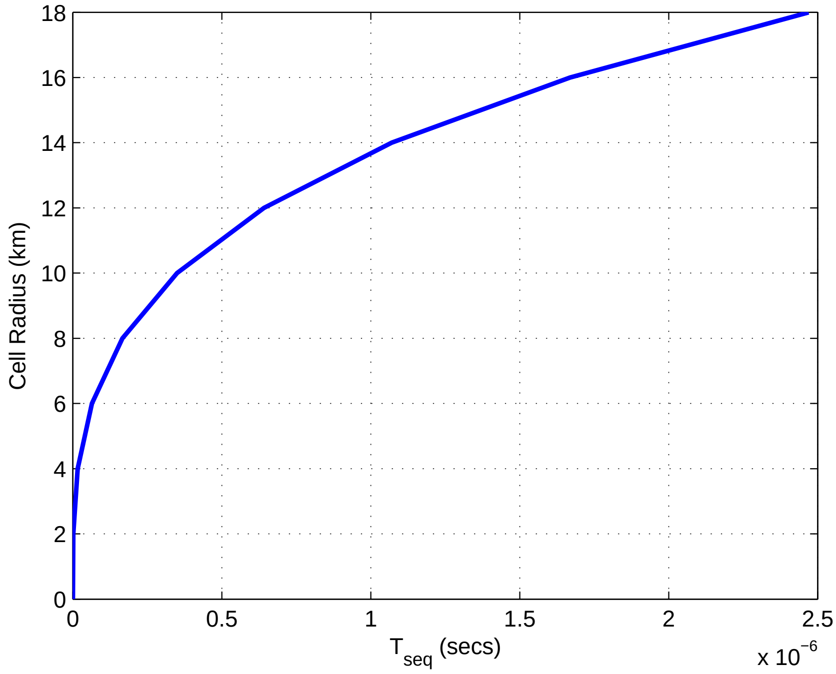

Figure 4.1: PRACH coverage performance against sequence duration

The implication of this is that, transmissions from outside the specified potential coverage radius are not decodable by the eNodeB considering the maximum round trip time and the value of the RA subframe specification in use. Therefore, for this preamble format, this maximum radius can be considered as the safe distance outside of which preamble reuse and guaranteed RA information exchange between eNodeB and machine devices is possible, preambles can then be reused outside of the safe distance.

In this work, we focus on the preamble format 0, which has a safe distance of about 14 km. Other preamble formats can be implemented in a similar fashion, although, the potential gain in terms of the number of preambles per cell would be lower when compared to the aforementioned format. The advantage of reusing preambles is that it frees up more Zadoff-Chu sequences which can then be used in generating more preambles. This will invariably lead to an increase in the number of preambles from its present value of 64 per cell to a number that is determined by the cell radius amongst other factors.

Our analysis explores two scenarios namely:

- An homogeneous network, consisting of cells of the same coverage radius covering the safe distance.
- An heterogeneous network, consisting of a mixture of small cells and macro cells of varying cell radii covering the safe distance.

If we assume that the cells cover a circular radius, then the safe coverage area A_{sc}, within which preamble reuse is not possible can be obtained by using:

$$A_{\text{sc}} = \pi D_{\text{s}}^2 \tag{4.12}$$

where D_{s} is the safe distance for the cell being considered.

We can then estimate the required number of cells to cover the safe coverage area (SCA). If A is the area of the cell with coverage radius r, the number of cells required will be given by:

$$N = ceil(\frac{A_{\text{sc}}}{\pi r^2}) \tag{4.13}$$

The number of cells in an SCA is important as it is one of the factors that determines the potential increase in the number of preambles per cell. Depending on the layout of the network, it is possible to fully cover the SCA without leaving out some areas uncovered, as two cells might cover distances greater than the safe distance, in this case, preamble reuse would be inadvisable. However, small cells are introduced to help enhance coverage and also reach uncovered areas in a network.

If we consider a bullish situation where all 838 Zadoff-Chu sequences can be allocated within the safe distance and reused outside of it, there will certainly be huge performance improvements (increase in the number of preambles). For example, consider a scenario in which a cell can only generate preambles from 32 Zadoff-Chu root sequences, using equation 4.2 to obtain the number of cyclic shifts required per cell and equation 4.3 to calculate the number of preambles per root sequence, it is easy to compute the performance increase in the number of preambles obtainable.

Table 4.2: Performance improvements with preamble reuse

Number of cells required to cover SCA	Cell Radius (km)	Number of preambles per Zadoff-Chu sequence	Maximum number of preambles obtainable per cell
1	14	7	224
10	4.5	22	704
8	6.5	20	640

In Table 4.2, the maximum number of preambles obtainable per cell is the product of the number of preambles per Zadoff-Chu sequence and the maximum number of root sequence per cell. The preamble increase, L, can easily be calculated as the maximum achievable rate of change between the maximum number of preambles obtainable per cell and the existing number of preambles per cell (64). Mathematically, this can be expressed using equation 4.14 below

$$L = \frac{(M - N)}{N} \tag{4.14}$$

where M is the maximum number of preambles obtainable per cell (see Table 4.2) and N is the initial number of preambles which defaults at 64, according to current 3GPP implementations.

Cells with a smaller radius of coverage in blanketing the safe coverage area produce a significantly higher increase in the number of preambles per cell than cells with larger radii. The increase begins to diminish as the radius of the cell is increased but there is a still a massive performance improvement when compared to the existing number of preambles.

The same process can be repeated in a network consisting of small cells and macro cells. Small cells can be strategically or randomly positioned to cover the SCA. This can improve the network coverage performance. If we assume a fixed number of small cells are introduced into the network and that the 838 Zadoff-Chu sequences are to be distributed across eNodeBs within the SCA, we will see a massive increase in the number of preambles that can be generated per cell, this number can be calculated using equations 4.12, 4.13, and 4.2.

Considering the safe distance of $14Km$, a performance graph, as shown in Fig. 4.2, can then be plotted to see the potentials of preamble increase in an homogeneous scenario and a heterogeneous one consisting of additional ten small cells each of 200 m radius. A macro cell is assumed to be the controller of some or all of the small cells and allocates Zadoff-Chu sequences first to the small cells in such a way that they are able to generate the required minimum of 64 preambles, we note that for small cells one Zadoff-Chu sequence is sufficient to generate this number. the remaining sequences are then allocated to macro cells in such a way that no more than 32 are allocated to a single cell.

As can be seen from Fig. 4.2, both scenario seem to have similar potential preamble increase when cell radii are smaller, up until about 2.5 km. However, as the cell radii increases we witness significant preamble increase. This is due to the effect of adding small cells in the heterogeneous networks. Whilst the small cells only require very few numbers of ZC root sequence for preamble generation (one, in this case), the same cannot be said of macro cells where the required number is determined by the cell radius amongst other factors, in fact, there is a direct relationship between the number of root sequence required for preamble generation and cell radius as described in equations 4.2 and 4.3. This means that there are fewer numbers of Zadoff-Chu sequence available for allocation to small cells in heterogeneous networks with large cell radius. Also, considering an homogeneous scenario and from equation 4.13, as cell radius increases the number of cells required to cover the safe distance reduces, this means more root sequences would be required. The addition of small cells in networks gives a greater

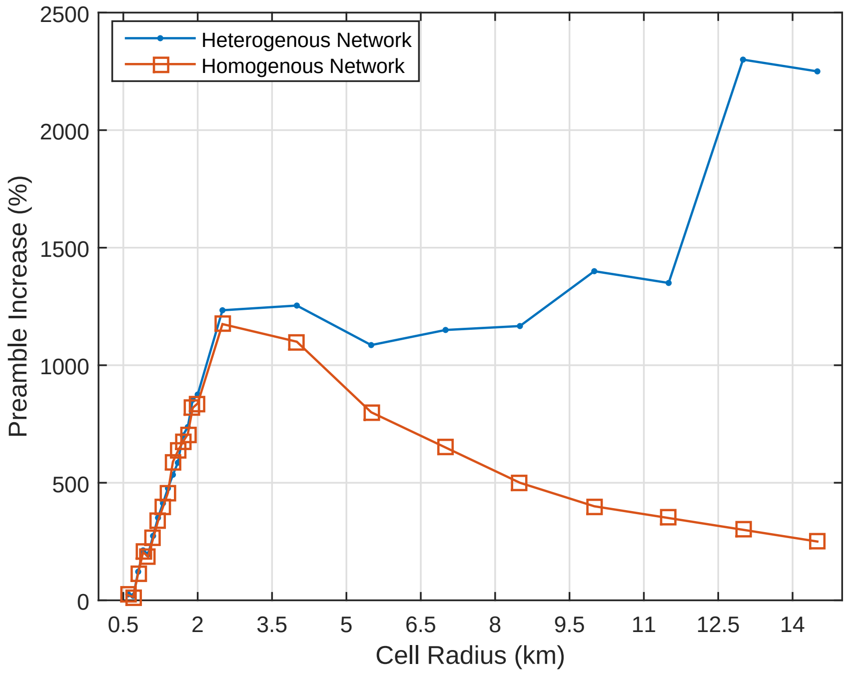

Figure 4.2: Preamble increase against cell radius

benefit than the use of macro cells alone, although macro cells also give a substantial amount of benefit, in terms of the increase in number of preambles, to the network.

4.2.3 Preamble detection performance

In order to detect transmitted preambles, a correlation operation between the received signal and all available preamble signatures is performed at the eNodeB end. A sequence as generated from equation 4.1 will have periodic cross correlation function, $corr_{\mathrm{m}}$, as defined in equation 4.15.

$$corr_{\mathrm{m}} = \frac{1}{\sqrt{L}} \sum_{n=0}^{L-1} p_{\mathrm{n}} q^{*}{}_{\mathrm{(n\text{-}m) mod\, L}} \tag{4.15}$$

where p and q are two ZC sequences which the same prime length L but generated from different root sequences, '*' is the complex conjugate and mod the modulo operator.

Detection occurs when a correlation peak whose value is greater than a certain threshold is obtained. The peak position is then used to estimate the uplink propagation delay and in determining the cyclic shift corresponding to the transmitted preamble [36] [83]. The threshold value is designed such that the false alarm probability requirements specified by the 3GPP is met. The presence of varying environmental factors such as noise or preambles generated from other root sequences increases correlation peak value. Therefore, threshold value is dynamically determined at the eNodeB end. Consequently, as the number of root sequences used to generate preambles increases, there are changes in the peak and threshold values. Using a higher number of root sequences (32) will definitely have an impact on these values. Consider an hypothetical scenario where each preamble is generated by a Zadoff-Chu sequence and another in which two preambles are generated from one root sequence. Figures 4.3 and 4.4 shows the variation in correlation and detection threshold values obtained from detecting a preamble in the presence of 32 and 64 root sequences respectively. We can observe higher values of correlation peaks and threshold in the scenario where we have 32 root sequences (Fig. 4.3) and there is a significant reduction in these values when the number of root sequences is doubled. This reduction is due to inter-preamble interference. However, there is a higher probability of successful preamble detection at the eNodeB end when 32 root sequences are used compared to 64. This further lends credence to the implementability of our proposal. Whilst we have used 32 preambles to demonstrate the effectiveness of our proposed scheme, it is not required to always use this number of preambles, infact, having a single additional root sequence to the existing number of root sequences in any cell will produce significant increase in the number of preambles generated.

This lends credence to the implementability of our proposed scheme. The design of preamble detection algorithms is not within the scope of our work, however, a number of algorithms [84] [85] exist in literature, that have done some work relating to preamble detection algorithm designs.

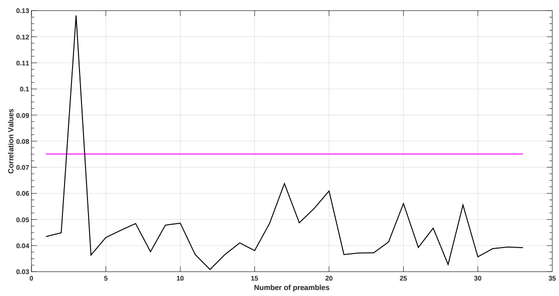

Figure 4.3: Correlation and threshold detection values using 32 root sequences

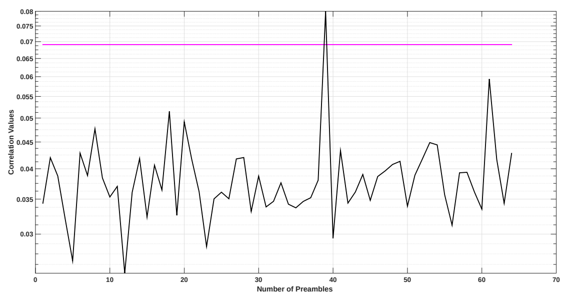

Figure 4.4: Correlation and threshold detection values using 64 root sequences

4.2.4 Implications on LTE system operation

The operation of RACH is linked to many different interfaces in the network. Therefore, changes that are proposed to be made has to take into consideration the potential effects on these other interfaces. Here, we discuss the potential effect of the proposed preamble reuse scheme on the existing LTE network. Allocation of resources to the RACH is determined by the PRACH configuration index which is defined in 3GPP standards. This index determines exactly when and where a UE is supposed to send RACH amongst other things. A full list of the various indices available is given in table 5.7.1-2 of [86]. The PRACH configuration index is specified in the system information block (SIB2) parameter and it has more to do with the preamble format, system frame and the subframe number in which a preamble should be sent rather than the total number of preambles available to a cell.

Although, there are 64 integers specified in the PRACH configuration index table, they bear no correlation with the existing default number of preambles. Thus, increasing the number of preambles does not constitute a change in the PRACH configuration index table as is presently constituted in the standards. This makes the proposed preamble reuse compatible with the existing scheme without requiring any changes. Also, in order to maintain backward compatibility with legacy LTE technology, the preambles could be grouped at the enodeB, such that, the existing 64 preambles are for the exclusive preserve of legacy devices whilst the additional preambles could be reserved for the use of newer devices. In this way, no compatibility issue would arise between legacy devices and newer ones.

Due to the complex nature of their construction, the generation and detection of Zadoff-Chu sequences in real time could be quite challenging but this is not the focus of our work. However, one of the ways of managing this challenge is by pre-computing preamble sequences offline and storing them in memory. For example, assuming 8-bit quantization, 2.5 Mbits of storage space is required to store preamble sequences in a 3-sector cell where 64 preambles are to be allocated per sector [87]. The proposed increase in the number of preambles would correspond to an increase in memory allocation for preamble sequences. It is safe to say that the generation of additional preambles is not an issue of serious concern in practical implementation of the preamble reuse scheme.

4.3 Preamble Allocation Algorithm

A practical implementation of this technique would require an algorithm for the allocation of preambles in the network. The traditional frequency reuse algorithm could be thought of as a natural candidate for the preamble allocation algorithm but due to the evolving heterogeneous nature of cellular networks, this would not be suitable. For example, a small cell (e.g femtocell) is dynamic in nature and could change location as frequently as the user likes, this variation of location would mean a change in the root sequence assigned to it every time. Therefore, a more robust algorithm to solve this challenge is required.

Preamble allocation can be thought of as a graph colouring problem, in which nodes (cells) are assigned a particular colour (root sequence(s)) which is different from its neighbours. Recall that preambles are generated from root sequences. However, this challenge is distinct from the graph colouring problem in the sense that, colours may only be reused at a certain safe distance. The graph colouring problem, and by extension the preamble allocation, is a non deterministic polynomial time (NP) problem where any given solution can be verified in polynomial time [88]. There is no known quick and efficient way to locate a solution to them. Also, as the size of the problem grows, the time required to solve it increases exponentially. In a cellular network, each cell has a set of neighbours which are adjacent to it.

A neighbour cell is said to be any cell within the coverage range of the safe distance (A_{sc}), whilst cells outside this range are considered to be safe cells. In a network, any non-neighbour cells, could be considered to be safe cells provided they are beyond the safe distance. Thus, for every cell, there are neighbour cells and safe distance cells. With this unique relationship, a preamble allocation algorithm inspired by the degree of saturation (DSATUR) colouration algorithm [89], is proposed. We note that the neighbour/safe cell relationship is mutually exclusive as you cannot have a cell that is both a neighbour to a cell and also at safe distance to that cell. The preamble reuse algorithm consists of a series of steps that helps to allocate root sequences to enBs using as few Zadoff-Chu sequences as possible. It is described in 1.

Algorithm 1: Preamble Reuse Algorithm

1 Identify the cell with the maximum number of neighbours that are not yet assigned root sequence indices. In case of a tie, randomly select any one cell that fulfils this criterion.
2 Assign the selected cell with the lowest available root sequence ID (RSID).
3 Assign the same root sequence ID to cell(s) that are at safe distance provided the safe cell's (under consideration for assignment) neighbours are not members of the initial chosen cell's (cell assigned an RSID in step 1 above) safe cells.
4 Select the next cell by identifying the cell with the highest number of neighbour cells that have unique RSID assigned to them. In case of a tie, select the cell with the highest number of neighbors and if there is a further tie, select a random cell that has fulfilled the stated criteria.
5 Assign the cell with the lowest possible available RSID, else, increase the number of RSIDs by 1 and assign same to the cell.
6 Assign the same root sequence ID to cell(s) that are at safe distance provided the safe cell's (under consideration for assignment) neighbours are not members of the initial chosen cellâĂŹs (cell assigned an RSID in step 4 above) safe cells.
7 Repeat steps 4 to 6 until all cells have been allocated RSIDs.

For heterogeneous networks, we can assume that at every point in time the macrocells are aware of small cells within their domain and can assign RSIDs to them. Since the macrocells in a network are aware of what RSIDs they have been individually assigned, it follows that the allocation of RSIDs to small cells can easily be done in a way that no two neighbouring macrocells' small cells have the same RSID allocated to them.

4.3.1 Implementation of the Preamble Reuse Algorithm

Consider a network that consists of a number of cells that are assumed to be randomly positioned in a 50 by 50 km space. For simplicity, the coverage range can be assumed to be free to overlap, this constraint does not affect the outcome of the algorithm, as it will only serve to help identify cells that are neighbours or at safe distances.

The results of implementing the preamble reuse algorithm when applied to a network with 30 cells is shown in Fig. 4.5. As can be seen from the figure, root sequences are represented using indexes $1, 2$, and so on. Also, no neighbour cells have the same preamble sequence assigned and the sequences are reused in cells that have fulfilled the criteria previously described above. Similar results are obtainable irrespective of the number and position of nodes used. However, the preamble reuse algorithm works with any number of cells, with a cost of an increase in the number of RSIDs as the number of cells increases.

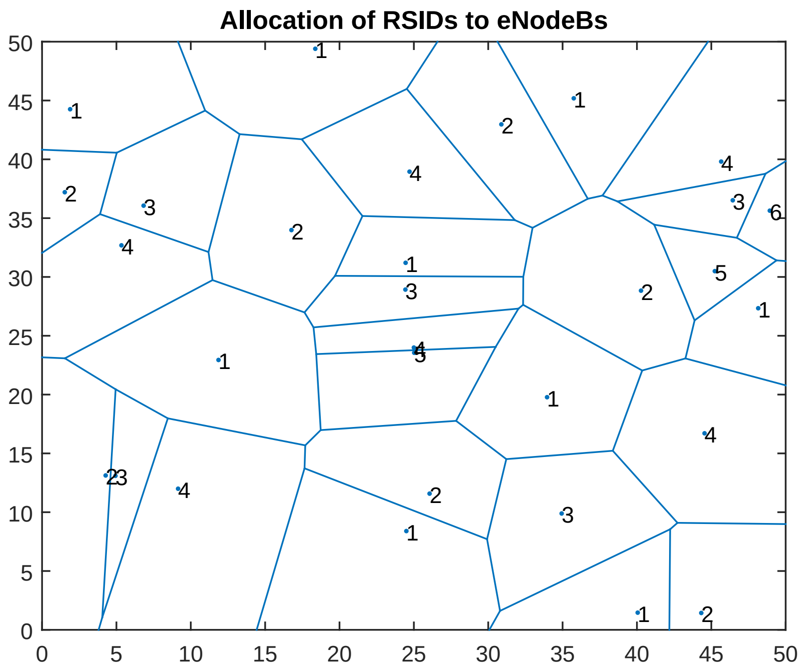

Figure 4.5: Position and allocation of 30 enodeBs

4.4 System Model and Numerical Results

4.4.1 System Assumptions and Settings

We consider homogeneous and heterogeneous networks that consist of MTC devices served by a number of cells covering the SCA. Arrival of random access attempts follows a Poisson distribution. The homogeneous network consists of a varying number of cells of 4 km, 8 km and 12 km radius as determined by equation 4.13 whilst ten small cells, each of 200 m radius, are added to the network to make it heterogeneous. As stated in previous sections, the cell radius determines the potential increase in the number of preambles that can be generated using the proposed scheme. We assume that all the preambles are available for use by the MTC devices and that the eNB will always have enough resources to deal with requests for devices that successfully pass through the steps of the random access procedure. The number of RA slots available per frame remains consistent with the 3GPP standards. Table 4.3 lists the basic parameters that are used in this study.

Table 4.3: Simulation Parameters Values

Parameter	Value
MTC device arrival distribution	Poisson distribution
Maximum # of MTC devices	10000
RA slot per frame	2
Arrival rate of RA attempts, λ	20 attempts/sec
Maximum # of preamble transmissions per device	10
Access Probability for ACB scheme	0.5
Backoff Indicator	20 ms

Three performance indicators, access delay, throughput and collision probability, have been chosen to evaluate the performance of the schemes considered in this study. The choice of these metrics is informed by the crucial roles they play in measuring the health of a cell or network. The access delay is a measure of how long it is likely to take a machine device to access the network. It is the time period that elapsed between the time when a device initiates the random access procedure to the time when it is successfully connected to the network. However, it does not include the time period when a device successfully transmits data either to a server or communicate with another MTC device. The collision probability helps the cell to estimate how well it is responding to the load offered.

The outcome of a preamble selection in the RACH process could be considered as one and only one of these:

- a preamble is selected by a device and successfully decoded by the base station;
- a preamble is picked by more than one device, in which case we assume there is a collision;
- a preamble is not chosen by any device.

The collision probability, P_{c}, is given by:

$$P_{\mathrm{c}} = \frac{N_{\mathrm{c}}}{N_{\mathrm{p}} - N_{\mathrm{u}}} \tag{4.16}$$

where N_{c} is the number of failed preambles, N_{u} is the number of unselected preambles and N_{p} is the total number of preambles in the cell. A preamble is said to be failed if it is chosen by two or more MTC devices from the same cell.

Throughput is considered to be the number of successful RA attempts that is served by the system per second. In the simulations, all MTC devices are of the same class and have equal priority when attempting network access. The cases where different classes of devices are present in a network is left for our future studies.

Four schemes are considered in this work namely, the default RA scheme, the ACB , the spatial group based reusable preamble allocation scheme (SRS) based on [47] , all of which have been described in the relevant section of this work, and the preamble reuse scheme (PRS).

Our simulation follows the RCA procedure described in previous sections. Device arrivals follow a poisson distribution and every device randomly selects a preamble which is sent to the receiver for detection. For ease of analysis, if two or more devices pick the select the same preamble, a collision is said to have occurred. We also assume that the eNodeB has enough resources to deal with devices that successfully complete the RCA procedure. Two hundred rounds of simulations were conducted at each point and the result is obtained by averaging.

4.4.2 Analytical Model

For a Poisson arrival process, the RCA throughput, G, is dependent on the arrival rate of RA attempts, γ, and the number of RCA opportunities, S [39]. We assume the number of RCA opportunities to be equivalent to the number of preambles:

$$N_{\mathrm{p}} = S \tag{4.17}$$

The per opportunity throughput, G_{o}, can be obtained as a function of S and γ as described in equation (4.18)

$$G_{\mathrm{o}} = \frac{\gamma}{S}\exp(-\frac{\gamma}{S}) \tag{4.18}$$

Thus, from equations (4.17) and (4.18), G, can be modelled, (4.19), as:

$$\begin{aligned} G &= N_{\mathrm{r}} G_{\mathrm{o}} \\ &= \gamma \exp\,(-\frac{\gamma}{S}) \end{aligned} \tag{4.19}$$

In our analysis, RCA throughput is dependent on factors including the number of devices seeking network access, the arrival rate and the number of RCA opportunities present at that instant. Thus, for a number of MTC devices in any network, N, and an arrival rate per device of λ, equation (4.19) can be further developed into (4.20), taking into account the total amount of traffic generated by the MTC devices over an arrival period.

$$G = N\lambda \exp\,(-\frac{N\lambda}{S}) \tag{4.20}$$

Using equation (4.16), we are able to estimate the probability of collision under varying load conditions (number of nodes). Computer simulations are conducted to evaluate the effectiveness of the proposed channel access scheme in terms of supporting multiple devices over LTE cellular networks, and verify the accuracy of the model used in calculating throughput as stated using equation (4.19) and (4.20). An AP value of 0.5 was used in simulations, this value only affects the ACB algorithm. There exists an inverse relationship between the value of AP and the number of devices that can start the RA procedure. Whilst this would mean that more devices can start the RA procedure if the value of AP is low, the probability of collision would increase. Thus, the choice of AP can be dynamic to reflect the load condition of the eNB.

4.4.3 Results

Performance results of the schemes in homogeneous and heterogeneous networks are presented in terms of throughput, access delay and collision probability. Figs. 4.6 and 4.7 shows throughput results obtained using simulation and numerical analysis while Figs. 4.8, 4.9, 4.10 and 4.11 show collision and access delay results obtained using simulations. It should be noted that in Figs. 4.6 to 4.11, the "Number of Nodes" on the x - axis refers to the number of devices attempting to access the network. At each point in the figures, this number is increased.

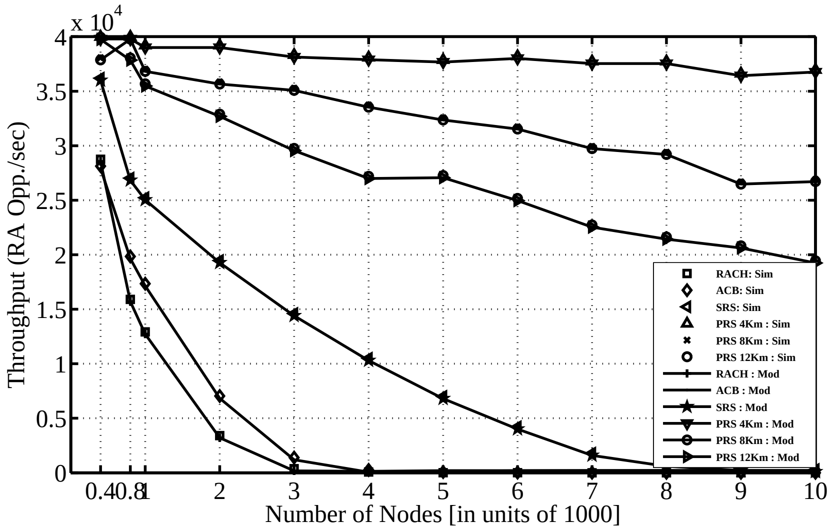

Figure 4.6: Throughput of representative RA schemes in homogeneous network

In terms of throughput, G, it is evident that an increase in the number of preambles would provide more successful RA opportunities per second and this will lead to an increase in the number of devices that can be supported. From Fig. 4.6 and Fig. 4.7, it can be seen that as the cell radius reduces, the number of successful RA opportunities per second increases and this means that more devices were able to access the network per subframe. PRS 12, 8 and 4 km radius cells shows this increase with PRS 4 km having the greatest number of RA opportunities per second followed by those of 8 km and 12 km implements the preamble reuse scheme but with cells of radius 12 Km. Whilst there is a steady decrease in throughput as the number of MTC devices attempting to access the network increases, this is unsurprising as the value of throughput is expected to decrease for large network load. The SRS scheme performs better than the default RA and ACB schemes, its grouping mechanism helps it to be able to cope with

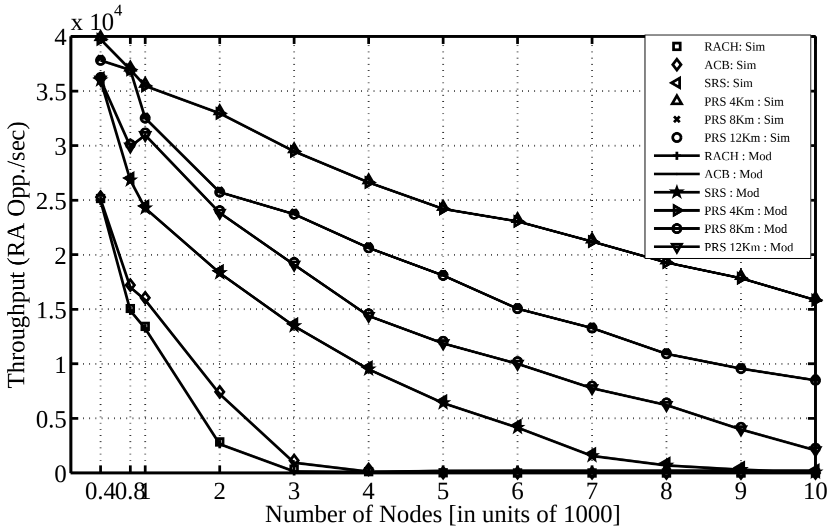

Figure 4.7: Throughput of representative RA schemes in heterogeneous network

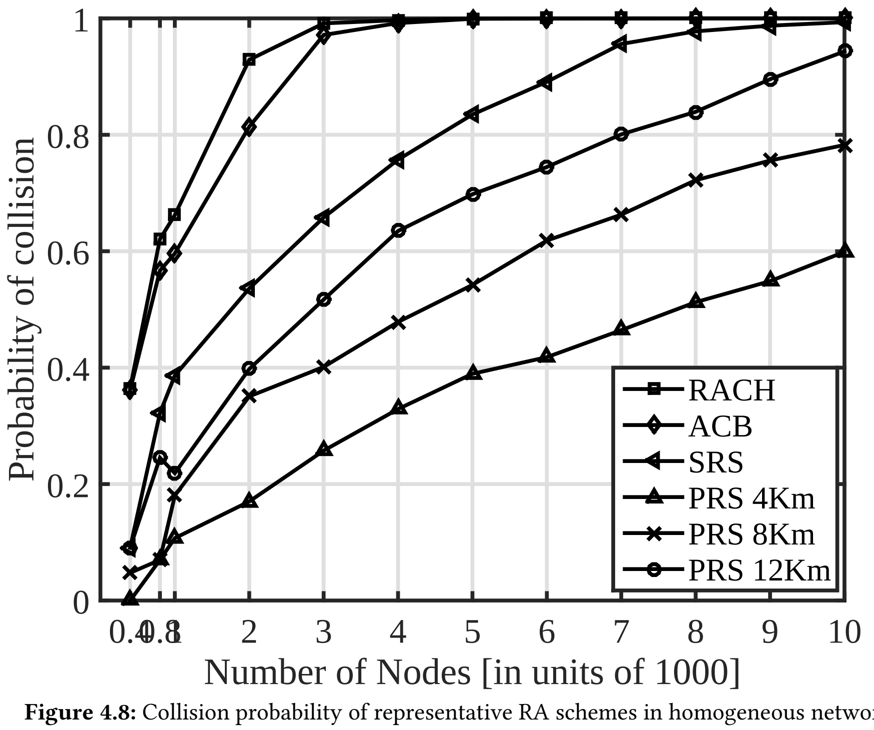

Figure 4.8: Collision probability of representative RA schemes in homogeneous network

some heavy RA loads but as the load increases, its performance begins to diminish. The default RA and ACB schemes are clearly unable to cope with the increase in load as both cannot provide RA opportunities for the huge number of devices.

Considering the probability of collision within the network, it is expected that as the number of devices attempting to access the system increases, the probability of collision also does as well. Fig. 4.8 and Fig. 4.9 confirms this expected result and whilst the ACB and default RA schemes (RCA) quickly get to saturation, the proposed scheme is able to cater for far more devices than both schemes under a reasonable probability of collision. However, the performance

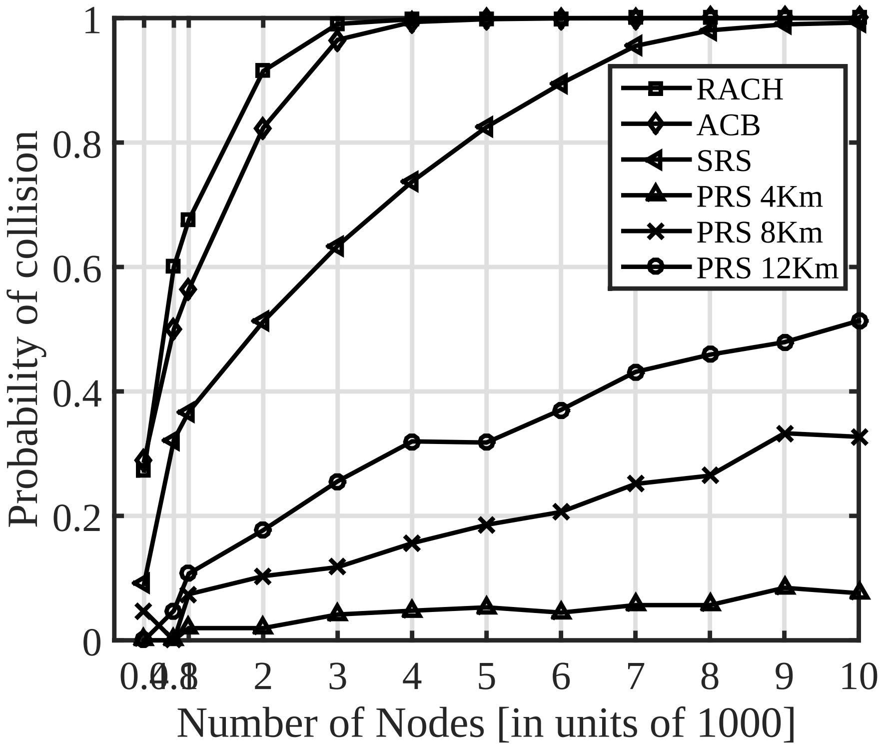

Figure 4.9: Collision probability of representative RA schemes in heterogeneous network

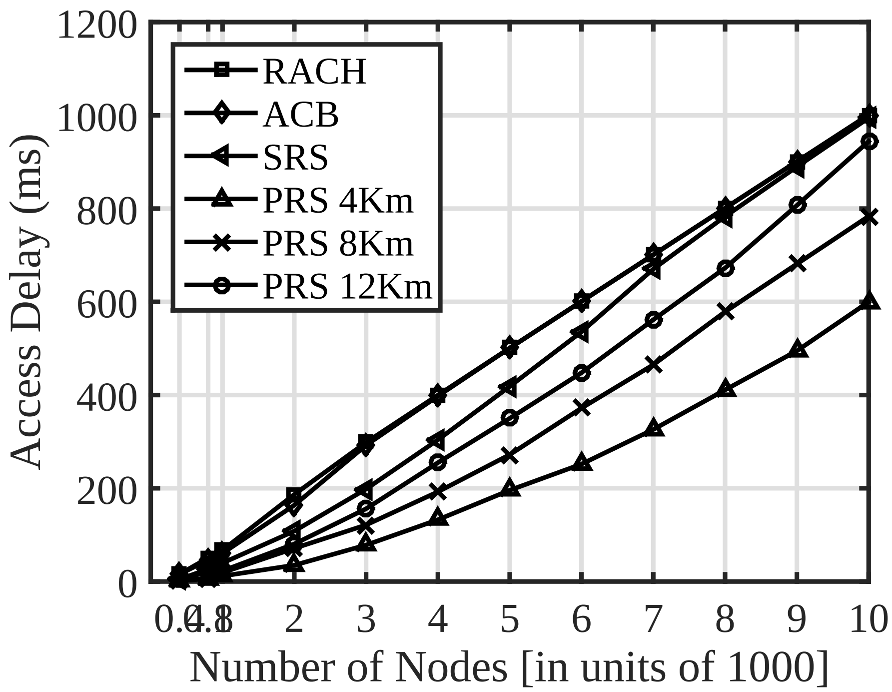

Figure 4.10: Access Delay of representative RA schemes in homogeneous network

of the proposed scheme is dependent on the number of preambles used, the greater this value, the lower the probability of collision. Even at maximum load considered in our simulation (10000 devices), the effect of an increase in the number of preambles can clearly be seen as the

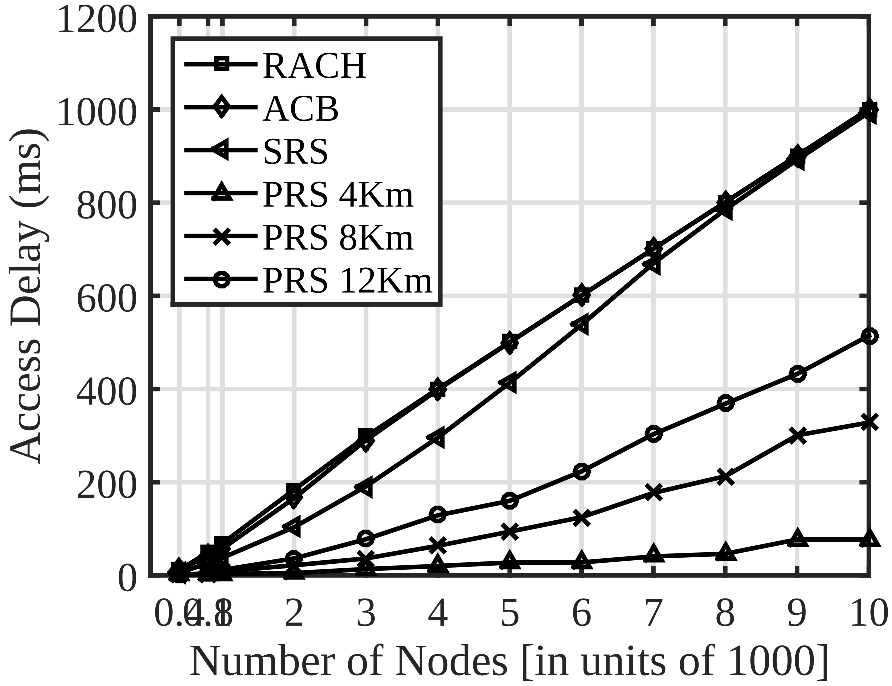

Figure 4.11: Access Delay of representative RA schemes in heterogeneous network

probability of collision is reduced by 50% and throughput performance increases by about 100%. However, the simulation can easily be extended to any amount of load. In terms of access delay, the same trend observed in the previous two metrics still follows. Figs. 4.10 and 4.11 shows the delay observed in homogeneous and heterogeneous networks respectively. From these figures, it is visible that the time spent waiting to access the network in an heterogeneous scenario is lesser than in an homogeneous network. This clearly demonstrates one of the advantages of small cell introduction into a network. Again, the spatial grouping of devices helps to achieve some level of performance improvement but this is not on the same level as to our proposed scheme. The drawbacks of SRS has also been stated in relevant sections. In terms of performance, the ACB and default RCA scheme have higher waiting times when compared to the proposed preamble reuse scheme. The waiting time is also reduces as the size of the cell, with a smaller cell radius having a lower waiting time. Thus, PRS 4 km has a lower wait time than PRS 8 km and PRS 12 km respectively.

4.5 Conclusion

In this chapter, we proposed the reuse of preamble sequences by cells within a certain distance and this has made it possible to increase the number of preambles available per random channel access slot for multiple machine devices. We derived a safe distance for preamble reuse while maintaining acceptable preamble detection probability by preventing excessive interference on preambles from neighbour cells. For practical deployment of our proposed preamble reuse scheme, a fast preamble allocation algorithm that can be used to allocate sequences in a cellular network was developed. Simulation and analytical model results demonstrated the effectiveness of the proposed preamble increase with an appreciable increment in throughput, corresponding reduction in the collision probability and access delay.

In practice, there are a number of ways in which the total number of devices attempting to access the network simultaneously can be minimised. One way is by aggregation, in which an MTC or specialized device is chosen as the sole intermediary between other devices and the network. Thus, rather than each device having to access the network, it sends its request to the intermediary which collates all requests and sends all as a single request to the network. For example, 10000 devices can be made into groups of a hundred and this would drastically reduce simultaneous RA attempts from 10000 to 100, a 99% reduction. Combining aggregation and our proposed scheme will further improve significantly the system performance considered under any metric. Practical implementation of our proposal can effectively tackle the problem of multiple devices trying to access the network simultaneously at any point in time by the provision of more RA opportunities to devices and a reduced collision probability. Our proposal opens a whole new vista of potential applications, in a wide variety of industries, that can be used in conjunction with it. In future, we intend to explore the coexistence of MTC and H2H devices and the effect of our proposal on them.

5 Random Channel Access Schemes for Massive Machine Devices in LTE

The ubiquity of cellular networks have made it an attractive option for emerging future communication paradigms. These advancements come with potential exponential increase in the number of devices needing network connectivity and would put a strain on the existing network, especially the RACH. Therefore, ways must be explored to cater for this projected increase especially in a way that guarantees satisfactory quality of service and throughput for both service providers and the end users amongst other metrics. This congestion in the RACH due to massive machine device access, is still an issue open to further research, with several solutions proposed. These have already been discussed in chapter 2 of this thesis.

From our investigations, it is clear that the fundamental challenge of the RACH when combating massive device access lies in the fact that there is only limited number of slots available for the process to take place and a fixed number of preambles within every available slot, from which devices have the opportunity of selecting. These resources are not sufficient for a massive number of machine devices. Also, the original design of cellular networks was for human to human communications, however, extending its use to machine communications poses a special kind of challenge especially when considering the issue of which of these type of devices would have priority when using the network.

In this chapter, we propose two ways in which the congestion in RACH challenge can be solved. The first is by allocating more frequency resources (subframes) for the random access procedure whilst the other is in using smaller cells enhancements to support RA especially to MTC devices.

The remainder of this chapter is organized as follows. In Section 5.1, we extensively discuss the idea of adaptive subframe allocation scheme. Specifically, the problem is formulated as an optimization problem with objective to maximize the system utility, which is defined as a function of RACH throughput and the subframe resource used by RACH. We first solve the optimization problem to find the best configuration of the number of subframes for a known RACH load (i.e., number of machine devices). Then we propose a method to estimate the RACH load according to the observations of historic RACH outcome at eNodeB to choose the number of subframes for the next round of random access. Section 5.2 presents the performance evaluation of the proposed subframe resource optimisation scheme. Section 5.3 investigates the effect of small cells in combating the congestion in RACH challenge including a discussion on Zadoff-Chu sequence and their configurations in heterogeneous networks. Section 5.4 discusses in details the performance of the proposed small cell enhancements in combating the RACH overload challenge. The chapter is then concluded in section 5.5.

5.1 Subframe Resource Optimization Scheme

In addition to the research efforts to improve random access protocols, the potential RACH congestion problem can be alleviated by allocating more frequency resources (subframes) for machine devices communications. However, this comes at a cost to the network, therefore, ways must be sought to dynamically allocate resources such that they are optimally utilized.

According to the 3GPP specification, there are ten subframes in one frame and by default, two subframes are allocated for the RACH procedure while the remaining are used for data communications. However, as the number of devices accessing the network varies per time, fixing the number of subframes that should be used to carry out random access would not be effective especially when the network is under heavy loads. Therefore, the network should be able to adjust to variation in load by varying the number of subframes with a view to providing effective service to end users. Presently, there is no specification from 3GPP or research reported on how to efficiently allocate the subframes for RACH to deal with the dynamic loads while also leaving sufficient resource for human devices.

In this section, we propose an adaptive subframe allocation scheme to address the above problem. The RACH challenge is formulated as an optimization problem with the objective of maximizing the system utility, which is defined as a function of RACH throughput and the subframe resource used by RACH. We first solve the optimization problem to find the best configuration of the number of subframes for a known RACH load (i.e., number of machine devices). Then, we propose a method to estimate the RACH load according to the observations of historic RACH outcome at eNodeB to choose the number of subframes for the next round of random access. This adaptive scheme produces an improvement in utility of about 75% when compared to non-adaptive resource allocation schemes over the default RACH access protocol.

5.1.1 LTE Frame structure

The resource block (RB) is the minimum unit of resource scheduling in LTE and one RB consists of 12 sub-carriers in the frequency domain and one subframe in the time domain. A subframe has a time duration of 1 ms and a group of ten subframes is referred to as a frame. For Frequency Division Duplex (FDD), this arrangement is shown in Figure 5.1. Subframes could either be used

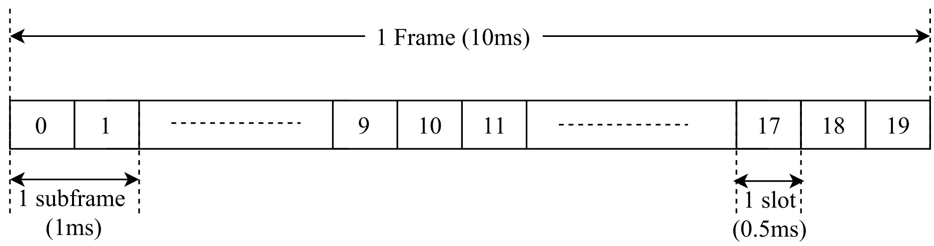

Figure 5.1: 3GPP Frame structure

to carry data or for Random Channel Access (RCA). The 3GPP specifies a number of subframes in a frame that can be used for random access and this is based on the physical random access channel (PRACH) configuration index. However, the number of subframes used for random access affects the number of subframes that can be used to carry data. The more subframes used for random channel access, the fewer the subframes that can be used to carry data in that frame. Within each RCA subframe, multiple users may access the channel concurrently using unique preambles. A problem will arise when a massive amount of users attempt network access via one RCA slot, as there is high probability of users selecting the same preamble and therefore not gaining access to the network.

5.1.2 Utility Function

The RACH is the first point of call for devices seeking to communicate in an LTE network. A number of steps, known as the RACH procedure, have to be successfully completed before devices are granted network access. This procedure is fully described in chapter 2.

Based on the RACH procedure, it is evident that a number of variables affect the RACH performance. Let N_d denote the system load in terms of the number of devices seeking to access RACH. Also, let N_s and N_p denote the number of subframes and the number of preambles allocated for the RACH process respectively. We can set N_s to a minimum value of 2 (the default) and a maximum of 8. Considering a fixed system load, increasing N_s will clearly have a positive effect on the number of devices that will be able to successfully complete the RACH process, even as devices will have more unique transmission opportunities, thus, having a positive effect on the system throughput, which is defined as the number of successful RA attempts measured in devices per frame. Clearly, the performance of any system, in terms of throughput, is affected by two main factors, which are, the number of devices seeking network access, N_d, and the number of subframes, N_s.

Let η denote the actual number of successful devices per frame, measured in devices per frame and let α be the benchmark number of devices per subframe. α is a variable under the control of the network operator , measured in device per subframe, that regulates the RACH procedure within a subframe. The effects of α will be discussed further in the following section. Considering these parameters, η, α and N_s, we can derive a utility function with the objective of determining the ideal number of subframes to be used to achieve maximum utility.

Consequently, we define a simple utility function to assist in the subframe allocation, which is shown in Equation 5.1 below:

$$U = \eta - \alpha N_\mathrm{s}, \tag{5.1}$$

where U denotes the system utility. It should be noted that the system utility has the same unit as throughput, however, more variables that affect the RACH procedure have been put into place and in order to improve on system performance, solving this utility function will be crucial.

5.1.3 Optimization Problem and Solution

Based on the utility function, a simple optimization problem can be formulated as follows: Given a RACH load (N_d), find the best allocation of subframes (N_s) to maximize the system utility U. An approach to solving this optimization problem is discussed next.

The throughput, η, can be expressed as a function of N_d, N_s and N_p as shown in 5.2 [39]:

$$\eta = N_\text{d}\text{exp}(-\frac{N_\text{d}}{N_\text{s}N_\text{p}}). \tag{5.2}$$

Substituting (5.2) into (5.1) becomes

$$U = N_\text{d}\text{exp}(-\frac{N_\text{d}}{N_\text{s}N_\text{p}}) - \alpha N_\text{s}. \tag{5.3}$$

Differentiating (5.3) produces:

$$\frac{dU}{dN_\text{s}} = -\frac{N_\text{d}^2}{N_\text{p}N_\text{s}^2}\text{exp}(-\frac{N_\text{d}}{N_\text{s}N_\text{p}}) + \alpha \tag{5.4}$$

At maximum value, $dU/dN_\text{s} = 0$. Equation (5.4) could be solved to yield and obtain the theoretical optimal value N_s^opt for N_s,

$$N_\text{s}^\text{opt} \approx -2N_\text{p}\text{F}_\text{w}(-N_\text{d}\alpha N_\text{p}/N_\text{d}^2)^{1/2}/2N_\text{p} \tag{5.5}$$

where F_w is the Lambert W function.

According to the above formula, the number of subframes to be allocated, N_s, can easily be calculated for a given number of devices, N_d and a fixed value of α. Figure 5.2 shows the number of subframes against various loads with different values α.

The effect of α cannot be overemphasized, as it can be used by the network operator to regulate the allocation of subframes for random access. Note that a frame is divided into subframes, some of which are used to carry data whilst the remainder is used for the RACH procedure. Increasing or reducing one, has a direct effect on the other.

From Figure 5.2, a low value of α quickly increases the number of subframes used for RACH procedure whilst a higher value gradually increases the number of subframes being allocated. For example, at $\alpha = 2$, there is a sharp rise in number of subframes allocated, from 2 subframes required for about 10 devices, to 6 for 70 devices, and then to the maximum of 8 even at just 100 devices. This is in contrast with an α value of 25, which does not use the maximum number of subframes until about 550 devices. It is also worthy of note that at $\alpha = 50$, we see the default number of subframes (2) being allocated until the system load reached 400 devices, before a gradual increase afterwards. Consequently, at high values of α, say 100, the default number of subframes will be used to achieve maximum utility whilst at very low value or even zero value ($\alpha = 0$), the maximum number of subframes (8) will be required to achieve maximum utility irrespective of the number of devices. However, for the remainder of this work, we will use a

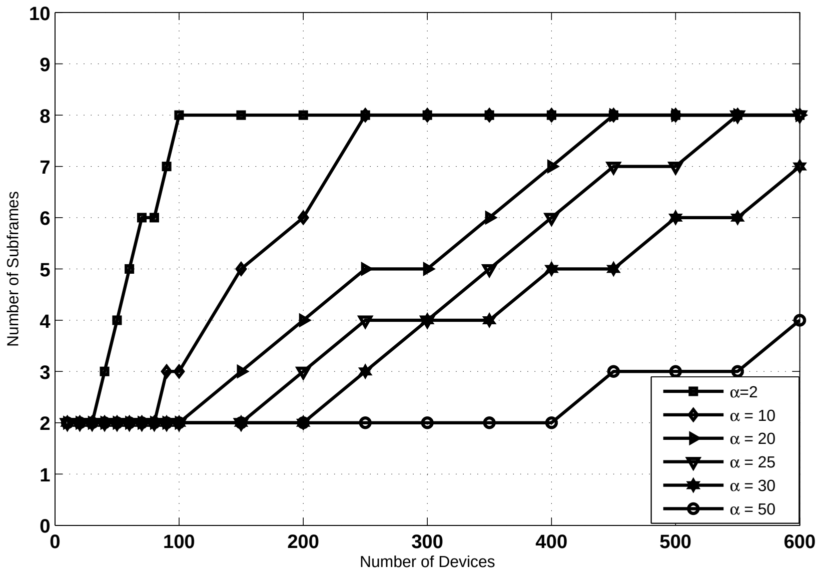

Figure 5.2: Number of subframes under varying system load and varying α.

value of $\alpha = 25$.

Consequently, as can be seen from Figure 5.2, we can easily determine the optimum number of subframes to be used by the network in providing effective service based on existing load. For values that are not within the range shown in the figure, we stick to the maximum number of subframes (8) in this case. In practise, based on the results obtained here, look up tables can be made offline and used to determine the optimum number of subframes for online operations.

Next, we consider the problem of estimating network loads by the eNodeBs, which is unknown but is required for the eNodeBs to determine the number of subframesN_s for RACH as discussed before. It is noted that the estimation can be performed on a very short time frame (e.g. for every frame period) to a long time frame (e.g. a few hours). Due to the randomness of the network loads it is extremely difficult to accurately estimate the network loads. In this work, we apply equation 5.2 to estimate the number of devices attempting to access the RACH for the last round of channel access. As the eNodeB knows the throughput η and the values of N_s and N_p, we can obtain an estimation of N_d by solving equation 5.2 which has only one unknown variable. It is noted that equation 5.2 can have two solutions for N_d corresponding to the light and heavy network load conditions, respectively.

$$N_\mathrm{d} \approx \begin{cases} -2N_\mathrm{p}N_\mathrm{s}\mathrm{F_w}(-(\alpha N_\mathrm{p}N_\mathrm{s}^2)^{1/2})/2N_\mathrm{p}N_\mathrm{s}\,; \\ -2N_\mathrm{p}N_\mathrm{s}\mathrm{F_w}((\alpha N_\mathrm{p}N_\mathrm{s}^2)^{1/2})/2N_\mathrm{p}N_\mathrm{s}\,. \end{cases} \tag{5.6}$$

where $\mathrm{F_w}$ is the Lambert W function and N_p is the number of preambles in an eNodeB.

In order to predict which of these is to be used, we use the preamble detection outcomes at the eNodeBs. The outcome of a preamble selection in the RACH process could be considered as one and only one of these:

1) a preamble is selected by a device and successfully decoded by the base station;
2) a preamble is picked by more than one device, in which case we assume there is a collision;
3) a preamble is not chosen by any device.
Unselected preambles could help distinguish when the system is under heavy load or not. Whilst the choice of preamble selection by devices is entirely random, we infer that if there is heavy load on the system, there would be fewer numbers of unselected preambles and vice versa. This helps estimate the numerical performance of the proposed scheme. Then the eNodeB can use an average of the estimated number of devices as the estimated network load to be used for subframe allocation.

5.2 Performance Evaluation of the Subframe Resource Optimization Scheme

In this section, we evaluate the performance of the proposed subframe resource optimization scheme in terms of utility and system throughput. We also verify the reliability of the system load prediction equation proposed in 5.6 both by using simulations and numerically.

5.2.1 Simulation results and Analysis

An homogeneous network with changing system load over time, served by a single eNodeB is considered. The arrival of devices following a poisson distribution with the mean arrival rate increasing linearly first and then decreases linearly, as shown in Figure 5.3. We assume that all 64 preambles are available for use by the devices and simulate the RACH procedure with the assumption that the eNodeB has enough resources to accommodate devices that successfully complete the RACH procedure. Performance metrics used in this study are the system throughput and the utility as obtained from (5.3).

It should be recalled that the default RACH process only uses a fixed number of subframes (2), and is denoted by RACH in our figures.

Next, we test the effectiveness of the method to estimate network loads by solving Equation 5.2. Known values of network load, as displayed using Load_sim in Figure 5.3, is used to predict the potential network load for the next round of access. For instance, under time 5ms, the network load assumed under simulations is 300 devices, however, for numerical estimations, the last estimated value (number of devices estimated at 4ms) is used to predict (using equation 5.2) the potential number of devices for 5ms and in this case, there is an exact match between the simulated and predicted values. However, cases exists where there is a mismatch between the simulated and numerical load values, (such as, at 11ms) and this could be attributed to the way in which the system predicts whether the network is under heavy or light loads. As stated earlier, in our calculations, unselected preambles are used to distinguish the load condition of the network and this have proven to be accurate to a large extent, with more than 95% of simulated values matching numerical calculations. Despite this rare mismatch occurrence, the proposal was able to adjust and correctly predict loads for the next round of random access and beyond, supporting the belief that this blip does not invalidate our proposed methodology. Conclusively, Figure 5.3 shows a close match between the simulated and expected network load. Load_sim and Load_num are the values of load in simulation and numerical analysis respectively. This further lends credence to the derivation in Equation 5.6.

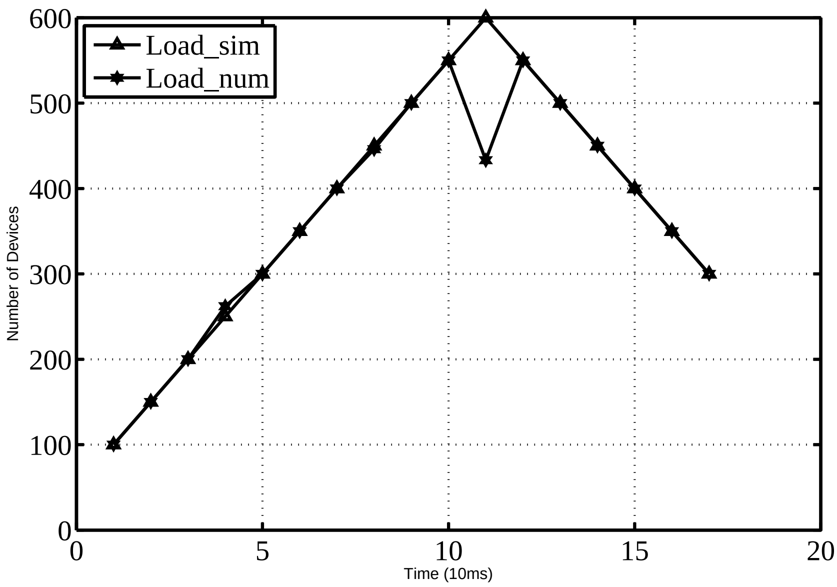

Figure 5.3: Network load estimation over time.

Fig. 5.4 presents the utility obtained by the adaptive and fixed allocation schemes with the varying network loads. Four performances are represented in the figure including, the utility performance of our proposed subframe adaptation scheme, both in simulation (FA_sim) and by numerical analysis (FA_num), the default RACH procedure and the ACB scheme. Using network loads shown in Figure 5.3, we evaluate the performance of these schemes in terms of utility. Again, very close matches between utility values obtained via simulations and numerical calculations are observed, further supporting the validity of our derivations. In terms of performance, the benefit of the adaptive scheme is clearly evident as it outperforms both ACB and the default RACH process, with more than 100% increase in utility in some instances. Both RACH and ACB are not able to obtain high values of utility and this will in effect translate to them not being able to cope with heavy network loads.

Further analysis of these schemes in terms of access delay, collision probability and throughput is carried out in the following sections. These metrics would be able to provide an interesting insight into the performances of these schemes and their impact on the QoS delivered to the network.

However, as we have proven from Figure 5.4 and Equation 5.1, the cost at which an increase in throughput is provided might not always be optimal. In cases where all subframes are used to provide RA opportunities, the subframes for data are sacrificed. Thus, fixing the subframes to a maximum would be detrimental to data services, whilst using the minimum number of subframes (RACH) would lead to degradable QoS under very heavy loads.

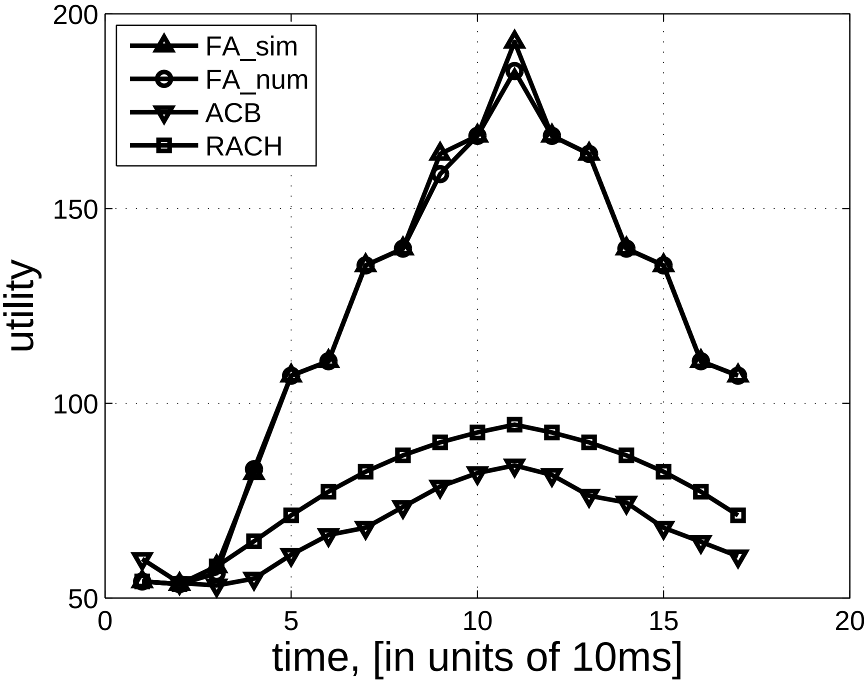

Figure 5.4: Utility comparison of the fixed and adaptive subframe allocation scheme with time.

5.3 Small cells in combating RACH congestion in LTE

In this section, we investigate the idea of using smaller cells to support RA to machine type communication (MTC) devices. Small cell deployment seem to be an attractive option for providing improved QoS to users partly due to the closeness of the cells to devices, making environmental conditions fairly stable [90]. Its use is gaining ground in industry, but to the best of our knowledge, using them to enhance random access and machine type communications has not been well studied in literature. One of the characteristics of machine type devices is that, it is expected that a large number of these devices will individually transmit little data [91], therefore, it makes sense to envision that small cells will be ideal in providing network access to these devices. In addition to using the existing small cells that are deployed for data traffic offloading purpose, additional small cells can be deployed on demand mainly to handle RA traffic loads.

This proposed implementation has the potential advantage of providing more random channel access opportunities for the machine devices and there could also be significant increase in random access throughput.

5.3.1 Small cells in LTE

Small cells have become increasingly important as they have the ability to provide increased system capacity. Present networks are gradually migrating to an heterogeneous phase where there is a mix of small cells and macrocells as shown in Figure 5.5. Expectedly, this kind of networks still suppose the communication interfaces and protocols of the present homogeneous cellular network with MTC devices or user equipments (UE) communicating via the air interface, inter MME communications via the S1 protocol and inter-enodeB or small cells to enodeB communications occurring using the X2 protocol.

Small cells could either be pico-cells or femtocells, the difference being that the latter is controlled and managed more closely by the operator network whilst the former are semi-autonomous which can decide what the optimum operating parameters are for the environment they are being used. Small cell also have the advantage of flexibility with regards to deployment and control. They could either be deployed indoors or outdoors and with or without macro-cell coordination. Other advantages of small cell deployment include significant increase in user throughput both in uplink and downlink, reduction in access delays and their proximity to user devices means that there is a potential reduction in energy consumption on the UE side. A comprehensive survey of the benefits and issues arising with the use of small cells and some operator views is fully discussed in [92].

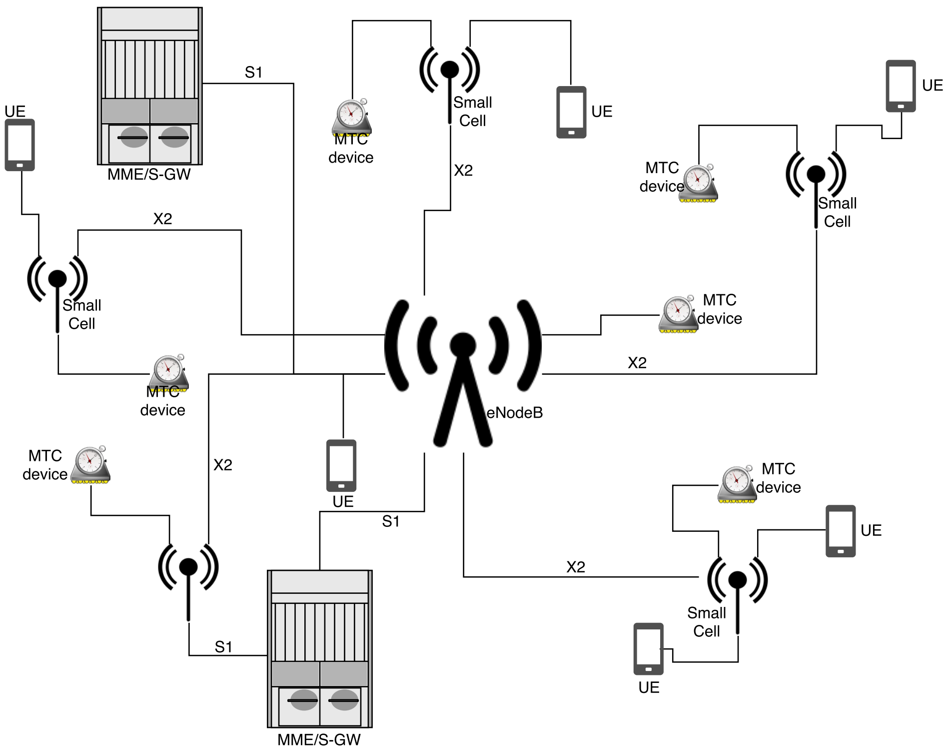

Figure 5.5: Heterogeneous 3GPP network Architecture

Next, we consider the impact of using small cells to enhance random access especially for massive MTC devices. Although, existing small cells are being primarily deployed for human device traffic, we propose that their use be extended to support the random access procedure especially for MTC devices.

The strategy employed in small cell deployment with regards to MTC should be different from those of the existing human devices in terms of cell coverage and spectrum resource demands.

At present, small cells are typically deployed in areas of high data traffic loads but this could further be extended to locations with heavy RA loads.
Whilst small cells have obvious advantages, issues such as the traditional spectrum allocation is still a challenge being actively researched in industry and academia. However, with respect to implementing small cell support for MTC, we focus on the configuration of Zadoff-Chu sequences and the generation of preambles for random access of these devices.

5.3.2 Zadoff-Chu Sequence and Random Access Preambles

As explained in previous chapters, random access preambles in LTE are generated from Zadoff-Chu sequences and they have specific properties that allows them to be detected by eNBs. Specifically, the constant amplitude and zero autocorrelation properties coupled with periodic and simple detection mechanism of Zadoff-Chu sequences makes them the prime candidate in LTE. With 838 root Zadoff-Chu sequences from which preambles can be generated and a root sequence length of 839, it is possible for several preambles to be generated by cyclically shifting one root sequence [19]. These cyclically shifted Zadoff-Chu sequences still possess constant amplitude which nis critical in maintaining the low power to average (PAR) ratio and ensuring efficient power amplifier utilization. Also, sequences generated from the same root have a cross correlation of zero between them, provided that the value of cyclic shift used when generating the preambles is larger than the maximum round trip time in the cell including the maximum delay spread of the channel [19]. The implication of this is that the ideal cross-correlation property of these sequences ensures that there is no intra-cell interference from multiple random access attempts using different preambles obtained from the same root sequences.

5.3.3 Root Sequence Configurations for Heterogeneous Networks

For small cells to be used for random access, the allocation of root sequences and subsequent preamble generation plays an important role. The number of sequences to be used in generating the default number of preambles per cell as required by the 3GPP standards is influenced by the cell radius amongst other parameters. Furthermore, the cell radius is used to determine the cyclic shift required by the cell and this relationship is given by equation (5.7).

$$N_{\text{cs}} \geq \Big[\Big(\frac{20}{3}r - T_{\text{ds}}\Big)\frac{N_{\text{zc}}}{T_{\text{seq}}}\Big] + n_{\text{g}} \tag{5.7}$$

where r is the cell radius (km), T_{ds} is the maximum delay spread, $N_{zc} = 839$ and T_{seq} is the preamble sequence length (in μs) and n_g is the number of additional guard samples due to receiver pulse shaping filter [19].

Usually, the value of the required cyclic shift is signalled as part of the system information and as can be deducted from equation (5.7), there is a direct relationship between cell radius and the value of cyclic shift, therefore, smaller cells will require a smaller cyclic shift value. This means that the required 64 preambles for a small cell can be generated by cyclically shifting a single Zadoff-Chu root sequence. However, in larger cells, larger cyclic shift values are required and consequently, multiple root sequences would be needed to generate the required number of preamble for the cell.

According to the 3GPP standards, there are 16 PRACH configuration indexes that can be used to configure a cell. These are shown in Table 5.1. The table shows the configuration index and the number of root sequence required for a stated cell radius.

Table 5.1: PRACH configuration index table

Configuration Index	Number of required root sequences per cell	Cell radius [km]
0	64	118.8
1	1	0.7
2	2	1
3	2	1.4
4	2	2
5	2	2.5
6	3	3.4
7	3	4.3
8	4	5.4
9	5	7.3
10	6	9.7
11	8	12.1
12	10	15.8
13	13	22.7
14	22	38.7
15	32	58.7

In order to allocate root sequences to eNBs, a simple centralized process is considered. This process ensures that each eNB within the network has sufficient root sequences required to generate the default 64 preambles. The possibility of generating more than the deafult number of preambles have been considered in chapter 4. The allocation process envisaged here is run at two levels. Firstly, the total available root sequences are divided into two pools which are sequentially numbered, the first for macrocells and the other for small cells. The pool sizes can be set as inversely proportional to the average radius of macrocells and small cells. Subsequently, allocation of sequences from the respective pools can easily be carried out.

The allocation of sequences could first be done at the macrocell level, starting from the macrocell at the center of the network being allocated the first sequence and so on until all macrocells are covered. The number of root sequences required can either be calculated from equation 5.7 or looked up in Table 5.1. Then, either in a clockwise or anti-clockwise manner, the remainder of sequences from the macrocell pool are allocated to neighbour macrocells from the first tier. However, upon reaching the largest sequence number in the pool, allocation would start again from the very first sequence.

Then, sequences from the macrocell pool are allocated to neighbor macrocells from the first tier in either clockwise or anti-clockwise way with the unused root sequences outwards until the last macrocell in the network. If the largest sequence number in the pool for macrocells is reached at any time of the allocation process then the sequences starting from 1 are reused sequentially. The process is easily extended to small cells but this time under the control of macrocells. In addition to this allocation process, a geographical based segement allocation can also be used, particularly, for the small cells. The idea here is to assign a block of sequentially numbered root sequence from the pool of sequences for small cells to each segment with regular shape and size. In this way, a small cell under the control of a macrocell receives root sequences assigned based on the location of the small cell and less or more sequences can thus be requested according to the needs of the small cells. However, it is the responsibility of the macrocells to ensure that small cells that are relatively close to each other do not end up using the same root sequences. The possibility of this occurence is very low, even as small cells usually require a minute number of root sequence for preamble generation. Therefore it is expected that interference from the same preambles transmitted from neighbor cells is negligible and zero cross correlation between sequences is also maintained.

An example of a distribution of root sequences is shown in Figure 5.6. In this figure, we take assume that each macrocells consists of three sectors and five root sequences are used to generate the default number of preambles in macrocells whilst small cells use two.

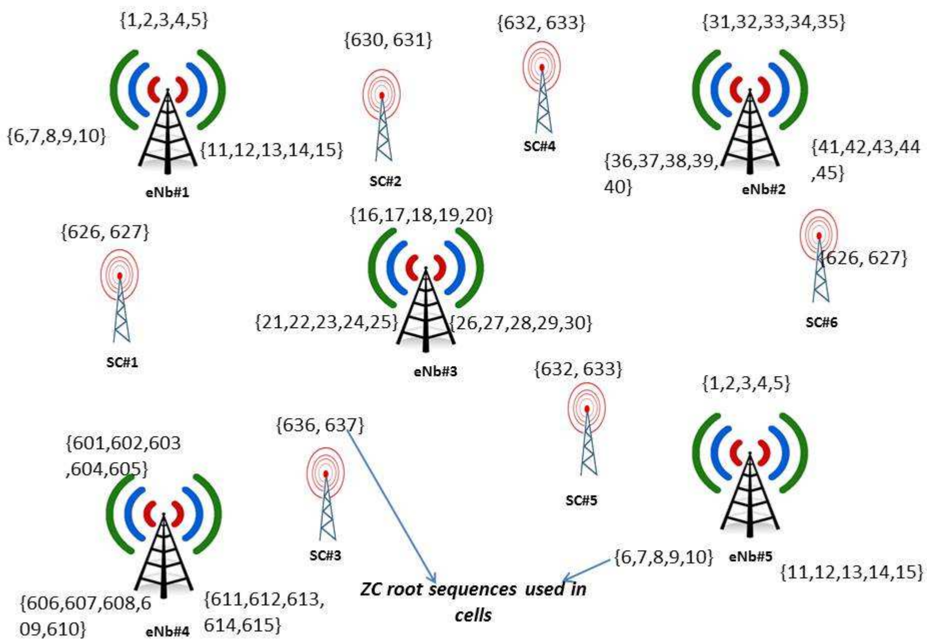

Figure 5.6: Sequence distribution in an heterogeneous 3GPP network Architecture

5.3.4 Analytical Model

The definition of collision probability, P_{c}, from chapter 4 still hold in our numerical analysis and with the outcome of preamble selection being either one of unselected, selected by a single device and selected by more than one device, our calculations for this metric remains the same. This is stated in equation 5.8.

$$P_{\mathrm{c}} = \frac{N_{\mathrm{c}}}{N_{\mathrm{p}} - N_{\mathrm{u}}} \tag{5.8}$$

where N_{c} is the number of failed preambles, N_{u} is the number of unselected preambles and N_{p} is the total number of preambles in the cell.

A preamble is said to be collided if it is chosen by two or more MTC devices.

Similarly, the throughput is considered to be the number of RA attempts that can be served by the system per second. Again, the calculation of throughput is also stated in chapter 4 but in this chapter we push the bounds a bit further. Equation 5.9 gives the throughput of the wireless system.

$$\begin{aligned} G &= (no.of\ RACH\ oppt.) \cdot (per\ opporunity\ throughput) \\ &= S \cdot [\frac{\gamma}{S} \cdot \exp(-\frac{\gamma}{S})] \\ &= \gamma \cdot \exp(-\frac{\gamma}{S}) \end{aligned} \tag{5.9}$$

Thus, for a number of MTC devices in an homogeneous network, N, and an arrival rate per device of λ, equation (5.9) can be further developed into equation (5.10) below, taking into account the total number of traffic generated by the MTC devices over an arrival period.

$$G = (N \times \lambda) \cdot \exp(-\frac{N \times \lambda}{S}) \tag{5.10}$$

However, for an heterogeneous network, the throughput is taken as the mean of individual throughputs generated in each small cells that make up the network. Thus, for m small cells,each with N_{sc} MTC devices, the overall network throughput is given as

$$G = \frac{\sum_{N_{sc}=1}^{m} (N_{sc} \times \lambda) \cdot \exp(-\frac{N_{sc} \times \lambda}{S})}{m} \tag{5.11}$$

Using equations (5.9), (5.10) and (5.11), estimating the value of collision probability and RACH throughput under varying load conditions (number of nodes) and small cells is easily done.

5.4 Performance Evaluation of proposed schemes

In this section, we evaluate the performance of the proposals discussed earlier in this chapter. These include the use of small cell enhancements and frame adaptation in supporting massive machine device access in LTE networks.Also included in this evaluation are legacy schemes including the default RACH protocol and the ACB scheme.

5.4.1 System Assumptions and Settings

For the simulations, an heterogeneous network served by a macro-cell and variable number of small cells is assumed for the small cell scenario whilst the frame adaptation assumes there is no small cell in the network. System load distribution is entirely random, however, in analysis, this randomness is taken care of by observing the mean of the cells within the network being simulated. The Poisson distribution is used to model the arrival of random access attempts. With an assumed small cell radius of less than $1Km$, by cyclically shifting a single Zadoff-Chu root sequence, the required 64 preambles are generated for each small cell. By using unique root sequences, there is minimal chances of inter-preamble interference from cells using the same sequences.

We assume devices have the freedom to select any other preamble and that macro cells or small cells have enough resources to cater for requests from devices that successfully scale through the first three steps of the random access process. Table 5.2 lists the basic parameters that are used in carrying out our analysis.

Table 5.2: Simulation Parameters

Parameter	Value
MTC device arrival distribution function	Poisson
Maximum number of MTC device	10000
RA slot per frame	2
Arrival rate of RA attempts (λ)	20
Number of small cells	15
Total number of preambles	64
Maximum number of preamble transmission	10
Access probability (AP)	0.5
Backoff indicator	20ms

Again the metrics used to evaluate the performance of the schemes are collision probability, system throughput and access delay. As expected, the access delay indicates how much time it takes a device to successfully complete the random access procedure, which is the first step in gaining network access. This time does not include the time it takes for resources to be allocated to the device or how long its communication with either the base station or other MTC devices take. The access delay gives an estimation of the wait time of the device for network access. Our

Simulations also consider that MTC devices are all of the same class and have the same priority in accessing the network, cases where different classes of devices are present in a network is to be considered in our future studies.

5.4.2 Results

Computer simulations and numerical results are presented in this section. In order to visualise the effects of the addition of small cells into a network, we consider a scenario in which the number of small cells is varied with 5 being the least and 15 being the most. We also compare the performances of the proposed frame adaptation scheme with small cell enhancements and other existing schemes including RACH and ACB. The number of RA slots available per frame remains consistent with the 3GPP standards. Figures 5.7, 5.8 and 5.9 show the results of our comparison in terms of collision probability, throughput and access delay. The performance of small cells represented by (SC-5) for five small cells, (SC-10) for ten and (SC-15) for 15 while FA represents the frame adaptation scheme discussed earlier in this chapter, RACH represents the performance of the default random access scheme and ACB represents the access class barring scheme.

Considering collision probability, a direct relationship is observed to exist between this metric, the number of small cells and the number of devices used. Thus, as evident in Figure 5.7, increasing the number of devices have the effect of reducing the overall collision probability of the system. However, we see the effects of the number of small cells with $SC - 15$ having the lowest value of collision probability when compared with (SC-10) and (SC-5). In majority of the cases considered, the frame adaptation algorithm (FA) does better than having five small cells within the network ($SC - 5$), however, as the number of small cells is increased, there is marked improvements in the performance for small cell enhancements i.e $SC - 10$ and $SC - 15$ outperforms the FA algorithm. This performance improvement comes at a cost, which is, increasing the number of small cells within the network leads to an increase in capital expenditure for operators. Operators, then, have a choice between reducing the collision probability at no direct cost and having a potentially similar impact on the system by adding small cells to the network.

The flaws of the default RA and ACB schemes are seen clearly under heavy network loads. Considering the whole system, increasing the number of small cells increases the number of random access slots and since there are fewer devices competing for a constant resource (preambles), there is bound to be a reduction in collision probability. Also, adjusting the number of subframes used for RA as done by the FA scheme, produces a similar effect with the number of random access slots being dynamically adjusted to suit the prevailing network load. The limited resources (64 preambles) from which devices have to contend before gaining network access is still a major flaw in the design of both the RACH and ACB schemes.

In terms of system throughput, Fig. 5.8, shows the performance of the schemes under comparison. Again, there is an improvement in the average number of RA opportunities available

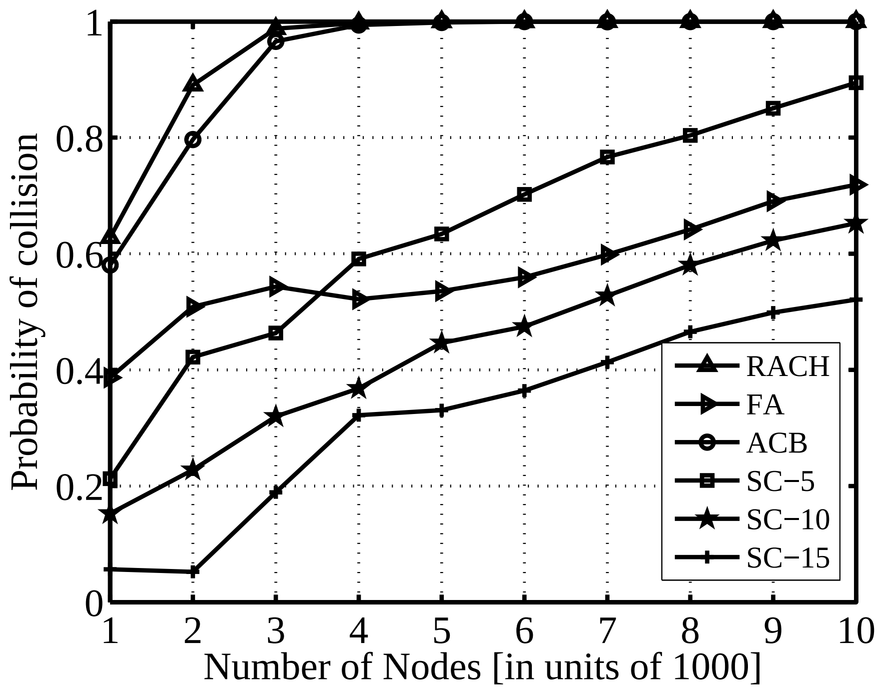

Figure 5.7: Collision probability of representative RA schemes with variable number of small cells

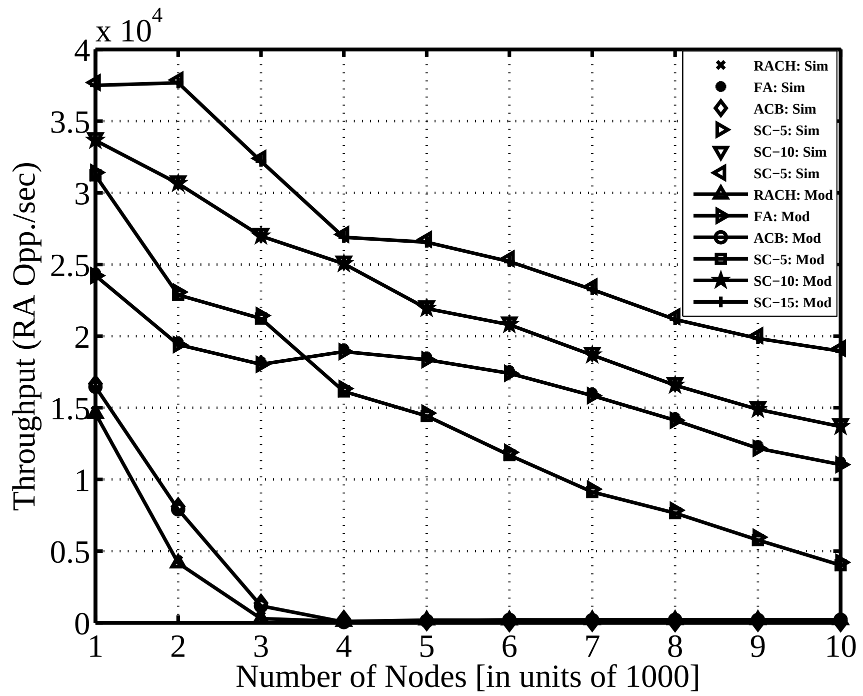

Figure 5.8: Throughput performance of representative RA schemes with variable number of small cells.

per second due to either an increase in resources attained by adding more small cells to the system (small cell enhancements) or by dynamically adjusting resources to suit existing demand. Adding more small cells has the effect of providing more slots for devices to perform the random access procedure as each cell would provide subframes for devices that may seek network

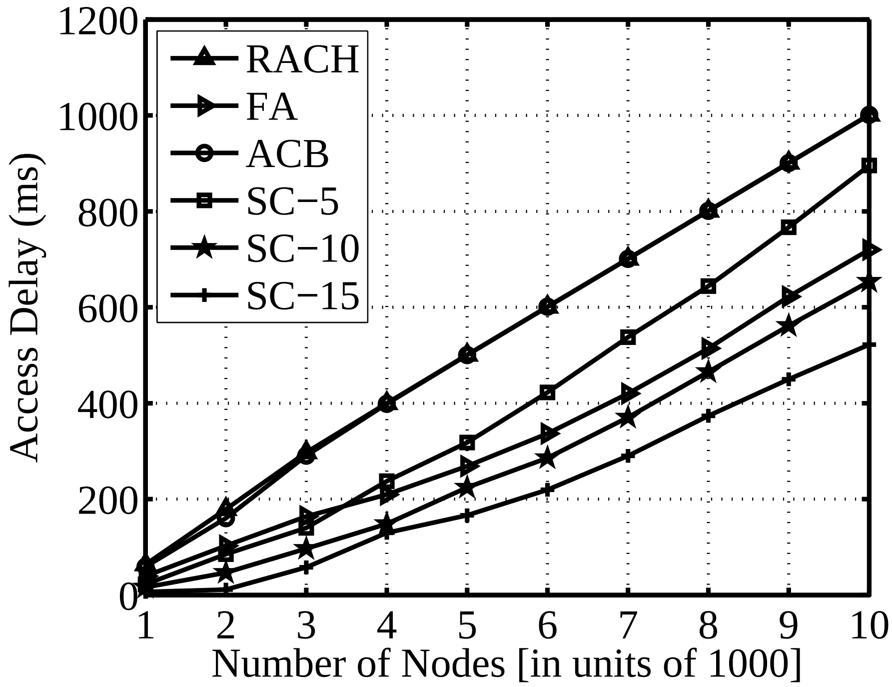

Figure 5.9: Access delay of representative RA schemes with variable number of small cells

access. The FA, on the other hand, provides more random access opportunities by dynamically adjusting the number of subframes used for the process. The effect of this can easily be seen in the figure, as the FA scheme even performs better than the $SC - 5$ proposal in most cases, proving it is better to adopt the FA scheme rather than add 5 small cells into the network considered. However, it is more beneficial to add more than 5 small cells to the network than to use the FA scheme. The decision on which of these to adopt is left to the discretion of the network operator.

Looking at performance in terms of access delay, Fig. 5.9, there is a marked improvement, with MTC devices being able gain network access much faster than they would were they to use either the default random access or the ACB schemes. In the case of small cell enhancements, the improvement in access times is dependent on the number of small cells used while FA schemes' method of dynamically adjusting the number of subframes also pays off, as it can be seen to outperform the $SC - 5$ scheme. Both proposals (FA ans small cell enhancements) performs better than both ACB, RACH offering devices faster connection times to the network.

In all metrics considered, both frame adaptation and small cell enhancement proposals offer considerable improvement when compared to legacy schemes such as the default RACH and ACB. However, as earlier stated, it is the decision of network operators, as to which of these schemes would be preferred, even as the FA would be more cost effective and still offer improved service when compared with small cell enhancements.

5.5 Conclusion

In conclusion, we have approached the challenge of congestion arising from massive device access in the RACH of LTE networks from two perspectives. The first proposed a frame adaptation scheme that is capable of handling massive machine device access in LTE networks. We define a utility function upon which an optimization problem is formulated and a solution is found to determine the optimal number of subframes to be allocated to the machine devices over RACH. Furthermore, we proposed a method to estimate the number of devices attempting to access the RACH. Simulation results have verified the effectiveness of our proposal over a fixed subframe allocation schemes such as the default RACH and in some cases over 100% increase in utility has been observed. Evaluation has also further confirmed the potentials of the proposed scheme in terms of collision probability, throughput and access delays. A potential issue for consideration in future is to investigate if there would be significant standards changes in implementing dynamic subframe optimization.

In the second part, we have proposed the use of small cells as an efficient way in combating the RA overload problem caused by massive machine device access in LTE cellular networks. By evaluating the effect of small cells on networks using simulations and analytical means, we have observed significant performance improvement. This proposal has also been compared with the frame adaptation scheme proposed earlier in this chapter. The addition of small cells to networks comes at a financial cost, the frame adaptation scheme comes at the cost of data frames which are being dynamically converted for the random access process. Also, we intend to study the inter small cell communications and intra cell (small and macro cell) coordination to provide efficient resource allocation, joint transmission and sufficient robustness for mobility and handover.

6 Rate Adaptation Algorithms in IEEE 802.11

In chapter 3 of this thesis, one of the major challenges of wireless networks, rate adaptation, was extensively discussed. The interest in vehicular networks lie in the fact that there are a huge number of applications that are presently being developed and many more will still be developed in the nearest future. There is also a lot of interest in this area both in academia and industry. Google's research and implementation of self driving cars is an obvious example.

In this chapter, we propose two algorithms to combat the challenge. The performances of existing and proposed algorithms are examined, both in, static and vehicular networks. A system level simulator for wireless network environments was developed and used in carrying out evaluations. This simulator takes into consideration, critical metrics in any wireless environment and is further discussed in the relevant section of this chapter.

The rest of this chapter is organised as follows. A context aware rate adaptation algorithm is presented in section 6.1, in section 6.2, we discuss performance evaluations of the proposed context aware rate adaptation algorithm. In section 6.3, we present a throughput based algorithm together with the ideal SNR based algorithm. Performance evaluation of the proposed throughput based and the ideal SNR algorithms are done in section 6.4 and finally, we conclude the chapter in section 6.5.

6.1 A Context Aware Rate Adaptation Algorithm for Vehicular Networks

Vehicular networks are very unpredictable, even as environmental conditions such as changes in weather, amongst others, affect the system performance. Node mobility at varying speed also introduces a measure of uncertainty. Since vehicular networks cover a range of distance, it then follows that nodes could spend a short amount of time within the range of access point. It is very difficult to determine accurately the prevailing channel conditions and then select an optimum transmission rate because the channel conditions can change drastically within very short interval of time. Thus, a rate adaptation algorithm for vehicular networks must be able to take care of the varying situations that can occur and respond effectively. Such an algorithm should:

- be able to predict the prevailing channel conditions and make rate transmission decisions as appropriate.
- be able to prevent sporadic rate changes due to spikes in channel conditions.

- be responsive to the dynamic environmental situation and make rate changes as appropriate.

In developing an algorithm for use in vehicular networks, the first step was to study existing ones used in static networks and then based on performance, make certain changes that would adapt the algorithm to their new environment. Based on this, a modified ONOE algorithm was developed.

6.1.1 The modified ONOE algorithm (mONOE)

In order to cope with the challenges above, contextual information is used to try and estimate the prevailing channel conditions. There are a number of metrics that can be used to provide an indication of the present conditions. However, in this study, the packet transmission statistics are used to determine the prevailing channel conditions.

mONOE is quite similar to the existing ONOE algorithm and after a careful study of the operating mode of ONOE, the following modifications were made to adapt it for use in vehicular networks:

- Reduction in the rate change determination period from 1 to 0.5. This is to guard against sporadic rate changes and create some form of stability.
- Reduction in the credits threshold from 10 to 3 with the aim of optimizing system throughput.

mONOE is a transmission statistics based algorithm that takes into account the number of packets transmitted, number of successfully transmitted packets, the number of retries per frame and issues credits based on the performance of the present transmission rate.

The current data rate is first evaluated at the expiration of a two second timer. At the start of the timer, the number of frames transmitted, N, the average number of retries, R, the number of successful packets, S, and the amount of credits are all initialized to zero.

Transmission starts at the highest possible rate, depending on the standard in which it is employed, for this evaluation, 54 Mbps is the highest data rate. If there is a successful packet transmission and the number of frames transmitted is greater than or equal to five with R greater than one, then the rate is decreased.

If not, if ten percent of the frames required retires, the credit threshold is decremented, else it is incremented. If the number of credits gathered is greater than or equal to three, the rate is increased and the variables are re-initialised.

A flowchart of the mONOE is presented in Fig. 6.1.

6.2 Performance Evaluation of the Context Aware algorithm

In this section, the performance of the context aware rate adaptation algorithm along with some other existing algorithms are presented.

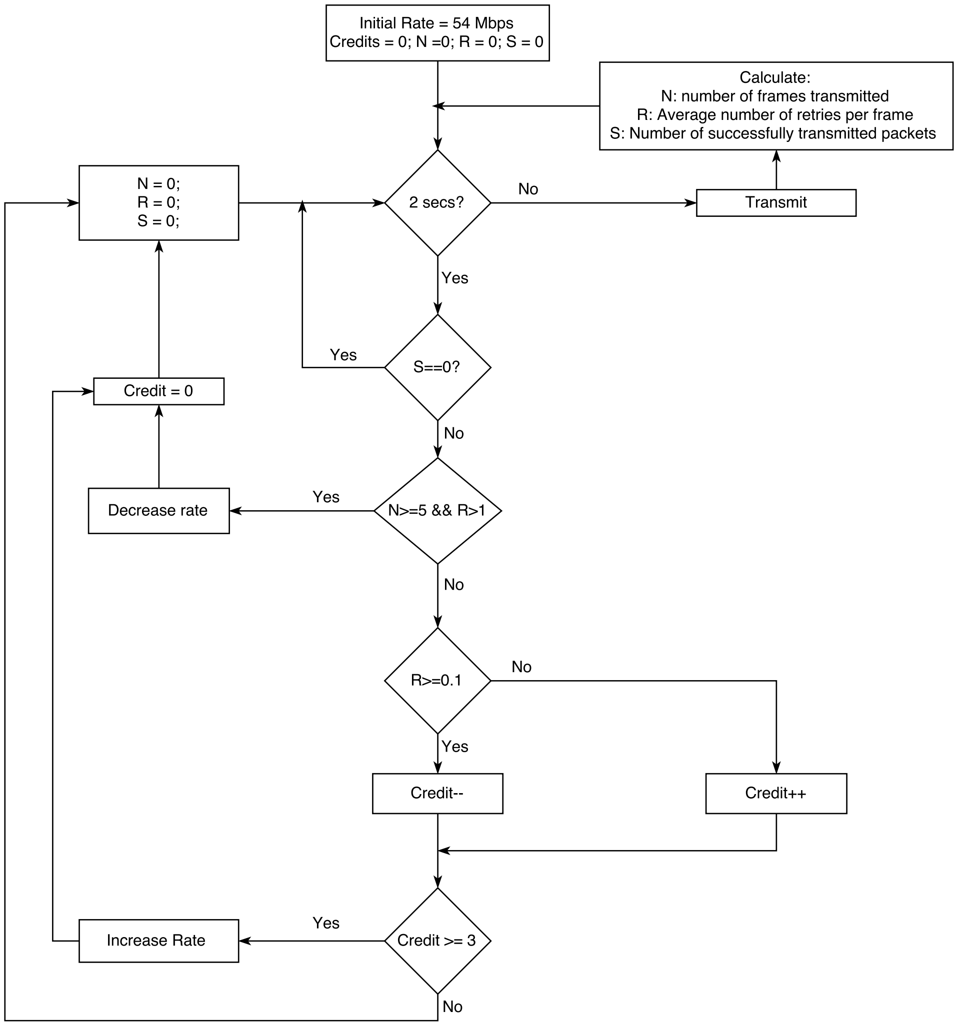

Figure 6.1: Flowchart of mONOE algorithm

6.2.1 Simulation design and considerations

The simulation scenario was a vehicle to infrastructure network consisting of a varying number of nodes and an access point to cover the a range of distance. The value of the distance covered was 100 meters. In the network, it is assumed that nodes pass through the coverage range at varying speed between $0 - 10m/s$ and transmission is only carried out when the nodes are within the range of the access point. Measurements made in simulations were based on the 802.11a standard. Both the 802.11a and 802.11p standards have the same modulation and coding schemes as well as training sequence [93]. Thus, the results generated by the simulations remain valid. Table 6.1 summarizes some of the major parameters used in this work.

A randomly generated value of the contention window and the truncated binary back-off procedure specified in the 802.11 standard are used. This means that the node with the least contention window size is given the preference to transmit in that contention round. As with any wireless system, the outcome of packet transmission could either be successful, collided or considered unacceptable due to errors, in which case the packet is discarded. The amount of data that a receiver can successfully get over a link is based on the number of bits it can decode

Table 6.1: Simulation Parameters

Parameter	Value
DIFS time in seconds	$34 * 10^{-6}$
SIFS time in seconds	$16 * 10^{-6}$
Slot time in seconds	$9 * 10^{-6}$
Header packet Length (bits)	464
ACK packet length (bit)	304
Contention window size (bits)	32
Average packet length (bits)	1500
Node Velocity(m/s)	0-10
Power distance gradient	2-3
Transmit Power	40mW

correctly. The Bit Error Rate (BER) is given in equation 6.1.

$$BER = B_{\text{err}}/B_{\text{T}} \tag{6.1}$$

where B_{err} is the number of bits received in error and B_{T} is total number of bits transmitted over the link.

Since frames are made up of bits, equation 6.3 can easily be extended to obtain the frame error rate (FER) in situations where the number of bits in error and the total number of bits sent during a transmission. A variety of forms of modulation can be used on each of the 802.11a subcarriers. BPSK, QPSK, 16-QAM, and 64 QAM can be used as channel conditions permit. For each set data rate there is a corresponding form of modulation that is used and based on this the condition of the frame received can be estimated. Simulations make use of an SNR-FER plot shown in Figure 6.2 which estimates channel errors as a function of modulation and coding schemes (MCS) which are shown in table 6.2.

.

At any point in time, the SNR value is a function of the received power P_r, transmit power, denoted by P_{tx} and the distance, denoted by d of the node from the access point. P_r is calculated using the path loss formula stated in (6.2) where alpha is the path loss gradient.

Table 6.2: MCS table

MCS index	Modulation type	Coding Rate	Data Rate (Mbits/s)
1	BPSK	1/2	6
2	BPSK	3/4	9
3	QPSK	1/2	12
4	QPSK	3/4	18
5	16-QAM	1/2	24
6	16-QAM	3/4	36
7	64-QAM	1/2	48
8	64-QAM	3/4	54

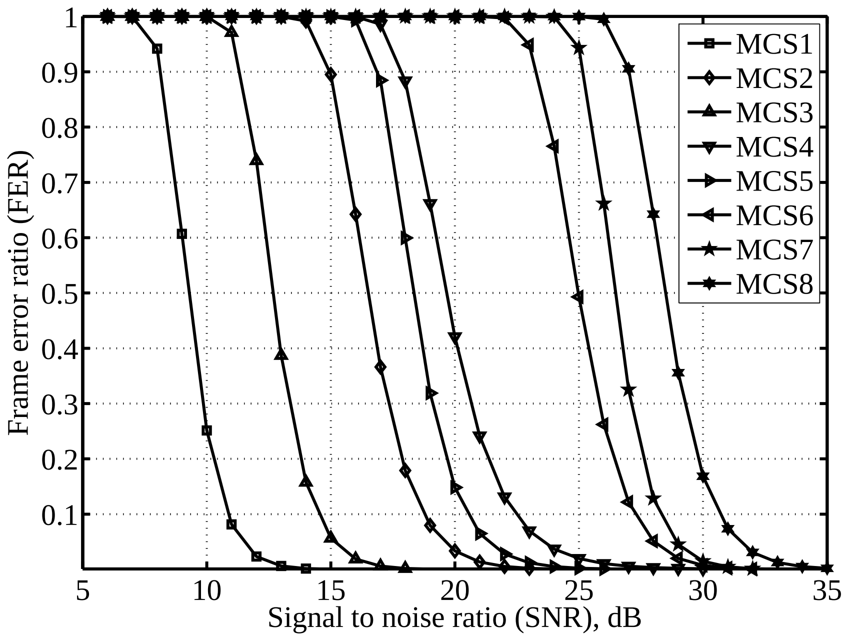

Figure 6.2: Frame Error Ratio versus SNR

$$P_{\text{r}}(dB) = 10log_{10}(P_{\text{tx}}/d^{\alpha}) \tag{6.2}$$

Factoring in the effect of shadowing and fading which are expected in vehicular environments, the log-normal distribution takes care of this. A mean of zero and standard deviation of 4 db is used to calculate the SNR (in db) as stated in (6.3), where X_a is the value due to shadowing.

$$S_{db} = P_{\text{r}} - P_{\text{noise}} + X_a \tag{6.3}$$

where S_{db} is the value of SNR in db and P_{noise} is power due to noise.

6.2.2 Results

Performance evaluation was carried out in two phases. The first involves evaluating the capability of AARF and ONOE in static environment, after which, the performance of both algorithms including the proposed modification in a vehicular environment is then undertaken.

Static Environment

The scenario depicted by a static environment is one in which network nodes are stationary. It is assumed that the nodes all transmit at line of sight to the access point. Situations in which there are obstructions to the transmission lines are left for future consideration due to the fact that a performance analysis of both algorithms considered here. In particular, the outperforming algorithm would be further modified to suit a vehicular network is the objective of this initial simulation.

The SNR gives an indication of the environmental condition in which a network operates and as such, the algorithms are benchmarked in varying SNR values. Figure 6.3 shows the performance of AARF and ONOE algorithms under an SNR value of 15 and 25 dB respectively.

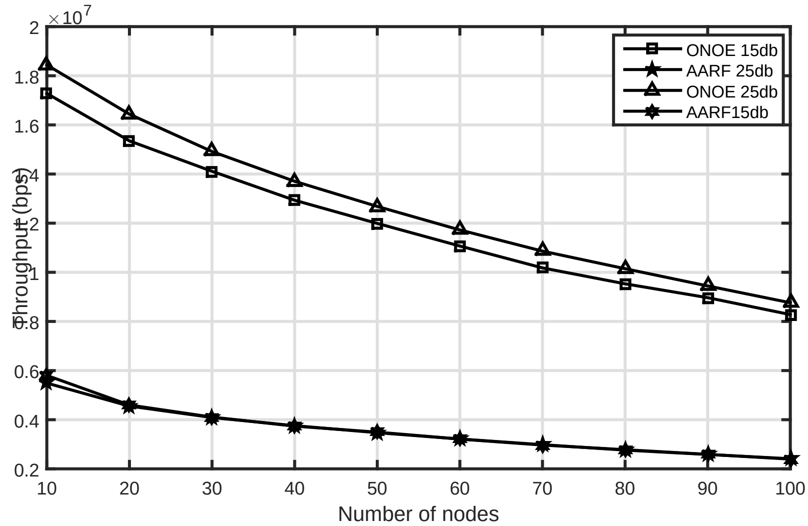

Figure 6.3: Performance of legacy algorithms in static network with SNR of 15db

As explained in previous chapters, the metrics used in making rate changing decisions play a crucial role in its performance. Both algorithms have some common metrics, however, the way they are measured and used is part of what makes the difference in performance. The number of successful transmission and period of observation is quite important for both rate adaptation algorithms.

The simulation results shows that, simply basing rate changing decisions on the number of successful transmissions, as in AARF, is not sufficient if the aim is to maximise throughput. Other performance indicators do need to be considered. For instance, in ONOE, the number of retries and failed transmissions are put into consideration. Also, an important element of any rate adaptation algorithm is its ability to guard against sporadic rate changes. This is usually implemented using timers or a collision period after which a data rate or system wide performance may be evaluated. AARF uses time-out whilst ONOE gathers statistics over a collation period before making decisions. The importance of this difference is that, whilst, a sort of time period is good, it is more important to evaluate occurrences within the network during that period, using various metrics, as this could help determine the network state. Rate decisions can then be made based on this state.

Notice that the maximum throughput obtained in Figure 6.3 is much higher for ONOE than AARF irrespective of the SNR value. It can be posited that a direct relationship exists between throughput and SNR, if all other factors remains constant. Furthermore, there is a decrease in throughput as network load (number of nodes) increases and this is expected as competition between nodes reduces the throughput. Consequently, in both SNR values considered, ONOE still does better than AARF.

Summarily, under the scenario considered, ONOE provides a higher throughput than AARF even under increasing network loads. The number of nodes, signal to noise ratio and the rate adaptation algorithm used will always affect the network throughput. A higher value of SNR could imply better channel condition and expectedly, the throughput at this values are higher compared to a lower SNR value.

Vehicular Environment

With ONOE outperforming AARF in static networks, modifications were made to the algorithm in order to improve its performance in vehicular networks. Nodes in this scenario are capable of moving at random speeds between 0 and 20 m/s within the access point coverage distance and this is only where communication is possible. The scenario is depicted in Figure 6.4.

Figure 6.4: Simulation scenario for vehicular network evaluation

With the knowledge of the speed at which nodes are travelling at any point in time, it is possible to estimate their position within the network at any time during the simulation.

$$d = s * t \tag{6.4}$$

where d is the node position, s is the speed at which the node is travelling and t is the simulation time at which the node position is being sought.

Using equation 6.4, it is easy to determine the node position relative to the access point. This is quite important because, a node closer to the AP is expected to have a higher probability of transmission success due to its proximity to the AP. This relative position can easily be used to calculate the SNR of a node using equations 6.2 and 6.3.

An initial evaluation of the algorithms (mONOE, ONOE and AARF) was done under varying path loss and a fixed coverage distance of 100m. The path loss is used to reflect the environmental condition of the network with a high value denoting harsh conditions. Results obtained are shown in Fig. 6.5 and 6.6.

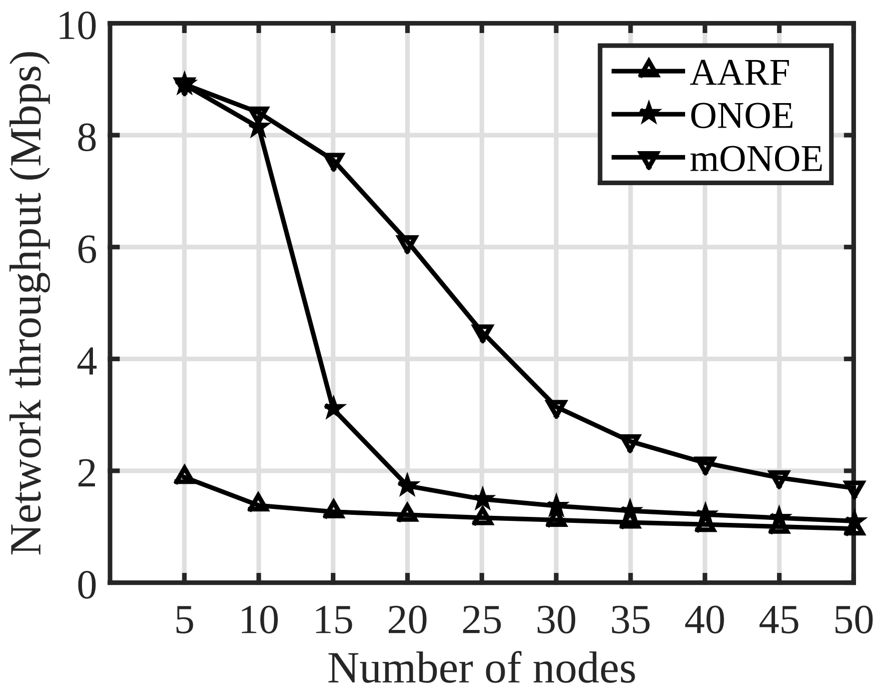

Figure 6.5: Performance at coverage distance of 100m and path loss exponent of 2

Expectedly, the throughput values of all the rate adaptation algorithms was higher at a lower value of path loss gradient. However, a general decrease in the value of throughput is observed as the number of nodes increases due to the contention mode in which IEEE 802.11 wireless networks operate. At a path loss value of 2, it is observed that the modified ONOE outperforms both AARF and ONOE, this is due to its modus operandi which has been explained in previous section of this chapter. Whilst quite similar performances was obtained between ONOE and mONOE under low network load, the impact of changes made to the ONOE algorithm becomes more visible as the network load is increased. Furthermore, in a relatively harsher environ-

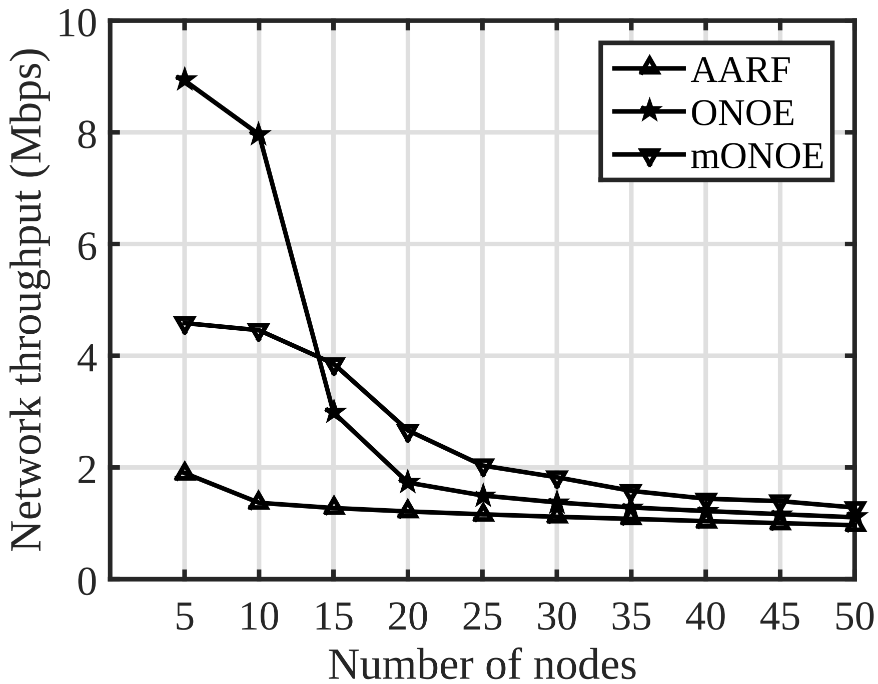

Figure 6.6: Performance at coverage distance of 100m and path loss exponent of 3

ment, depicted by a pathloss value of 3, we see that ONOE performs better than mONOE under relatively low network load, infact, up until when there were 10 nodes in the system, ONOE performed better but as the load increased, mONOE begins to outperform ONOE. In all these scenarios, the performance of AARF was still below par when compared to ONOE and mONOE. However, the performance of all three algorithms in the scenarios considered still leaves much to be desired especially considering the improvements that have been made with regards to metrics used in making rate changing decisions. This has led to further research and the proposal of a throughput based rate adaptation algorithm.

6.3 Throughput based Rate Adaptation Algorithm for Vehicle Networks

According to what has been studied in the previous section, it is obvious to see that there is still work to be done in the development of rate adaptation algorithms for vehicular networks. As state of the art in rate adaptation progresses, more algorithms are being developed with performances much better than the ones considered in the previous chapter. Using throughput as a metric, a number of algorithms have been proposed in literature, albeit, for different scenarios. For example, [94] proposed an algorithm for use in dense networks.

Whilst the nature of challenges facing vehicular networks have not changed, the methods used in trying to mitigate them has improved over the years. In this section, we consider an algorithm for use in vehicular network that uses the effective throughput as its metric for decision making. The contributions presented here are as follows:

- Design an efficient algorithm for vehicular networks based on state of the art algorithm designs.
- Design an ideal SNR based algorithm and compare performances with representative algorithms including the proposed one in order to determine how far away performances are from the ideal.

The identified performance gaps could help understand the spaces that future rate adaptation algorithm may have in terms of performance improvement. Also, future research can use the ideal algorithm to evaluate the performance of their designs especially in vehicular networks.

6.3.1 The Throughput based rate adaptation algorithm

A rate adaptation algorithm based on throughput would have the advantage of estimating the effect of channel conditions on the successful delivery of packets to its destination. The proposed algorithm guards against making sporadic transmission rate decisions by sampling at certain intervals. The process of sampling is explained later in this section. A flowchart of the proposed algorithm is presented in Figure 6.7. As can be seen from the flowchart, at the beginning of packet transmission, data is sent at the highest possible bit rate. For the 802.11p standards, this rate is 27 $Mbps$. During this transmission, if three successive failures occur ($N_f < 3$) and no packets are acknowledged ($N_a > 0$), a temporary rate index (R_{temp}), which is the highest bit rate that has not had three consecutive failures, will be calculated. If this condition is not fulfilled by any bit rate, transmission is done at the lowest bit rate.

Packets are also periodically sent at rates other than the current transmission rate in order to determine their average throughput; this is referred to as sampling. To guard against sporadic changes and to limit the amount of time spent on sampling as against actual effective data transmission, sampling is undertaken for 5% of the average transmission time, T_{avg}, of the current rate.

Thus, when sampling is carried out, the throughput, G_s, obtained at the sampled rate (R_s) is compared with the throughput (G_{curr}) obtained at the current rate, R_{curr}. The rate that provides the higher throughput is chosen to be the current transmission rate.

Periodic sampling is implemented using two functions, $apply\_rate()$ and $proc\_feedback$, similar to [69]. The $apply\_rate()$ function assigns a bit rate to a packet while the $proc\_feedback()$ processes outcome statistics of transmission including number of packets acknowledged and the number of retries. The $apply\_rate()$ function does the following:

- If no acknowledgement, transmit at highest bit rate that has not had three consecutive failures.
- Increment the number of packets transmitted.
- If sampling time (T_s) is less than average transmission time (T_{avg}), select a random bit rate that has not had three consecutive failures and have higher throughput than the current transmission rate.

- Else, send packet at bit rate with highest throughput.

The $proc\_feedback()$ function recalculates throughput for a particular bit rate and updates information that tracks the number of samples. It also carries out the following operations

- Throughput calculation for a packet based on the data rate and number of retries using equation 6.6.
- In case of a successful packet transmission, update the number of successfully transmitted packets accordingly.
- In case of failure of packet transmission, update the number of successively failed packet transmission attempts, else, reset it.
- Recalculate throughput for the bit rate.
- Set current bit rate to the one with highest throughput.

The determination of throughput takes into consideration, the number of retries,length of frame being transmitted, and the number of frames successfully transmitted by the rate, as well as protocol overhead. The time taken to transmit a unicast packet, (T_u), for a number of retries ,N_r, and a transmission rate, R, is calculated using equation 5.1.

$$T_u = T_{difs} + T_{slot} \sum_{i=0}^{N_r} \frac{w_i}{2} + (N_r + 1)(T_{sifs} + T_{ack} + T_{head} + \frac{8 * L_{frm}}{R}) \qquad (6.5)$$

The throughput,G_r, is obtained using the formula below

$$G_r = \frac{L_{frm} * 8}{(T_u)} \qquad (6.6)$$

where T_{difs} is differential Interframe spacing, T_{sifs} is short interframe spacing, T_{ack} is the acknowledgement duration, T_{head} is header duration, L_{frm} is number of bytes in the data frame ,and G_r is throughput. w_i is the back off window for the ith retry. The values of T_{difs}, T_{sifs}, T_{ack} and T_{head} are based on the 802.11 standards.

Spontaneous variations in channel conditions could lead to sporadic rate changes, however, to address this, the throughput based rate adaptation algorithm uses the mechanism described above to adapt to channel conditions.

6.3.2 An Ideal rate determination algorithm

The idea behind an ideal rate determination algorithm is that it would be useful in demonstrating how far off research studies are to the maximum throughput that can be obtained in the wireless system being modelled. The thought of an ideal scenario is quite interesting, although, in practice, no ideal scenario exists. However, with defined values of data rate and SNR,

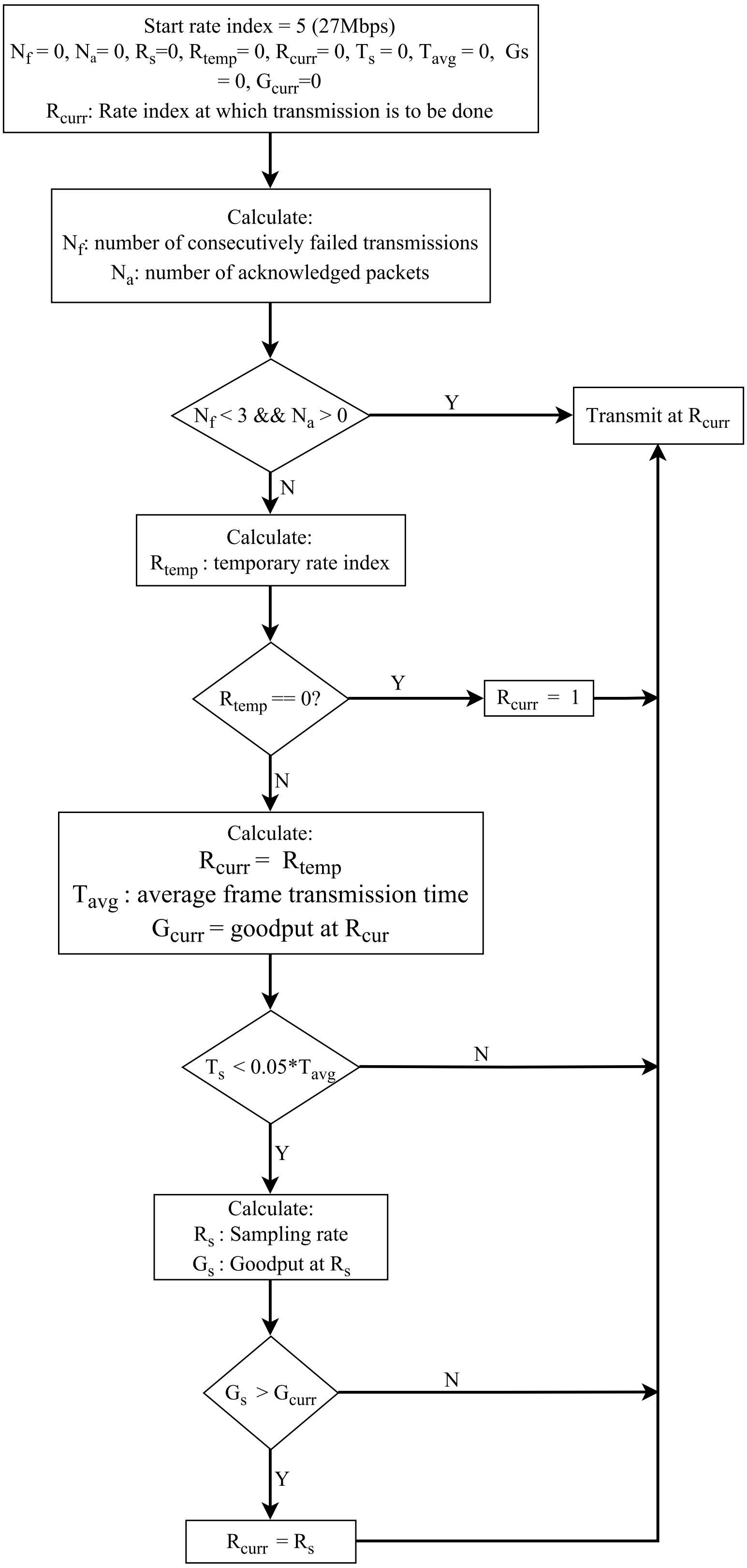

Figure 6.7: Flowchart of proposed throughput based algorithm

it is possible to envisage what the throughput of a system would be. Therefore, an ideal rate determination algorithm based on these metrics is developed.

By using the SNR-FER plot shown in Figure 6.2, it is possible to estimate the Frame Error Ratio (FER) of a particular SNR value. The goodput at any transmission rate, R_i, can then be easily calculated.

Let S_i denote the SNR value at a particular interval i. At S_i, the goodput of transmitting at R_i can be calculated using equation 6.7.

$$G_i = R_i \times (1 - P_i) \tag{6.7}$$

where G_i is goodput obtainable at a transmission rate, R_i, and P_i is the frame error rate at S_i. Thus, at any particular value of SNR, it is possible to obtain P_i, from which we can easily get the value of good put using equation 6.7 above. Both R_i and G_i are measured in $Mbps$. A look-up table consisting of SNR values, the transmission rates and the associated goodput can then be drawn up. For any SNR, S_j, in the range of $[S_j, S_{j+1}]$, there is always an associated optimum goodput, $G_{tx}(s_j)$. Thus, with n number of available rates, we can say,

$$G_{tx}(S_j) = max(G_{j,1}, G_{j,2}, ...G_{j,n}) \tag{6.8}$$

where $G_{tx}(S_j)$ is maximum goodput at (S_j) .
In an ideal situation, transmissions at (S_j) should be done using the data rate that provides the optimum goodput. For verification purposes, SNR values between -40 and $50db$ with steps of $5db$ was used to obtain a goodput-snr curve as shown in Figure 6.8. The steps and range of SNR values can be changed to produce a curve based on desired values, however, for simulations, we have used the values stated earlier. Evidently, from the curve, the transmission rate that can guarantee maximum throughput can easily be obtained.

As noted earlier, achieving an ideal scenario in real world is very difficult, if not impossible. Rapid variation in SNR values is still a possibility in today's vehicular networks, and also ideal situation in which the effects of hidden nodes and idle stations on overall system throughput is not obtainable in the real world. However, the effect of hidden or idle stations is not considered in our simulation. This is left for future studies.

6.4 Performance Evaluation of the Throughput based algorithm

In this section, the performance of the throughput based algorithm along with some other existing algorithms are presented.

6.4.1 Simulation design and considerations

The scenario depicted by Figure 6.4 was used to evaluate the performance of the various rate adaptation algorithms considered. It consisted of an infrastructure vehicular wireless network with a maximum of 50 nodes. The access point/Road Side Unit (RSU) covers the entire stretch of its coverage distance and is stationed in the middle covering equal distances on both sides

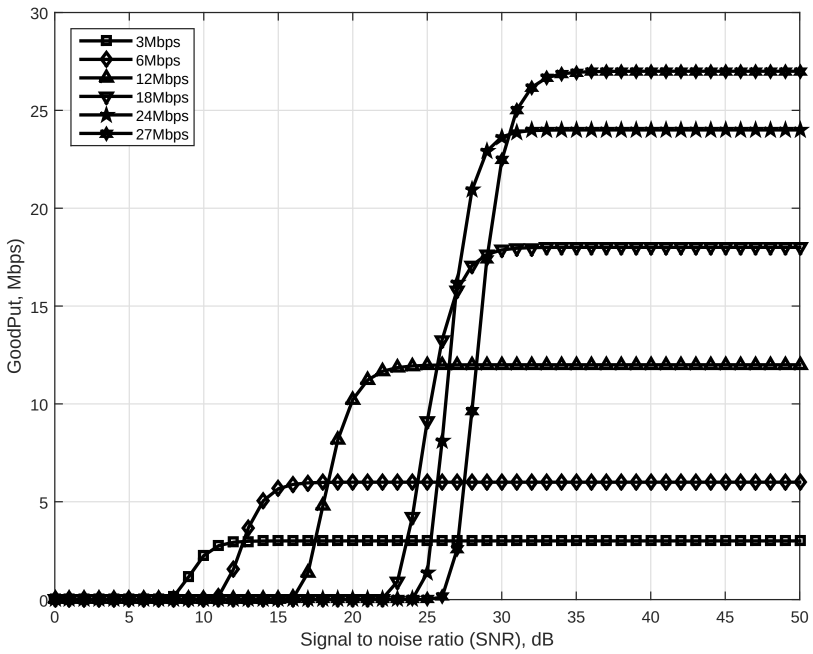

Figure 6.8: Goodput versus Transmission Rate for IEEE 802.11p

of its coverage range. The nodes move with variable speed between $0 - 20m/s$ and again communication with the RSU is only possible when nodes are within its coverage range. The data rate and other parameter values are based on the IEEE 802.11p standards. Table 6.1 is still valid in carrying out the evaluations because the same environment, as the one previously described, is being simulated.

The log distance path loss model is used to estimate the relationship between received and transmitted power with respect to distance. The value of SNR is dependent on the estimated received power and the noise level. If transmitted power, P_{tx}, travels a distance d in meters, then the received power P_r could be estimated using the equation 6.2. Furthermore, the SNR is determined by the estimating the received power signal in relation to the noise power used as stated in equation 6.3.

Variation in the value of path loss could help simulate performances in different scenarios. For example,a path loss value of 2 depicts a propagation model that could be considered as the free space path loss model, whilst a value of 4 portends the two ray ground reflection model. A path loss value of 3 describes a propagation environment that could be expected in an urban area. In order to simulate a scenario that is as close to reality as possible, the effects of shadowing and fading are also considered. This is modelled in simulations by using the log-normal distribution which considers the effect shadowing has on a large number of measurement locations that have the same transmitter-receiver separation but varying levels of clutter in the propagation path [95]. Thus, the value of SNR is calculated to include a mean distant dependent value, X_a, with zero mean and standard deviation of four (in decibel) as described in equation 6.3.

A packet is deemed to contain errors if at least one bit in the packet is an error. Thus, mathematically, the packet error probability for a packet containing M bits by the equation below

$$P_{\mathrm{p}} = 1 - (1 - P_{\mathrm{e}})^M \tag{6.9}$$

where P_p is the packet error probability, P_e the bit error rate and M the number of bits in the packet.

6.4.2 Performance results

With three existing algorithms namely AARF, ONOE , Samplerate, the proposed throughput based algorithm was evaluated in terms of the throughput it can provide in various scenarios. Also, the performance of the ideal SNR based algorithm was included to help visualise how far away from ideal, results are. At any particular number of nodes, the throughput obtained is the average of individual results gotten during fifty rounds of simulations. This helps increase the reliability of our results.

Path loss = 2

As earlier described, it is expected that this scenario would be the most friendly, under the scenarios considered during simulation, in terms of channel conditions. A path loss gradient of 2 and an antenna coverage distance of $100m$ is used. The performance is presented in Figure 6.9.

The performance of any algorithm is hugely dependent on the way it works. Considering the three existing algorithms, it is easy to see that AARF is the worst performer of them all, whereas, ONOE performs best under considerably lower loads (<10 nodes) but is quickly overtaken by samplerate as the number of nodes increases (>10 nodes). Again, the performance of AARF restates the fact that it cannot be used effectively in vehicular networks. The complex nature of this network when compared to the static implies that the number of successfully transmitted packet cannot be used as the sole determinant of performance. With ONOE taking its time to study the channel conditions before making rate changing decisions, there is a form of stability that is obtained even in dynamic channel conditions, as the data rate is still maintained for a particular period before determining whether to change it or not based on performance. However, the obvious disadvantage of this is that, there is a potential for delay in changing rates especially when there is a change in channel conditions. Rather than respond swiftly, ONOE has to wait a maximum of its rate determination period. This conservative approach would not be ideal in maximising efficiency considering that at any prevailing channel condition, there is always an optimum transmission rate which may not be the data rate in use. Samplerate's rate changing metrics, such as ,average frame transmission time, including retransmissions, and an evaluation of the performance of other data rates at certain periods, gives it a measure of reliability and makes it efficient in allocating the best possible transmission rate at the existing channel conditions. The algorithm performs better than the ONOE and AARF at higher

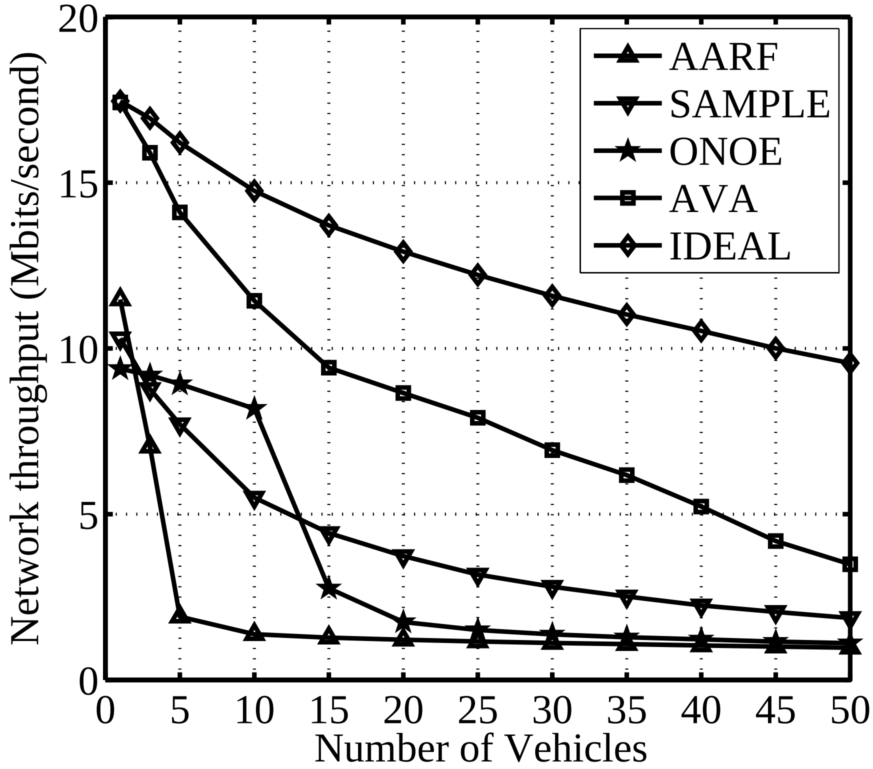

Figure 6.9: Performance at coverage distance of 100 m and path loss 2

nodes. Overall, the proposed throughput based algorithm performs better than all other three algorithms whose performances we have just discussed. Its throughput (represented by AVA in Figure 6.9) is higher than that of ONOE, AARF and samplerate (SAMPLE). The decision to measure channel conditions using throughput and not clog the network by performing sampling for a specific period seems to pay off in terms of increasing throughput. Changing channel conditions are quickly responded to and an efficient data rate is selected.

Path loss = 3

In a more challenging channel condition, as implied by equating the value of path loss to 3, the system performance is once again studied. At first glance, the effect of this change is seen as the maximum value of throughput obtainable is reduced when compared with Figure 6.9. In this scenario, the overall network performance is shown in Figure 6.10. The value of path loss affects the successful delivery of packets and in turn the throughput. Delving into specific algorithm performances, we notice that the trend observed here is very similar to those obtained when the path loss value was 2, albeit, with a drop in the value of throughput.

Comparing throughput values at maximum number of network nodes(50) considered in our simulation, with results as shown in Figures 6.9 and 6.10, there is a reduction in throughput of about 45% specifically when considering the performance of the proposed throughput based algorithm. This drop is also witnessed in the ideal SNR based algorithm.

This expected fall in throughput is attributed to the harsh channel conditions in which the

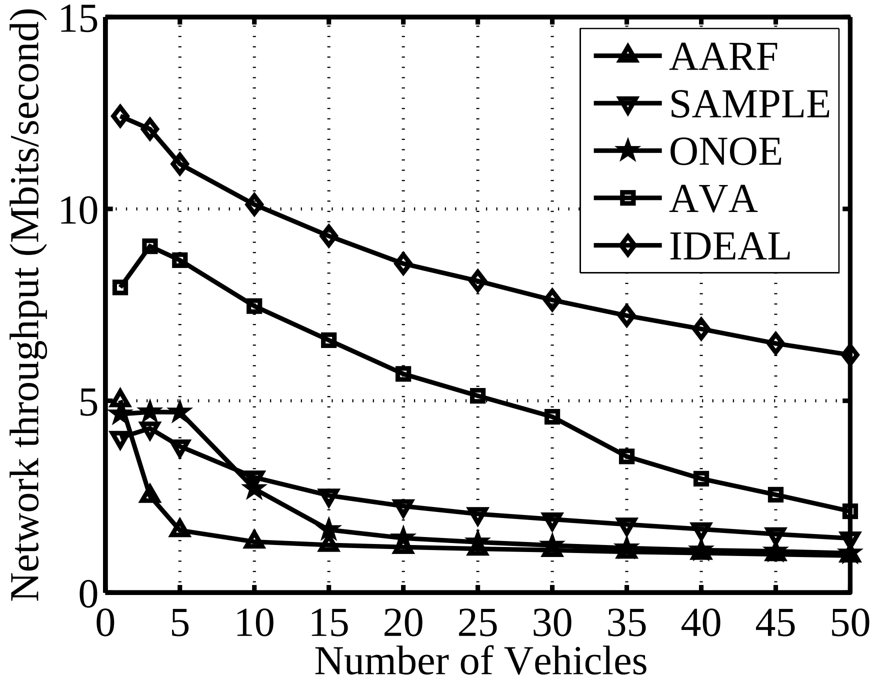

Figure 6.10: Performance at coverage distance of 100 m and path loss 3

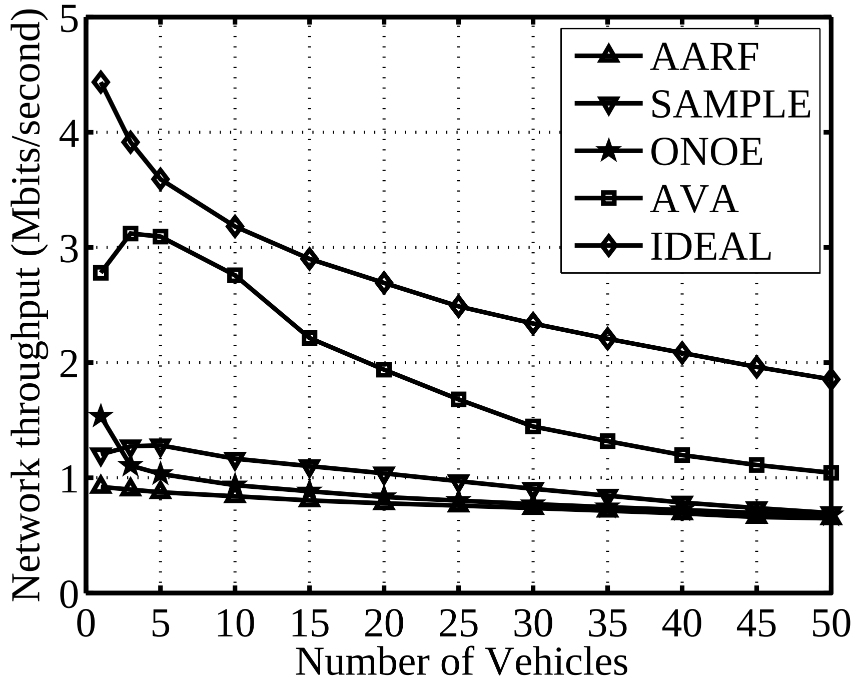

Figure 6.11: Performance at coverage distance of 100 m using the two ray ground reflection model

system performance was tested. Although, throughput reduction was witnessed across board, there are still performance gaps between the ideal algorithm and the others in which it has been compared with.

Further tests of the algorithm was undertaken using the two ray ground reflection channel model. Results obtained is shown in Figure 6.11. Consistent with previous trends, there is a much more reduction in throughput. However, the proposed algorithm still outperforms all others with which it was benchmarked.

6.5 Conclusion

In an attempt to tackle the rate adaptation challenge faced in vehicular networks, two algorithms have been proposed in this chapter. The first was an adaptation of ONOE whilst the other was a based on throughput. Also, in order to visualise how far away performances are from the ideal scenario, another algorithm based on SNR was proposed. By starting out with the well studied static environment, we are able to establish a base from which performances of algorithms could be studied before delving into the dynamic nature which vehicular networks are known for. The problem with this kind of dynamic network is that channel conditions can change swiftly leaving existing algorithms with the decision on how best to respond to this changes. As a result, the impact of path loss, which is an indication of channel condition, was extensively studies.

Using simulations, the performances of three well known algorithms (AARF, ONOE and Samplerate) including proposed algorithms (mOnoe, Throughput based and Ideal SNR based algorithm) was studied in vehicular networks under various channel configurations. In practice, implementation of the proposed algorithms would be relatively easy as all that would be required will be the updating of device drivers in wireless network interface controllers (NIC). Also, proposed algorithms would be power efficient since huge and complex computations are not required.

In practise, the use of these algorithms could benefit a lot of industries especially the transportation industry. Application of these algorithms would enhance the provision of services such as autonomous driving, maps downloading and other applications, where a effective QoS is a requirement, through wireless networks.

7 Conclusion and Future Work

There is no doubt that the IoT would be one of the key shapers of future technological innovations. Therefore, in developing applications for the IoT, the communication medium between devices will play a crucial role. Whatever the means devices choose to communicate with, it must be able to accommodate the exponential increase in system load without sacrificing its QoS. Whilst this is a tough conundrum, hybrid wireless technologies is seen as a viable solution to this challenge. In particular, the LWA could be of utmost benefits in alleviating potential problems that may arise. However, the individual technologies that make up the LWA could suffer severe performance issues. On the LTE side, there is a challenge of accommodating massive device access to the network and on the Wi-Fi side, rate adaptation is still an issue of concern.

This thesis presents a systemic study of the performance issues of LTE and IEEE 802.11 networks with regards to supporting IoT applications. For LTE, majority of the devices would transmit small data, thus, a critical challenge is how to deal with network access for these massive amount of devices. As these data could be aggregated and passed on through the Wifi network, an interesting area of interest is how it (Wifi network) would adapt its rate to provide optimum transmission of data. We view this within the lenses of a vehicular adhoc network, which is also an application area of the IoT.

It is observed that the present RACH of LTE would be unable to cope with the exponential increase in access traffic and that existing rate adaptation algorithms would not be able to provide optimum throughput for devices transmitting data. Therefore, novel ways must be sort to address this issues. The major research activities and outcomes are summarised in this chapter and a proposal of future work is also presented.

7.1 Conclusion and summary

In this section, a brief summary of the research work carried out within this thesis and the main achievements are summarised. In chapter 1, we introduce the IoT, a new technology that is taking the world by storm. We start by discussing application areas of this technology which could be found in agriculture, utilities, transportation and logistics, smart homes amongst others. We also consider candidate technologies which could be used to drive communications within the IoT. Some of these, such as, Zigbee, Bluetooth and Wi-Fi are better suited for short range IoT devices while cellular technologies have an obvious advantage of ubiquity and are most commonly used for long range IoT communication implementations.

In chapter 2, we discuss IoT technologies in 3GPP networks. Firstly, we consider the state of the art in 3GPP cellular networks and give a detailed insight into its most recent implementation, the LTE. We discuss the system architecture of this technology which included a description of its functional nodes such as E-UTRAN, MME, S-GW and HSS. Interfaces, protocols and channels that are operational in LTE are also discussed. We then delve in to the network architecture for IoT communications as envisaged by the 3GPP. With each new release, the 3GPP has always made some provision for IoT communications, for example, in release 12 and 13, enhancements were being considered to make the network capable of being used both for human communications and IoT as well. Also, we discuss LAA which is being developed to aid coexistence and cooperation between Wi-Fi and LTE technologies. Next we considered the evolving 3GPP standards for IoT which are under consideration and development by the member organisations of 3GPP. These three are eMTC, NB-IoT and EC-GSM-IoT. A comparison of these technologies including their physical layer specifications, deployment, bandwidth and duplexing modes are also presented.

However, in order to implement this technology within the 3GPP, a number of challenges were identified including device addressing and identification, device monitoring, power consumption and overload and congestion control. Whatever technology is to be used to implement the IoT, it must be able to provide solutions to these challenges. For example, the sheer number of devices that would seek network access would lead to potential overload in initial device access and signalling, thus, ways must be sought to effectively curb this challenge. We then discuss the random access procedure used in LTE as this would be affected directly as a consequence of massive device access and also review proposals that have been presented in literature to combat the massive device access challenge.

In Chapter 3, communication technology in IoT from the viewpoint of the IEEE 802.11 was discussed. We start with an in-depth insight into the system architecture of IEEE 802.11 technology and delve into analysing the physical architectures encompassing the topologies such as the BSS and ESS and review the logical architecture where we briefly mentioned the OSI model and the functions of each layer. We then talk about the PHY and MAC layers which are crucial to the operation of the IEEE 802.11. In the MAC layer, a contention based access scheme called CSMA/CA is used during transmission and also that there is a specialised protocol used for non-contention based access, this is called the (RTS/CTS).
Next we discuss the frame format and explore the functions of the MAC layer which included collision avoidance and interframe spacing amongst others. We complete our expose on the MAC layer by examining the present challenges this layer faces including time varying channel and bursty channel errors. The physical layer is then explored and our examination of this layer included salient points such as the modulation and coding schemes used in IEEE 802.11 networks, a discussion on channel errors as well as popular 802.11 standards.

IEEE 802.11ah which was specifically designed for IoT and three use cases were adopted during its development. Areas in which IEEE 802.11ah is expected to be used include in sensor networks, as a backhaul aggregator and a range extender for cellular traffic offloading. We then mention the key challenges faced by IEEE 802.11 standards in IoT including the likes of physical layer design consideration, path loss models and rate adaptation. We then expand more on the rate adaptation challenge, examining and critiquing solutions that have been proposed in literature.

In an attempt to tackle RACH congestion in LTE, chapter 4 presented a root index allocation scheme. The proposal can also be considered as a preamble reuse scheme where preambles are reused in cells that are at some distance to one another. This safe distance was derived from link budget calculations, considering factors such as interference and the 3GPP preamble formats and can be implemented in cells of various layouts and sizes. In order to maintain the required detection probabilities, the effects of intra- and inter-preamble interference from neighbouring cells was also examined. Using throughput, collision probability and access delay as metrics, simulations and analytical modelling was used to evaluate the performance and demoinstrate the effectiveness of the proposed scheme. In deriving an algorithm for preamble allocation, the challenge is viewed as a classical graph colouring one in which the nodes are the cells and the sequences are the colours. From this, we developed a robust sequence allocation algorithm that can be used in practical network deployments.

Further attempts at proposing solutions to the RACH congestion challenge yielded two other proposals. Chapter 5 presented two schemes aimed at combating the issue. The first looked at the RACH problem from the view point of resource allocation and proposed a subframe resource optimization scheme as a solution. Due to the dynamic nature of the network load, an adaptive resource allocation scheme is proposed in which subframes are considered resources and an optimized amount of subframe required for a given or predicted system load was derived. The network load is estimated according to observations of historical RACH outcome at the eNodeB end and this makes it possible to be able to determine the number of subframes to be used for the next round of random access. The performance of the adaptive scheme was evaluated in terms of utility, throughput, collision probability and access delay and it showed appreciable increase against other non-adaptive schemes.

The second proposal investigated the use of small cells to combat the overload challenge. Existing networks are now moving towards being heterogenous in nature and there is a lot of potential in its ability as a potential solution to the challenge. Using simulations and numerical analysis, the extent of possibility of this happening was presented. Other issues that may arise in implementing this proposal in practical network deployments, such as, root sequence indexing was also discussed. Results show this is a promising proposal but also comes at the cost of installation and management of small cells being deployed.

Chapter 6 investigates a budding application of the IoT, the vehicular networks. In particular, the rate adaptation challenge in these networks were considered and two attempts at solving them presented. The first was a context-aware rate adaptation algorithm that was simply a modification of an existing algorithm used in static networks. The performance was evaluated with respect to the system throughput. The second delved further into solving this problem, albeit from the angle of using contextual information, such as, the rate of successful packet delivery amongst others. Thus, a throughput based rate adaptation algorithm was proposed. Also, an ideal SNR based algorithm was presented to show how far off we are from achieving an ideal system performance. Extensive simulations in vehicular environments show that our proposed throughput based algorithm outperforms existing rate adaptation algorithms under various system and channel configurations.

7.2 Future Work

In this thesis the main problems that were studied was on challenges of implementing wireless and cellular communication media in IoT, especially the overload and congestion in LTE and rate adaptation in IEEE 802.11 networks. Extensive Simulations and numerical analysis have been carried out to understand the problem and several proposals have been presented to mitigate the adverse consequences of the challenge.

However, in future, there are several problems deserving further research and they are discussed as follows.

- The analytical and simulation tools have been designed to model the first phase of network access. The results obtained can be used to get insights into the potential issues that may exist at the first mile of network access. However, the analytical and simulation tools can be further extended to the other parts of the network, modelling an end-to-end scenario. This will help provide a complete understanding of the potential effects of this problem within the entire network.
- Future works could also extend the simulation and analytical tools to model coexistence of Wi-Fi and cellular technologies in order to ascertain the state of the art with regards to combating existent and potential challenges that plague communications within the IoT. Performance evaluation could be done to evaluate end-to-end implementations of this coexistence.
- In future works, it will be interesting to extend the simulation designs to model the effect of potential signalling congestions within the LTE network. Also, evaluating the performance of proposed algorithms with other 3GPP standards specially designed for the IoT namely, NB-IoT, EC-GSM-IoT and eMTC would be quite fascinating especially as practical IoT applications are increasingly being developed.
- In the development of rate adaptation algorithms a limited number of parameters were taken into consideration in order to ascertain the state of the network. These parameters can be increased to effectively estimate the channel conditions more accurately. It is believed that with more parameters and scenarios, there can be an improvement in the design of algorithms that address rate adaptation in IEEE 802.11 wireless networks.
- Whilst two rate adaptation algorithms were designed in this thesis to combat the rate adaptation challenge in vehicular infrastructure networks, in future works it will be quite

intriguing to explore alternative approaches to investigate and address the rate adaptation challenge and develop protocols specifically for the IEEE 802.11ah standards targeting other low data rate IoT applications.

Bibliography

[1] Ji Chun Zhao, Jun Feng Zhang, Yu Feng, and Jian Xin Guo. The study and application of the IOT technology in agriculture. *Proceedings - 2010 3rd IEEE International Conference on Computer Science and Information Technology, ICCSIT 2010*, 2:462–465, 2010.

[2] Yi Liu, Shengli Xie, Yan Zhang, Rong Yu, Maziar Nekovee, and Stein Gjessing. Cognitive Machine-to-Machine Communications: Visions and Potentials for the Smart Grid. *IEEE Network*, 26(June):6–13, 2012.

[3] Jayavardhana Gubbi, Rajkumar Buyya, Slaven Marusic, and Marimuthu Palaniswami. Internet of Things (IoT): A vision, architectural elements, and future directions. *Future Generation Computer Systems*, 29(7):1645–1660, 2013.

[4] Ru Xue and Huan Sheng Song. Study of ubiquitous transportation environment based on IOT. *2011 International Conference on Computer Science and Service System, CSSS 2011 - Proceedings*, pages 2603–2605, 2011.

[5] Christian Prasse, Andreas Nettstraeter, and Michael Ten Hompel. How IoT will change the design and operation of logistics systems. *2014 International Conference on the Internet of Things, IOT 2014*, pages 55–60, 2014.

[6] Chengen Wang, Zhuming Bi, and Li Da Xu. IoT and cloud computing in automation of assembly modeling systems. *IEEE Transactions on Industrial Informatics*, 10(2):1426–1434, 2014.

[7] Zhuming Bi, Li Da Xu, and Chengen Wang. Internet of things for enterprise systems of modern manufacturing. *IEEE Transactions on Industrial Informatics*, 10(2):1537–1546, 2014.

[8] a Zanella, N. Bui, A Castellani, L. Vangelista, and M. Zorzi. Internet of Things for Smart Cities. *IEEE Internet of Things Journal*, 1(1):22–32, 2014.

[9] Jose A Galache, Juan R Santana, Verónica Gutiérrez, Luis Sánchez, Pablo Sotres, and Luis Muñoz. Towards Experimentation-Service duality within a Smart City scenario. In *2012 9th Annual Conference on Wireless On-Demand Network Systems and Services (WONS) Towards*, pages 175–181, 2012.

[10] The Mksmart Project.

[11] Marica Amadeo, Claudia Campolo, Antonio Iera, and Antonella Molinaro. Information Centric Networking in IoT scenarios: The case of a smart home. *IEEE International Conference on Communications*, 2015-Septe:648–653, 2015.

[12] Luca Catarinucci, Danilo De Donno, Luca Mainetti, Luca Palano, Luigi Patrono, Maria Laura Stefanizzi, and Luciano Tarricone. An IoT-Aware Architecture for Smart Healthcare Systems. *IEEE Internet of Things Journal*, 2(6):515–526, 2015.

[13] SIG Bluetooth. Inc.(2008). *Bluetooth–Technical Specifications*, 2008.

[14] Kuor-Hsin Chang. Bluetooth: a viable solution for iot?[industry perspectives]. *IEEE Wireless Communications*, 21(6):6–7, 2014.

[15] ZigBee Alliance et al. Zigbee specification, 2006.

[16] Ericsson AB. Cellular networks for Massive IoT âĂŞ enabling low power wide area applications. Technical Report January, 2016.

[17] Min Chen. Machine-to-Machine Communications: Architectures, Standards and Applications. In *KSII Transactions on Internet and Information Systems*, volume 6, pages 480–497, 2012.

[18] B Jorn a B Jerke. Lte-Advanced and the evolution of Lte deployments. *Ieee Wireless Communications*, 18(October):4–5, 2011.

[19] Stefania Sesia, Issam Toufik, and Matthew Baker. *LTE - The UMTS Long Term Evolution: From Theory to Practice.* 2011.

[20] Puneet Jain, Peter Hedman, and Haris Zisimopoulos. Machine Type Communications in 3GPP Systems. *IEEE Communications Magazine*, (November):28–35, 2012.

[21] Tarik Taleb and Andreas Kunz. Machine Type Communications in 3GPP Networks : Potential , Challenges , and Solutions. *IEEE Communications Magazine*, (March):178–184, 2012.

[22] Javier Gozalvez. New 3GPP Standard for IoT [Mobile Radio]. *IEEE Vehicular Technology Magazine*, 11(1):14–20, 2016.

[23] Hamidreza Shariatmadari, Rapeepat Ratasuk, Sassan Iraji, Andrés Laya, Tarik Taleb, Riku Jäntti, and Amitava Ghosh. Machine-type communications: current status and future perspectives toward 5g systems. *IEEE Communications Magazine*, 53(9):10–17, 2015.

[24] 3GPP. Standardization of Machine-type Communications V0.2.4 (2014-06). Technical report, 2014.

[25] 3GPP. Service requirements for Machine-Type Communications (MTC) Stage 1 (Release 12). Technical Report Release 12, 2013.

[26] Qualcomm. LTE MTC : Optimizing LTE Advanced for Machine-Type Communications LTE Advanced is evolving and expanding into new frontiers. *Optimizing LTE Advanced for Machine-Type Communications LTE Advanced is evolving and expanding into new frontiers*, (November), 2014.

[27] 3GPP. Tr 36.888 study on provision of low-cost mtc ue based on lte, v. 12.0.0. Technical report, 2013.

[28] Alberto Rico-Alvarino, Madhavan Vajapeyam, Hao Xu, Xiaofeng Wang, Yufei Blankenship, Johan Bergman, Tuomas Tirronen, and Emre Yavuz. An overview of 3gpp enhancements on machine to machine communications. *IEEE Communications Magazine*, 54(6):14–21, 2016.

[29] Nokia. LTE evolution for IoT connectivity. 2016.

[30] Huawei. NB-IOT White Paper. Technical report, 2015.

[31] Xingqin Lin, Ansuman Adhikary, Yutao Sui, Johan Bergman, and Hazhir S Razaghi. A Primer on 3GPP Narrowband Internet of Things (NB - IoT). pages 1–8, 2017.

[32] Xingqin Lin, Ansuman Adhikary, and Y. P. Eric Wang. Random Access Preamble Design and Detection for 3GPP Narrowband IoT Systems. pages 1–4, 2016.

[33] Robert W Heath, Michael Honig, Satoshi Nagata, Stefan Parkvall, and Anthony CK Soong. Lte-advanced pro: part 2 [guest editorial]. *IEEE Communications Magazine*, 54(6):12–13, 2016.

[34] 3GPP. Cellular system support for ultra-low complexity and low throughput Internet of Things (CIoT) TR 45.820. Technical report, 2015.

[35] Andreas Maeder, Peter Rost, and Dirk Staehle. The challenge of m2m communications for the cellular radio access network. In *Proc., Würzburg Workshop on IP: Joint ITG and Euro-NF WorkshopâĂİ Visions of Future Generation NetworksâĂİ(EuroView2011)*, 2011.

[36] F. Javier López-Martínez, Eduardo Del Castillo-Sánchez, Eduardo Martos-Naya, and J. Tomás Entrambasaguas. Performance evaluation of preamble detectors for 3GPP-LTE physical random access channel. *Digital Signal Processing: A Review Journal*, 22(3):526–534, 2012.

[37] 3GPP. Study on RAN Improvements for Machine-type Communications. Technical Report Release 11, 2011.

[38] ZTE. 3GPP TSG RAN WG2 #71: MTC simulation results with specific solutions. Technical report.

[39] Ki-Dong Lee, Sang Kim, and Byung Yi. Throughput comparison of random access methods for M2M service over LTE networks. In *2011 IEEE GLOBECOM Workshops (GC Wkshps)*, pages 373–377. Ieee, dec 2011.

[40] Shao-Yu Lien, Tzu-Huan Liau, Ching-Yueh Kao, and Kwang-Cheng Chen. Cooperative Access Class Barring for Machine-to-Machine Communications. In *IEEE Transactions on Wireless Communications*, volume 11, pages 27–32, jan 2012.

[41] Nuno K. Pratas, Henning Thomsen, Cedomir Stefanovic, and Petar Popovski. Code-expanded random access for machine-type communications. In *2012 IEEE Globecom Workshops*, pages 1681–1686. Ieee, dec 2012.

[42] Yu Chen and Wei Wang. Machine-to-Machine Communication in LTE-A. *2010 IEEE 72nd Vehicular Technology Conference - Fall*, pages 1–4, sep 2010.

[43] Kaijie Zhou, Navid Nikaein, Raymond Knopp, and Christian Bonnet. Contention Based Access for Machine-Type Communications over LTE. In *2012 IEEE 75th Vehicular Technology Conference (VTC Spring)*, pages 1–5. Ieee, may 2012.

[44] Zehua Wang and Vincent W.S. Wong. Joint access class barring and timing advance model for machine-type communications. In *2014 IEEE International Conference on Communications (ICC)*, pages 2357–2362. Ieee, jun 2014.

[45] Jen-Po Cheng, Chia-han Lee, and Tzu-Ming Lin. Prioritized Random Access with dynamic access barring for RAN overload in 3GPP LTE-A networks. In *2011 IEEE GLOBECOM Workshops (GC Wkshps)*, pages 368–372. Ieee, dec 2011.

[46] Anthony Lo, Yee Wei Law, Martin Jacobsson, and Michal Kucharzak. Enhanced LTE-Advanced Random-Access Mechanism for Massive Machine-to-Machine (M2M) Communications. In *27th World Wireless Research Forum (WWRF)*, pages 1–7, 2011.

[47] Taehoon Kim, Han Seung Jang, and Dan Keun Sung. An Enhanced Random Access Scheme with Spatial Group Based Reusable Preamble Allocation in Cellular M2M Networks. *IEEE Communications Letters*, 19(10):1714–1717, 2015.

[48] Han Seung Jang, Su Min Kim, Kab Seok Ko, Jiyoung Cha, and Dan Keun Sung. Spatial group based random access for M2M communications. *IEEE Communications Letters*, 18(6):961–964, 2014.

[49] Xiaobin Yang, Abraham Fapojuwo, and Emeka Egbogah. Performance Analysis and Parameter Optimization of Random Access Backoff Algorithm in LTE. *2012 IEEE Vehicular Technology Conference (VTC Fall)*, pages 1–5, sep 2012.

[50] IEEE. Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. Technical report, 2009.

[51] Anand R. Prasad and Neeli R. Prasad. *WLANs and IP Networking: Security, QoS, and mobility*. Artech House, London, 2005.

[52] P. Roshan and J. Leary. *802.11 Wireless LAN fundamentals*. Cisco Press, 2004.

[53] IEEE. IEEE Standard for Wireless Access in Vehicular Environments (WAVE)âĂŤ Multi-channel Operation IEEE Vehicular Technology Society. Technical Report February, 2011.

[54] Jeongwook Seo, Younsung Lee, Hyunsik Kim, Kyungwon Park, Jaemin Kwak, and Dongku Kim. Simulation performance of ieee 802.11 p wave system in small-scale fading models. *Yonsei Mobile Communication*, 2008.

[55] Anand Prasad and Neeli Prasad. *802.11 WLANs and IP Networking: Security, QoS, and mobility*. Artech House, London, 2005.

[56] Arafet Ben Makhlouf and Mounir Hamdi. Design and Experimentation of Rate Adaptation for IEEE 802.11n WLANs. *IEEE Transactions on Wireless Communications*, 12(2):908–916, 2011.

[57] Stefan Aust, R Venkatesha Prasad, and Ignas GMM Niemegeers. Ieee 802.11 ah: Advantages in standards and further challenges for sub 1 ghz wi-fi. In *2012 IEEE International Conference on Communications (ICC)*, pages 6885–6889. IEEE, 2012.

[58] Ali Hazmi, Jukka Rinne, and Mikko Valkama. Feasibility study of ieee 802.11 ah radio technology for iot and m2m use cases. In *2012 IEEE Globecom Workshops*, pages 1687–1692. IEEE, 2012.

[59] Weiping Sun, Munhwan Choi, and Sunghyun Choi. Ieee 802.11 ah: A long range 802.11 wlan at sub 1 ghz. *Journal of ICT Standardization*, 1(1):83–108, 2013.

[60] Toni Adame, Albert Bel, Boris Bellalta, Jaume Barcelo, and Miquel Oliver. Ieee 802.11 ah: the wifi approach for m2m communications. *IEEE Wireless Communications*, 21(6):144–152, 2014.

[61] Evgeny Khorov, Andrey Lyakhov, Alexander Krotov, and Andrey Guschin. A survey on ieee 802.11 ah: An enabling networking technology for smart cities. *Computer Communications*, 58:53–69, 2015.

[62] Stefan Aust, R Venkatesha Prasad, and IGMM Niemegeers. Advances in wireless m2m and iot: Rapid sdr-prototyping of ieee 802.11 ah. In *the Proceedings of the 39th IEEE Conference on Local Computer Networks (LCN), Demo Abstract*, pages 1–3, 2014.

[63] V Ponampalam, J Wang, and R Porat. Tgah outdoor channel models-revised text, 2011.

[64] Stefan Aust and Tetsuya Ito. Sub 1ghz wireless lan propagation path loss models for urban smart grid applications. In *Computing, Networking and Communications (ICNC), 2012 International Conference on*, pages 116–120. IEEE, 2012.

[65] Orod Raeesi, Juho Pirskanen, Ali Hazmi, Toni Levanen, and Mikko Valkama. Performance evaluation of ieee 802.11 ah and its restricted access window mechanism. In *2014 IEEE International Conference on Communications Workshops (ICC)*, pages 460–466. IEEE, 2014.

[66] Mathieu Lacage, Mohammad Hossein Manshaei, Thierry Turletti, and Sophia Antipolis. IEEE 802.11 Rate Adaptation : A Practical Approach. In *ACM international symposium on Modeling, analysis and simulation of wireless and mobile systems*, pages 126–134, 2004.

[67] MadWifi. The MadWifi Project.

[68] Saiid Biaz and Shaoen Wu. Rate Adaptation Algorithms for IEEE 802.11 Networks: A Survey and Comparison. In *IEEE Symposium on Computers and Communications*, pages 130–136, 2008.

[69] John C Bicket. *Bit-rate Selection in Wireless Networks.* PhD thesis, Massachusetts Institute of Technology, 2005.

[70] Jianhua He, Wenyang Guan, Lin Bai, and Kai Chen. Theoretic Analysis of IEEE 802.11 Sample Rate Rate Adaptation Algorithm SAMPLE RATE. *IEEE Communications Letters*, 15(5):524–526, 2011.

[71] Duy Nguyen, J. J. Garcia-Luna-Aceves, and Cedric Westphal. Throughput enabled rate adaptation in wireless networks. *2013 International Conference on Computing, Networking and Communications (ICNC)*, pages 1173–1178, jan 2013.

[72] Z Li, A Das, A K Gupta, and S Nandi. Full auto rate MAC protocol for wireless ad hoc networks. *IEE Proc.-Commun.,*, 152(3):311–319, 2005.

[73] Ioannis Pefkianakis, Starsky H Y Wong, Hao Yang, Suk-Bok Lee, and Songwu Lu. Toward History-Aware Robust. *IEEE Transactions on mobile computing*, 12(3):502–515, 2013.

[74] Yang Song, Xiaoyan Zhu, Yuguang Fang, and Hailin Zhang. Threshold Optimization for Rate Adaptation. *IEEE Transactions on Wireless Communications*, 9(1):318–327, 2010.

[75] Ting-Yu Lin, Ching-Yi Tsai, and Kun-Ru Wu. EARC: Enhanced Adaptation of Link Rate and Contention window for 802.11 Multi-Rate Wireless Networks. *IEEE Transactions on communications*, 60(9):2623–2634, 2012.

[76] Pravin Shankar, Tamer Nadeem, Justinian Rosca, and Liviu Iftode. CARS: Context-Aware Rate Selection for vehicular networks. *2008 IEEE International Conference on Network Protocols*, pages 1–12, oct 2008.

[77] Ce Liu, Siyuan Liu, and Mounir Hamdi. GeRA: Generic rate adaptation for vehicular networks. *2012 IEEE International Conference on Communications (ICC)*, pages 5311–5315, jun 2012.

[78] Oscar Punal and James Gross. RFRA: Random Forests Rate Adaptation for vehicular networks. In *2013 IEEE 14th International Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM)*, pages 1–10, jun 2013.

[79] Xi Chen, Prateek Gangwal, and Daji Qiao. Practical Rate Adaptation in Mobile Environments. In *IEEE International Conference on Pervasive Computing and Communications*, number 978, 2009.

[80] 3GPP. Standardization of Machine-type Communications V0.2.0 (2013-06). 0.2.0:1–44, 2013.

[81] Alberto Rico-Alvariño, Madhavan Vajapeyam, Hao Xu, Xiaofeng Wang, Yufei Blankenship, Johan Bergman, Tuomas Tirronen, and Emre Yavuz. An overview of 3GPP enhancements on machine to machine communications. *IEEE Communications Magazine*, 54(6):14–21, 2016.

[82] Carlos Úbeda, Salvador Pedraza, Miguel Regueira, and Javier Romero. LTE FDD physical random access channel dimensioning and planning. *IEEE Vehicular Technology Conference*, (ii):1–5, 2012.

[83] José Ramón Gállego, Ángela Hernández, Israel Guío, and Antonio Valdovinos. Performance evaluation of nonsynchronized initial random access for mobile broadband systems. *Telecommunication Systems*, 43(3-4):279–294, 2010.

[84] Felipe A P De Figueiredo, Fabbryccio A C M Cardoso, José A Bianco F, Rafael M Vilela, Karlo G Lenzi, and Fabrício L Figueiredo. Multi-Stage Based Cross-Correlation Peak Detection for LTE Random Access Preambles. *Revista Telecomunicaoes*, 15(2):21–27, 2013.

[85] Sungbong Kim, Kyunghwan Joo, and Yonghoon Lim. A delay-robust random access preamble detection algorithm for LTE system. *RWW 2012 - Proceedings: IEEE Radio and Wireless Symposium, RWS 2012*, pages 75–78, 2012.

[86] 3GPP TS 36.211 V13.2.0. Evolved Universal Terrestrial Radio Access (E-UTRA); Physical channels and modulation (Release 13). Technical Report Release 13.

[87] Mohammad M Mansour. Optimized architecture for computing zadoff-chu sequences with application to lte. In *Global Telecommunications Conference, 2009. GLOBECOM 2009. IEEE*, pages 1–6. IEEE, 2009.

[88] Tommy R Jensen and Bjarne Toft. *Graph coloring problems*, volume 39. John Wiley & Sons, 2011.

[89] Daniel Brélaz. New methods to color the vertices of a graph. *Communications of the ACM*, 22(4):251–256, 1979.

[90] Dusit Niyato, Ping Wang, and Dong In Kim. Performance Modeling and Analysis of Heterogeneous Machine Type Communications. *IEEE Transactions on Wireless Communications*, 13(5):2836–2849, may 2014.

[91] Antonis G Gotsis, Athanasios S Lioumpas, and Angeliki Alexiou. M2M Scheduling Over LTE. *IEEE Vehicular Technology Magazine*, (September):34–39, 2012.

[92] Takehiro Nakamura, Satoshi Nagata, Anass Benjebbour, Yoshihisa Kishiyama, Tang Hai, Shen Xiaodong, Yang Ning, and Li Nan. Trends in small cell enhancements in LTE advanced. *IEEE Communications Magazine*, 51(February):98–105, 2013.

[93] Bodhisatwa Chakravarty. *Rate Control Algorithms for IEEE 802.11 wireless networks.* PhD thesis, The University of Texas, Arlington, 2007.

[94] Angeles Santos, Jose Villalon, and Luis Orozco-Barbosa. Rate Adaptation Algorithm for 802.11 Networks: a Dynamic Decision Approach. *Wireless and Mobile Networking Conference (WMNC)*, pages 75–80, 2012.

[95] Bao Hua Liu, Brian Otis, Subhash Challa, Paul Axon, Chun Tung, and Chou Sanjay. On the Fading and Shadowing Effects for Wireless Sensor Networks. In *IEEE Mobile Adhoc and Sensor systems*, pages 51 – 60, 2006.