Intelligent RACH Access strategies for M2M Traffic over Cellular Networks

Lawal Mohammed Bello

PhD

University of York

Electronics

April 2015

Abstract

This thesis investigates the coexistence of Machine-to-Machine (M2M) and Human-to-Human (H2H) based traffic sharing the Random Access Channel (RACH) of an existing cellular network and introduced a Q-learning as a mean of supporting the M2M traffic. The learning enables an intelligent slot selection strategy in order to avoid collisions amongst the M2M users during the RACH contest. It is also applied so that no central entity is involved in the slot selection process, to avoid tampering with the existing network standards.

The thesis also introduces a novel back-off scheme for RACH access which provides separate frames for M2M and conventional cellular (H2H) retransmissions and is capable of dynamically adapting the frame size in order to maximise channel throughput.

A Frame ALOHA for a Q-learning RACH access scheme is developed to realise collision-free RACH access between the H2H and M2M user groups. The scheme introduces a separate frame for H2H and M2M to use in both the first attempt and retransmissions. In addition analytical models are developed to examine the interaction of H2H and M2M traffic on the RACH channel, and to evaluate the throughput performance of both slotted ALOHA and Q-learning based access schemes.

In general it is shown that Q-learning can be effectively applied for M2M traffic, significantly increasing the throughput capability of the channel with respect to conventional slotted ALOHA access. Dynamic adaptation of the back-off frames is shown to offer further improvements relative to a fixed frame scheme. Also the FA-QL-RACH scheme offers better performance than the QL-RACH and FB-QL-RACH scheme.

Contents

List of Figures

Acknowledgements

I would firstly like to thank Tertiary Education Trust Fund (TETFund) Nigeria for funding this work and also thank Bayero University Kano for providing me with the opportunity. I would like to express my deepest gratitude to Professor Hafiz Abubakar for his constant support in many issues related to the TETFund funding.

I am grateful to my two supervisors Dr Paul D. Mitchell and Dr David Grace for their exceptional encouragement, invaluable guidance and motivation throughout my PhD studies.

I would like to thank my parents, brothers, sisters and my entire family for their encouragement and prayers. My special thanks go to my brothers Engineer Abu-sufyan Muhammad and Dr. Abubakar Muhammad for their selfless love and support. I am very grateful to them for being part of my foundation, for all of the advice and the incredible strength they have forced me to see in myself. I am indeed lucky to have such supportive brothers.

Finally the success of this work depends greatly on the encouragement and guidance from many others. I therefore wish to thank all my friends and the entire members of communications research group for their help in many different ways and tolerance of all my annoying questions. It has been nice working with all of you and I enjoyed every bit of it especially the Thursday boring football game!

Declaration

This work has not previously been presented for an award at this, or any other, University. All contributions presented in this thesis as original are as such to the best knowledge of the author. References and acknowledgments to other researchers have been given as appropriate.

Some of the research presented in this thesis has resulted in an initial patent filing and number of publications as follows:

Publications and Patents

L. M. Bello, "Application of Q-Learning for RACH Access to Support M2M Traffic over a Cellular Network", *Patent filed, no. 1405908.3*

Conference Presentations

L. M. Bello, P. Mitchell and D. Grace, "Application of Q-Learning for RACH Access to Support M2M Traffic over a Cellular Network", in *20th European Wireless Conference; Proceedings*, 2014, pp. 1-6.

L. M. Bello, P. Mitchell and D. Grace, "Frame based back-off for Q-learning RACH access in LTE networks", *in Telecommunication Networks and Applications Conference (ATNAC), 2014 Australasian,* 2014, pp. 176-181.

L. M. Bello, P. Mitchell, D. Grace and T. Mickus, "Q-learning based Random Access with Collision free RACH interactions for Cellular M2M", accepted for publication and is to be

presented in the *9th International Conference on Next Generation Mobile Applications, Services and Technologies (NGMAST'15) Cambridge, UK.*

Journal Articles

L. M. Bello, P. Mitchell, and D. Grace, "Throughput Performance of Intelligent RACH Access Techniques to Support M2M Traffic in Cellular Networks", submitted to *the Transactions on Emerging Telecommunications Technologies (ETT) journal*

Chapter 1. Introduction

1.1 Motivation and Purpose of the thesis

Machine-to-Machine (M2M) communication is communications between machines or devices with little or no human intervention [1]. This is due to the exponential increase in the number of devices that require connection to wireless networks [2] which will cause a large difference in the ratio of the number devices to that of humans. This means that direct control of the devices will be difficult by humans, and they will need to communicate autonomously. M2M devices may be connected using wired or wireless access networks. Although wired networks are highly reliable and secure, they are very expensive to roll out and not flexible. Therefore wired M2M will not be considered further in this work. On the other hand, a wireless network provides good coverage, flexibility and mobility, hence wireless networks which may be short range or long range are considered as the most suitable option for M2M communications [3].

Existing cellular networks are considered in this work to support long range M2M communication. This is because of their existence architecture, ubiquity, and roaming support. Some applications of M2M are already supported by GSM and GPRS [4], also as described in [5] LTE and next generation cellular are potential networks for cellular M2M. However, cellular networks are primarily designed for human-centric communication (communications initiated by human) – Human-to-Human (H2H) communication. They primarily support voice calls, video streaming, web surfing and social networking, all of which require basic service requirements like mobility, high data rate, direction of transmission etc. On the other hand, M2M communication has different characteristics from a

potentially huge number of devices, device originated communication, small data payloads, one way traffic (mostly uplink). These distinct features of M2M are expected to cause a congestion problem to the Random Access Channel (RACH) of a cellular network. This is as a result of a large increase in the rate at which the devices access the Base Station (BS) or Evolved NodeB (eNB) by sending simultaneous uplink (UL) signalling messages through the RACH channel [6]. Therefore it is clear from the above discussion that, RACH overload is potentially a critical issue of cellular M2M. Hence supporting M2M on an existing cellular network requires an effective RACH congestion control mechanism.

This thesis considers and investigates the coexistence of H2H and M2M in sharing the RACH of the existing cellular network and mainly focuses on the problem of congestion and overload of the RACH channel. Therefore the purpose of this work is to enhance the RACH access scheme of the existing cellular network to support M2M traffic without negatively affecting or changing the mode of operation of H2H users.

1.2 Overview

Third Generation Partnership Project (3GPP) and other different cellular network standardisation organisations have started working on developing global standards for supporting M2M over existing cellular networks [7, 8]. 3GPP has identified RACH overload as the critical problem in cellular M2M [9]. RACH is the first channel initially used by Mobile Station (MS) or User Equipment (UE) to accessed cellular network (from 2G through to LTE) [10]. Also 5G networks are expected to have similar RACH structure and access techniques to its predecessors. In its effort to realise cellular M2M, 3GPP has suggested different possible solutions (see chapter 6). The solutions proposed by the 3GPP will require a direct involvement of a central entity (BS or eNB). In order not modify the existing cellular

network standards; this work takes a different approach which has a benefit of not involving a central entity to solve the RACH overload problem.

RACH, as the first link to the cellular network, is a very important channel which is structured into frames and slots where access attempts are only allowed in slots. Therefore despite its poor throughput performance, traditional slotted ALOHA (s-ALOHA) has been the popular random access scheme to use for the RACH access in all cellular network standards [10, 11]. This is because the RACH is not a heavily loaded channel since the channel always has limited activity in comparison with the voice/data traffic requirements on the main transport channel. Also the s-ALOHA is a simple protocol and has ability to handle multiple spatially distributed nodes accessing a single channel. Therefore s-ALOHA (being a simple protocol) has been effectively adopted for the RACH access by the conventional H2H users. However, supporting M2M traffic (in addition to the existing H2H traffic) on the existing cellular network will affect the RACH access performance of both M2M and the existing H2H users. Hence the existing cellular network requires some adjustment or the RACH access strategy needs to be altered in order to control the traffic while supporting M2M. The latter option is potentially more straightforward as it could just apply to new M2M devices, without the need to alter the existing cellular infrastructure and standards.

This thesis considers and investigates coexistence of H2H and M2M during RACH access and proposes novel schemes that control the M2M traffic in order to reduce the impact to the existing H2H traffic. As a key contribution, we examine the behaviour of s-ALOHA (as the existing RACH access scheme) as applied to H2H as a single user group and when it is used with the additional M2M traffic. This establishes the instability behaviour of the s-ALOHA scheme which results from the impact of M2M traffic and risks RACH overload. The instability here refers to pushing the system beyond s-ALOHA capacity which causes the channel throughput to degrades and tends towards zero.

In order to maintain existing H2H standards and also not involve a BS or an eNB in controlling the RACH access, this thesis proposes an intelligent RACH access scheme which has the ability to force the M2M users self-organise themselves during the RACH contest. Q-learning which is a model free or an off-policy Reinforcement Learning (RL) [12] is used to control the M2M traffic during the RACH access. RL is preferred to other machine learning methods because of its unsupervised learning nature [12, 13] and its ability to learn from direct interactions with the environment make it suited to distributed RACH access scenario. In addition the Q-learning is chose here because its learn action approximates and simplifies the RL algorithm that enables early convergence.

One of the key contributions of this work is the use of the Q-learning in an intelligent slot selection strategy to realise a novel scheme called Q-Learning for RACH access (QL-RACH). Results show that the QL-RACH scheme has the potential to improve the performance over s-ALOHA when used as a single user group and also controls the RACH overload problem when combined with the conventional s-ALOHA scheme. Another scheme is introduce to improve the performance of the QL-RACH by enhancing the back-off strategy adopted by the H2H traffic. This reduces the random effect of the uncontrolled H2H traffic by implementing a separate frame for H2H retransmissions and no back-off for M2M, to realise what is called Frame-based Back-off for QL-RACH (FB-QL-RACH) scheme. The scheme is improved further by introducing a dynamic frame size adaptation method to the H2H back-off frame. In another effort to simplify our schemes, we introduce another novel scheme that enables collision free RACH access between H2H and M2M users. Separate frames are proposed to be used by H2H and M2M RACH request transmissions to realise a Framed-ALOHA for QL-RACH (FA-QL-RACH) scheme.

This thesis additionally provides analytical model for thorough investigation of H2H and M2M traffic behaviour in sharing the RACH access. This allows us to predict the throughput

performances of our proposed schemes and also validates the simulation models. Finally the thesis examines the possibility of coexistence of Poisson traffic distribution (typical for H2H) [14-16] and Periodic traffic distribution (typical for M2M), by controlling the M2M traffic using the QL-RACH scheme. Figure 1.1 presents a summary of the conventional RACH access scheme and our proposed schemes.

Our research resulted into a number novel contribution and their details are presented in chapter 10. This work is intended to be applicable to all current and future cellular network standards; this is because they have similar RACH signalling channels, functional structures and access schemes.

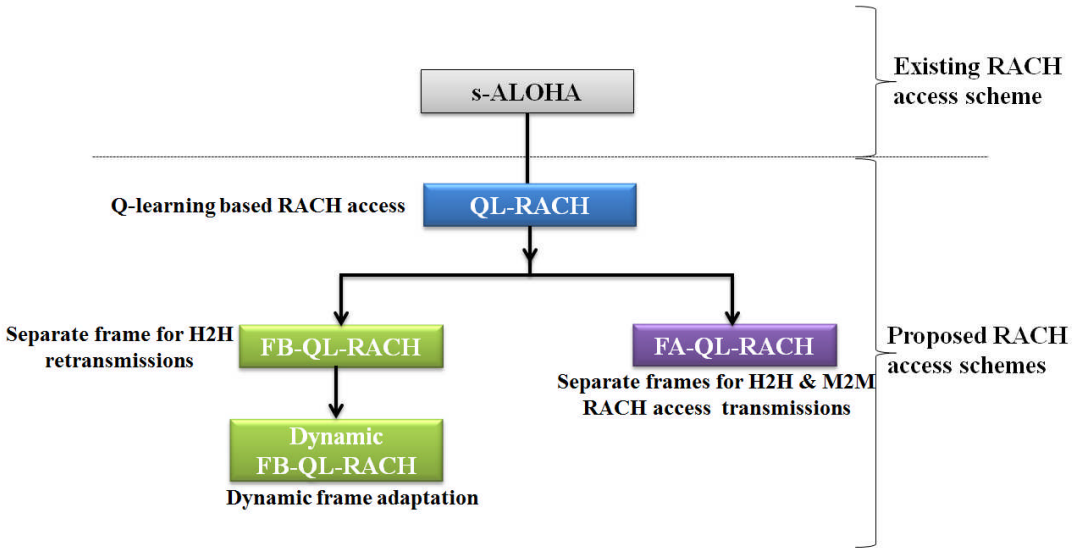

Figure 1.1 Block diagram presentation of existing and proposed RACH access schemes

1.3 Structure of the thesis

The remainder of thesis is organised as follows:

Chapter 2 provides an overview of the concept and motivation for M2M communication with classifications of M2M based on their supporting access networks. The chapter also provides

some examples of different M2M application scenarios that will foresee the realisation of M2M communications.

Chapter 3 presents background information on different types of cellular systems and their resource management. Evolution of cellular network generations is discussed followed by the various multiple access techniques used in coordinating access to the network resources. Different types of cellular network channels, their structure and functions are also presented with emphasis on the RACH channel due to its relevance in this work.

Chapter 4 is reserved for simulation techniques and methods of validation. The motivation and importance of using simulation to evaluate the designed communication system is provided first. Different steps follow to evaluate the system as well as techniques used to examine the accuracy of the introduced steps. The general assumption on the traffic model adopted in this work is presented here.

Chapter 5 introduces the basis of the proposed schemes in this work, where the impact of additional M2M traffic in the existing cellular network is investigated. The basics of the two ALOHA (slotted ALOHA and pure ALOHA) schemes are introduced and their performances compared. The cause of RACH instability is illustrated by analysing the performance of slotted ALOHA as an existing RACH access scheme. RACH-throughput, average end-to-end delay and blocking probability performances are compared using retransmission limits and back-off interval windows. An analytical model to develop RACH-throughput performance is also presented.

Chapter 6 describes the implementation of a novel scheme called QL-RACH which is proposed to control M2M traffic using Q-learning. A scenario of a combined RACH access scheme in which H2H traffic used s-ALOHA for RACH access (SA-RACH) and M2M traffic controlled by the QL-RACH is presented. New RACH-throughput equations for the

combined access schemes are developed using an analytical model and their performances are compared with a simulation model. Part of this work has been published in IEEE proceedings of European Wireless Conference 2014 presented in Barcelona, Spain. Also the same work has been filed for patency.

Chapter 7 introduces a novel back-off scheme called FB-QL-RACH to enhance the performance of the QL-RACH scheme. A separate frame is provided for H2H traffic in retransmission to minimise the effect of the uncontrolled H2H users. Also an analytical model is developed to predict the RACH-throughput performance of the scheme. Both simulation and the analytical models are used to present and compare the performances of the FB-QL-RACH with that of QL-RACH scheme. To make the scheme more efficient, a FB-QL-RACH with dynamic Back-off Frame size (BFZ) is introduced to enable an Evolved Node B (eNB) as a central entity to update the required BFZ value automatically based on a defined threshold of probability of blocking. Part of this work has been published in IEEE proceedings of Telecommunication Network and Applications Conference 2014 presented in Melbourne, Australia.

Chapter 8 presents further development of the QL-RACH scheme by providing complete separate frames for both H2H and M2M to realise what is called FA-QL-RACH scheme. This scheme eliminates collisions between H2H and M2M user groups during the RACH combined access. The cutting-edge scheme also includes dynamic frame size which adjusts to match the H2H traffic condition. Part of this work has been accepted and to be published in IEEE proceedings of Next Generation Mobile Applications, Services and Technologies, 2015 to be presented in Cambridge United Kingdom.

Potential further work to extend this research is presented in chapter 9, followed by an overall summary and conclusion of the work in this thesis provided in chapter 10.

Chapter 2. Background

2.1 Introduction

The purpose of this chapter is to lay the foundations by providing important background information related to this thesis. The concept and motivation for M2M communication is introduced in section 2.2. Types of M2M and possible supporting networks are both presented in section 2.3. Section 2.4 presents some of the application scenarios of M2M. Section 2.5 present reinforcement learning, its area of application and methods of solution, as well as Q-learning algorithm. Finally the chapter is summarised in section 2.6.

2.2 M2M communications

M2M communication is an emerging field in modern wireless communications. It is envisioned to enable various electrical/electronic devices (M2M devices) to be connected and operate autonomously with little human intervention. This is becoming necessary as it is forecast in [17-19] that industrial and domestic M2M applications will encompass about 50 billion devices that require access to wireless networks by 2020. The difference in the ratio of the number of the M2M devices to that of the estimated human population of 8.3 billion will be significant. Therefore direct control of the machines by humans will be difficult and hence the need for them to communicate among themselves.

There is a close relationship between M2M communication and Internet of Things (IoT) with a clear distinction between the two. M2M can be considered as an early form of IoT where the devices shared information and make decisions among themselves without human

intervention. On the other hand IoT is the advanced interconnection between various devices of M2M (with the same or different application) and information technology [20-22].

M2M communications have been defined in [23] as ‘‘the communication between computers, embedded processors, smart sensors, actuators and mobile devices without or with limited human intervention’’ M2M uses devices like sensors and meters to capture another machine’s status and pass on through either wired, wireless or even a combination of the two to the target destination [24, 25]. Most present and emerging applications of M2M communications often involve sensors as the M2M devices which are located in both accessible and non-accessible locations that necessitate them to be locally connected for end-to-end communication. Furthermore, Radio-Frequency Identification (RFID) technology has been foreseen as another potential technology that will play a very important role in M2M communication [26].

2.3 Types of M2M and the Access Networks

An important aspect of M2M communication is the access network which can either be wired or wireless. Even though wired M2M communication can be highly reliable and can provide high data rates due to the dedicated cabling, it will often be rather difficult to implement. This is because it will be expensive to roll out due to the high number of the devices. Based on this, wired M2M communication is not considered further in this work. On the other hand, wireless M2M communication can be implemented using either short range or long range networks or the combination of the two (hybrid). The types of access networks that can be used as a basis in the classification of M2M systems are described later in this section. Authors of [27] classified M2M as fixed or mobile, high traffic or low traffic where the application area determines the category and the type of network to support M2M communication.

The European Telecommunications Standards Institute (ETSI) in its ICT standards specify M2M network made from the following five structures shown in figure 2.1 [1].

i. **M2M devices**: mostly autonomous in receiving and sending data.
ii. **Gateway**: serves as a link between the M2M devices and the outside world, i.e. when the devices are connected using different networks.
iii. **M2M area network**: the network that connects the M2M devices locally and sometimes to the gateway.
iv. **Communication network**: the network that links up devices with the other side of M2M application mostly via the gateway.
v. **M2M application**: the main purpose of using the devices.

This research focuses on the M2M access networks and this is directly related to the **M2M area network** and the **communication network** in **iii** and **iv** above. The former include Wireless Personal Area Network (WPAN) technologies (like IEEE 802.15, Ultra-wideband (UWB), Zigbee and Bluetooth) or Wireless Local Ares Network (WLAN) (e.g. Wi-Fi) and the latter include (but is not limited to); Global System for Mobile (GSM), General Packet Radio Service (GPRS), Worldwide Interoperability for Microwave Access (WIMAX), Long Term Evolution (LTE) and LTE-Advanced [1].

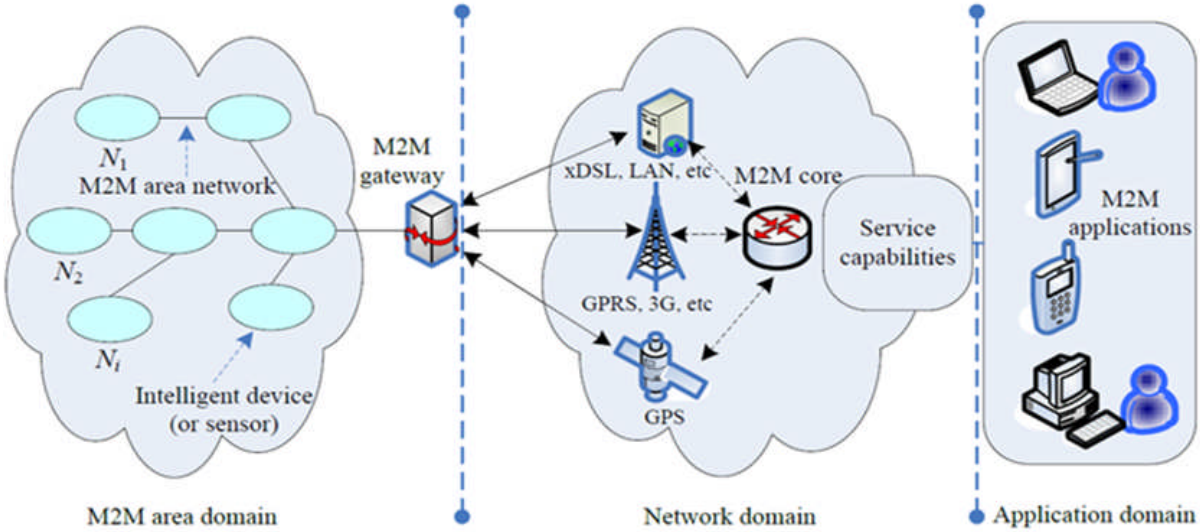

Figure 2.1 Illustration of M2M architecture domain (directly reproduced from [1])

Based on the supporting access network classification different types of M2M systems can be realised and discussed in the next three sections:

2.3.1 Short range M2M Communications

Areas of applications of M2M communications are diverse in nature and require a means of information collection at various places that are not easily accessible. Nowadays, in industrial settings, sensor nodes are placed at different locations to collect real time machine status information which helps in monitoring the system operation. The same procedure can be applied to any area of M2M application where nodes are deployed as sources of information and linked/connected using various technologies that support short range communication which realises what is called Short range M2M communication. Also Near-Field Communication (NFC) a good communication technology where different consumer devices can talk to each other using RFID technology will be a good potential to serve as a short range M2M interface [26]. The sensor nodes are mostly chosen for such tasks because of their small size and the fact they require less physical monitoring. Therefore M2M devices need to be power efficient because of their independent nature of operation, i.e. they should

have low energy consumption capability and one of the ways to make this practical is to provide short range connectivity.

Most existing M2M devices therefore communicate within a short range. This is possible either through an ad-hoc (infrastructure-less) network, e.g. Wireless Sensor Network (WSN) which is used in supporting most of the industrial applications. Local networks that use existing technologies like: Zigbee (IEEE 802.15.4), Bluetooth (IEEE 802.15.1), UWB and the infrastructure based networks that apply Wi-Fi (IEEE 802.11) technology [28-30] are also viable networks for short range M2M. These wireless communication technologies have different modes of operation, data rate capabilities, power requirements and of course protocols and operate in unlicensed spectrum bands. Depending on the type of application or scenario, one or more of these technologies may be required for communication and that has been identified as one of the challenges of M2M communication using devices with different modes of operation [29]. Therefore there is a need to find a good way in which the devices with different network technologies can share the unlicensed spectrum. Authors of [31] presented details of spectrum sharing policy under cognitive radio technology. TV white space bands [32], obtained from the process of digital television transition, is foreseen as a potential for M2M communication. Authors of [20] proposed a standard called Weightless for the deployment of M2M in TV white space. On the other hand there are many standardisation activities reported in [32, 33] on how to coordinate the sharing of these TV white spaces among short range communication devices that use higher frequencies mostly in the GHz range.

Capillary (a smart way to get things connected) M2M communication together with the M2M gateway form what is called the M2M area domain in [1]. This domain may comprise a large number of devices (sensors, meters, RFID etc) with the same or different functions that may be connected together and also (when required) to the gateway.

2.3.2 Long range M2M Communications

Long range M2M is required where the communication range needs to be extended beyond short range coverage or where the M2M devices need to move around whilst communicating. Depending on the area of application, sometimes the data (information) captured by the individual devices needs to be forwarded to the outside world (area different from the current devices' location). Therefore for a long range M2M implementation, ubiquitous network coverage is required. Cellular networks (like GSM, GPRS, LTE/LTE-Advanced and WIMAX) are available networks that satisfies the above requirement and are likely to become the long range network for M2M [1]. In addition cellular networks can be considered as an appropriate network for M2M since most of the applications cannot justify the installation of a private radio network. Cellular networks, because of their availability and ubiquity, are already used in some M2M applications like vending machine monitoring and fleet management [20].

In cellular M2M the devices connect to the cellular network directly, possibly by equipping each M2M device with its own Subscriber Identity Module (SIM) card to allow the connectivity [34].

Authors of [35] proposed an architecture shown in figure 2.2 that simplifies the cellular network by introducing an additional element called an M2M facilitator, which transforms the M2M system where by the communication between M2M devices is through the facilitator. Another alternative that can be considered for a long range M2M which may be suitable to the indoor system is linking the devices using Wi-Fi/Ethernet to the wider network (internet).

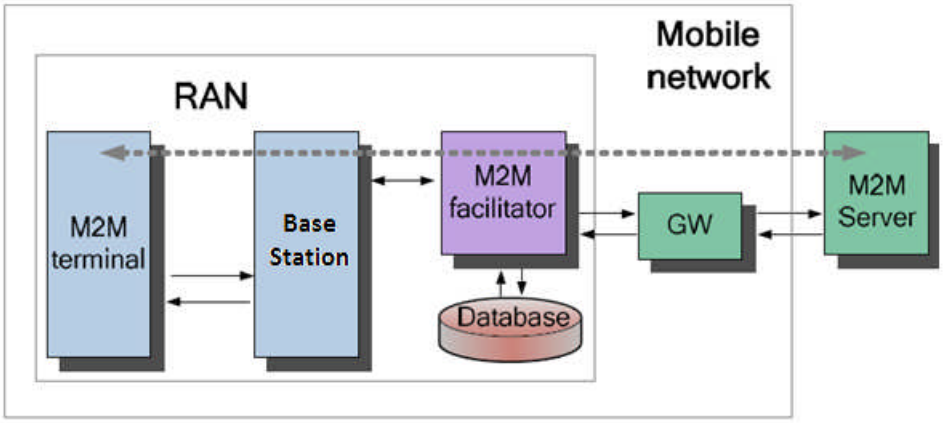

Figure 2.2 Proposed cellular M2M architecture (directly reproduced from [35])

2.3.3 Hybrid M2M Communications

It is assumed that the two types of M2M communications discussed above operate in a different way, i.e. short range is restricted to a confined area, and long range or cellular covers a large area, and both have different standards and technology. However, some applications will require the combination of the two types in order to achieve efficiency in terms of ubiquity, power and portability. This can be achieved in what is called hybrid M2M communication. Therefore depending on the application scenario a cellular network can be used to forward the content of a short range M2M system to a different location. Here the communication is initially on short range and then aggregated to the cellular network (or the other way round) in order to provide end-to-end M2M communication. Due to the non-homogeneity of the network, hybrid M2M communication will to be more complex than the other two. This is because two or more networks with different features need to be aggregated to achieve the hybrid M2M.

Wireless network integration is the key factor to achieve hybrid M2M communication. For example WLAN may be used to carry the usual short range communication using either infrastructure (access point) or an ad-hoc network and depending on the application and location of the end-point, the system may require access to the cellular network. This concept

and examples of various architectures using heterogeneous network architectures can be found in [36].

2.4 M2M Applications

There are many existing applications of M2M communications and more are being realised every day. In fact as mobile communication transformed itself from a luxury service to a necessity, M2M communication is expected to do very much the same. A wide range of applications of M2M have been reported in [1, 20, 37-39] which can be either personal, environmental, or domestic, such as; surveillance, health care, smart grid, billing utilities, traffic control, transport, smart homes, smart farming, industrial automation, smart cities to mention but a few.

Applications of M2M communication are too numerous to mention, and a few have been presented to provide an overview and also to emphasize the need for end-to-end networks which will require both long and short range communications and also that depend more or less on the type of application. Also the applications provide a clear view of the nature of the communication. For example, some applications require periodic data transmission and some are continuous. Also some contain a small amount of data while some are high data based applications. In addition some applications are on huge scale involving numerous nodes while some are on a much smaller scale.

Some specific examples are chosen to illustrate typical application scenarios. Figure 2.3 presents images of some applications of M2M which are briefly discussed below:

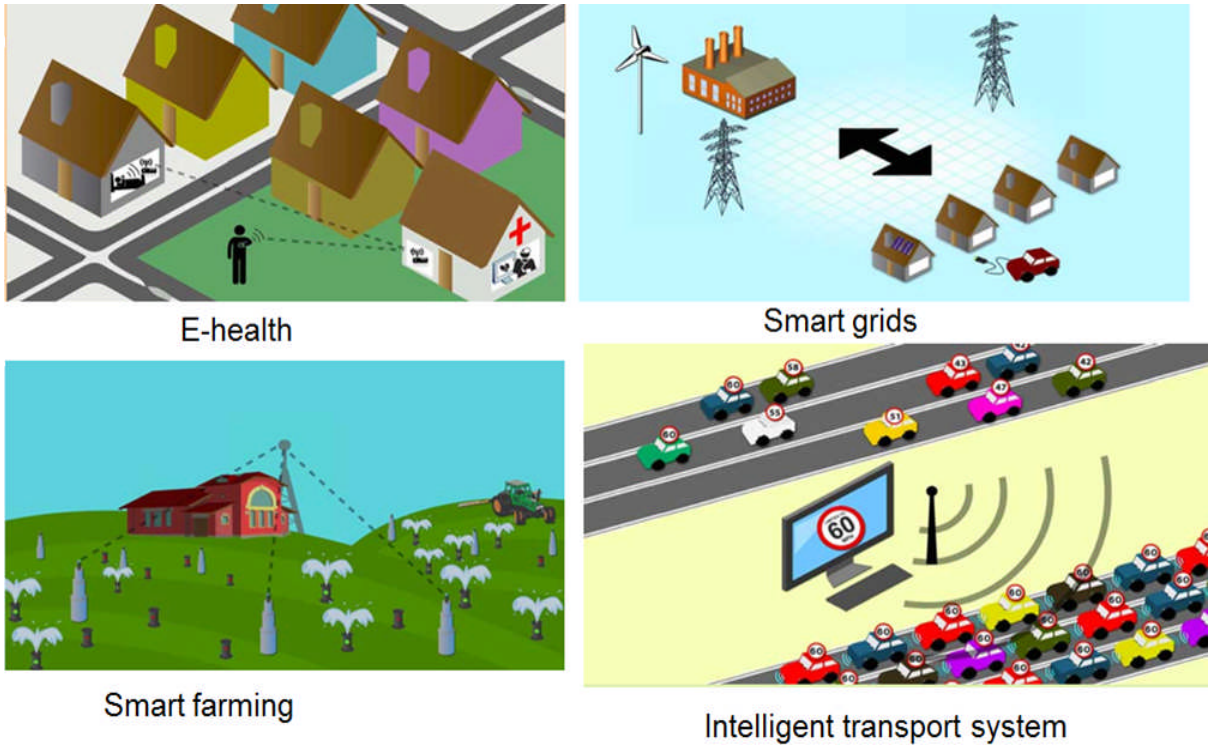

Figure 2.3 Examples of M2M application scenarios (images copied from [40])

2.4.1 E-health

Electronic health (E-health) is a process that enables remote data communication between patients and health care centres. M2M has a significant role to play in this application by providing a means of remotely monitoring patient health (sometimes fitness), activating warning alarms on detection of critical condition and to offer required solutions [21]. Wireless sensors are mostly used as the M2M devices and attached to patients' bodies as the monitoring devices. Readings of parameters like: blood pressure, heart rate, medicine levels in the pill box etc, can all be sent to the monitoring centre when required. Figure 2.3 illustrates an example of a remote health monitoring process, where the wearable sensors send the patients status to the central medical data base and an intelligent pill box triggers an alarm to the patient's mobile in case of tablets not been taken [41].

2.4.2 Smart Grids

This is a process of intelligently managing, controlling and using energy resources through communication between the central producing unit and the end consuming sites. The process will provide energy efficiency by balancing the generation and demand, cost effective to both production and consumption, easy fault detection, etc. Therefore as presented in [38] smart grids are emerging as a convergence of information technology and communication technology with power system engineering. M2M is envisaged as a solution here where the power consumption of different domestic devices can be monitored using M2M devices (sensors) called smart meters. The energy billing, consumption and generation can be managed and controlled in real time by using the collected data from the smart meters [25]. Figure 2.3 shows an example of two way communications between the generation plant and the domestic consumers for the establishment of an intelligent controlled power system.

2.4.3 Intelligent Transport Systems

This category has a wide range of applications which include (but is not limited to): fleet management services, traffic control, smart parking etc. The fleet management services for example can be used to track a vehicle location or its status, or even acquire some information regarding its engine such as; speed, brake, mileage using Digital Tachograph (DTG). M2M devices will be attached to the required part of the vehicle to collect the information and send it to the central unit for necessary action. An example of traffic congestion monitoring is illustrated in Figure 2.3 where the information is sent to the control unit for regulation. This could be achieved by attaching M2M devices to the wheels of the vehicle in order to measure the speed and send data to a central traffic control system which will automatically impose a variable speed limit to reduce the congestion [41].

2.5 Reinforcement Learning

Reinforcement Learning (RL) is an area of machine learning which is a multitask field that discovers the design and study of algorithms that learn from some actions and improve the performance with experience [12]. Therefore RL is a sub-field of Machine learning which is widely applied in various areas such as; psychology, medical diagnosis, neuroscience, informatics, cybernetics and control theory. The idea of RL was developed from the understanding of how humans/animals interact with their environment and use their experience to choose the best action(s) [42]. The distinct feature of RL that makes it differs from other forms of machine learning is its learning independency (unsupervised learning). The learner here has the ability to discover which actions produce the best outcome by trying them through trial-and-error from the direct interaction with the environment. Hence the individual learning capability of RL where the learner learns only on local observation makes it suitable to distributed network.

The following three approaches have been developed to solve different RL problems [12]:

i **Dynamic Programming (DP)**: these methods involve collection of algorithms define as a Markov Decision Process (MDP) used in computing optimal learning policy from perfect model of the environment. This is not widely use in solving RL problems as they require assumption of the perfect model of the learning environment and therefore is not considered in this thesis.

ii **Monte Carlo**: even though Monte Carlo methods don't require a model but they are not suitable for step-by-step incremental computation. We intend to use learning in this work in slot selection strategy which will require an incremental update of the learning process. Therefore Monte Carlo is also not considered in this thesis.

iii **Temporal Difference (TD)**: these methods require no model and are fully incremental. TD methods have the ability to learn directly from the experience of the learning process without prior knowledge of the model of the environment. Features of TD methods make them suitable in our application slot selection and therefore considered in thin thesis. We adopt an off-policy TD control algorithm known as Q-learning [12, 13] which directly approximate the optimal action independent of the followed policy. This is effect simplifies the analysis of the algorithm and enabled early convergence.

Q-learning is considered as an efficient algorithm of TD methods used to solve RL problem in this thesis. The algorithm has been applied in the design of MAC protocol of the RACH channel for additional M2M traffic in the existing cellular network. We consider sharing of the RACH as a single channel (structured in time slots) amongst the M2M users. We then apply the Q-learning as an intelligent slot selection strategy. The detailed Q-learning implementation is presented in chapter 6.

2.6 Summary

This chapter provided background information related to the research presented in this thesis. A general overview of M2M communication has been given with its various classifications which are mainly based on the M2M applications. The classifications also define the type of access network required by the system which is the purpose of this work. Some of the applications of M2M were also introduced. Concept of Reinforcement Learning has been introduced where three different solution methods are briefly described. The reason for choosing Q-learning algorithm of temporal difference methods has been explained.

Chapter 3. Resource Management for Cellular Systems

3.1 Introduction

This chapter provides background information on different types of cellular systems and their resource management. Firstly, evolution of various cellular network generations from the analogue first generation towards fourth generation is briefly discussed in section 3.2. Section 3.3 describes the various multiple access techniques used in coordinating access to the cellular network resources. Cellular network channel classifications, structure and functions are discussed in section 3.4. The RACH is relevant in this work and therefore details of its structure, function and access protocol is provided in section 3.5 and finally the chapter is summarised in section 3.6.

3.2 Cellular Systems

Cellular networks represent the most accepted and widely deployed wireless network technology. This is because of their market impact on society as a result of various services offered as well as the ubiquity of the network. Technological development influences users' high expectations (high capacity requirement, ubiquitous communication etc.) from cellular service providers and has made the system pass through many evolutionary trends. The evolution started from first generation (1G) analogue cellular systems which have limitations in several ways such as; low capacity, lack of roaming capabilities among others [43]. Lack of these features made 1G give way to the Second generation (2G) cellular systems which have a number of advantages such as; voice quality improvement, advanced roaming capability and many more as described in [43]. The Global System for Mobile

communications (GSM) is a 2G standard and the most popular cellular system worldwide because of its promising features of various services (in addition to the traditional voice) and ubiquity. However, it is still found to be inefficient in some applications especially data communication due to its lower spectral efficiency that makes it unable to transmit at high data rates. An interim step has been taken between 2G and 3G (actually before activating 3G) to realise 2.5G which is basically 2G with improved spectral efficiency that introduced higher throughput capability for data services [43]. However the desire for higher data rate improvement motivated the development of Third Generation (3G) systems which appeared to be the problem solver by improving the spectral efficiency. It enables wide area wireless voice telephony, as well as integrating the features of wireless communication and an internet protocol (IP) based network that supports voice and data communications [43, 44]. 3G has of course, provided a lot of additional capacity and much faster data rate per user, therefore data communication is mobile. Consumers' desire for high data capacity and speed is increasing everyday which brings the introduction of the Fourth Generation (4G) standard. 4G has been introduced based on promise to give high data rates at vehicular speeds which according to [43] will enable cellular and wireless network technology to be merged. In contrast to 3G, 4G plans to integrate the existing cellular technologies. Long Term Evolution (LTE) and LTE-Advanced (LTE-A) have so far shown to be a popular candidate for the 4G standard.

3.3 Multiple Access Techniques in Cellular Networks

In physical sciences, radio spectrum is the term referring to the block of channels used for wireless communications which can be considered as a natural resource required in the field. Even though from its features the radio spectrum can be classified as non-exhaustible, it is considered as scarce resource due to the way it is being used [45]. Therefore because of this

effect, radio spectrum could be limited and to maximise communication system capacity the limited available spectrum has to be shared among the numbers of users accessing the system. Various access methods are employed in wireless systems to coordinate the successful operation of numerous users (terminals) over the wireless medium (air) using what is termed a multiple access technique. In cellular networks, a multiple access technique is used as a means of dividing up the capacity for simultaneous access by the multiple users in the system. Different protocols are used to achieve this, some use a Fixed Assignment (scheduling) technique while others are Random Access (contention) based. An example of the former is when users' transmissions are separated in frequency, time or with sets of orthogonal codes. Depending on the technology employed, the following four main schemes are used in cellular systems: Time Division Multiple Access (TDMA), Frequency Division Multiple Access (FDMA), Code Division Multiple Access (CDMA) and Orthogonal Frequency Division Multiple Access (OFDMA) [46].

The schemes are implemented by a Medium Access Control (MAC) sub-layer located in the lower half of second layer (data link layer) of the International Standard Organisation-Open systems Interconnection (ISO-OSI) reference model [47].

3.3.1 Frequency Division Multiple Access (FDMA)

FDMA is a technique where users are separated by dividing the frequency spectrum into blocks of non-overlapping frequency bands and each user is assigned a fixed band. Once assigned, the channel will be occupied by the user until the end of the transmission. A guard band is inserted between the frequency blocks to avoid interference from adjacent frequencies. This looks simple since each user has a unique channel for transmission without any coordination. FDMA was implemented by the first generation analogue cellular networks one of which is the Advanced Mobile Phone System (AMPS) [43, 48], also 2G and 4G to

some extent apply FDMA. On the other hand, FDMA can become inefficient since the assigned channels cannot be used by other users even if they are idle.

3.3.2 Time Division Multiple Access (TDMA)

In TDMA, a given block of frequency (spectrum) is divided into time slots in which a user is allocated regular time slot intervals for contention free transmission. Therefore in this scheme users are allowed to transmit on a common frequency. This scheme maximises the efficiency of the channel usage and increases capacity as well, since a user occupies a block of frequency just for a given period of time. A temporal guard interval is required between successive time slots to solve the problem of inter-symbol inference as a result of transmission delays from a Mobile Station (MS) to a Base Station (BS). TDMA is used in some cellular standards like GSM and Digital Enhanced Cordless Telecommunication (DECT) where speech is digitised and sent as a short packet of data quantified by units called bursts [49]. In addition and relevant to this work, the same technique is used to divide the RACH resource of cellular network. Conventionally in a cellular system, a repeating frame structure containing some number of time slots (the number varies depending on the design specification) is designed within the block of frequency. As introduced earlier, the focus of this work is to design MAC protocols for cellular network RACH access and this is transmitted as bursts on a repeated time slots of frame in a given block of frequency. Figure 3.3 (appears in later pages where it is discussed in more detail) is a representation of TDMA describing time slots on a repeating frame.

3.3.3 Code Division Multiple Access (CDMA)

In CDMA techniques, users transmit simultaneously on the same frequency using different spreading codes. Each user will be assigned its own unique pseudorandom codeword (to be used for transmission) which is orthogonal to all other codewords in the system. The receiver

correlates the received signal with the codeword of the desired user to decode the transmitted information [49, 50] and considers other codewords as random noise. This provides an increase in the channel capacity since the scheme allows multiple users to transmit their information simultaneously on the same frequency.

3.3.4 Orthogonal Frequency Division Multiple Access (OFDMA)

Understanding OFDMA requires basic knowledge of Orthogonal Frequency Division Multiplexing (OFDM) which is a multiplexing technique where by, a block of frequency is divided into multiple orthogonal frequency sub-carriers. Transmission is done by dividing the signal in to parallel sub-streams and by transmitting each on a separate orthogonal sub-carrier [51]. The OFDM technique is used to alleviate the problem of channel impairments (like frequency fading due to multipath) and it also provides high spectral efficiency [48, 52]. OFDM is utilised as a multiple access technique to realise OFDMA which allows different users to transmit simultaneously over different sub-carriers. OFDMA is used as the multiple access scheme in the 3GPP LTE downlink and Single Carrier OFDMA (SC-FDMA) for the uplink. The main issue for different schemes is the power requirement, for the details see [52] as this will not be discussed further in this thesis.

3.3.5 Random Access Techniques

The multiple access schemes discussed above are preferably used when the system traffic is fixed or at least slowly changing. In other words, a contention-free multiple access scheme is more effective in a system with a small population of known users with heavy and regular load [53]. In this situation the system's traffic will agree with the technique without building an unacceptable queue as well as resource wastage. However, for a system with a large number of users having bursty traffic and unknown (unsteady) topology, a random access scheme is more practical. Random access techniques allow a user to decide when to transmit

on the channel. In a cellular system, multiple users are spatially distributed within the cell coverage and are all unknown to the central entity i.e. BS for GSM or eNodeB (eNB) which is a central entity that controls network access and other functions in LTE. Hence a random access scheme is used here by the users to send initial requests, where the central entity will become aware and start communicating to enable contention-free resource assignment. Based on this approach, random access is another scheme relevant to this work since it is the protocol used by existing H2H users to access the RACH.

ALOHA schemes were the first random access protocols [54, 55], with two standard techniques; pure ALOHA (p-ALOHA) and slotted ALOHA (s-ALOHA). In the former technique users transmit on to the channel as soon as a packet arrives in their queue. On the other hand s-ALOHA's transmission starts at the beginning of a slot (mostly a time portion obtained from a TDMA scheme). Difference in the performances of p-ALOHA and s-ALOHA are discussed in chapter 5. All cellular network standards used TDMA frames (containing slots) mapped on to the physical channel as RACH channels. The RACH access is restricted to the slots i.e. the access is on a slot basis which makes s-ALOHA or s-ALOHA-like to be the appropriate RACH access scheme. Therefore s-ALOHA is used as the basis of all our proposed schemes in this work. Details of the RACH channel structure and access procedure is presented in section 3.5.

3.4 Types of Channel in Cellular Networks

This section presents different types of channel and the access techniques with reference to especially GSM and LTE/LTE-A. This is because of the close integration in their technologies, with the former being seen as still relevant because of its ubiquity and

affordability, the latter is foreseen to have a lot of promise in supporting future M2M applications.

3.4.1 GSM Network and Channel Arrangement

GSM uses both FDMA and TDMA as multiple access techniques and operates in the 900 MHz and 1800MHz bands to realise GSM 900 and GSM 1800 respectively. In the former, the 900MHz band is used as the primary band and includes two sub-bands of 25 MHz each with the forward link (BS to MS) allocated 935-960 MHz and reverse link (MS to BS) allocated 890-915 MHz [56-58]. On the other hand GSM 1800 uses 1.8 GHz as the primary band with 1805-1880 MHz for the forward link and 1710-1785 MHz for the reverse link. Each of the allocated bands of frequencies is then planned by dividing into block of frequencies. For example in GSM 900, the band of 25 MHz is divided into 124 carrier channels of 200 kHz each which allows one carrier channel (the first 200 kHz) as a guard band between GSM and other lower frequency related services.

As shown in Figure 3.1, GSM has three important interfaces that link various components of the network. **A-interface** is used in connecting the Base Station Controller (BSC) to the Mobile Switching Centre (MSC). The **A_{bis}-interface** is the interconnection between the Base Transceiver Station (BTS) and the BSC. The **U_m-interface** (radio interface) connects MS to BTS and this is considered as the relevant interface in this work (because it carries the RACH channel) and will be discussed further. However, for the details and functions of the different components of the GSM network and the remaining interfaces see [57, 59, 60].

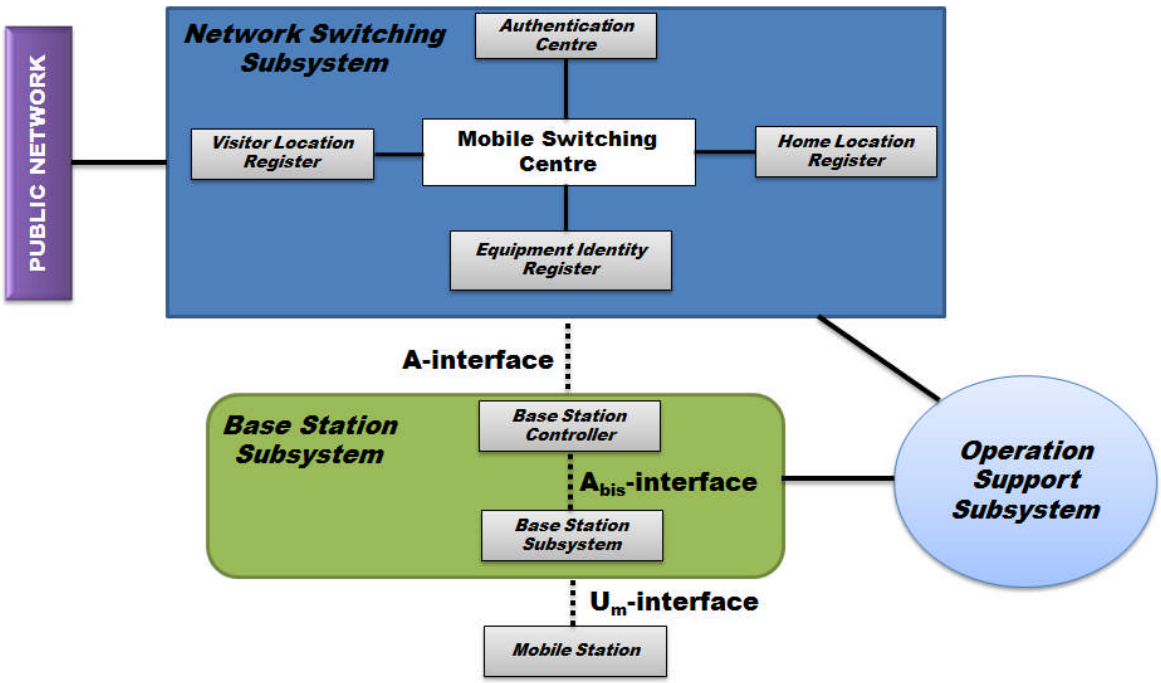

Figure 3.1 Representation of GSM network architecture

The U_m-interface is the air interface (radio) that connects MS to BS and is an important interface in a GSM network that determines the system's capacity. The interface is designed in a way to enable interoperability of MSs within different GSM networks. The Physical layer (layer 1 in the seven-layer OSI model) is used as a physical interface that link the data link layer (layer 2 of the OSI) and the radio resource management sub-layer of the MS and the BS [57, 59]. Therefore the physical interface is made up of sets of physical channels that combine both FDMA and TDMA [58]. As shown in Figure 3.3 (appears in later pages where it is discussed in more detail) each of the 200 kHz carriers described above is divided (in time) into 8 time slots of period 0.57 ms each. The 8 time slots form a frame which is 4.615 ms duration.

In GSM, a channel is defined by the recurrence of time slots in a given TDMA frame and that is called a physical channel. Logical channels are other type of channels classified in GSM which are used to carry data and signalling information and are defined based on the

information carried within the physical channel. There are two types of logical channels; **control channels** and **traffic channels.** The traffic channels are used to send data services and in GSM is the final stage of call setup. The focus of this work is not on the traffic channel and it will not be discussed further, see [56] for the details. On the other hand control channels are used in sending or receiving command messages between an MS and BS. The first thing an MS does when switched ON is to scan and search for a strongest carrier frequency from its operator's BTS and find out if the located carrier is a control channel. Some of these channels are very relevant and important in this work and are broadly divided into three categories presented below:

i **Broadcast Channel (BCH)**

ii **Common Control Channel (CCCH)**

iii **Dedicated Control Channel (DCCH)**

RACH is a control channel class belongs to the CCCH channel set, which is for uplink point-to-point use by the MS for call initiation. Its functionality, structure and format will be discussed in section 3.6 because of its relevance in this work. The remaining control channels are not the focus of our work and therefore not discussed in this thesis. However a summary of their classifications is presented in Figure 3.2 and also the details for their classifications, structure and function are provided in [56, 58, 61]. Furthermore, knowledge of all the control channels is important and required to model the MAC layer of a GSM system. This is because some of the control channels are shared not only by frame but slot numbers.

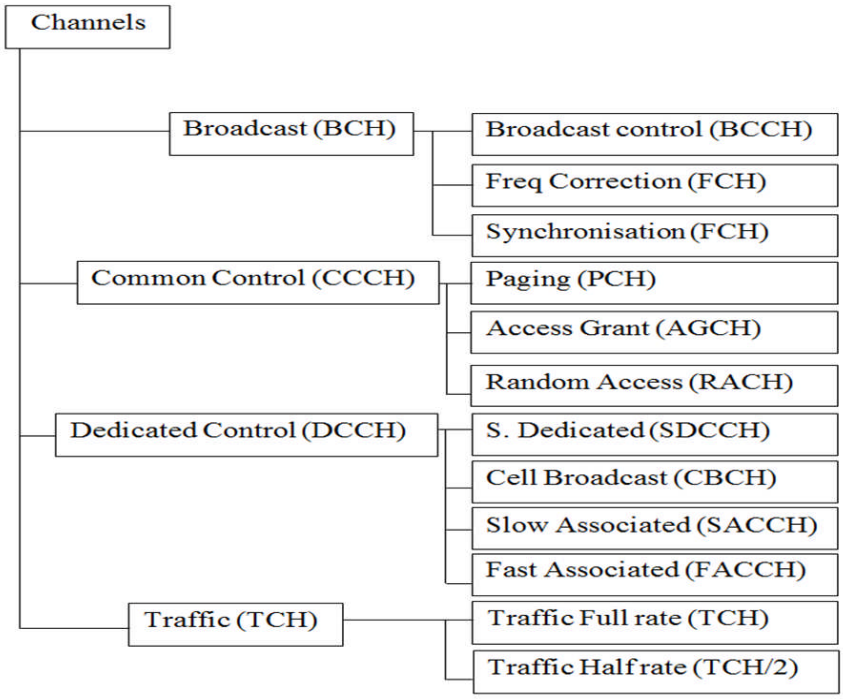

Figure 3.2 Types of GSM logical channel (directly reproduced from [58])

3.4.2 GSM Channel Mapping

As explained above, and also described in [61], the contents of logical channel are transmitted on a physical channel using the method called **Mapping** ("the method of placing logical channel on physical"). Some of the logical channels discussed above require only one time slot to transmit whereas some use more depending on the application and the amount of data needed to be transmitted. For the application that requires one slot, the same time slot is used in the successive TDMA frames. To understand this clearly it is important to know how a frame is structured into multiple frames (multi-frame).

Understanding the basics of the channel mapping is a key requirement to model the MAC layer of the GSM network. In the GSM's channel configuration, a separate frame arrangement is available for traffic and control channels as described below:

- **26-Multiframe**: this consists of 26 TDMA frames with a total period of 120 ms (4.612×26) and mainly carries the traffic channel plus the Slow Associated Control

Channel (SACCH) and Frequency Association Control Channel (FACCH). The control channels are inserted from time to time in between the traffic channel for control signalling transmission during a phone call (data transmission).

- **51-Multiframe**: this consists of 51 TDMA frames with a total period of 234.5 ms and carries the following control channels; FCCH, Synchronisation Channel (SCH), Broadcast Control Channel (BCCH), CCCH, Stand-alone Dedicated Control Channel (SDCCH), SACCH. Therefore, the RACH channel (relevant to this work) which is a class belongs to CCCH is located within the 51-Multiframe.

According to [56] the multi-frames can further be combined to form a super-frame consisting of 51 of the 26 TDMA frames i.e. 1326 frames (having 51 traffic Multi-frames or 26 control Multi-frames) with a duration of 6.12 s or Hyper-frame consisting of 2048 super-frames with a duration of 3 hrs, 28 min and 760 ms.

In general, as illustrated in figure 3.2, GSM assigns a TDMA frame with a specific number (Frame number) that can only be available again after every 3 hrs, 28 min and 760 ms in the hyper-frame arrangement. There is a counter within the hyper-frame where every time slot has a unique sequential number containing frame number and slot number, all these are formed from the multi-frame and the super-frame with the distinctions described above [61]. This arrangement is used to maintain synchronisation of operations like **frequency hopping** and **encryption** that are schedule within the GSM frame structure.

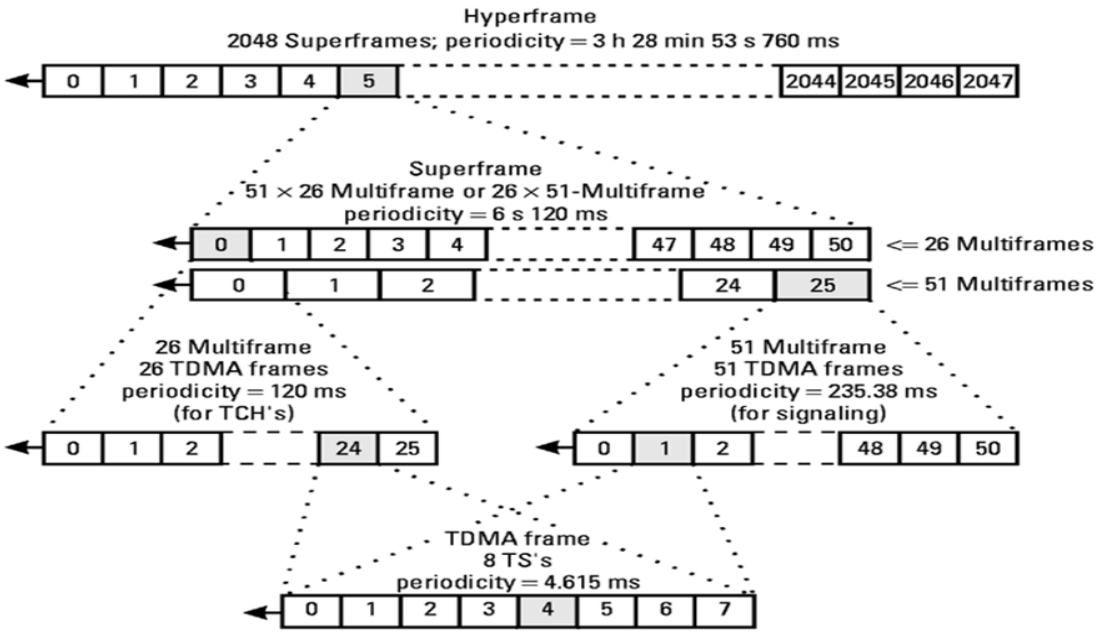

Figure 3.3 Representation of a GSM frame structure (directly reproduced from [61])

There are six different ways to combine the GSM channels, from channel combination 1 to 6 (CC1-CC6) as described in the GSM 05.02 standards. The channel combination to use depends on the cell capacity and some rules regarding the expected traffic [62]. Based on the classification, a cell can be low capacity, medium capacity or high capacity with one, four or 12 transceivers respectively. In each of these classifications different traffic and control channels arrangements are used. Figure 3.4 represents one configuration of the channel arrangement using both 51-multiframe and 26-multiframe. As shown here two slots are used in each frame to carry the control channel and the remaining six slots are used for the traffic channel. The RACH channels (which are our focus) are located in the first slot (TS 0) starting from frame 14 (FN 14).

In our first simulation presented in chapter 5 we consider one of the channel combinations of a cell with medium capacity, where the RACH is located in the first slot of each frame.

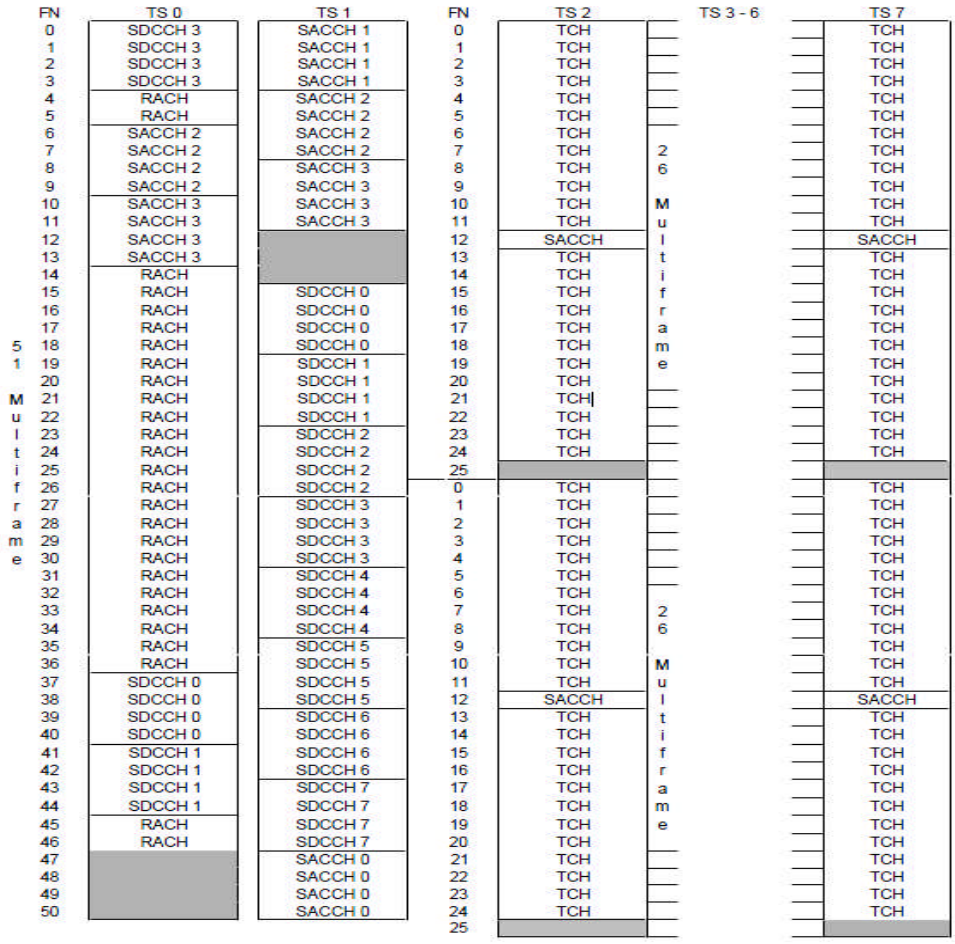

Figure 3.4 Uplink full rate channel configuration (directly reproduced from [61])

3.4.3 LTE Network and Channel Arrangement

LTE uses OFDMA and SC-FDMA as the radio multiple access techniques for the uplink (UL) and downlink (DL) respectively. This is according to the 3GPP design specification to reduce interference due to multiple channel effects and to improve network capacity because subcarriers can be allocated to different users within a transmission interval [63].

Different access techniques are used for the UL and DL since SC-FDMA has the ability to reduce the signal's Peak-to-Average Power Ratio (PAPR), thus providing power efficiency (better battery life) to the User Equipment (UE) [63, 64]. In addition, LTE has the flexibility of supporting both Frequency Division Duplex (FDD) and Time Division Duplex (TDD) to

separate UL and DL traffic which allows it to accommodate various channel bandwidths in the available spectrum. Also LTE gives network operators options to provide different services based on spectrum where it enables operation in scalable bandwidths from 1.4 up to 20 MHz [52]. With this, LTE is capable of supporting peak data rates of up to 100 Mbps on the DL and 50 Mbps on the UL when using the 20 MHz bandwidth with a single transmit antenna at UE and two receive antennas at the eNB. Furthermore, when LTE adopts the Multiple Input Multiple Output (MIMO) system, the 20 MHz can provide user data rates of up to 150 Mbps using 2x2 MIMO, and 300 Mbps with 4x4 MIMO at the DL with UL peak data rate of 75 Mbps [52].

3.4.4 LTE Generic Frame Structure

LTE transmissions are segmented into frames in which all of their timing units are specified as a factor of T_S which has been used to determine the frame time (T_f) as shown below [65]:

$$T_S = \frac{1}{15000} \times 2048 \text{ seconds} \tag{3.1}$$

Therefore the radio frame for UL and DL transmission is:

$$T_f = 307200T_S = 10ms \tag{3.2}$$

Since LTE supports both FDD and TDD the frame structure can be in FDD or TDD format called type 1 or type 2 respectively [66].

The type 1 frame described in Figure 3.5 consists of 10 subframes of duration 1ms with each subframe having 2 slots of 0.5ms. Therefore in total, a type 1 frame consists of 20 slots with slot duration of 0.5ms. On the other hand, the type 2 (TDD format) frame shown in Figure 3.6 consists of two half-frames 5ms long and each half frame consist of five subframes 1ms long. Since the same frequency is used for UL and DL transmission in a TDD system, a large guard band is provided during which a switch between transmission and reception is made.

Hence based on the UL-to-DL switch point periodicity, a half-frame is divided into four subframes and a special subframe or five subframes without the special subframe [65, 67]. For example if a 5ms UL-to-DL switch point periodicity is adopted, the special subframe exists in both halves of the frame. A special subframe exists in the first half-frame only in case of 10ms UL-to-DL switch point periodicity. As shown in Figure 3.6 the guard period is created by splitting the special subframe into three fields [65, 67]: a downlink part (DwPTS), a guard period (GP) and an uplink part (UpPTS). There are seven different configurations of TDD frame structure which allows different mode of operation for the UL and DL transmission with different UL-to-DL switch point periodicity arrangement. The details of LTE frame structure and mode of operation can be found in [65].

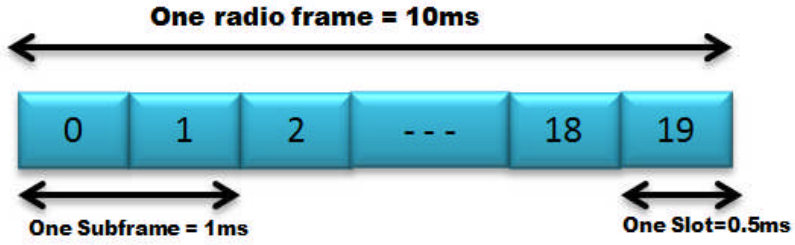

Figure 3.5 FDD frame structure

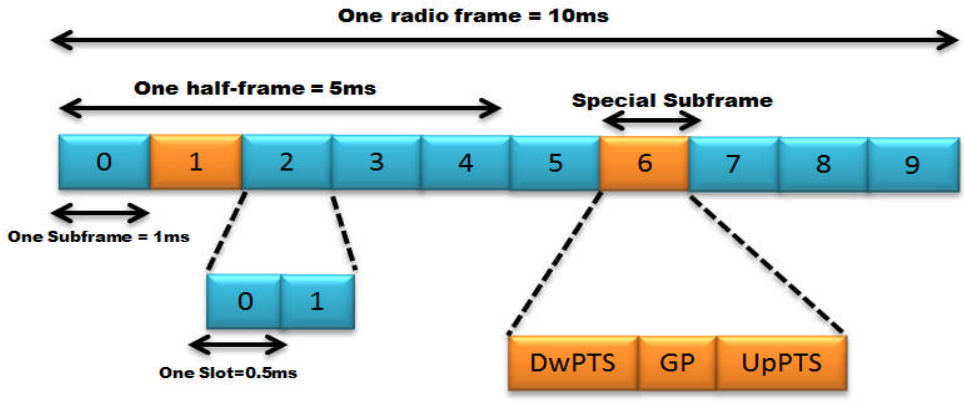

Figure 3.6 TDD frame structure with 5ms switch point periodicity

A resource element represented by one symbol is the smallest modulation structure in LTE which is equal to one 15 kHz subcarrier in the frequency domain. Resource elements are aggregated into Resource Blocks (RB) which is the minimum allocation unit where twelve consecutive subcarriers assemble an RB with a bandwidth of 180 kHz in frequency and 6 to 7 symbols in the time domain [63, 64]. Therefore as shown in Figure 3.7 a single RB is allocated in a slot period and consists of 12 subcarriers in the frequency domain and 7 to 6 symbols (depending on the type of the cyclic prefix used) in the time domain [65].

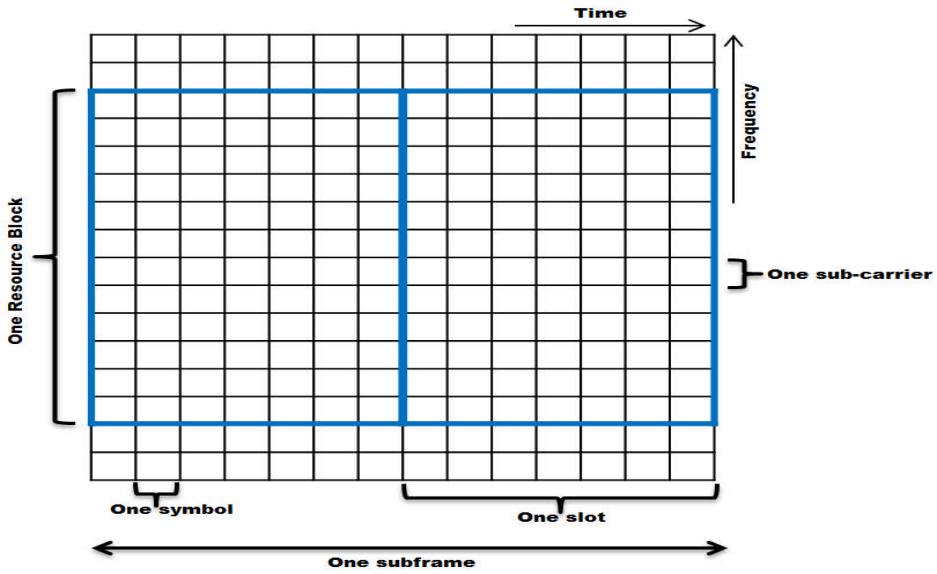

Figure 3.7 LTE resource block structure

3.4.5 LTE Channel Configuration

Channels are used to transport data across the LTE air interface where three different types of channels are engaged. The channels are distinguished based on the kind of information they carry and process and also provide interfaces to the higher layer within the LTE protocol structure (see Figure 3.8) as well as enabling a logical and distinct segregation of the data in transmission [68].

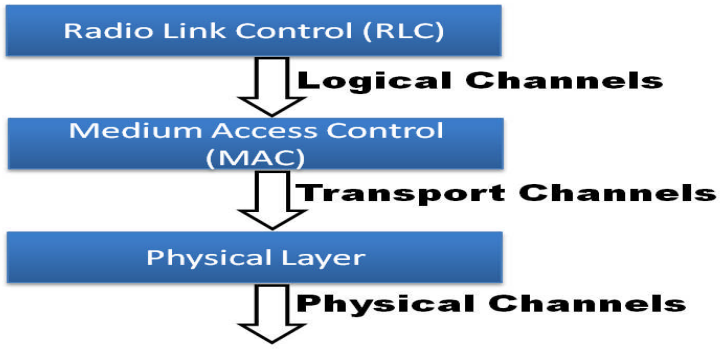

Figure 3.8 LTE channel architecture

The channels are grouped into: **Logical channels**, **Transport channels** and **Physical channels** having different functions and provide services to MAC layer and Physical layer. A summary of the LTE channel classification is presented in Figure 3.9. For the actual transmission over the air, the transport channels are mapped onto the physical channel as shown in Figure 3.10

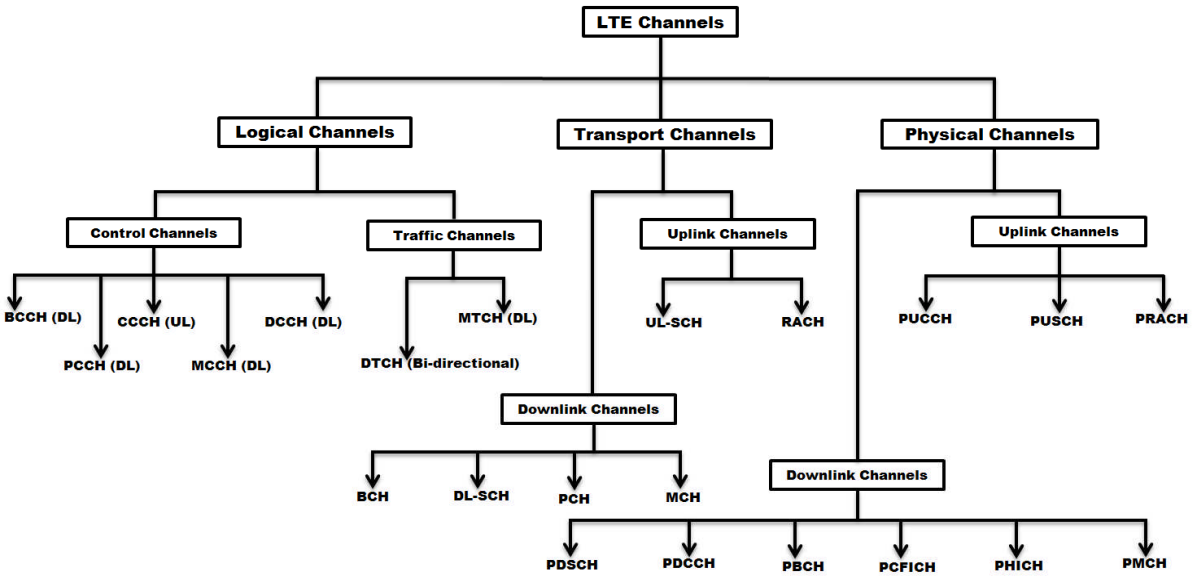

Figure 3.9 Types of LTE channel

The details of the functions and the categories of the above channels will not be provided in this thesis as most of them are not relevant to this work, however the details can be found in [69, 70]. CCCH is an uplink control channel of a logical channel which is used for random

access information by UE without RRC connection, is the relevant channel in this work. Figure 3.10 shows how the CCCH is mapped to its uplink transport channel counterpart RACH via UL-SCH and then finally the RACH is mapped directly to PRACH to transmit the connection request over the air to the eNB. More information on the LTE RACH structure and access procedure is provided in section 3.5.

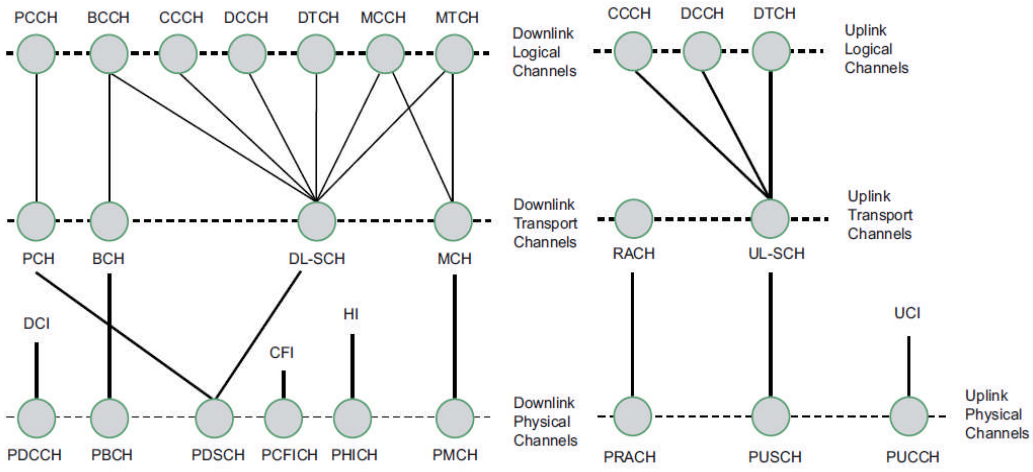

Figure 3.10 LTE channel mapping (directly reproduced from [69])

3.5 Random Access Channel

The RACH is the initial channel through which a user is linked with a cellular network. It is an uplink channel and carries control information to the central entity (BS or eNB) using a random access technique since users are dispersed in a cell and need some initial access to the system. Being the initial channel the RACH is very important and needs to be protected against congestion otherwise the channel will be overloaded and cause a bottleneck which will affect the system performance. Therefore, the focus of this work is on the RACH channel because is the key bottleneck in the support of M2M traffic over a cellular network. Evidence to support this is provided in chapter 5 where the impact of additional M2M traffic on the RACH channel is studied and it is found that, the RACH channel is unstable if supported by the existing protocol. Based on that, we then propose a better scheme that controls the

additional M2M traffic in order to effectively share the RACH channel with existing H2H users.

3.5.1 GSM RACH Structure and Access Procedure

In GSM, RACH is the channel used by MS to request a dedicated channel (SDCCH) and is mainly used for; **call set up**, **authentication**, **location update** and **SMS point-to-point**. In addition, RACH is also used by MS to acknowledge paging from a downlink Paging Channel (PCH) [49, 56]. As described in section 3.4, GSM uses a mixture of FDMA and TDMA where a transmission is based on a slot in a TDMA frame and the RACH uses the 51 multi-frame structure in the uplink. Figure 3.4 illustrates one of the possible configurations in which the RACH is mapped onto TS0. There are up to five different RACH configurations that can be used depending on the GSM cell size.

Since the transmission in GSM is restricted into slots of the TDMA frame and the RACH as the initial channel is accessed using a random access procedure, slotted ALOHA (s-ALOHA) is the scheme used in the RACH access. This is because of its simplicity and ability to handle multiple spatially distributed nodes with bursty data accessing a single channel [71, 72]. Also a transmission in a TDMA system is quantified by a unit called a burst i.e. a TDMA frame (8 time slots) is equal to 8 bursts. Therefore GSM transmission in a time slot is transmitted in bursts [56]. A burst is the actual content of the time slot with a period of approximately 577 μs and it carries in addition to the information bits, other control information like tail bits and even guard period where nothing is transmitted. The modulating bit rate for a GSM carrier is 270.833 kbit/s and this means that, the above time slot of burst will contain 156.25 bits [73]. GSM defined five types of burst, however our interest in this work is the RACH access burst shown in figure 3.11. The access burst is used for mapping RACH information in which it

uses only a part of the length of the slot and the remainder is used for Tail Bits (TB), a synchronisation sequence and Guard Period (GP) [10].

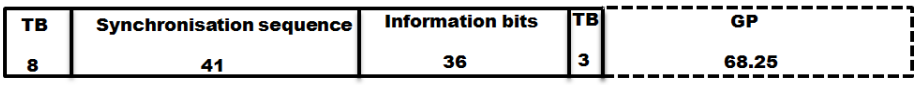

Figure 3.11 GSM RACH access burst

The GSM RACH access procedure is described in [10, 72] where an MS is schedule to transmit in the next RACH slot once a request is generated, and selects a random slot for a possible retransmission in case of collision (when 2 or more MSs send the request in the slot). A cut-off retransmission strategy with fixed uniform back-off [74] is the common collision resolution mechanism applied by the standards. The back-off interval as suggested by the GSM standard is an integer within the range of 3 to 50 RACH slots. A higher interval provides better throughput performance but at the expense of high delay. RACH-throughput performances and access delays using different retransmission intervals are presented and discussed in chapter 5. 14 RACH slots are chosen as the back-off interval with the number of retransmissions limited to 7 in [10].

3.5.2 LTE RACH Structure and Access Procedure

Similar to the GSM, the LTE network is also accessed initially through the RACH in what is called the Radom Access (RA) procedure [75-77]. As mentioned in [34, 78] the RA procedure in LTE is performed for one of the following reasons: (a) initial access of an idle mobile, (b) re-establishing a radio link after failure, (c) handover, (d) out of time-synchronisation downlink data transmission i.e. by paging acknowledgment and (e) out of time-synchronisation uplink data transmission.

In LTE, the RA procedures are of two types; **contention based** and **contention free**. The latter is used by UEs in connecting mode (example of (c) to (e) given above) and will not be

considered in this work. Contention based RA, is applied to UEs in the idle state in order to establish a connection with an eNB (example of (a) and (b) giving above) through messages exchange (summarised in Figure 3.12) using the following four steps [34, 37]:

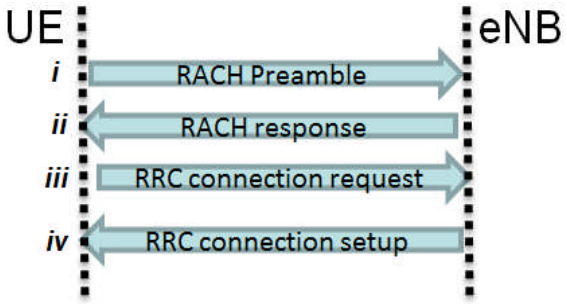

Figure 3.12 Representation of RACH access procedure

i. one out of the available preamble sequences is selected at random by the UE and transmitted on the next RA-slot of the PRACH to the eNB.
ii. the eNB responds to the detected preamble sequence by sending an RA Response (RAR) on the PDSCH.
iii. the UE transmits its resource request using the assigned address received through the RAR in the message of step ii on the PUSCH.
iv. the request sent in step iii will be answered by the eNB here on PDSCH.

Up to 64 preambles are assigned in each cell of LTE, some of which are reserved for the contention-free RA and the remaining are for the contention based RA [37]. Each preamble represents an RA Opportunity (RAO) on a dedicated RA-slot. The RA-slots are the time-frequency resource definitions of the PRACH, which shares resources with PUSCH and PUCCH, located and repeated periodically within the PUSCH region [37, 79]. The PRACH is configured in a conventional LTE frame structure. The number of the RA-slots per frame depends on the PRACH configuration index adopted which also depends on the cell size. In LTE type 1 frame structure, only one PRACH resource can be configured into a sub-frame where the eNB broadcasts the periodicity of the RA-slots that varies between a minimum of 1

RA-slot in every other frame to a maximum of 1 RA-slot per sub-frame [34, 52]. Up to 6 PRACH configurations are presented in [34] as shown in Figure 3.13.

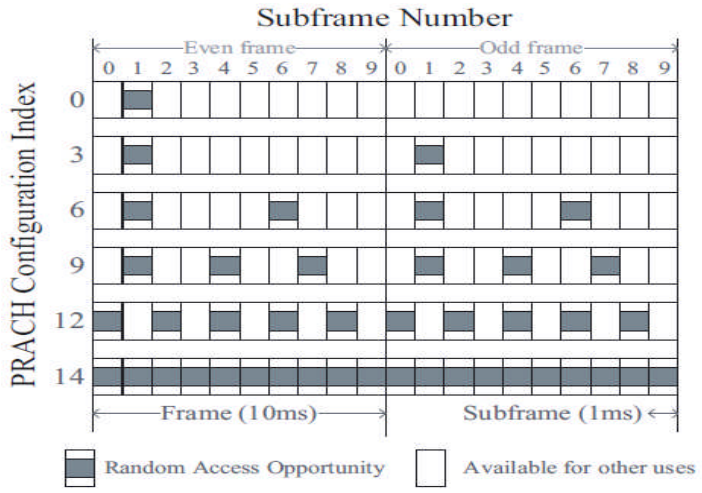

Figure 3.13 Representation of PRACH configuration index (directly reproduced from [34])

Due to the nature of the PRACH arrangement where the RACH requests are restricted to RA-slots, LTE also adopts similar protocol to GSM where s-ALOHA is used to control RACH access. Collisions also occur here when two and above UEs select the same preamble and send in the same RA-slot. A similar collision resolution mechanism to GSM is adopted by LTE as well, where a UE is allowed some certain number of trials to retransmit within a fixed back-off window. As mentioned earlier, a preamble is used as the RAO in LTE which according to [70] is made up of a Zadoff-Chu sequence (root sequence). For the details of how the RACH access of both GSM and LTE is implemented using both existing scheme (SA-RACH) and the proposed scheme (QL-RACH) to control M2M traffic see chapters 5 and 6 respectively.

Figure 3.14 presents four different preamble formats for the LTE FDD frame structure.

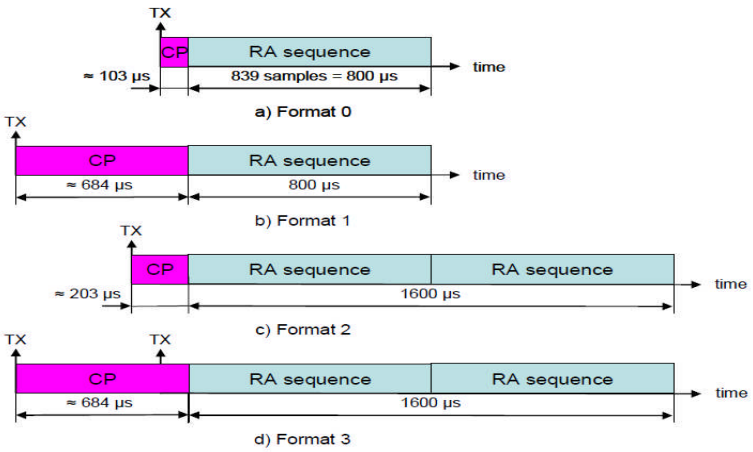

Figure 3.14 Representation of LTE preamble formats for FDD (directly reproduced from [78])

Different preambles formatted with different lengths of Cyclic Prefix (CP) and some with a double preamble sequence are used in the LTE standard. This is to compensate for path loss and UL timing uncertainty. For example the CP length is increased with increase in the cell size and the repeated preamble are used to compensate for path loss [52, 80].

3.6 Summary

This chapter presented a brief overview of the cellular network resources and also emphasized the RACH channel which is relevant in this work. It is clear that the RACH is a very important channel in cellular networks and therefore needs protection from overload else the network will have poor performance due to a bottle neck at the RACH. Structure and access protocol of the RACH has been presented and similarities in GSM and LTE RACH channel have been established.

Chapter 4. System Modelling and Performance Evaluation Method

4.1 Introduction

The purpose of this chapter is to introduce the methods used to develop a MAC layer simulation model for an uplink cellular system architecture using a software simulation tool. This is the main technique used in this work to evaluate the performance of the schemes proposed to support M2M on existing cellular networks.

The importance of using simulation in the realisation of a communication system will be described first in section 4.2 followed by the steps followed to evaluate the system in section 4.3. Section 4.4 outlines the techniques used in checking the accuracy of the method followed in the realisation of the system. Finally communication traffic models will be introduced in section 4.5 and section 4.6 summarises the chapter.

4.2 Communication System Simulation

Any system irrespective of its area of application undergoes various stages of development before finally being put into use. Even though the stages may vary perhaps due to different approaches, the aim remains the same. For example in engineering, a communication system can be viewed to undergo a design stage which needs to be evaluated before the final implementation stage. In some cases an existing system may be altered or modified (in the design stage) in order to realise a new one that has better performance. The work in this thesis can be described as an example of such a case where MAC protocols of an uplink control channel of a cellular network are re-designed to accommodate M2M traffic. It may be argued in the above examples that the design stage is the most important because it determines the

final behaviour of the system; however the evaluation is equally important since it authenticates the design procedure. In developing a communication system, validation of its performance is required where feasible using a combination of the following methods:

i Analytical Models

ii Hardware measurements

iii Simulation Techniques

Analytical models (mathematical analysis) are conventional method used when the system is at its simplified (early) design stage and the models are often based on a simplified model. A popular example is the Erlang-B formula used to determine the call blocking probability for a circuit-switched networks [63]. Modern technological developments in communications are making the system become more complex to the extent that it is difficult to effectively model the system analytically, and a simulation based approach is more appropriate [81]. Therefore the simulation method proved to be the best evaluation option or techniques in this work. The method has been adopted to model the temporal behaviour of the MAC layer of the uplink RACH in a cellular network. Alternatively a hardware prototype of a system may be used to evaluate the communication system by measurement methods. However because of its inflexibility and costly nature it will not be used in this work and therefore the method will not be discussed further.

Simulation has been defined by the famous Industrial Engineer Dr Robert E. Shannon as ‘‘the process of designing a model of real system and conducting experiments with this model for the purpose of either understanding the behaviour of the system or evaluating various strategies (within the limits imposed by a criterion or set of criteria) for the operation of the system’’. From this definition it is clear that the starting point of simulating any system is developing a model that represents the entire or some interesting part of the system that needs

to be tested. The model should be simplified but contains all the necessary system information that is required to develop the simulation model and this can be logical, mathematical or structural [82]. This will enable the designer to predict the system performance before the implementation stage. Good prediction realises a system with better performance and of course the converse will be realised as a result of poor prediction. In this work a model that represents the MAC protocol for RACH access in a cellular network is developed on which the simulation model is built. Another importance of modelling (in the case of modifying an existing system) is to give assurance to the existing system that the modification would be worthy when implemented. For example in the scenario of this work, a cellular network provider will not risk accommodating additional traffic from M2M without any source of confirmation that there will not be negative effect to the existing system.

Discrete event simulation (DES) is the type of simulation used in this work, where events are captured at discrete time intervals, unlike the time-continuous approach in which the events are monitored at all times. The former method is actually used because it is the standard approach and much simpler to implement as well as less time consuming [83]. For example to simulate a communication network it is first considered that users are generated and each user will also generate demand for the network resources that need to be allocated in a control (protocol) manner [84]. All the activities described above (i.e. generating users, users' demands and protocols) can be described as events. An event executes an activity in DES which happens at a specific time.

4.2.1 Monte-Carlo simulation

Most communication simulations are based on a random process and this is due to the nature and the complexity in designing a system that works in all possible situations. Mostly a system is designed (based on prediction) by characterising various required components and

parameters that capture various random scenarios and operating conditions and these are represented by an accurate probabilistic distribution. Based on this it is therefore impossible to analyse our system using a deterministic algorithm but rather a Monte-Carlo simulation [52, 67].

A Monte-Carlo simulation is a method that is based on a probabilistic distribution and depends on repetitive random sampling of the modelled system parameters to generate the statistical results [52]. The accuracy of this type of simulation depends on the number of trials (length of the simulation) and a reasonably large number is required in order to push the system far enough out of the transient phase. Alternatively the accuracy of a result can be verified using what is called ***confidence interval*** [85]**.** Here a researcher constructs the intervals by setting a desired confidence level and observe the performance. The simulation accuracy is determined if the proportion of the intervals that contain the correct result of the experiment match the confidence level. The value of the confidence interval is represented by a percentage and the confidence interval defined within a given range. The higher the number of trials (length of the simulation), the smaller the confidence interval for a desired confidence level. In this work we developed analytical models and compare the results with the simulation models to validate the accuracy of our simulations.

4.2.2 Simulation Software tools

There are a number of software tools available for simulating communication systems. Some are programming packages that can be used for multiple purposes and require prior programming skills. Among which are Pascal, FORTRAN, C, C^{++}, Visual Basic etc. These categories of tools are flexible in such a way that the programmer can model any system of his/her wish. On the other hand some tools are simulation packages with built-in models designed for specific use, example include; $OMNeT^{++}$, OPNET, Network Simulators (NS1,

NS2, NS3) etc. Traditionally, choosing a simulation software tool depends on the complexity of the desired task, in other words different protocol layers (e.g. of the ISO stack) of a communications system have different ideal simulation tools and techniques. For example it is easier to develop a model (non-programming) to model upper layers since multiple users here undergo the same procedure concurrently. On the contrary, to model a physical layer which has different features that require lots of manipulation to develop the algorithm, will preferably be modelled using a programming package that will allow user's flexibility. There is actually a trade-off (in using the different categories) between cost, simplicity, time and accuracy. Where some packages with no programming are easier and quicker to use if contain the desired model, the programming package could be used for almost any modelling at the expense of good programming skills, time and effort [82]. A third category that has both built in model (customise simulator) and programming package that give user a flexibility to develop his/her desired model can be defined. Matlab that is used as a simulation tool in this work is among this category, a user has the option of either choosing the customised simulator (if it exists) or being more flexible by developing the algorithm according to the desired task.

Matlab, an acronym for matrix laboratory is a popular numeric computing environment and high performance-language for technical computing [86]. Matlab is developed by the MATHWORK Inc and was designed to provide easy access to matrix evaluation. It has been modified and extended over a period of time through the help of many users in the industries and universities to an important tool that can be used in various fields of Mathematics, Engineering, and Science etc. [86]. Matlab is similar to C programming with the availability of number of functions and tool boxes as well as user friendly environment. It is adopted as a simulation tool in this work because of the following reasons: First it provides good structure with simple arrangement for designing and simulation of transmission algorithm in wireless

communication system. In addition, the Matrix oriented nature of Matlab allow us to mimic a scenario of different user groups of H2H and M2M where a matrix is used to represent multiple users in each group with their packet generation time, arrival time, transmission time.

4.3 Performance Measures

After describing the methods of simulation used in this work the next is to explain the steps followed to evaluate the system performance by defining some important parameters used. The parameters will be used to quantify our system performance in terms of capacity measurement and system QoS. Normally the schemes used by the system determine the parameters required to measure the system performance. For example a scheme that organises packet transmission will require throughput and delay for the performance measurement. In this thesis, a few important parameters (discussed below) have been selected to measure the system performance.

4.3.1 Offered Traffic

Offered traffic is defined as the average number of users being served by a system over a period of time. In this thesis, offered traffic is defined as the number of RACH request trials in the system in a given time interval. This is similar to the definition given by Harada and Prasad [47] as the total quantity of active packets that include new generated packets and retransmission packets in a given time interval. We introduced **packet** and **retransmission** here because the RACH channel considered in this work is a signalling channel that carries bursts as information and are therefore treated as packets. For the retransmission when an access fails due to a collision, retransmission is allowed and that will add up to the existing offered traffic within the system. Based on trunking theory developed by a Danish

Mathematician, offered traffic can be measured in a unit called Erlang [49], which describes the portion of a line (channel) occupied in time by user(s). For example 0.5E means 50% of the line is occupied at the time. In addition offered traffic has been defined by John Little in what is called Little's Law and is expressed mathematically as follows [49, 87]:

$$G_U = \lambda\tau, \tag{4.1}$$

where G_U is the offered traffic level of an individual user, λ is the average number of service requests per unit time and τ is the average service duration which is normally represented by its inverse, called the departure rate (μ).

For a system with N users the total offered traffic is the product of the individual offered traffic and the number of users :

$$G = G_U \times N \tag{4.2}$$

and from 4.1 and 4.2 the total offered traffic G is:

$$G = \lambda\tau N \tag{4.3}$$

Finally offered traffic can also be measured in bits/sec which is another way of measuring data related services. Therefore both Erlang and bits/sec can be used as unit (depending on the area of application) in communication engineering. In general Erlang is more of a fundamental measure of a system to understand the proportion of its capacity in use. On the other hand, bits/sec gives more of a specific measurement of a system, i.e. it gives absolute information about the system's data rate. In this work Erlang is preferred to the bits/sec since we are not dealing with the useful data (signalling information) and our focussed in on measuring the channel usage.

4.3.2 Throughput

Throughput is a very important parameter that is commonly used to measure the performance of a system that offers data services. It has been defined as the successful information that has been delivered over a certain link in a communications system [65]. In this thesis, the throughput is measured based on the successful RACH request send to the BS or eNB by the nodes or users in the system. Ideally throughput obtained should be equal to the channel capacity; however this is only achievable in perfect scheduling protocol (not in a random access used in this work) where overhead is not considered. In this thesis, a single RACH channel is considered to examine the throughput performance where a maximum of 1E is the channel capacity and the throughput is calculated as follows:

$$S(E) = \frac{R_S T_{trx}}{t_{end}} \tag{4.4}$$

S(E) is throughput in Erlang, *Rs* is the number of successful RACH requests which are defined in bits since it is a signalling message, T_{trx} is the transmission time of each packet (burst of the controlled signalling message). In this work, all packets have the same transmission time because they have the same length and the transmission time is calculated from the packet length as shown in equation 4.5, t_{end} is the period over which the throughput is calculated which is the simulation time in this work.

$$T_{trx} = \frac{l}{C} \tag{4.5}$$

l is the packet length in bits and *C* is the data rate in bits/sec.

Similar to offered traffic, throughput can also be obtained in bits/sec as follows:

$$S\left(\frac{bits}{sec}\right) = \frac{R_S l}{t_{end}} \tag{4.6}$$

4.3.3 Delay

End-to-end delay is a time dependent important parameter that is also use in data communication system to evaluate the time from which a packet is generated until it is successfully received. On the other hand transmission delay can also be evaluated by considering just the time between the initial transmission start and the final transmission (successful) end time. Based on the fact that, delay values in a simulation are collected over a long period of time after the initialisation, end-to-end delay is calculated as an average delay of the individual packets successfully transmitted over considered period of time. Therefore if t_s refers to time of successful RACH request transmission and t_g as the time of RACH request generation then the average delay D of the i^{th} delay (in the range of the values of delays) is expressed as:

$$D = \frac{1}{L}\sum_{i=1}^{L}\left(t_s - t_g\right) \tag{4.7}$$

L is the number of successful transmission from which the range of the delays is calculated.

4.3.4 Blocking Probability

Grade of Service (GOS) has been traditionally associated with tele-traffic engineering which was originally developed in the design of telephone networks but it is also applicable to cellular networks [88]. A GOS defines the probability that all servers will be busy when the call attempt is made and can be described using probability of blocking and dropping which are the probability of an incoming call (service request) being disallowed and the on-going being disturbed respectively [89]. The interest of this work is the Probability of Blocking (P_B) and the blocking here is not referring to the call directly but it is use to define the RACH access failure. The request is said to be blocked when collisions occur and the number of allowed retransmissions are exhausted. A GOS is used by communication engineers to design

and dimension a communication system with the maximum usage demand. See [90, 91] for more details and some numerical examples on GOS.

The P_B is described mathematically by the following equation:

$$P_B = \frac{R_b}{(R_b+R_s)} \tag{4.8}$$

Alternatively, P_B can be expressed as;

$$P_B = 1 - (\frac{S(E)}{T_L}) \tag{4.9}$$

where R_b and T_L are the number of blocked RACH requests and generated traffic load respectively.

4.4 Validation of Results

A very important and necessary step in any system design is providing reliable way(s) of checking the accuracy of the method followed to realise the system. In a situation where the developed model has preceding work of similar or the same system that has been proved to be from reliable source, the validation is done by comparing the former with the latter. In addition some models can be described mathematically; therefore approach to validation is to develop an analytical model as a base-line (standard) of the realised system performance. Matching the simulation results of the research work with the developed analytical model will provide confidence on the accuracy of the models and performance evaluation.

In this work both methods are used in validating the simulation results of the realised models. The model of a cellular RACH access scheme is developed and compared with the model provided in [10]. This is because the authors investigate RACH channel performance of the s-ALOHA random access protocol using existing GSM standards. Since our intention is to use

existing cellular network we therefore adopt similar approach to realise the effect of additional M2M traffic on the existing GSM network. This can be extended for other cellular standards since they have similar RACH structure and access protocol. In addition to the simulation method, a comprehensive analytical model that predicts the performance of s-ALOHA, QL-RACH and FB-QL-RACH schemes have been developed to validate the simulation scenario in chapters 5, 6 and 7.

4.5 Traffic Models

It was made clear in section 4.2 that building a simulation model for any system requires proper representation of the target by a good model. Traffic modelling is a very important element in designing and analysing communication network. It provides an ability to capture the statistical characteristics of the traffic behaviour and enables predicting the effect of any change that might be required in realising the network [83]. In order to properly evaluate the performance of MAC of cellular network designed in this work to support M2M traffic an accurate traffic modelling is necessary. This is achieved as suggested in [92] by capturing (in the model) the exact features of the actual traffic.

Traffic modelling has been originated based on the traditional telephone networks which is a fixed network (i.e. wire line network) [93]. This has been modelled by Agner K. Erlang based on queuing and prediction of allowed probability of blocking to dimension the network. This is expressed mathematically by what is called Erlang B formula (also called Erlang loss) as shown below [14, 93]:

$$P_b = \frac{\frac{G^m}{m!}}{\sum_{i=0}^{m} \frac{G^i}{i!}} \tag{4.10}$$

where P_b here is the predicted allowed blocking probability (which the system can tolerate) m is the number of servers (e.g. telephone lines). G as defined in equation 4.3 is a function of λ which is the average call/packet arrivals in a given time using predicted traffic distribution. In this work λ refers to the average RACH request arrivals in a given period of time.

In the communication area different types of traffic distributions are required to model arrival of information. The nature of the application of the information determines the traffic model to be used. We based our model on this fact, and due to wide applications of M2M, different traffic models [84] may be relevant in this thesis. However as a first step in this work a common traffic model is considered.

A Poisson distribution model is one of the most important models used in queuing theory and has been widely applied to model and analyse the sequence of times at which the telephone calls are originated in the traditional telephone network traffic. It is characterised in [84, 92] as a renewal process having exponential inter-arrival time and also assumed to have the following characteristics;

i Infinite number of sources

ii Random traffic arrival pattern

Even though the traffic modelling, especially the Poisson distribution, was originated to model fixed networks it is being used as well in modelling cellular networks. In this work a mix of H2H and M2M traffic is considered and various different applications of M2M will have different traffic distributions. In the first part of the work it is assumed that both H2H and M2M traffic is Poisson. This is because Poisson is well established as an accurate model of the call arrival process at the telephone exchange/eNB which resembles typical human behaviour. On the other hand, to consider different M2M application scenarios, a periodic traffic distribution is used to model the M2M traffic as presented in Chapter 7. The

probability of either call or packet arrivals using a Poisson distribution as presented in [47, 94] is as follows:

$$P(x\ arrival\ in\ a\ number\ of\ trials) = \frac{\lambda^x e^{-\lambda}}{x!} \tag{4.11}$$

Since the average arrival rate is Poisson distribution the inter arrival time (*t_inter*) between the successful calls or RACH requests will be defined by the following density and probability distribution function as follows:

- ✓ the cumulative density function (CDF) is define when the inter arrival exists i.e. ***t_inter*** $\geq$ **0**

$$F(t) = P(t_{inter} \leq t) = 1 - P(t_{inter} > t) \quad \text{for} \quad (x \geq 0) \tag{4.12}$$

if ***t_inter*** > ***t*** there will be no other calls or packet arrivals within **t** time and hence

$$1 - P(t_{inter} > t) = 1 - P(zero\ arrival\ in\ t\ time) \tag{4.13}$$

Therefore from 4.12;

$$F(t) = 1 - \frac{(\lambda t)^0 e^{-\lambda t}}{0!} = 1 - e^{-\lambda t} \qquad for\ t \geq 0 \tag{4.14}$$

- ✓ the probability function is obtained by taking the derivative of 4.14 and this is as follows;

$$f(t) = F'(t) = \lambda e^{-\lambda t} \qquad for\ t \geq 0 \tag{4.15}$$

Therefore the inter arrival time between successive calls or packet arrival of a Poisson distribution is an exponential.

4.6 Summary

An overview communication simulation system has been discussed in this chapter. Discrete event type and Monte-Carlo methods are used in this work and have been discussed. Various types of simulation software tools were briefly introduced with emphasis on Matlab (the simulation software used in this work). Different parameters used in measuring our system performance were also discussed. Finally the Poisson distribution as a common model for the call arrival process at the telephone exchange which is widely used to model typical human behaviour was introduced. It was also mentioned that a periodic traffic model used for some M2M application scenarios is considered later in this thesis.

Chapter 5 Slotted ALOHA for RACH access

5.1 Introduction

This chapter introduces and motivates our proposed schemes presented in chapter 6 to 8, by showing the impact of additional M2M traffic in existing cellular networks. The chapter starts in section 5.2 with an introduction to the basic of two ALOHA schemes (slotted and Pure ALOHA) and comparison of their performances. Section 5.3 presents the performance of SA-RACH for a single user group and analysis based on which the cause of RACH instability as the impact of additional M2M traffic is established. The performance (throughput, delay and blocking probability) of the SA-RACH scheme is presented. Different results are compared using: a) the retransmission limit and b) the back-off interval window, as the two variable parameters. Section 5.4 presents an analytical model to predict the RACH-throughput performance, where new equations are developed using the SA-RACH scheme with retransmissions. Finally the chapter is summarised in section 5.5.

5.2 The ALOHA schemes

This section provides a general description of ALOHA (p-ALOHA and s-ALOHA) with more emphasis on the s-ALOHA scheme because of its relevance in this work.

ALOHA was the first multiple-access data communications network in the world which was developed and named after the system (ALOHA) in 1979 at the University of Hawaii by Norman Abramson and his colleagues in their effort to connect users on remote islands to the main system (computer) [95]. The system uses a simple MAC strategy that allows users to share a single channel without involving any central entity to control the sharing [54]. Their

work has been extended by many researchers since then and the protocol is very useful in today's wireless communication. There are two standard methods of ALOHA: **p-ALOHA** and **s-ALOHA**.

1) **p-ALOHA**: in p-ALOHA users transmit on a shared channel whenever they have packets (data) to send. There is no coordination with other users sharing the channel and hence a high chance of collision and consequently data loss is expected. To provide reliability, positive acknowledgement from each receiver is often sent back to the transmitter for all correctly received packets. On the other hand if the transmitter waits for a specified time period without receiving an acknowledgment from the receiver it assumes that the packet is lost and retransmits, this will keep repeating until a packet is successfully transmitted or for a maximum number of trials.

   The efficiency of p-ALOHA is briefly analysed by looking at the fraction of transmitted packets that escape collision. The same concept will be extended to analyse s- ALOHA which is of interest in this work. Both analyses are based on the standard analysis where a packet generation follows a Poisson distribution.

To analyse the efficiency of p-ALOHA we make the following assumptions with regards to users' nature in number, arrivals and packet lengths [47, 95]:

- All packets have the same length and the same time of transmission (*t_trans*).
- Each user (or node) has at most one packet ready to transmit at any moment.
- There is a large number of users (nodes) ***n*** approaching infinity.
- All lost packets are as a result of collision.
- Packet generation follows a Poisson distribution.
- Collisions occur when any part of two packets overlap and this results in loss of both.

Since packet generation is assumed to follow a Poisson distribution then the probability of generating ***n*** packets in the period *t_gen* is similar to what is presented in (4.11) as follows:

$$P_n(t_{-gen}) = \frac{e^{-\lambda_{t-gen}}(\lambda_{t-gen})^n}{n!} \quad , \tag{5.1}$$

where λ is the expected number of packets to be generated in a given time

If *G* is the system's offered traffic then as defined in chapter 4;

$$G = \lambda t_{trans} \tag{5.2}$$

Figure 5.1 represents p-ALOHA where three users are transmitting on a shared channel. User 1 starts transmission at time t_o and takes *t_trans* time to finish the transmission, therefore user 1 will finish transmission at t_o+ *t_trans*. Now it is clear from the figure that a collision will occur if another user (users 2 and 3 in this case) transmits between t_o-*t_trans* to t_o+ *t_trans*. In other words to transmit a packet successfully no other packet should be generated within what is termed in the figure as the vulnerable time which is 2 x *t_trans* (i.e. 2 times the duration of the transmission).

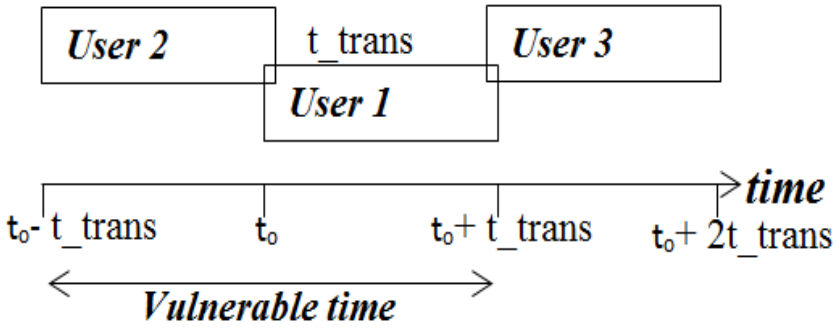

Figure 5.1 Representation of p-ALOHA showing vulnerable period of transmission

Therefore the probability of successful transmission in p-ALOHA is equal to the probability of no packets being generated within the period 2 x *t_trans.* From equations (5.1) and (5.2) this can be presented as:

$$P_{success} = P_0(2t_{trans}) = \frac{e^{-2\lambda t\_trans}(2\lambda t\_trans)^0}{0!} = e^{-2G} \quad (5.3)$$

The throughput of p-ALOHA ($S_{p\text{-}ALOHA}$) is the successful fraction of the traffic (packets transmitted) offered to the channel:

$$S_{p-ALOHA} = GP_{success} = Ge^{-2G} \quad (5.4)$$

To obtain the value of offered traffic (G) at which p-ALOHA offers maximum throughput we differentiate equation (5.4) with respect to G and equate to zero, this occurs when *G=0.5*.

Therefore substituting for G in (5.4) gives the maximum throughput as:

$$S_{p-ALOHA}{}^{*} = 0.5e^{-1} \approx 0.184 \quad (5.5)$$

The above analysis shows that the best channel utilisation of p-ALOHA is 18% which is quite low and not encouraging. This was the motivation for s-ALOHA which was introduced to improve the channel utilisation as described below [95, 96].

2) **s-ALOHA:** It was clear in the above analysis that p-ALOHA has a low efficiency and high risk of collisions because the probability of interrupting other users' transmissions is high. Lawrence Robert was able to solve this problem in 1972 by reducing the vulnerable period shown in figure 5.1 which doubled the capacity of p-ALOHA [95]. Here the transmission time is divided into discrete time intervals of equal duration (corresponding to a packet length) called slots and a user transmits only at the beginning of a slot. Therefore contrary to the p-ALOHA protocol, a user with a packet to transmit has to wait until the beginning of a slot. This process presented in figure 5.2.

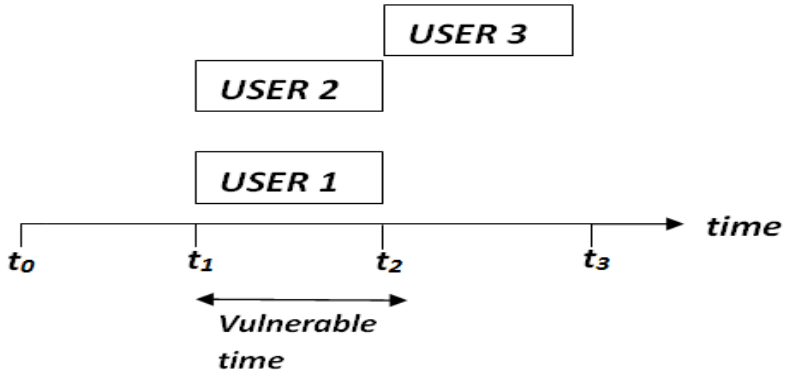

Figure 5.2 Representation of s-ALOHA showing the reduced vulnerable period

Looking at figure 5.2 if a packet is generated by a user before beginning of time slot the transmission will be delayed. For example users 1 and 2 must have generated their packet before t_1 and delayed their transmission until the beginning of t_1. Also user 3 generated its packet after t_1 and starts to transmit at the beginning of t_2. As shown in the figure user 1 and user 2 will collide while user 3 will have a successful transmission. If the time interval is set equal to the transmission time (the packet length in our model) then as shown in figure 5.2 the vulnerable period of s-ALOHA is now only halved that of the p-ALOHA i.e. *t_trans*.

Therefore analytically the probability of successful packet transmission using the s-ALOHA protocol is equal to the probability that no any other packet is generated within the period when another packet is already generated. This is expressed mathematically from equation 5.1 as:

$$P_{success} = P_0(t_{trans}) = \frac{e^{-\lambda t\_trans}(\lambda t\_trans)^0}{0!} = e^{-G} \tag{5.6}$$

The throughput is obtained as;

$$S_{s-ALOHA} = Ge^{-G} \tag{5.7}$$

Similar to the analysis of p-ALOHA, the maximum throughput for s-ALOHA is obtained at G=1, and is;

$$S_{s-ALOHA}{}^{*} = e^{-1} \approx 0.368 \quad (5.8)$$

It is therefore clear that the channel throughput of s-ALOHA doubled that of the p-ALOHA but of course at the expense of design complexity. The performance of both p-ALOHA and s-ALOHA are compared in figure 5.3.

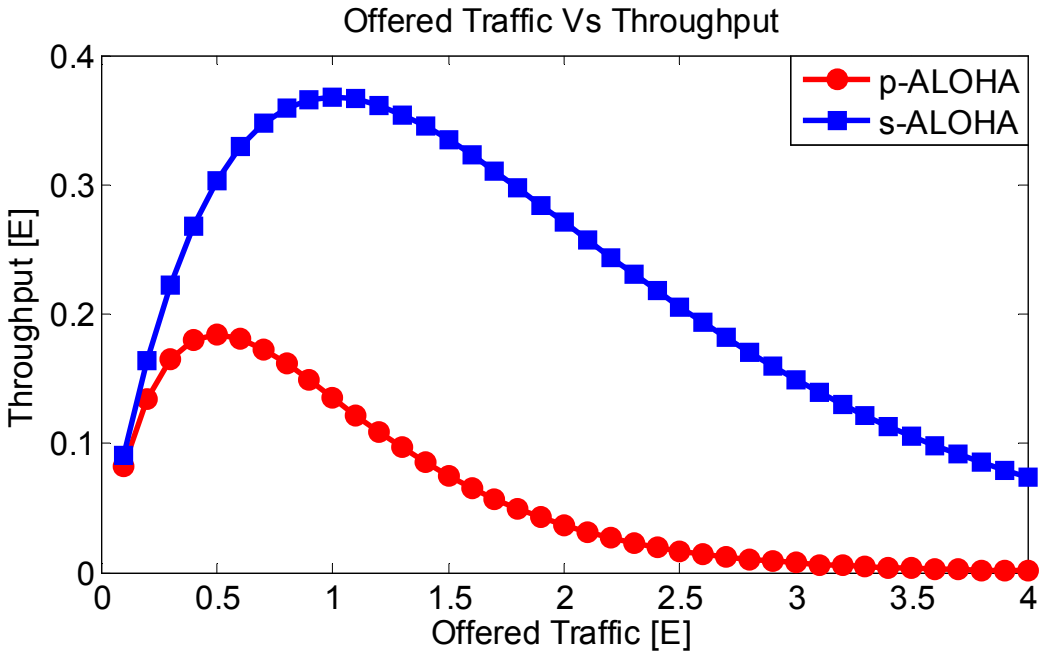

Figure 5.3 Channel throughput of p-ALOHA and s-ALOHA protocols

As shown in the analysis, it is also clear from the graph of figure 5.3 that p-ALOHA peaks at an offered traffic of 0.5 Erlang with a throughput of about 18%. Also it is shown in the figure how s-ALOHA doubled that performance to about 37% at peak offered traffic level of 1 Erlang. On the other hand for both protocols it is shown on the graph how their channel utilisation decreases exponentially with the increase in the number of attempts from the users (offered traffic) after reaching their peak levels. This is as a result of having more users attempting to transmit at the same vulnerable time and results in more collisions than successful transmissions.

It is obvious that the probability of success presented in equations (5.3) and (5.6) for p-ALOHA and s-ALOHA respectively is nothing but the probability of users avoiding

collision. To appreciate the above analysis we can find the possible number of collisions at any given attempt as expressed in the following two equations:

$$P_{c-p-ALOHA} = 1 - e^{-2G} \tag{5.9}$$

$$P_{c-s-ALOHA} = 1 - e^{-G} \tag{5.10}$$

where $P_{c-p-ALOHA}$ and $P_{c-s-ALOHA}$ are the probabilities of collision in p-ALOHA and s-ALOHA respectively and are presented in the graph of figure 5.4.

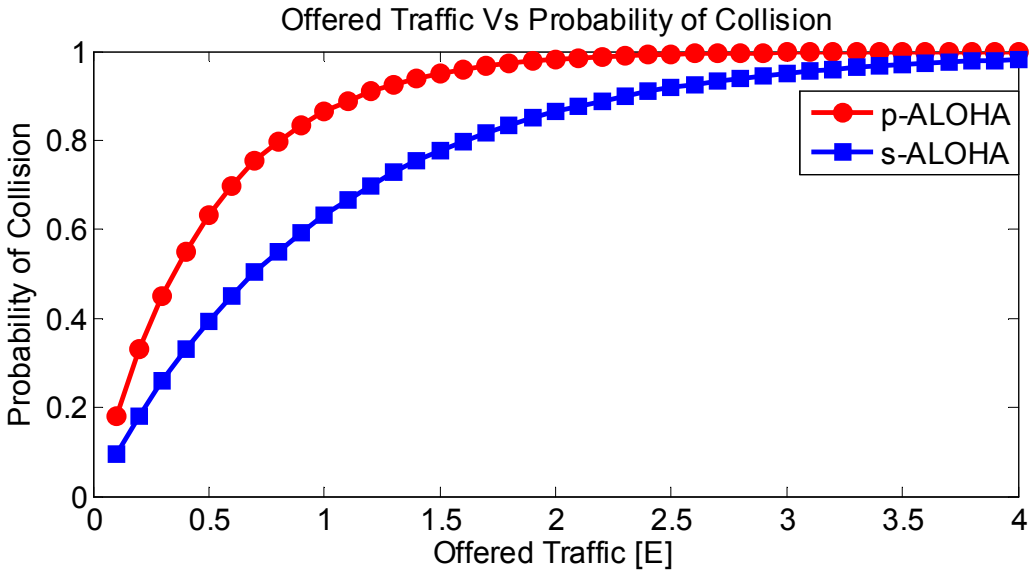

Figure 5.4 Collision probabilities of p-ALOHA and s-ALOHA protocols

The graph above shows how the collision probabilities of the two protocols increase exponentially with increase in the channel traffic. Of course p-ALOHA has a higher collision possibility.

In summary, p-ALOHA is a simple protocol (just transmit whenever you wish) but has a very low channel throughput while s-ALOHA improve the performance with up to double the channel throughput using a more complex protocol. However the increase in the efficiency justifies the trade-off and this made s-ALOHA more useful than p-ALOHA in a system where time synchronisation is straightforward. s-ALOHA protocols were not popular since

there realisation in the early ‘70s until when the need to share a channel among multiple users of the internet access over the cable arose [95]. Since then, the protocol has been widely used in the field of communication engineering and it is relevant in our work.

5.3 Impact of M2M on existing cellular users

It has been explained in chapter 3 that s-ALOHA is the protocol used to access the RACH in existing cellular systems, and this is because of the nature of the RACH channel in which the access is restricted to slots. Also, s-ALOHA has the ability to handle multiple spatially distributed nodes accessing a single channel without involving a central entity. In this work, we look at the possibility of supporting M2M traffic in the existing cellular network. We start by studying s-ALOHA as an existing RACH access protocol, analysing the effect of additional M2M traffic. The issue of RACH instability when additional M2M traffic is supported using the s-ALOHA protocol is established.

5.3.1 SA-RACH scheme with retransmissions and RACH instability

The RACH access using the s-ALOHA protocol is called SA-RACH in this work. Due to the nature of the protocol, collisions during the RACH access are unavoidable. This leads to the users’ requests being lost, causing potentially poor throughput performance at high offered traffic loads. In order to maximise the chance of the RACH request getting through, a cellular system allows a certain number of retransmissions of the requests following collision. The retransmissions help to reduce the possibility of blocking a user out of the RACH contest by providing more chances. However, to make the system stable the retransmission procedure needs to be controlled to avoid more collision due to the same users retransmitting in the same slot. Figure 5.5 illustrates a simple flowchart of the RACH request process of a cellular system.

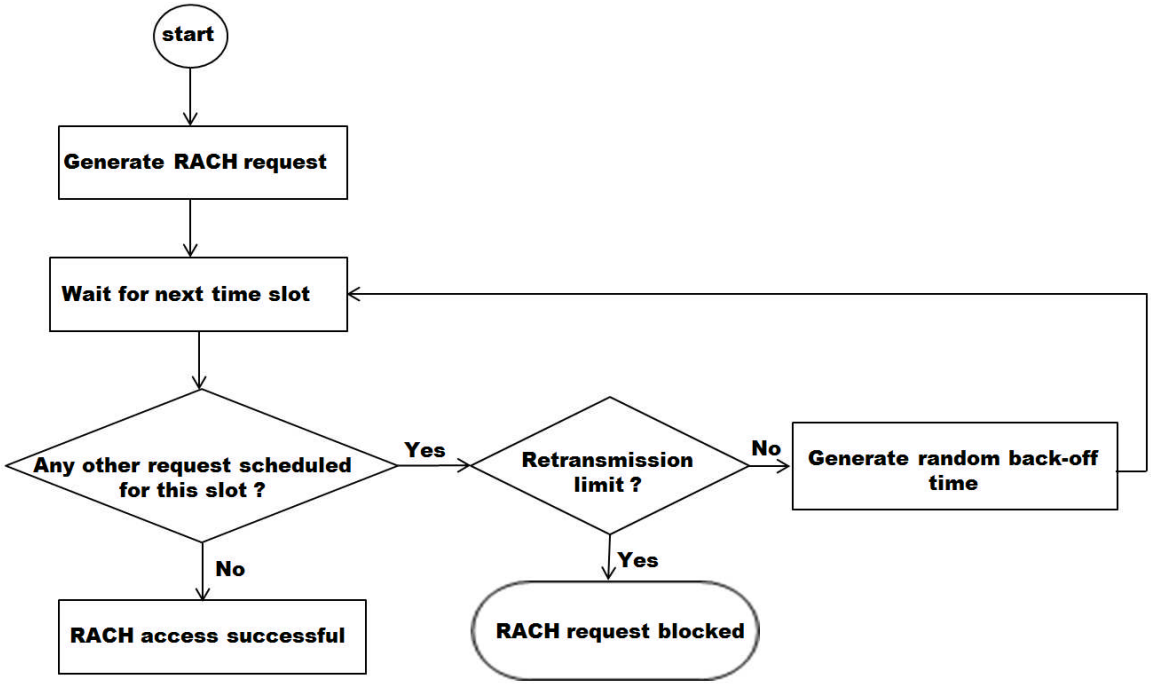

Figure 5.5 RACH request process of a cellular system

A cut-off retransmission strategy with fixed back-off [74] is the common collision resolution mechanism applied by cellular standards. For example as presented in [10], the back-off interval suggested by the GSM standard is an integer within the range of 3 to 50 RACH slots with the number of retransmissions limited to 7. We have used the GSM standard parameters shown in Table 5.1 to demonstrate the performance of RACH access using different settings of retransmission interval widths and allowable number of retransmissions.

Table 5.1 Simulation parameters based on GSM standards

Parameter	Value
Transmission data rate	271 kbps
Data packet length	156.25 bits
slot period	5.7657e-004 s
Conventional frame period	0.0046 s
Simulation length	≥ 150000 successful RACH requests
Retransmission limit	7
Back-off window	14 RACH slots

Figure 5.6 shows the RACH-throughput performance at the following range of number of retransmissions [0,1,2,4 and 7] with the back-off window fixed at 14 time slots within which a user selects a slot at random. The results show that the highest maximum number of retransmissions produced better throughput performance at the lower generated traffic, where the throughput increases with increasing traffic up to some point of generated traffic and then starts to drop. At the higher generated traffic levels the lower the maximum number of retransmissions the better the throughput performance. Therefore a maximum number of retransmissions of 7 at lower traffic levels will provide a high chance of transmissions getting through with almost negligible probability of blocking. On the other hand the maximum number of retransmissions is not recommended at higher generated traffic levels.

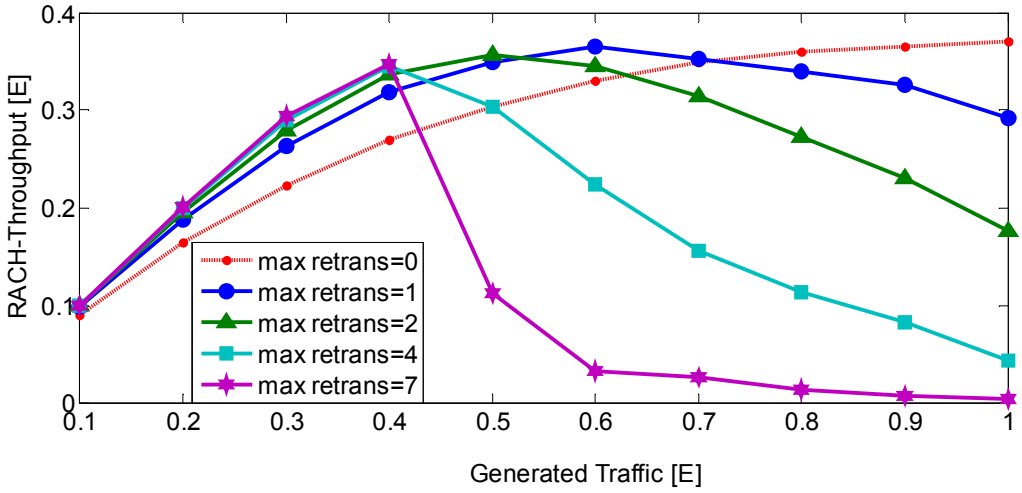

Figure 5.6 RACH-throughput against generated traffic at different maximum number of retransmission with fix retransmission interval of 14

The reason for the observed performance depends on the generated traffic and s-ALOHA capacity. At lower generated traffic (within the channel capacity) the system can accept more retransmission from the traffic. However, when the traffic exceeds s-ALOHA capacity (at higher generated traffic) higher numbers of retransmissions means injecting more traffic to the system which results in more collisions.

Average end-to-end delay is another parameter used here to describe the behaviour of users' RACH access at different allowed maximum numbers of retransmissions. This is presented in figure 5.7. As shown, the delay changes (builds up) when retransmission is introduced where it increases with increases in the generated traffic level. For example when there are no retransmissions (i.e. at max retrans=0) the delay is level throughout the range of generated traffic which shows that there is no queue building up in the transmission process by the users' requests. However when retransmissions are allowed, the delay increases with the increase in the maximum number of retransmissions, where a maximum of 7 retransmissions is the highest.

Therefore, we can see there is a trade-off between the RACH-throughput and the average end-to-end delay with regards to the maximum number of retransmissions allowed for the user trying to get access to the RACH.

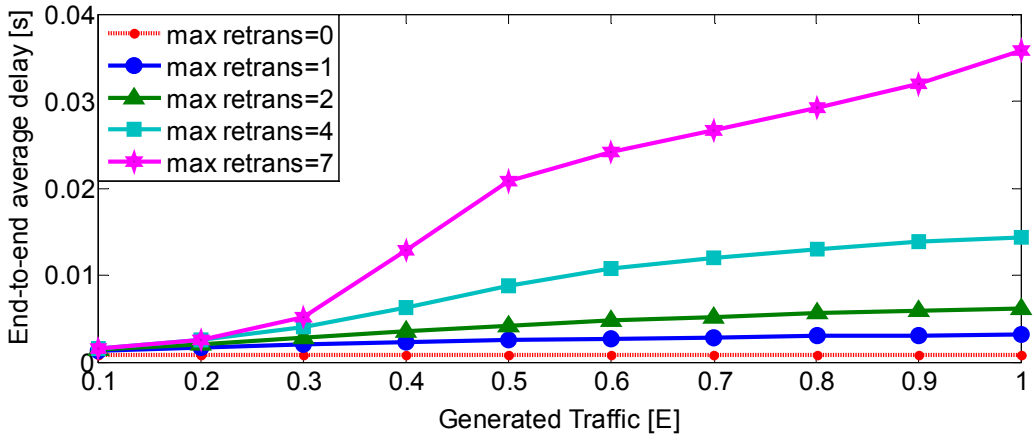

Figure 5.7 Average end-to-end delay against generated traffic at different maximum number of retransmission with fix retransmission interval of 14

Figure 5.8 shows the blocking probability which agrees with the behaviour of the throughput curves explain above. As shown, at low generated traffic levels, the highest maximum number of retransmissions has significantly reduced the blocking. However at the higher traffic level the highest maximum number of retransmission increases the probability of blocking. This is because at low traffic since the system is operating within the s-ALOHA capacity, users' requests get through within the allowed maximum retransmission limit. Of course as expected, the higher the maximum number of retransmission the lower the blocking. However, at the higher generated traffic level, when the system is pushed beyond s-ALOHA capacity, more collisions occur here which subjects more users to retransmission. A higher number of retransmissions therefore increase the probability of collision which as a result will cause more blocking.

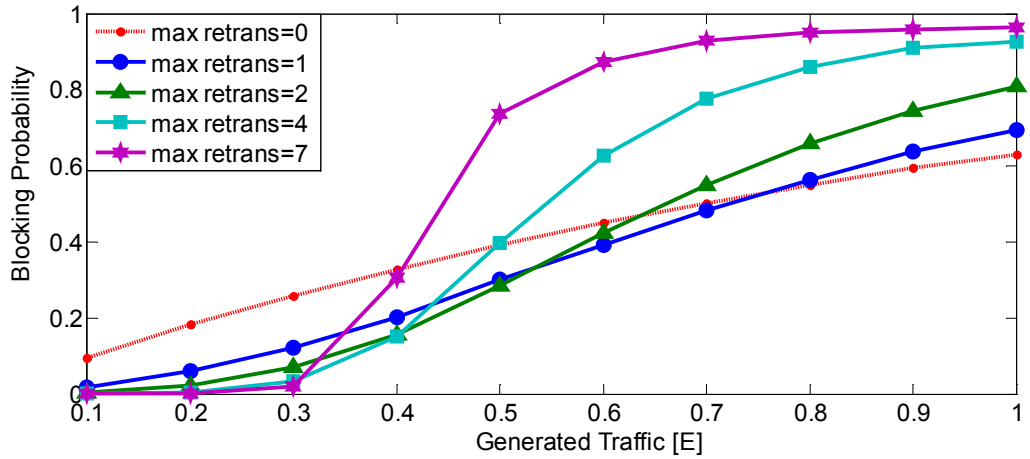

Figure 5.8 Blocking probability against generated traffic at different maximum number of retransmission with fix retransmission interval of 14

The retransmission interval width is the second parameter that can be used to determine the retransmission process of RACH access. A high interval reduces the possibility of more than one user retransmitting in the same time slot that could cause another collision. As shown in figure 5.9, at lower generated traffic levels the width of the retransmission interval does not matter since fewer users are engaged in the retransmission. However, at higher generated traffic levels, the possibility of collision due to high contention level has increased and therefore more users are engaged in the retransmission. As shown in figure 5.9, the larger the interval the better the throughput performance but at the expense of end-to-end delay.

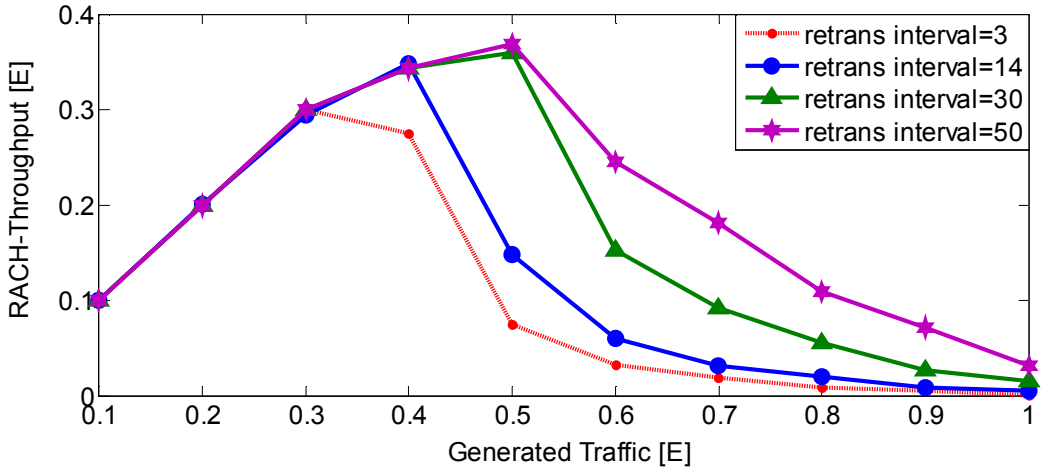

Figure 5.9 RACH-throughput against generated traffic at different maximum retransmission interval width and maximum number of retransmission fixed at 7

Figure 5.10 represents the average end-to-end delay at different retransmission interval widths. As shown, the delay builds up with increase in the retransmission interval which also explains the trade-off between the RACH-throughput and the delay. For example a high interval width is not required at low generated traffic levels since it will just delay the access and give the same RACH-throughput performance. On the other hand sometimes a little increase in the delay is worth taking provided there is a better throughput performance.

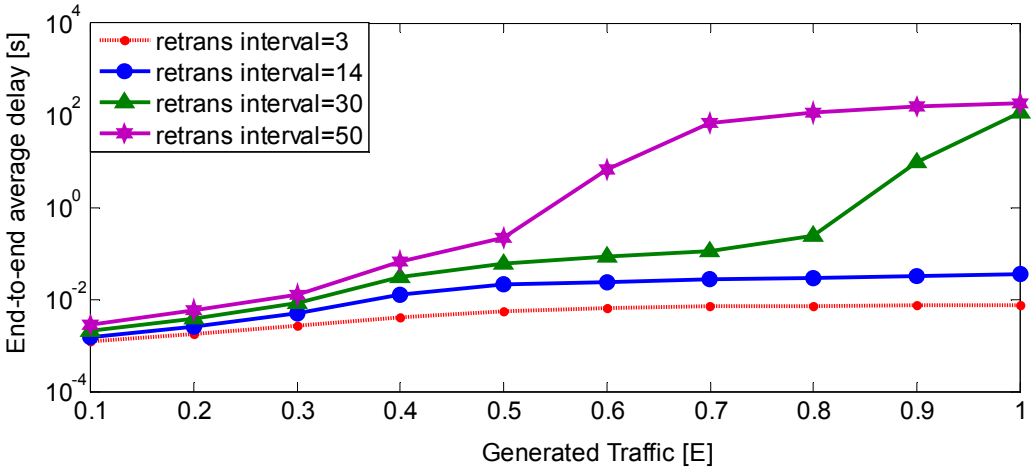

Figure 5.10 Average end-to-end delay against generated traffic at different maximum retransmission interval width and maximum number of retransmission fixed at 7

The throughput performance shown in Figure 5.6 and 5.9 showed that s-ALOHA is associated with low channel throughput, and at a certain point the aggregated (generated plus retransmitted) traffic exceeds the channel capacity. This renders the RACH channel useless and the system becomes unstable. For the details on the causes of s-ALOHA instability and the proposed solutions see [97, 98] where the authors defined the $\boldsymbol{e^{-1}}$ throughput achieved by s-ALOHA as a **critical point** which can be used to maintain the stability. Also detailed analysis of retransmission control and s-ALOHA stability are presented in [15, 16, 53, 99].

Despite the instability issue of the s-ALOHA, it works effectively as a protocol for RACH access in the current cellular network, which is designed for human-centric communication

which is termed as H2H communication. This is because of the dimensioning of the system and the regularity of the H2H traffic in the RACH requests being within the s-ALOHA throughput capability. However, when a cellular network supports M2M traffic, the activity of the RACH channel may be very high. This is because M2M systems are expected to have a huge number of devices with frequent access rate that will increase significantly the frequency of the RACH access and add extra load to the channel. This phenomenon will potentially drive the system beyond its capability, causing RACH overload and s-ALOHA will become inefficient. Supporting M2M on the existing cellular network will therefore require an effective RACH congestion control mechanism.

5.4 Throughput Analysis of SA-RACH with retransmissions

In this section an analytical model is developed to predict the throughput capability with basic SA-RACH when retransmissions are considered. We start the analysis by considering a single user group (where the RACH is assumed to be accessed by only single user group) and then consider dual user group RACH access (when H2H and M2M share the RACH). In both single and dual user group analysis both H2H and M2M use the SA-RACH scheme. The equations for the probability of successful packet transmission and throughput of s-ALOHA are defined (5.6) and (5.7) respectively, assuming no retransmissions. However, as presented earlier, cellular network standards allow retransmission in order to limit the number of RACH requests lost. In addition, the retransmission process is controlled using a retransmission cut-off strategy where the number of allowed retransmissions is limited. Retransmission is considered here to develop new throughput equations for a single user group as well as the dual user group considering their interactions in sharing the RACH channel and the probability of success or collision in a given period of a RACH slot. The retransmission cut-off strategy is applied, where a fixed window (number of slots) is used for the back-off. i.e.

when a collision occurs a user schedules to retransmit in a random slot within the back-off window. Note that both H2H and M2M users apply the same retransmission strategy here. In addition since the analytical model is based on basic s-ALOHA, similar assumptions to those presented in section 5.2 are applied as follows:

- All packets for the RACH request have the same length and transmission time τ which is also equal to the slot length.
- There are large number of users (i.e. $N \rightarrow \infty$) and therefore the probability of a user generating more than one request in a given small time period is negligible.
- New RACH requests arrive at the network at λ packets/slot according to a Poisson arrival process.
- The system is perfectly synchronised with every user only transmitting at the beginning of a slot with length equal to the RACH packet.
- All users share a single RACH channel and this corresponds to providing only one preamble per each RA-slot in LTE.

Figure 5.11 illustrates the retransmission cut-off strategy and also shows how the aggregated traffic offered to the system increases with an increase in the number of retransmissions from both H2H and M2M users contesting for the RACH slot or RA-slot.

From figure 5.11 the total aggregated traffic from both H2H and M2M ($G_{T_{total}}$) can be expressed as:

$$G_{T_{total}} = \lambda_{total} + r_{total} \tag{5.11}$$

where λ_{total} is the new request arrivals (per RA-slot) generated by both H2H and M2M and r_{total} is the total traffic in retransmission from both H2H and M2M.

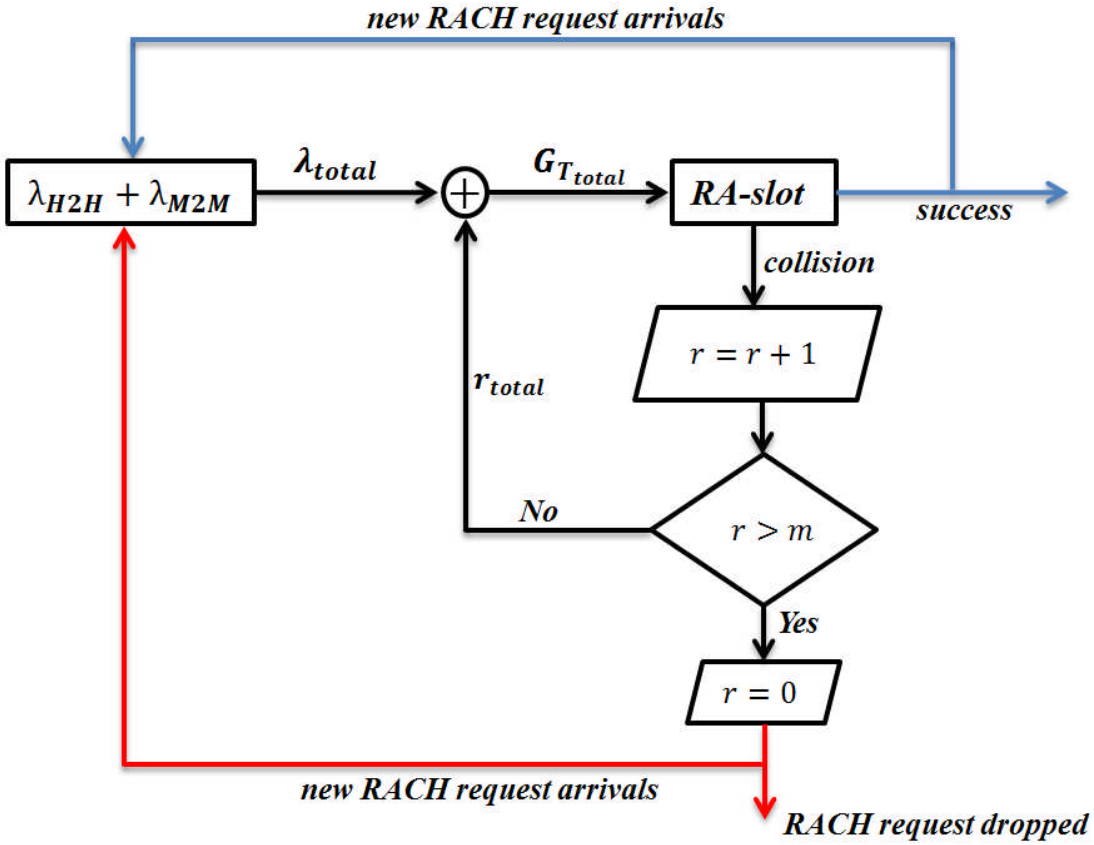

Figure 5.11 RACH request retransmission strategy

As shown in figure 5.11, when a collision occurs a user checks the retransmission counter, if the counter is still within the allowed limit the user will schedule to retransmit and increment its current number of retransmissions (r) by 1. This is compared with the set limit of number of allowed retransmissions (m). No further retransmission is allowed if r is higher than m and the RACH request is dropped and r is reset to zero, else the request is sent again.

In this analysis we assume the retransmitted RACH requests also follow a Poisson process. This can be true based on the fact that our random retransmission delay is up to 14 RA-slots, since a request is only retransmitted if a collision occurs i.e. if the request is not successful. Therefore considering the probablity of success and collision or failure (similar to equations (5.6) and (5.10)): $e^{-G_{T_{total}}}$ and $1 - e^{-G_{T_{total}}}$, the aggregated traffic can be obtained from the fraction of the traffic (from the request) being retransmitted for the i^{th} time as follows:

$$G_{T_{total}} = \lambda \sum_{i=0}^{m} \left(1 - e^{-G_{T_{total}}}\right)^{i} \tag{5.12}$$

and the throughput can be obtained as:

$$S_{retrans} = G_{T_{total}} e^{-G_{T_{total}}} \tag{5.13}$$

Considering a single user group and *m=2* the throughput will be:

$$\begin{aligned} S_{retrans} &= \left[\lambda + \lambda\left\{1 - e^{-G_{T_{total}}}\right\} + \lambda\left\{1 - e^{-G_{T_{total}}}\right\}^{2}\right] e^{-G_{T_{total}}} \\ &= 3\lambda e^{-G_{T_{total}}} - 3\lambda e^{-2G_T} + \lambda e^{-3G_{T_{total}}} \\ &= \lambda\left[1 - \left(1 - e^{-G_{T_{total}}}\right)^{3}\right] \end{aligned} \tag{5.14}$$

Therefore in general, the throughput equation of a single user group, using SA-RACH scheme with cut-off retransmission of limit *m*, is:

$$S_{SA-RACH_{retrans}} = \lambda\left[1 - \left(1 - e^{-G_{T_{total}}}\right)^{m+1}\right] \tag{5.15}$$

Now, the new throughput equations can be obtained for the combined access of H2H and M2M when both use the SA-RACH scheme with *m* allowed retransmissions as follows;

Considering H2H transmissions, the aggregated traffic offered to the system will also be obtained from the fraction of the failed traffic during the i^{th} attempt shown:

$$\lambda_{H2H}\left(1 - e^{-G_{T_{total}}}\right)^{i} \tag{5.16}$$

Therefore from equation (5.13) and (5.15) the throughput equation of H2H is:

$$S_{H_{SA-RACH_{retrans}}} = \lambda_{H2H} + \left[\lambda_{H2H} \sum_{i-1}^{m} 1 - e^{-G_{T_{total}}}\right] e^{-G_{T_{total}}}$$

$$= \lambda_{H2H}\left[1-\left(1-e^{-G_{T_{total}}}\right)^{m+1}\right] \quad (5.17)$$

Similarly:

$$S_{M_{SA-RACH_{retrans}}} = \lambda_{M2M}\left[1-\left(1-e^{-G_{T_{total}}}\right)^{m+1}\right] \quad (5.18)$$

Simulation results of SA-RACH for both single and dual user groups are developed and compared with the above analytical model in chapter 6. This is to clearly show the impact of the additional M2M traffic on the existing H2H users which is the motivation for the research. The results are also compared with our proposed scheme (QL-RACH) which is developed to control the additional M2M traffic and reduce the impact.

5.5 Summary

This chapter has presented an overview of the existing RACH access scheme. The differences between pure ALOHA and slotted ALOHA schemes are presented and comparisons of their throughput and blocking probability performance shows that s-ALOHA is the better scheme. The SA-RACH scheme is analysed and the causes of RACH instability has been observed, based on which it was realised that SA-RACH scheme has shown difficulties in supporting additional M2M traffic on the existing cellular network. This is because of its general low throughput capability. Analytical models that can be used to predict the RACH-throughput of SA-RACH for both single and dual user groups have been developed.

Chapter 6. Application of Q-Learning for RACH Access

6.1 Introduction

Chapter 5 presented and analysed the issue of RACH overload and instability which could occur when M2M is supported on an existing cellular network. Based on this it was realised that the existing RACH access scheme (SA-RACH) will not be capable of supporting the additional M2M traffic over the cellular network and therefore there is a need for a more effective scheme.

Ongoing work from different standardisation organisations is developing global standards for supporting M2M over a cellular network [7, 8]. Most of these standards organisations agreed that RACH overload could be one of the challenges of cellular M2M. For example, the Third Generation Partnership Project (3GPP) suggests RACH overload control as the first priority of cellular M2M [9]. In the standardisation report produced by the 3GPP [9], up to five possible RACH overload solutions have been identified: Access Class Barring (ACB) scheme, Slotted Access Scheme, Pull base scheme, M2M specific back-off scheme and Dynamic allocation of the RACH resource. The solutions proposed by the 3GPP require either involvement of the central entity (BS or eNB) or modification of the existing standards or signalling of the cellular system.

In this work, we also agree that RACH overload is a potential problem in supporting M2M on the existing cellular network. Hence we consider and investigate the coexistence of H2H and M2M (existing users in cellular network) users during RACH access and propose to control their interaction by controlling the M2M traffic without direct involvement of the central

entity. Q-learning is applied as an access scheme (QL-RACH) to self-organise the M2M traffic. This has the potential to improve the throughput performance over s-ALOHA when used as single user group (as shown in figure 6.5), and it controls the overload problem when combined with the SA-RACH scheme used by H2H.

Implementation details of the QL-RACH scheme are presented in section 6.2. An analytical model to predict the combined access of SA-RACH and QL-RACH is developed in section 6.3. Section 6.4 presents the simulation scenario used to model the combined access scheme. Also the performance of SA-RACH and QL-RACH using both analytical and simulation models is compared here. Finally the chapter is summarised in section 6.5.

6.2 Q-learning based RACH Access (QL-RACH) Scheme

The QL-RACH scheme is realised using Q-learning which is a simplified model of reinforcement learning with a simple algorithm that enables early system convergence [12]. In general, reinforcement learning can be described as a trial-and-error technique in which an action is decided through learning the system behaviour in a given environment. The action is determined based on prior experience that is built up using rewards and punishments [12, 13].

The QL-RACH scheme is implemented using what is described in [100, 101] as Frame ALOHA (F-ALOHA) where repeating frames are applied to the s-ALOHA scheme. Here user transmissions are restricted into a frame and each user can transmit only once in a frame. Each frame consists of a certain number of slots to which a user selects at random and transmits. If collision occurs, a user is scheduled to retransmit in a random frame within a given window where the width will be increased exponentially for each retransmission until the packet is successfully transmitted or dropped. The focus of this work is not on F-ALOHA, however it is mentioned briefly as a basis for QL-RACH implementation.

QL-RACH is a process that enables a user to determine its best slot (time slot for transmission) within a defined frame using its transmission history. The history is a record of weights from each user's transmission, having one per slot based on the number of slots in the repeating frame. The value of the weight will determine the user's choice of slot for transmission. In this way users are forced to self-organise by learning to avoid each other's slot during the contention process. At the end of the learning process (at convergence) each user has a unique dedicated slot and the contention is collision free with 100% throughput performance (neglecting overheads). A similar approach has been used to improve the throughput performance of s-ALOHA in a wireless sensor network by the authors of [102]. However, as described earlier, in this work we implement the learning by considering the existence of H2H users which maintain the existing SA-RACH scheme.

To implement QL-RACH, a frame (Q-Frame) is designed from the main cellular frame. To maximise the learning result we make the size of the Q-Frame (the number of slots per Q-Frame) equal to the number of the users. This is because if the Q-frame size is less than the number of the users, convergence will not be attained and Q-frame size higher than the number the users will drop throughput performance. The Q-Frame repeats in time within the main cellular frame with each slot having a weight (a Q-value) that keeps the transmission history. Each user has individual Q-values (initialised to zero) for every slot in the Q-Frame which is updated at every attempt using the following equation:

$$Q' = (1 - \alpha)Q + \alpha r, \tag{6.1}$$

where Q is the current *Q-value*, α is the learning rate (speed of the learning), r is the reward or punishment depending on the status of the sent request. For a normal situation, a negative number of equal value used for the reward is assigned as a punishment. The value of the learning rate determines the speed of the convergence and this has to be within the same low value range as that used as a reward. For example if +1 is adopted as a reward then learning

rate is chosen within the range of 0 to +1. The closer the learning rate is to +1 the higher the speed of the convergence but the lower the quality in terms of maintaining the convergence. The converse is the case when the learning rate is chosen close to zero. This process results in each user having different Q-Values for each slot, where a user always chooses a slot with maximum Q-value for transmission and if multiple slots have the same Q-values the user selects one at random. An example of the learning process of a user is shown in Figure 6.1. At the start of learning, all the slots have the same *Q-value* and the user selects the first slot at random and transmits it successfully. Using a learning rate (α) of 0.01 and a reward of $+1$, the *Q-value* of the slot is updated using equation 6.1. This first slot is the user's preferred slot (in the next Q-frame) since it has the highest Q-value. However coincidently the slot is selected by a different user (also learning) and therefore a collision occurs. This reduces the *Q-value* on the update with a punishment of -1. Therefore in the next Q-Frame, the user has two preferred slots (2 & 3) with the same highest Q-value and selects one at random. This process continues until every user finds a dedicated unique slot in the repeating Q-Frame and this is called a convergence state.

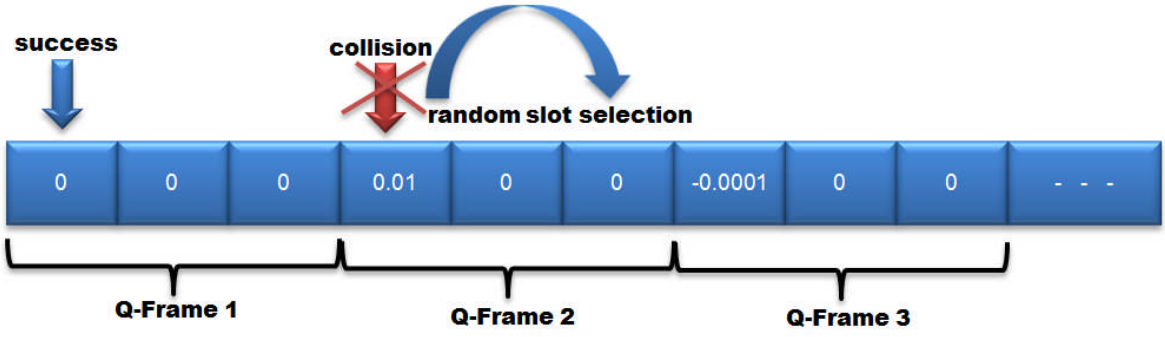

Figure 6.1 Presentation of slot learning process

6.2.1 SA-RACH and QL-RACH combined access schemes

QL-RACH has been described previously, considering only a single user group in the system. However as stated earlier, this work considers the coexistence of H2H and M2M user groups where the QL-RACH is used to control the M2M traffic. In addition, to avoid tampering with the existing standards of the cellular network, our approach assumes to have control (of the RACH access) of the M2M users only. Therefore we implement the slot learning here by considering the existence of H2H using the conventional SA-RACH scheme. The intuition is that combining QL-RACH with SA-RACH in the RACH access reduces the overall number of collision since there will be no collision between the M2M users after convergence.

Similar to the description of the single user QL-RACH scheme, the M2M users also learn their individual dedicated slots in a virtual frame called the M2M-Frame of size equal to the number of M2M users. The M2M-Frame is designed from the main cellular frame and carries RACH-slots or Random Access slots (RA-slots) as adopted in GSM and LTE standards respectively. The slot timing and length is mapped directly on to the control frame for the GSM standard or PRACH frame for the LTE standard. The information on frame timing and the number of active M2M users are broadcasted by the central entity (eNB or BS) via the downlink channel. The M2M-frame keeps repeating and is only considered by M2M users in which user transmissions are restricted to only one per M2M-frame. Similarly, the transmission history of each M2M user is recorded using the Q-value in each slot of the M2M-frame and is updated at every successful or failed transmission attempt using equation 6.1. Also the transmission decision is made (by an M2M user) based on the slot with the highest Q-value. At convergence, M2M RACH access becomes contention free amongst the M2M users, with H2H using SA-RACH and not being aware of the M2M-frame.

Since the QL-RACH scheme controls only the M2M traffic, in the steady state there is still disturbance from the H2H traffic due to its random access nature and the retransmission strategy which increases collisions. This causes more punishments than rewards in the affected dedicated M2M slots which at some point could make the M2M user to lose the slot. Figure 6.2 shows an example of how an M2M user (*M1*) loses a dedicated slot as a result of repeated collisions. The figure shows the transmission history (Q-values) of *M1* in an M2M-frame with 3 M2M users, where *M1* converged in slot 1 with a Q-value of *Q+2,* and *H1* transmits at random in the same slot which results in a collision. The *M1-slot* is punished and the Q-value reduces to *Q+1,* but the slot is still maintained for the next transmission since it has the highest Q-value. Unfortunately for *M1, H1* is using a fixed back-off window (W) of 3 and schedules to retransmit in *M1-slot.* Another collision occurs having the Q-value further reduced to *Q*. Now the Q-values of *M1* are the same in all the slots and hence there is no preferred slot and *M1* choses a slot at random where slot 3 is selected. However since the system is already in steady state, slot 3 belongs to *M3* (Q-values not shown). This results in another collision between *M1* and *M3* where the slot is punished. The effect of repetitive collisions in the dedicated M2M slots will force the users into another learning process and this affects the overall performance.

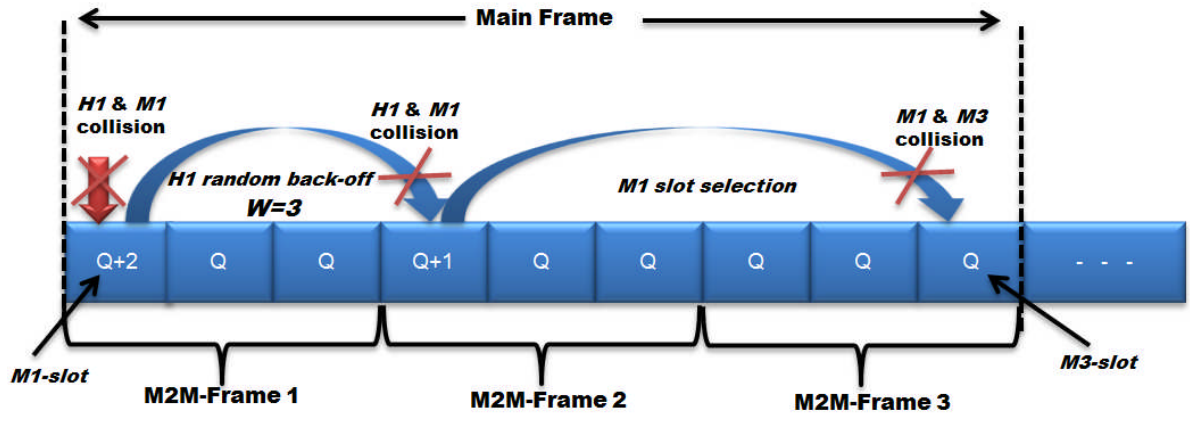

Figure 6.2 Example of H2H disturbances on the dedicated M2M slots

To avoid M2M users losing their dedicated slots, a different punishment strategy proposed by changing it from the fixed value presented in the example above (from equation 6.1). Here the punishment has become variable and reduced in such a way that it is less than the reward. The idea is, if a reward value is higher than the punishment value, collision from random H2H users will not make M2M users lose a slot. This is achieved by making the punishment value dependent of the number of collisions in the affected slot. Therefore when a collision occurs, the Q-value of the affected slot is updated (punished) using the following equation:

$$Q' = (1 - \alpha)Q - \alpha K, \tag{6.2}$$

and

$$K = \frac{1}{Nc} \tag{6.3}$$

where Nc is the number of collisions occur in the affected slot. Equations (6.2) and (6.3) serve to reduce the punishment value so that even though H2H traffic disturbs the steady state of M2M, the dedicated slots are maintained.

Note that, even though M2M users do not loose there dedicated slots due to the new punishment method, but collisions between H2H and M2M user groups still occur. It was mentioned that H2H use fixed uniform back-off window as presented in figure 6.2. On the other hand M2M users apply Binary Exponential Back-off (BEB) algorithm [103], which is a special case of exponential back-off. Each M2M user has an initial contention window of one frame which contains its dedicated slot, once it suffers collision with the H2H, the window size (number of frames) will be doubled. We use this method so that M2M users can back-off there traffic flow to reduce the impact of the existing H2H users. When an M2M user has more than one frame as the back-off one is selected at random to retransmit. The window of an M2M user is always set back to the initial contention window when the M2M RACH is successful or at the end of the allowed retransmission limit. The BEB reduce the

overall probability of collision between H2H and M2M user groups, however it affect the performance of M2M at high H2H traffic loads. This effect is considered and solution has been proposed in chapter 7.

Finally, based on the fact that all cellular standards have a similar RACH structure and configuration, it is important to note that our proposed scheme is compatible with all cellular standards.

6.3 Modelling and Performance Analysis

In this section the analytical model that predicts the RACH-throughput capability representing the interaction of H2H and M2M is presented. The section also presents the simulation model for the similar scenario mentioned above. Finally the results (simulation and analytical) are compared and the results are discussed.

6.3.1 QL-RACH Analytical Model Scenario

A similar method to that used in developing the analytical model presented in chapter 5 is adopted here, however the two user groups (H2H and M2M) are considered to have different schemes with H2H using SA-RACH and M2M using QL-RACH in sharing a PRACH resource. Here the H2H group maintains the existing SA-RACH scheme while the M2M group uses the QL-RACH access scheme. A steady state is assumed for the QL-RACH where every M2M user has a dedicated slot and there will be no more collisions amongst the M2M user groups. Therefore in this state QL-RACH works like a TDMA scheme, hence the interaction of H2H and M2M here is SA-RACH combined with TDMA. New equations for the RACH request collision probability are then obtained as well as the combined RACH-throughput performance. The same condition for collisions as presented in chapter 5 is

considered here. In addition, the same assumptions for the s-ALOHA analysis described in chapter 5 are also considered in this chapter.

6.3.1.1 Basic throughput analysis of SA-RACH and QL-RACH

In this combined access, the H2H user group maintains the existing SA-RACH scheme and the M2M traffic is controlled using the QL-RACH scheme. There are no retransmissions for either user group considered in this scenario. Since it is assumed that M2M users have converged to their dedicated slot that means the scheme is contention-free and therefore during the interaction there are zero collisions amongst the M2M user group. However collisions may occur between H2H and M2M or H2H with other H2H.

We start the analysis by assuming a H2H user is transmitting and obtains the probability of successful transmission in an RA-slot of duration τ which depends on the condition shown in Figure 6.3.

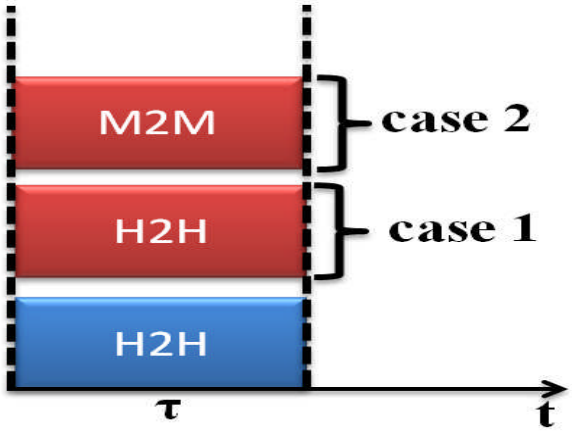

Figure 6.3 Collision conditions for H2H transmission

From figure 6.3, it can be seen that for a H2H transmission to be successful, case 1 and case 2 must not happen - i.e. the transmission will be successful if no other H2H and M2M users transmit in the same slot. The probability that no other H2H users transmit in the RA-slot (probability of no collision from H2H) is similar to what is presented in equation 5.6 of

chapter 5, and presented in terms of H2H here as the probability of no collision from H2H ($P_{no_{coll_{H2H}}}$):

$$P_{no_{coll_{H2H}}} = e^{-G_{H2H}} \tag{6.4}$$

On the other hand, the probability that the H2H user will suffer no collisions from M2M will be different since the M2M transmission is frame based with zero collision amongst the M2M user group. Therefore we need to consider the proportion of the M2M-frame in which an M2M user transmits. In the long term this is equal to the traffic load of the M2M **(G_{M2M})** offered in the frame for transmission. Therefore the probability that M2M transmits successfully is:

$$P_{M_{succ}} = G_{M2M}, \qquad G_{M2M} \leq 1 \tag{6.5}$$

and the probability that H2H suffers no collision from M2M user is:

$$1 - G_{M2M}\,, \quad G_{M2M} \leq 1 \tag{6.6}$$

Therefore from equations 6.4 and 6.6, the probability of a successful H2H transmission in time τ using the combined access scheme is:

$$P_{H_{succ_{Q_{Aloha}}}} = e^{-G_{H2H}}(1 - G_{M2M})\,, \qquad G_{M2M} \leq 1 \tag{6.7}$$

Hence the H2H throughput of the combined SA-RACH and QL-RACH ($S_{H_{Q_{Aloha}}}$) is:

$$S_{H_{Q_{Aloha}}} = G_{H2H}e^{-G_{H2H}}(1 - G_{M2M})\,,\ G_{M2M} \leq 1 \tag{6.8}$$

On the other hand, to obtain the combined performance of M2M, here M2M transmissions will only suffer collisions from the H2H users - i.e. there are no collisions between M2M users since every user has a dedicated RA-slot. This scenario is presented in Figure 6.4 where

the probability of a successful M2M transmission in an RA-slot of duration τ is equal to the probability that an M2M user suffers no collision from any H2H user.

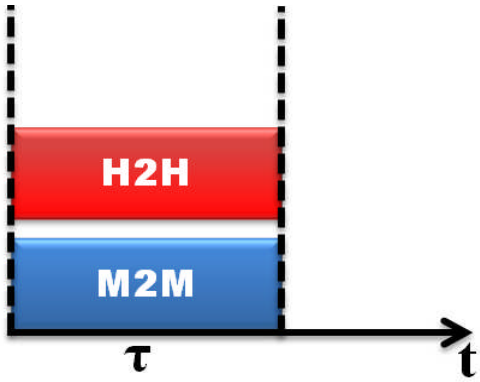

Figure 6.4 Collision conditions for M2M transmission

This is similar to (6.4) and the M2M combined RACH throughput ($S_{M_{Q_{Aloha}}}$) is:

$$S_{M_{Q_{Aloha}}} = G_{M2M} e^{-G_{H2H}} \tag{6.9}$$

Finally the overall throughput of both H2H and M2M in the combined access can be obtained as:

$$S_{T_{Q_{Aloha}}} = S_{H_{Q_{Aloha}}} + S_{M_{Q_{Aloha}}}$$

$$= e^{-G_{H2H}}(G_{total} - G_{M2M}G_{H2H}) \tag{6.10}$$

Note that G_{total} here is the total traffic generated by H2H and M2M without considering traffic in retransmissions.

6.3.1.2 Throughput analysis of SA-RACH and QL-RACH with retransmission

A similar method is adopted to develop the throughput equation for the combined access schemes, i.e. when H2H is using SA-RACH and M2M is using QL-RACH. The only difference here (since retransmission is considered) is that the probabilities of success for the individual user groups will be in terms of the aggregated traffic ($G_{T_{total}}$) presented in chapter

5 (see figure 5.11). Therefore putting the probability of H2H successful transmission for the combined access presented in (6.7) (note that the offered traffic in this case will be in terms of the aggregated traffic) into (5.13) gives the combined access throughput equation for H2H with retransmissions as follows:

$$S_{H_{Q_{Aloha_r}}} = \left[\lambda_{H2H} + \lambda_{H2H}\sum_{i=1}^{m}\left(1 - e^{-G_{T_{total}}}\right)^i\right] \times \left[e^{-G_{T_{H2H}}\left(1-G_{T_{M2M}}\right)}\right] \quad (6.11)$$

where $G_{T_{H2H}}$ and $G_{T_{M2M}}$ are the *H2H* and *M2M* aggregated traffic respectively (see Figure 5.11).

Note that:

$$G_{T_{total}} = G_{T_{H2H}} + G_{T_{M2M}} \quad (6.12)$$

Therefore;

$$e^{-G_{T_{H2H}}} = \frac{e^{-G_{T_{total}}}}{e^{-G_{T_{M2M}}}} \quad (6.13)$$

Substituting (6.13) in (6.11) and re-arranging gives:

$$S_{H_{Q_{Aloha_r}}} = \lambda_{H2H}\left[1 - \left(1 - e^{-G_{T_{total}}}\right)^{m+1}\right]\left[\frac{\left(1-G_{T_{M2M}}\right)}{e^{-G_{T_{M2M}}}}\right] \quad (6.14)$$

Similarly, the throughput equation of *M2M* in the combined scheme with retransmission can be obtained as follows:

$$S_{M_{Q_{Aloha_r}}} = \left[\lambda_{M2M}\sum_{i=0}^{m}\left(1 - e^{-G_{T_{total}}}\right)^i\right]\left[e^{-G_{T_{H2H}}}\right] \quad (6.15)$$

Substituting (6.13) in (6.15) and re-arranging gives:

$$S_{M_{Q_{Aloha_r}}} = \frac{\lambda_{M2M}\left[1-\left(1-e^{-G_{T_{total}}}\right)^{m+1}\right]}{e^{-G_{T_{M2M}}}} \quad (6.16)$$

6.3.2 Simulation Scenario

An event-based simulation is used to evaluate the performance of the interaction between H2H and M2M using Matlab. We consider the LTE standard in this simulation and assume that users (both H2H and M2M) are deployed within a cell coverage area with a single preamble sequence. All users share and access the channel randomly with equal access rights and different load ratios. All users have the same packet (preamble format) length which is equal to the slot length. Details of the random access have been modelled with initiation of the RACH request generation, which follows a Poisson distribution. Poisson is considered here because it is well established as an accurate model of the call arrival process at the telephone exchange/eNB which resembles typical human behaviour. However other traffic models are considered for M2M in this work (see chapter 7) due to their various application scenarios. The simulation is modelled with the initiation of the RACH request generation from each user group (H2H and M2M) having a mean inter-arrival time (τ_{ia}), determined by the traffic load distribution as:

$$\tau_{ia} = \frac{lN}{GC} \tag{6.17}$$

where l is the preamble format length in bits, N represents number of users in the system, G is the desired traffic load in Erlang and C is the transmission data rate in bits/sec.

6.3.3 Traffic load sharing for combined RACH access

Supporting M2M on the existing cellular network means that H2H users will share/contend for resources with the M2M users. Therefore a better way of coexistence should be devised by addressing the level of competition for the resource. The level of interaction between the users is determined by the individual (user group) load generated to make the total system's load traffic. Since H2H is the existing user group in the system, we consider predicting and

fixing its traffic load based on s-ALOHA (existing RACH access scheme) capacity. The critical point presented in chapter 5 is used as the baseline for the prediction where two limits of traffic level are defined. An upper limit (close to the critical point) and lower limit (away from the critical point) are defined as the peak hour load H2H could generate during the interaction. Note that different upper and lower limits may be considered by different individuals provided that they are within the s-ALOHA capacity and that will also depend on the cell size. However to demonstrate the performance of our schemes in this thesis, 0.3E and 0.1E are chosen as the upper and lower limits respectively. This is done in such a way that the system is not pushed to its upper and lower boundaries, so that the level of RACH contention between the H2H and M2M will be clearly understood. The load sharing is expressed below:

$$G_{total} = G_{H2H} + G_{M2M} \tag{6.18}$$

where G_{total} is the desired total traffic of the users, G_{H2H} is the fixed predicted H2H load and G_{M2M} is the M2M generated traffic, which is a variable and always generated on top of the G_{H2H} to complete the desired total traffic.

6.3.4 QL-RACH Results and Discussion

RACH-throughput and end-to-end delay are used here to evaluate the performance of the schemes using the simulation parameters shown in Table 6.1. Collection of simulation data starts after 150,000 successful RACH requests. This is done to run the simulation long enough to ensure that the system is out of the transient phase

Table 6.1 Simulation parameters based on LTE standards

Parameter	Value
PRACH configuration index	12
RA-slot period	1 ms
Preamble format duration	1 ms
Back-off period	28 ms
RAO	5 per frame
Retransmission limit	7
Learning rate	0.01
PB_{thr} (for chapter 7)	0.05

The results of RACH-throughput performance from analysis are compared with the simulation. Firstly, we compare throughput performance of the two SA-RACH and QL-RACH schemes for a single user group. In addition to demonstrate the instability behaviour of the SA-RACH scheme, we compare the performance with retransmissions (typical of the LTE standard) and without retransmission. The dual user group performance of the two schemes is then compared.

6.3.4.1 Single user group access

Figure 6.5 compares the single user group RACH-throughput performance of the SA-RACH (with and without retransmission) scheme and the steady state RACH-throughput of the QL-RACH scheme. The analytical equations for the SA-RACH are developed and presented in chapter 5.

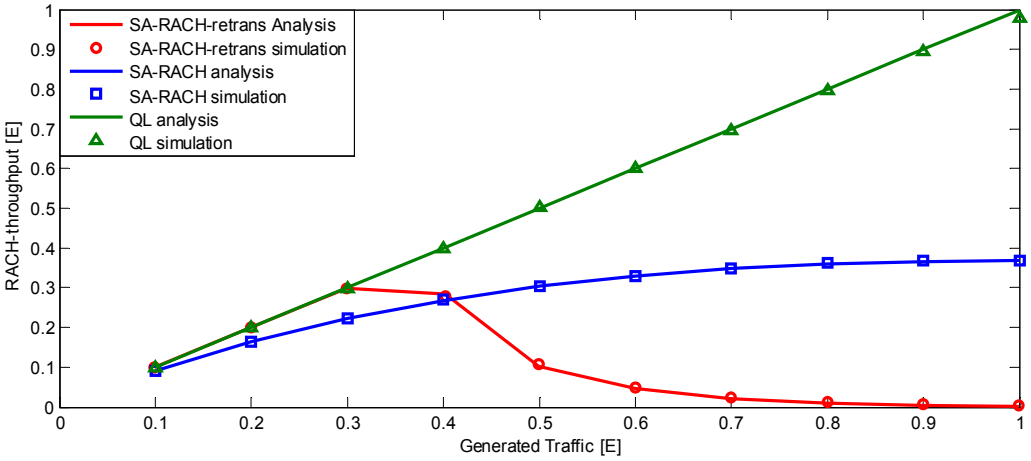

Figure 6.5 RACH-throughput comparisons of SA-RACH and QL-RACH schemes for a single user group

The results presented in figure 6.5 compare the analytical model from (5.7), (5.15) and (6.5) with the simulation. The analytical and simulation results are represented by solid lines and markers respectively. In addition, ***SA-RACH*** and ***SA-RACH-retrans*** represent the results of the performance of the SA-RACH scheme without and with retransmission repectively. It can be seen that the SA-RACH throughput increases with the increase in the generated traffic. The maximum throughput achieved is ~37% which is the maximum capacity of the s-ALOHA scheme at 1E of generated traffic. Also the result of SA-RACH shows that the channel is stable since the offered traffic does not exceed the s-ALOHA channel capacity. When retransmission is considered (*SA-RACH-retrans),* it can be seen that the scheme performs better up to the s-ALOHA limit. This happens because the scheme reduces the packet loss due the retransmissions and therefore (almost) all the packets generated get through. However, immediately after the channel throughput limit is reached, the aggregated traffic increases to the point that the s-ALOHA scheme can no longer support the traffic. This is why we can notice the throughput dropping with an increase in the traffic, to the extent that the channel becomes unstable. On the other hand, with steady state of the Q-learning, the *QL-RACH* scheme offers up to 100% throughput. This is because there are no collisions since

the scheme is contention free. Finally all of the results in figure 6.5 illustrate good agreement between the analysis and the simulation.

6.3.4.2 Dual user group access

All the results shown in this section represent the performance of dual user groups, i.e. when H2H and M2M coexist in the RACH contest. The results compare the performances of the SA-RACH and QL-RACH schemes. The following definitions are useful in interpreting the legend of the results:

- ***SA-RACH:*** this is the result of the combined access when both M2M and H2H users use the s-ALOHA scheme with a fixed back-off window for the retransmission.
- ***Q-L*:** this is when M2M uses QL-RACH with fixed punishment and retransmission using Binary Exponential Back-off (BEB) and H2H use SA-RACH with fixed back-off as the retransmission window.
- ***QL-RACH*:** this represents the results when M2M uses QL-RACH with varied punishment and Binary Exponential Back-off (BEB), and H2H use SA-RACH with fixed back-off window.

Figure 6.6 compares the RACH-throughput performances of the two different schemes (presenting both analytical and simulation results) at the upper limit considering retransmission. Note that we separate the results into the two different user groups where figure 6.6(a) and 6.6(b) represent M2M and H2H user group respectively. All the results illustrate a good agreement between the analysis and the simulation.

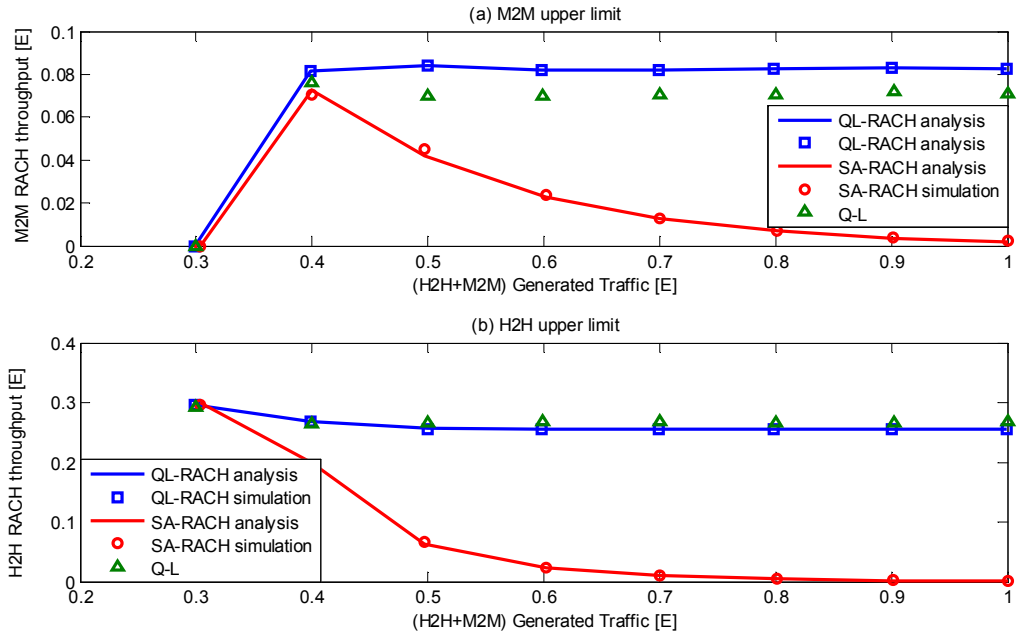

Figure 6.6 RACH-throughput comparisons of SA-RACH and QL-RACH schemes for dual user group at upper limit

Figure 6.6(a) illustrates the M2M RACH-throughput performance in the combined access scheme. As shown in the figure, SA-RACH exhibits the worst performance indicating the instability (the RACH-throughput tends to zero after the channel capacity limit) effect from the impact of the additional M2M traffic which renders the scheme inefficient. The impact is significant here because the H2H load is close to the s-ALOHA capacity which leads to a high probability of collision and the aggregated traffic increases due to the retransmissions. The result representing QL-RACH exhibits better performance which shows that the scheme has reduced the impact of M2M to some extent by providing the channel stability at some point of the total generated traffic. i.e. the RACH-throughput performance levelled at it peak throughpout the generated traffic. As can be seen, even though the generated traffic is beyond the s-ALOHA capcity limit, but the throughput deos not dropped to wards zero. However the learning has not introduced any improvement in the M2M RACH throughput here. This is because the performance is dominated by the random effect of the H2H users with their traffic contribution close to the critical point. In addition the Q-L result is provided to show

how the loss of convergence (M2M users losing slots) described in section 6.2 affects the throughput performance. It can be observed that the performance of M2M has slightly dropped and this is as a result of the back-off (BEB) used by M2M which serves as a sort of priority to H2H. Figure 6.6(b) illustrates the H2H performance in a similar arrangement, with *SA-RACH* giving the worst performance due to the same reason mentioned above. However the H2H RACH-throughput representing the QL-RACH scheme is better than that of the M2M scheme where the H2H performance stablised at arround 0.25E of RACH-throughput against the ~0.1E achieved by the M2M. This is because the H2H access is not restricted.

Looking at the RACH-throughput performance we could draw a conclusion on the operational limit of the QL-RACH scheme at the upper limit. In order to provide a maximum protection to the H2H (as the existing) users, we can determine the QL-RACH scheme operational based on the H2H performance. Therefore as can be seen the H2H RACH-throughput peformance is lower than generated traffic, hence it can be concluded that the H2H is not fully protected by the QL-RACH scheme at the upper limit.

Figure 6.7 shows the performance at the lower limit and similar to the upper limit results presented in figure 6.6, all the results from analytical model agree with those from the simulation model. In addition we also separated the M2M and H2H results into Figure 6.7(a) and 6.7(b) respectively.

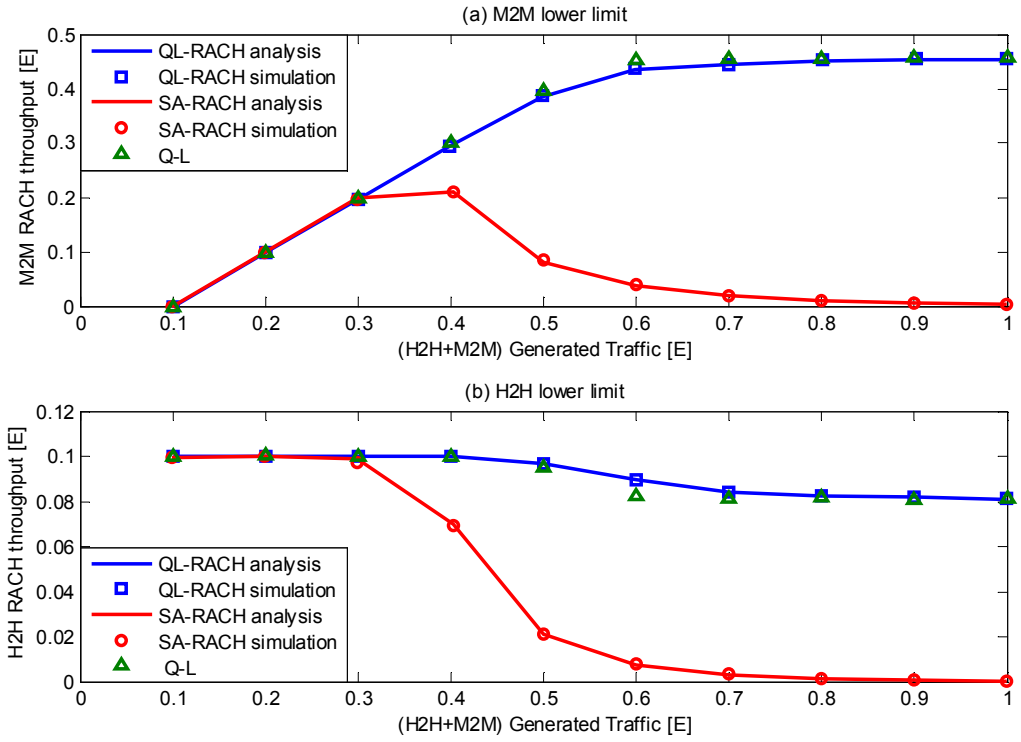

Figure 6.7 RACH-throughput comparisons of SA-RACH and QL-RACH schemes for dual user group at lower limit

As can be seen in both figures, the performance of M2M and H2H shows that the SA-RACH scheme has the worst throughput performance. Since the 0.1E (lower limit) is much further away from the critical point, the channel is able to support additional M2M traffic before reaching its capacity limit. However, similar to figure 6.6, when the generated traffic is beyond the critical point, the channel becomes unstable and the RACH-throughput degrades and tends to zero with further increases in the generated traffic level. On the other hand, QL-RACH in this scenario improves the performance (in addition to the channel stability) because of the gap between the critical point and the 0.1E. As shown in Figure 6.7(a) M2M RACH-throughput representing QL-RACH increases linearly with increasing generated traffic reaching a maximum of 0.45E. Finally the results representing the Q-L schemes for both M2M and H2H presented in figure 6.7 (a) and 6.7 (b) respectively show the same performance as QL-RACH. This indicates that there is no loss of convergence from the M2M users at the lower limit. Therefore it can be concluded that the M2M users only lose their

dedicated slots after convergence at the upper limit due to the high traffic levels of H2H, most of which stems from the retransmissions.

Similar to the upper limit, we can also check how much protection the QL-RACH scheme offers to the existing H2H users at the lower limit. As shown in the results of figure 6.7 the H2H performance has been fully protected from 0.1E up to about 0.5E of the total generated traffic. Hence based on protecting the H2H performance the QL-RACH scheme is operational up to this traffic level after which the H2H performance lost protection and dropped. On the other hand at the same 0.5E of the total generated traffic, the QL-RACH scheme produces M2M RACH-throughput performance of 0.4E which is the same as the M2M generated traffic and therefore this can be described as the operation limit of the QL-RACH scheme.

Figure 6.8 shows the effect of H2H load on QL-RACH by comparing the RACH-throughput performances at different H2H load levels from the lower to the upper limit. This is to see the performance of the scheme through various levels of traffic load combination of H2H and M2M. The idea here is when H2H load is known we can determine (from the performance of the scheme) how much M2M traffic load the system can support on top of the existing H2H traffic. Also as shown in the figure, there is agreement between the analytical and the simulation results.

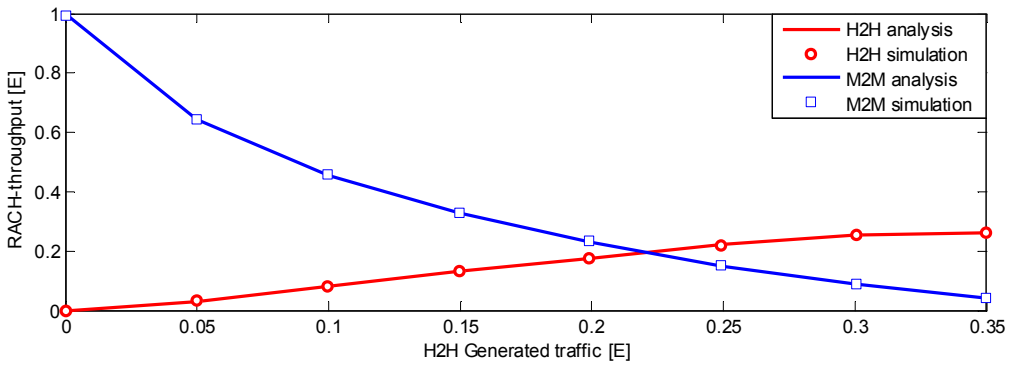

Figure 6.8 Operation level performance of QL-RACH scheme

Looking at the results shown in Figure 6.8, M2M obtained 100% RACH-throughput at zero H2H traffic. However the performance starts to drop with the introduction of the H2H traffic and keeps decreasing with increases in the traffic level. This is as the result of the random effect of the H2H traffic where its RACH-throughput increases with an increase in the H2H generated traffic.

Figure 6.9 presents the end-to-end delay by comparing the two schemes at the upper and lower limit of H2H fixed traffic levels.

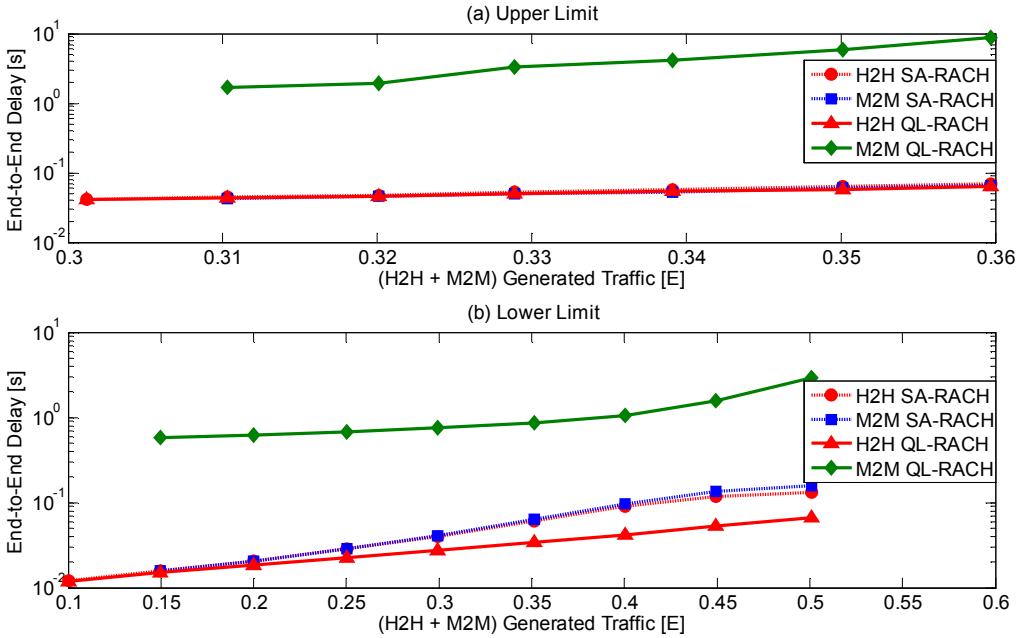

Figure 6.9 Average end-to-end delay comparisons for SA-RACH and QL-RACH scheme at upper and lower H2H traffic level limits

Figure 6.9(a) shows the delay performance at the upper limit where the three curves representing SA-RACH (both H2H and M2M) and H2H using QL-RACH exhibit similar performance. This is because they both used a fixed window of slots for the back-off. On the other hand, M2M with QL-RACH has higher delay since the users' transmissions are restricted to specific slots within the M2M-frame and also BEB is used as the back-off strategy. The delay performance is similar in figure 6.9 (b) with some improvement though. It

can be noticed that the delay performance of M2M with QL-RACH has significantly improved here compared to that at the upper limit. This is because there are fewer disturbances from the uncontrolled H2H traffic which reduces the level of collision between H2H and M2M user groups. The delay performance of H2H with QL-RACH is also shown to be better than the delay of H2H with SA-RACH by up to about 10%. In general for both upper and lower H2H traffic limits, the channel stability and improved RACH-throughput provided by QL-RACH is obtained without negatively affecting the delay performance of H2H.

The performance comparison of the SA-RACH and QL-RACH schemes described above represents steady-state (after M2M learning convergence) performance. However it is equally important to examine the RACH-throughput performance in the process of learning and that is presented in figure 6.10.

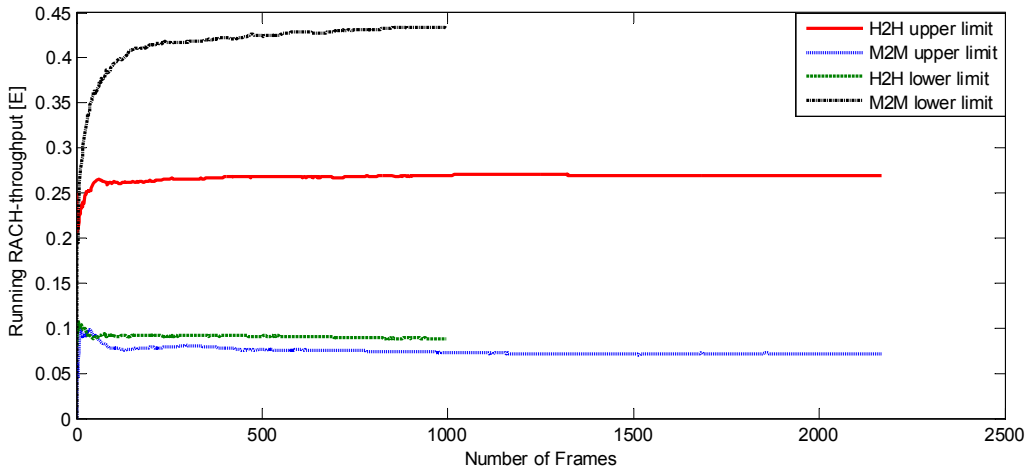

Figure 6.10 Running RACH-throughput during the M2M learning process

Figure 6.10 presents the running RACH-throughput of H2H and M2M against time (the number of frames). This is to show that even though the M2M users are in the process of learning to acquire dedicated slots, the system offers useful throughput from a very early stage. In addition, the learning convergence is achieved within a short period of time. It can

be noticed that in all cases, close to final throughput is reached within a small number of frames.

6.4 Summary

This chapter introduced a learning based RACH access scheme (QL-RACH) which enables M2M users to coexist with H2H users in the RACH contest in a cellular network. Q-learning has been used by M2M users to avoid collisions by learning dedicated slots within an M2M-frame. The scheme is generally applicable to all cellular network standards, thanks to the similarities in their RACH structure. Two limits of traffic load were predicted as the fixed existing cellular (H2H) load in our model. The capacity of s-ALOHA (e^{-1}) has been used as the base line for the traffic prediction where 0.3E and 0.1E were chosen as upper and lower limits respectively. From the simulation results is has been shown that at the upper load limit, the learning scheme only provides RACH stability, since the process is dominated by the random access procedure of H2H. On the other hand, at the lower H2H load limit a significant improvement in the RACH channel performance is obtained, that is, both the RACH-throughput and the RACH stability are simultaneously achieved. Notably, use of the learning scheme for M2M improved H2H delay performance as well.

In addition the chapter also presented an analytical model (based on the analysis presented in chapter 5) which is developed to predict the performance of the single and dual user groups for the QL-RACH scheme. There is an agreement between the analytical and simulation results of the SA-RACH and the QL-RACH schemes.

Chapter 7. Frame Based Back-off for H2H to Improve QL-RACH

7.1 Introduction

Chapter 6 demonstrated the ability of the QL-RACH scheme to reduce the impact of additional M2M traffic on the existing H2H users in a cellular network. However the uncontrolled H2H traffic affects the performance of the QL-RACH scheme.

This chapter introduces a novel back-off scheme to improve the performance of the QL-RACH scheme presented in chapter 6. The effect of uncontrolled H2H traffic on the QL-RACH scheme is presented in section 7.2. Section 7.3 presents a Frame based back-off scheme which is proposed to remove the effect of the uncontrolled H2H traffic. The performance of the FB-QL-RACH scheme is evaluated by comparing the RACH-throughput of the scheme with that of QL-RACH using both analytical and simulation models in section 7.4. FB-QL-RACH with dynamic Back-off Frame Size (BFZ) is a scheme proposed to update the required BFZ value automatically by an eNB and is presented in section 7.5. Section 7.6 presents and analyses the performance of FB-QL-RACH with dynamic BFZ and finally the chapter is summarised in section 7.7.

7.2 Effect of SA-RACH on QL-RACH

As described in chapter 6, depending on the H2H traffic load, the QL-RACH scheme provides channel stability and also improves RACH-throughput. In the combined RACH access, two schemes are interacting, with H2H using SA-RACH and M2M using QL-RACH. It was also made clear (in chapter 6) that, even though QL-RACH controls M2M traffic, the

level of disturbance due to collisions from the uncontrolled H2H traffic has a significant impact on the overall throughput (especially M2M) performance. The level of the interference increases with increasing H2H traffic load. This is because the closer the H2H traffic load is to the s-ALOHA capacity, the higher the probability of collision amongst the H2H user group as well as between the H2H and M2M user groups. Collisions between the two different user groups will increase the possibility of having idle slots in the M2M-frame. This happens because M2M transmission is frame based with a binary exponential back-off retransmission strategy. So whenever collision happens (between H2H and M2M) more than once, M2M backs off some number of M2M-frames leaving the slots empty, creating idle slots in the frames. Based on this, the FB-QL-RACH is introduced to reduce the probability of collision and eliminate the probability of idle slots.

7.3 Frame Based Back-off for QL-RACH (FB-QL-RACH)

In order to remove the effect of the uncontrolled H2H users and improve the overall performance of the QL-RACH scheme, the probability of collision between H2H and M2M user groups and the probability of idle slots in the M2M-frame need to be controlled as described in equation 7.1.

$$min \sum P_{collision} + P_{idle\_slot} \tag{7.1}$$

where $P_{collision}$ is the probability of collision between H2H and M2M users and P_{idle_slot} is the probability of an idle slot in the M2M-frame.

The FB-QL-RACH scheme introduces an improved back-off strategy to enhance the back-off process used in the QL-RACH scheme. A separate frame is provided for H2H back-off to redirect the H2H traffic in re-transmission where the users are restricted from re-transmitting

in an M2M-frame and also M2M is not allowed to transmit in the H2H back-off frame. In this way $P_{collision}$ is reduced since the collision occurs only during the first attempt. Also to remove the effect of the idle slot (P_{idle_slot}), no back-off is used for M2M and therefore there will be no idle slots in M2M frame. The frame timing of the FB-QL-RACH scheme is described in Figure 7.1 where two consecutive virtual frames of RA-slots are designed from the main LTE PRACH resource, making what we called a **global-frame**. The first frame is the **M2M-frame** which is similar to the one described in chapter 6. The second frame is the H2H back-off frame (**H2H B-frame**) which is used for H2H re-transmissions and can also be used by H2H for the first attempt when required.

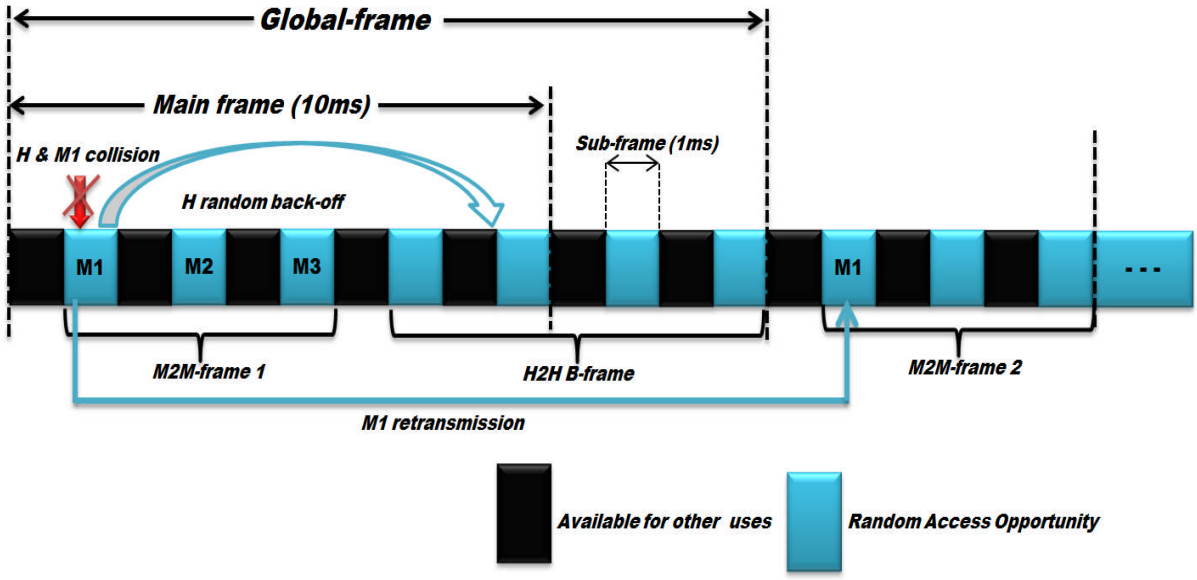

Figure 7.1 LTE PRACH configuration index 12 structure with the presentation of FB-QL-RACH scheme frame structure

Figure 7.1 illustrates the FB-QL-RACH operation where a steady state of three M2M users is assumed. Since H2H uses SA-RACH, a collision occurs in the first slot between a random H2H user (*H)* and an M2M user on slot 1 of the M2M-frame (*M1). H* is allowed to retransmit only in the **H2H B-frame** by selecting a random slot within a fixed back-off window. On the other hand, *M1* retransmits in its dedicated slot in the next M2M-frame without backing-off.

The information about **M2M-frame**, **H2H B-frame** and **H2H B-frame size** (BFZ) is made available by the eNB through the downlink broadcast channel to the users. In addition the eNB determines and controls the BFZ value based on the system's traffic load.

7.4 FB-QL-RACH Modelling and Performance analysis

This section presents the analytical model which is developed to predict the throughput performance of FB-QL-RACH scheme. A simulation model is also developed to evaluate the performance of the scheme.

7.4.1 Throughput analysis of the FB-QL-RACH scheme

In line with the FB-QL-RACH representation shown in figure 7.1, here (in order to develop the analytical model of the scheme) we describe how the H2H traffic in retransmission is redirected to the **H2H B-frame**. As shown in figure 7.2 the H2H aggregated traffic is directed to **H2H B-frame**. The process will reduce the probability of collision as well as the total aggregated traffic and significantly increase the overall performance.

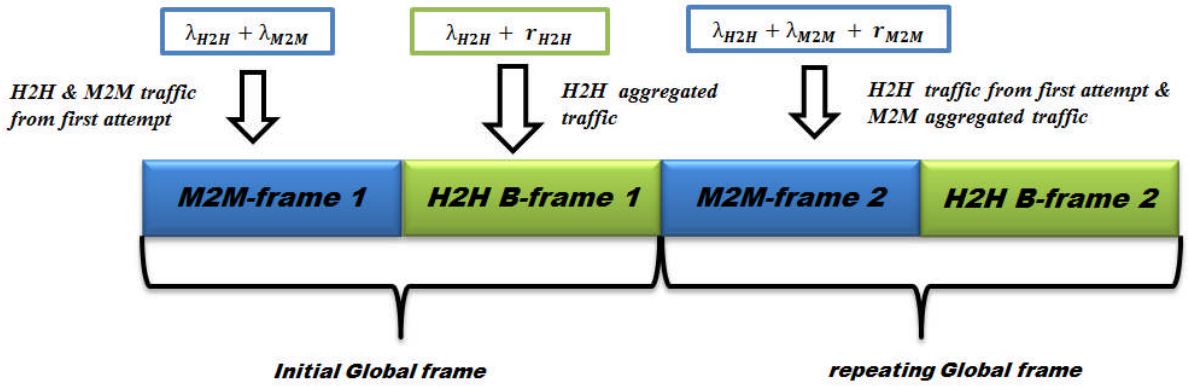

Figure 7.2 Traffic control representation of FB-QL-RACH scheme

To realise the RACH-throughput equation for H2H we consider the two different frames (M2M-frame and H2H B-frame) separately and obtain the aggregated traffic offered to the frames as follows.

In the M2M-frame the aggregated traffic is the first attempt traffic from both H2H and M2M user groups as well as the traffic from M2M in re-transmission as follows:

$$G_{T_{BF_{M2M}}} = \lambda_{total} + r_{M2M} \tag{7.2}$$

where $G_{T_{BF_{M2M}}}$ is the aggregated traffic offered to the M2M frame using the *FB-QL-RACH* scheme, λ_{total} is the total initial access traffic (from H2H and M2M) and r_{M2M} is the M2M traffic in retransmission.

Also in the H2H B-frame the aggregated traffic is the traffic from H2H first attempt and retransmissions as follows:

$$G_{T_{BF_{H2H}}} = \lambda_{H2H} + r_{H2H} \tag{7.3}$$

where $G_{T_{BF_{H2H}}}$ is the aggregated traffic offered to the H2H B-frame using the FB-QL-RACH scheme, λ_{H2H} and r_{H2H} are the H2H initial access and retransmission traffic respectively.

Therefore combining (7.2) and (7.3) using figure 5.11 gives the total aggregated traffic using the FB-QL-RACH scheme as:

$$G_{T_{BF_{total}}} = G_{T_{total}} + \lambda_{H2H} \tag{7.4}$$

Hence the new aggregated traffic is used to determine a new probability of successful transmission and the throughput equation which is similar to (6.14) with a different aggregated traffic level:

$$S_{H_{Q_{Aloha_{FB}}}} = \lambda_{H2H}\left[1 - \left(1 - e^{-G_{T_{BF_{total}}}}\right)^{m+1}\right] \tag{7.5}$$

For the M2M equation, since there are no collisions amongst the M2M user group (steady state of QL-RACH scheme) and collisions between H2H and M2M users occur only during first attempt, therefore the condition is similar to the analysis of basic SA-RACH and QL-RACH with no retransmissions presented in section 6.3. The throughput equation is similar to

(6.9) with a modification to the proportion of the M2M traffic which is now a total aggregated traffic as shown:

$$G_{T_{M2M}} = \lambda_{M2M} + r_{M2M} \tag{7.6}$$

and the throughput equation is:

$$S_{M_{Q_{ALOHA_{FB}}}} = G_{T_{M2M}} e^{-\lambda_{H2H}} \tag{7.7}$$

7.4.2 Simulation scenario of the FB-QL-RACH scheme

The simulation scenario in terms of users' deployment, number of preambles per RA-slot and access right as well as load ratio, considered here is similar to that considered in chapter 6. However, different to the method used in chapter 6 where both users in the system are assumed to have the same traffic model (Poisson), here we consider additional traffic models for M2M. This is because various M2M applications will have different traffic characteristics. Here we consider M2M applications with periodic traffic, divided into; **synchronised (sync)** and **un-synchronised (u-sync)**. In the sync periodic traffic model, the M2M users report their data periodically at the same time while for the u-sync traffic model, the users will have an offset in the timing of the periodic data transmission.

Therefore another important issue this research considers is the possible coexistence of different traffic models for the RACH access. The RACH request generation for H2H follows a Poisson distribution (due to the same reason mentioned in chapter 6) and M2M follows Periodic. To evaluate the performance of the proposed scheme we adopt the LTE simulation parameters presented in Table 6.1. Also traffic load sharing between H2H and M2M is as in chapter 6 with the estimation of H2H load fixed at upper and lower limits and M2M traffic varying to complete the desired system traffic.

7.4.3 FB-QL-RACH Simulation Results and Discussion

A single preamble is also used to examine the RACH-throughput to describe the performance of our simulation. To show the effect of the FB-QL-RACH scheme we compare its performance with the performances of SA-RACH and QL-RACH presented in chapter 5 and 6 respectively. In addition, we use both simulation and analytical results in the comparison in order to show the agreement between the two models.

Firstly we compare the RACH-throughput performance of a single user group where different traffic distributions are used and their performances using SA-RACH and QL-RACH schemes are compared, as shown in Figure 7.3.

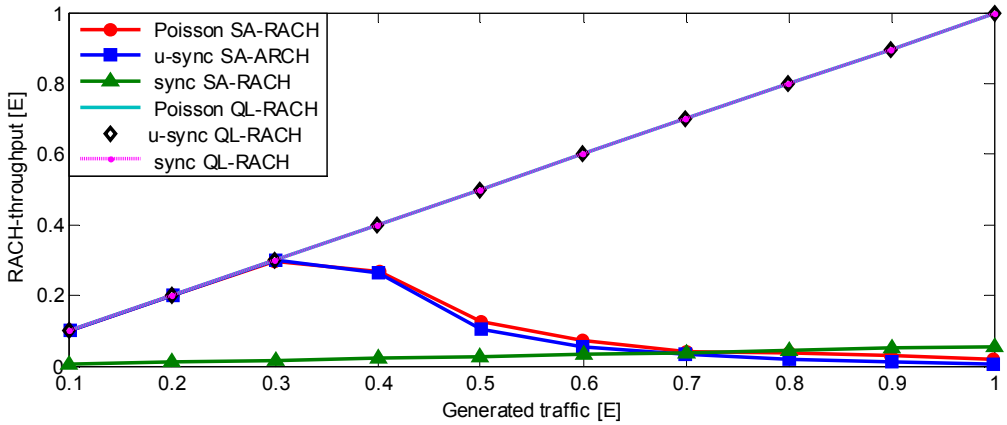

Figure 7.3 RACH-through comparisons of different traffic distributions using SA-RACH and QL-RACH

From the results it can be seen that when using the SA-RACH scheme, Poisson and Periodic (u-sync) traffic behave in the same way. Similar to what is presented in the previous chapters, since the SA-RACH is using a fixed back-off for the retransmission, the channel becomes useless beyond the s-ALOHA capacity when the RACH is overloaded.

On the other hand, when the requests follow a synchronised periodic traffic distribution, the results (from the RACH-throughput) show that the channel is not useful, hence the s-ALOHA scheme cannot support the traffic. The three traffic distributions behave in the same way with

the QL-RACH scheme which shows that both traffic models can be controlled using the Q-learning and 100% RACH-throughput is obtained at steady state. Therefore irrespective of whether a traffic distribution is Poisson or periodic (sync or u-sync) the QL-RACH scheme can be implemented.

The graphs in Figures 7.4 to 7.6 show the performances of the dual user groups with each graph representing a particular user group performance. The results compare FB-QL-RACH, QL-RACH and SA-RACH schemes. Different values of BFZ are used to show the relationship between H2H traffic load, BFZ value and M2M performance. Analysis and simulation results are used in the comparisons where solid lines and markers represent the analysis and simulation results respectively. In addition to the definitions provided in chapter 6 to interpret the legend of the results, the legend bearing ***FB-QL-RACH*** here represents results of the combined access when M2M users use QL-RACH with no back-off and H2H users use SA-RACH with a separate frame for back-off using a fixed window.

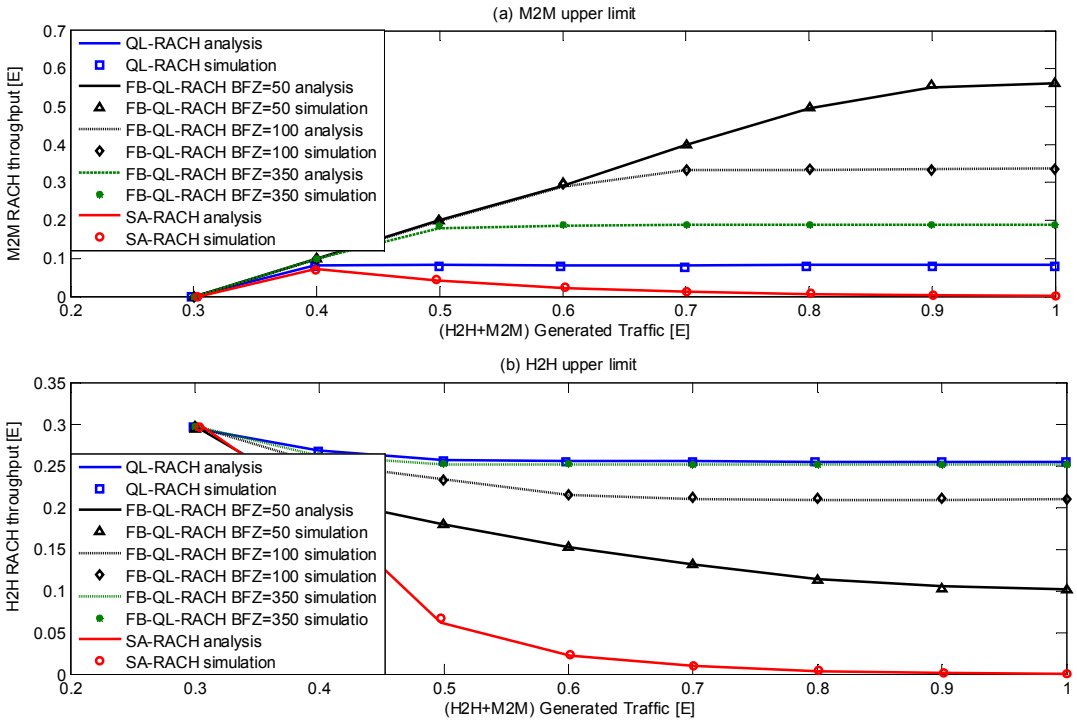

Figure 7.4 H2H and M2M RACH access throughput comparison of SA-RACH, QL-RACH and FB-QL-RACH schemes at the upper limit

Figure 7.4 compares the RACH-throughput performances of the three different schemes at the upper limit. All the results illustrate a good agreement between the analysis and the simulation. Figure 7.4(a) is the M2M performance in the combined access scheme where the SA-RACH exhibits worst performance. This indicates the impact of supporting additional M2M traffic in the cellular network as presented and explained in chapter 6. The impact has been reduced by the QL-RACH scheme and the performance is still low due to disturbance from the uncontrolled H2H traffic (especially at this limit). The FB-QL-RACH scheme significantly improves the performance depending on the BFZ value used. As shown in the results the lower the BFZ value the better the M2M performance, where up to about 0.56E RACH-throughput is realised at an M2M traffic load of 0.7E using a BFZ of 50. This is much better than what has been achieved by the QL-RACH at the same M2M load. Figure 7.4(b) shows the H2H performance in a similar pattern with the QL-RACH improving the poor performance of the SA-RACH scheme. On the other hand, for the FB-QL-RACH scheme, the higher the BFZ value the better the H2H performance. This is because H2H use fixed back-off as the retransmission window and a high BFZ value provides enough number of slots within the H2H B-frame. BFZ of 350 is having the highest performance which is similar to what the QL-RACH scheme offers.

Analysing the operation of FB-QL-RACH scheme based on protecting the H2H performance shows that the performances from the values of the BFZ used do not offer protection to the H2H. As shown and explained above the highest value of BFZ (350) used offers the same performance as the QL-RACH scheme, therefore to protect the H2H a higher BFZ value is required and that is at the expense of M2M performance.

Figure 7.5 shows the lower limit performance and similar to Figure 7.4, all the analytical results representing the three schemes agree with the simulation results.

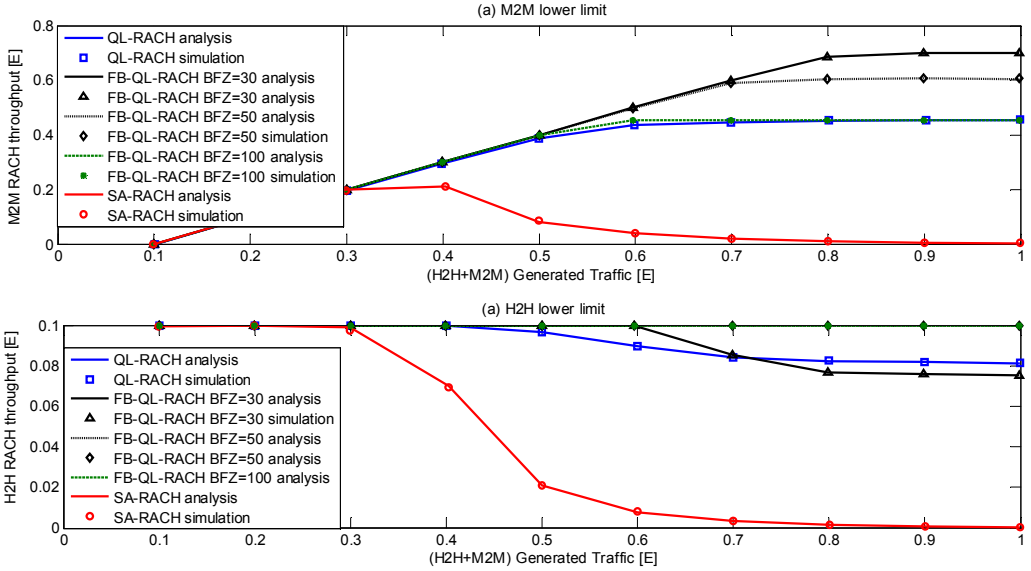

Figure 7.5 H2H and M2M RACH access throughput comparison of SA-RACH, QL-RACH and FB-QL-RACH schemes at lower limit

As shown in both of the results presented in Figure 7.5 (a & b) the SA-RACH scheme is again exhibiting the worst performance which is still improved by the QL-RACH scheme. For the FB-QL-RACH scheme, high BFZ is not required here since H2H generates only 0.1E of traffic load which is far below the s-ALOHA capacity. As shown in Figure 7.5 (a) from 0 to 0.5E of the total generated traffic, the change in the value of BFZ used has no effect on the performance. However, beyond 0.5E of the total traffic, the lower the value of the BFZ used, the better the M2M performance. For the H2H performance (in Figure 7.5(b)) the RACH-throughput increases with increases in the BFZ value.

In general, the proposed FB-QL-RACH scheme works better at the lower limit, up to 100% RACH-throughout performance is achieved for H2H with 50 and 100 BFZ. In addition M2M RACH-throughput increases significantly at a BFZ of 50. In addition there is a trade-off between H2H and M2M performances in terms of the BFZ value used.

In terms of protecting the H2H performance the FB-QL-RACH scheme at the lower limit offers protection to the H2H depending on the value of the BFZ used. As shown in figure 7.5,

BFZ values of 50 and 100 offer full protection to the H2H in all the traffic levels with the scheme operational limit at M2M traffic of 0.45E and 0.6E respectively.

Figure 7.6 compares the performance of the schemes at different H2H traffic levels from the lower to the upper limit. This is to show the practical operating limit of FB-QL-RACH scheme at different BFZ values.

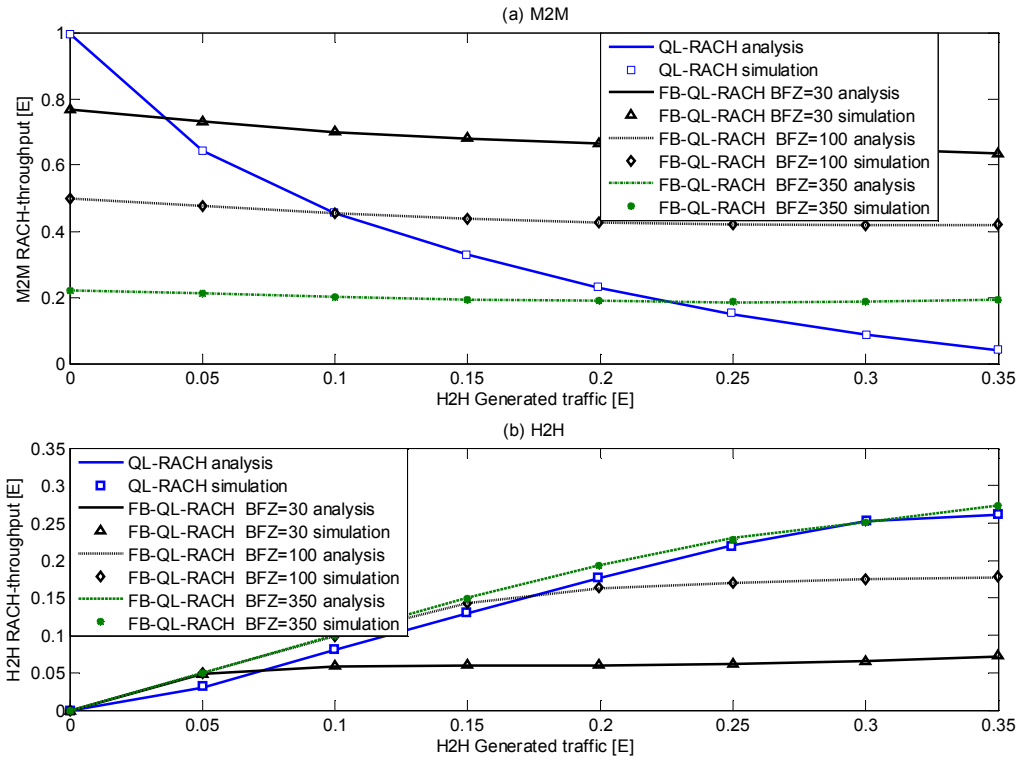

Figure 7.6 Practical operation level performances of QL-RACH and FB-QL-RACH schemes

Looking at the results shown in Figure 7.6(a), it can be seen that QL-RACH performs better from the start (i.e. from 0 to about 0.05E of H2H traffic) when H2H generates small amount of traffic and M2M has high traffic load. The performance here indicates that at this H2H load the QL-RACH scheme controls M2M traffic which is much higher and the uncontrolled H2H load has no effect on the QL-RACH. From H2H load of 0.05E and above, the FB-QL-RACH performs better than the QL-RACH depending on the value of the BFZ used. For example from 0.05 to 0.1E only a BFZ value of 30 is better than the QL-RACH which shows

that lower BFZ is required since H2H generates low traffic load and collisions are rare. In addition, at a H2H load of 0.1E to 0.23E and 0.23E to 0.35E, the FB-QL-RACH scheme with a BFZ value of 100 and 350 respectively supersedes the QL-RACH performance. The contrary performance is shown in Figure 7.6(b) where for the FB-QL-RACH scheme the higher the BFZ value used, the better the performance at higher H2H load. This explains why both 30 and 100 values of BFZ perform better proportionately than the QL-RACH scheme and BFZ of 350 at lower H2H load. On the other hand, the 30 and 100 values of BFZ perform lower than the QL-RACH scheme and BFZ of 350 at higher H2H load. The BFZ value of 350 has similar performance with the QL-RACH with slight difference at the low and high H2H traffic loads.

7.5 FB-QL-RACH Scheme with Dynamic BFZ

From the performances of the FB-QL-RACH scheme shown above, it is clear that the value of the BFZ used plays a significant role in the RACH-throughput performance of both H2H and M2M. It was observed that the value of the BFZ required increases with increase in the H2H traffic load. Therefore fixing the value of the BFZ irrespective of the load generated by the H2H users will not provide the best performance of the FB-QL-RACH scheme. FB-QL-RACH with a dynamic frame size is introduced here to provide better performance of the FB-QL-RACH scheme.

The dynamic frame size is implemented by an eNB with the help of prior information from H2H users as presented in Figure 7.7. We propose to use message 3 of the RACH request procedure (see chapter 3) for a H2H user to send (in addition to the information for the resource request) information of parameters needed to obtained blocking probability. Each H2H users sends its cumulative number of blocking and successful transmissions to the eNB

in a given window of frame time. The eNB uses the information from all the active H2H users within the window period to calculate H2H system blocking probability (P_B) which is compared with a define threshold (BP_{thr}). Note that the value of the $P_{B\,thr}$ depends on the acceptable blocking probability of the considered traffic and therefore different application may have different $P_{B\,thr}$. However we chose 5% (as shown in Table 6.1) in this work to demonstrate the performance of our scheme. A decision on the BFZ value is made (by the eNB) from the above comparison using the following algorithm:

$$\boldsymbol{If}\quad P_B > P_{B\,thr}$$
$$BFZ = BFZ + j$$
$$\boldsymbol{Else}$$
$$BFZ = BFZ$$
$$\boldsymbol{End}$$

Where j is an integer value use to increment the current BFZ value.

Blocking Probability is chosen (to control the BFZ value) because it is directly depends on the probability of collision which is controlled by the BFZ.

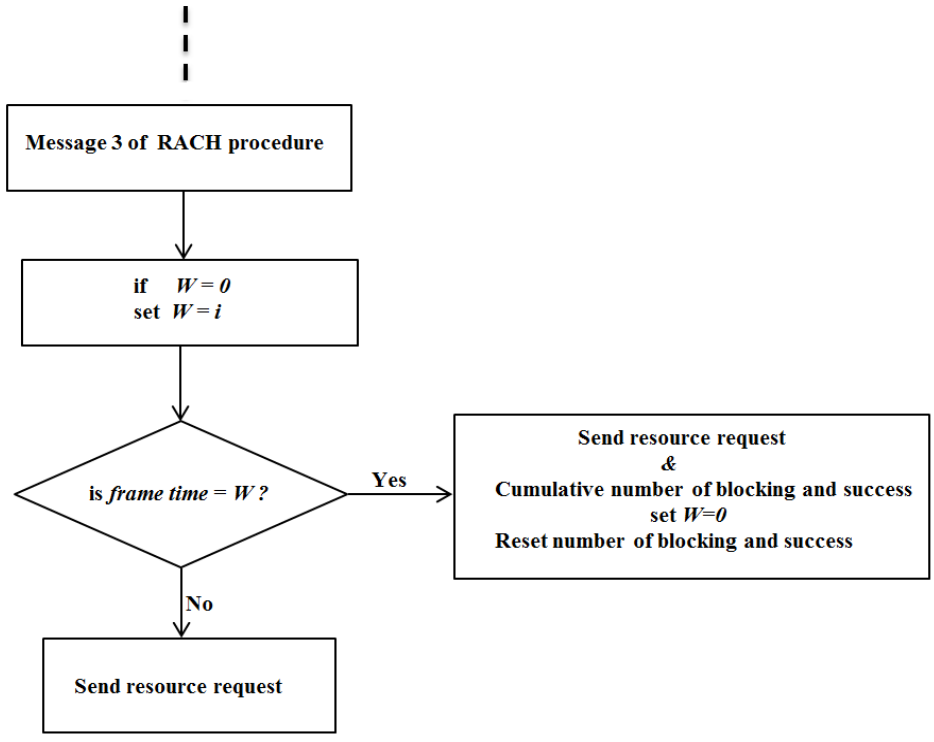

Figure 7.7 H2H user sending blocking probability parameters to the eNB

As shown in figure 7.7 above, a H2H user when sending message 3 checks and increases the value of a window (*W*). If the value of *W* is equal to a set frame time, the user sends a resource request and the parameters used to obtain blocking probability and reset the value of *W* to zero. On the other hand, if the value of *W* is not yet the frame time the H2H user sends a resource request only.

7.6 FB-QL-RACH with Dynamic BFZ, Results and Discussion

The first result presented in Figure 7.8 represents dynamic Back-off Frame size (BFZ) convergence time taken to achieve the maximum required BFZ value. Two separate total traffic loads are considered both at lower limit i.e. H2H load is fixed at 0.1E.

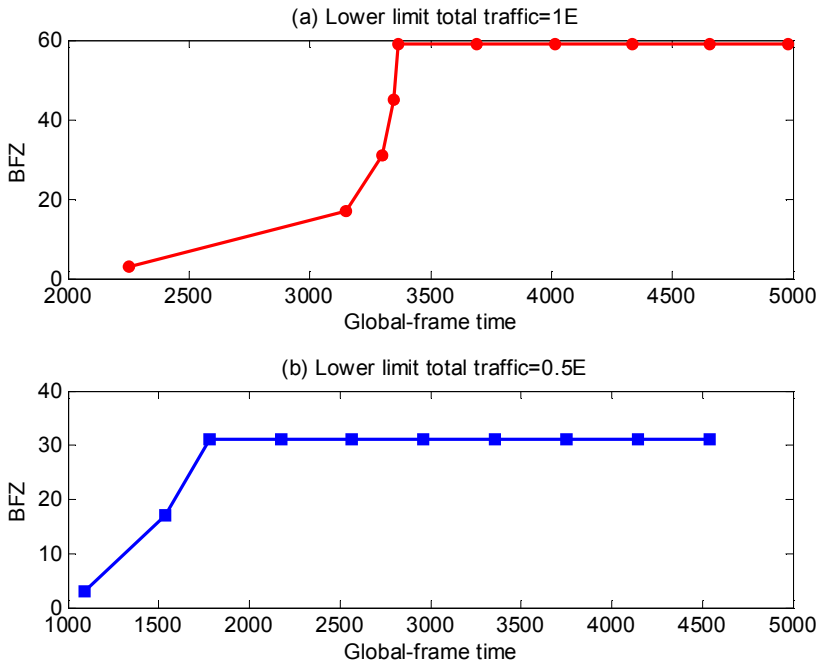

Figure 7.8 BFZ convergence time

In Figure 7.8(a) the total traffic is at 1E with H2H traffic fixed at 0.1E and M2M generates 0.9E. The result shows how the BFZ increases with increases in the number of **global-frames** (M2M-frame + H2H B-frame) used and system converges at about BFZ value of 60. On the

other hand, in figure 7.8(b) considers a total traffic of 0.5E with the H2H traffic still fixed at 0.1E and M2M generate 0.4E. It can be seen that, even though the H2H traffic here is the same as in figure 7.8(a) but the BFZ converged here at a value of 30. This is because M2M traffic load is lower here which reduces the collision between H2H and M2M user groups. This indicates that the value of BFZ required depends on the level of collision between the two user groups. In addition, for the same reason, Figure 7.8(b) has lower convergence time than 7.8(a).

As discussed earlier, 100% H2H throughput was achieved at a fixed BFZ of 50 in Figure 7.5(b). Since the same load is used in Figure 7.8(a) this shows that the converged dynamic BFZ of 60 is not the best. We check this situation by varying the steps at which the BFZ value is increase in the process of the dynamic BFZ implementation and the results are presented in Figure 7.9.

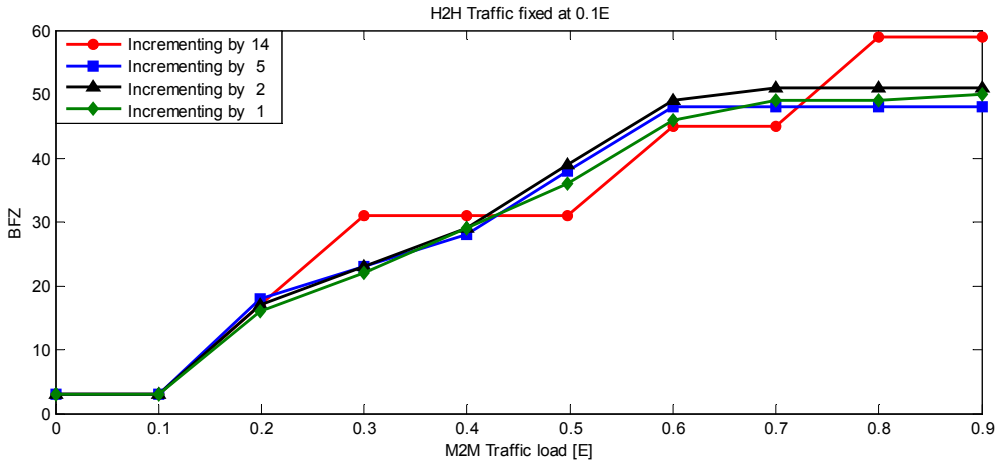

Figure 7.9 BFZ convergence at various incrementing steps of integer value

The lower limit (i.e. H2H traffic at 0.1E) is considered here and integer values of [14, 5, 2 and 1] used as the increment values (j) of the BFZ when required. As shown in the result, from 0 to 0.2E of M2M generated traffic, the step at which BFZ increases is immaterial because the total traffic is below the s-ALOHA capacity and therefore collisions are fewer and a lower BFZ is required. However, from 0.2E to 0.5E of M2M traffic, an incrementing

value of 14 exhibits a different behaviour from the other values where the BFZ increases and levels at 30. It can be seen that the value of 14 performs worst at 0.3E; this is because the load is at upper limit and collisions increase here which increases the rate of the BFZ increment. Hence the performance worsens since the value of 14 is higher. From 0.5E to 0.7E of M2M traffic, a value of the 14 has a better BFZ performance, because the system is steady (within the $P_{B\,thr}$) due to the higher incrementing BFZ value that makes the BFZ to have enough slots for H2H retransmissions. Finally the value of 14 converges at around BFZ of 60 similar to what is obtained in figure 7.8(a). On the other hand, integer values of [5, 2 and 1] have similar effect on the BFZ with 5 having the best effect at higher M2M load where it converged at around BFZ value of 48.

Figure 7.10 (a) and 7.10 (b) present the running throughput at upper and lower limits respectively. This is presented to show that the system is still useful on the process of obtaining the BFZ value. In addition, both results also agree with the fixed BFZ results presented earlier where M2M RACH-throughput decreases with the increase in the BFZ and the converse is obtained for the H2H RACH-throughput. Finally the results also confirm the convergence of the best BFZ value especially 7.10 (b) which presents the lower limit BFZ convergence at 48.

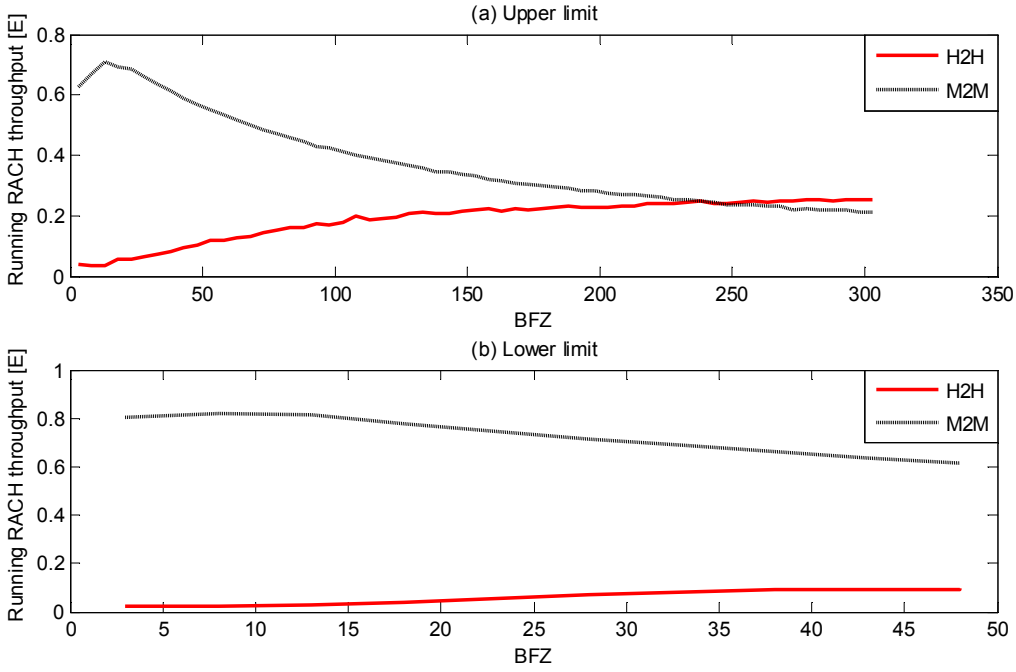

Figure 7.10 Upper and lower limit running RACH-throughput of dynamic BFZ

Results presented in Figure 7.11 compare the throughput performance of the two different schemes to analyse the effect of the dynamic BFZ on the FB-QL-RACH scheme.

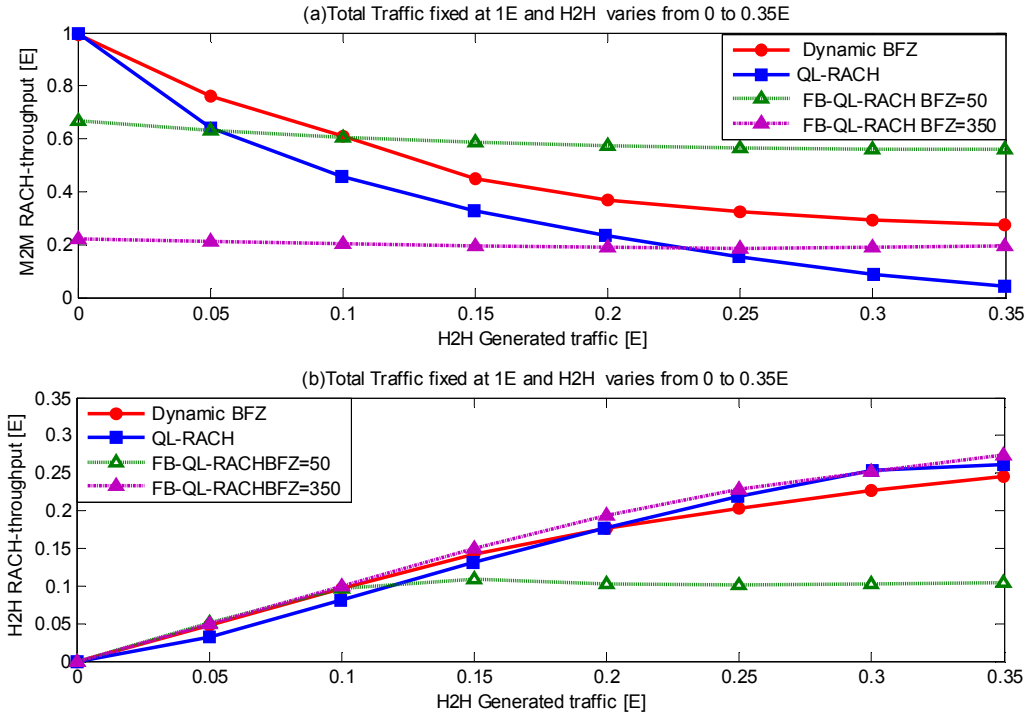

Figure 7.11 RACH-throughput comparisons of QL-RACH, FB-QL-RACH and Dynamic BFZ schemes

The dynamic BFZ scheme shows better M2M RACH-throughput performance, from 0 to 0.1E of H2H traffic load compared to other schemes as presented in Figure 7.11(a). This is because a lower BFZ is required here (as confirmed in figure 7.9) since the uncontrolled traffic (H2H) is generating less load that minimises the level of the collisions. Above 0.1E of H2H generated traffic, the dynamic BFZ scheme offers better M2M RACH-throughput performance than the other two schemes with the exception of FB-QL-RACH at BFZ value of 50. This shows that at the lower limit, the BFZ value of 50 provide the best performance for M2M; however that is at the expense of H2H performance. Therefore the FB-QL-RACH scheme can be implemented with a dynamic BFZ which has been shown to provide improved M2M RACH-throughput performance much better than the QL-RACH scheme. It also helps to understand the trade-off between H2H and M2M performances with respect to the BFZ value. On the other hand, the dynamic BFZ scheme shows a good effect to the H2H RACH-throughput where up to 100% performance is achieved at the lower H2H generated traffic as shown in Figure 7.11(b).

7.7 Summary

This chapter presented an enhanced back-off scheme to be used with QL-RACH that enables M2M traffic to be effectively controlled when sharing RACH channel with H2H. The new approach is designed to operate with the existing cellular network standards, significantly enhancing its usability. The scheme has improved the performance of the QL-RACH which is not efficient at higher H2H loads. FB-QL-RACH incorporates a separate frame for H2H back-off, and proposed no back-off for M2M. The scheme is shown to reduce the probability of collision between H2H and M2M user groups which normally happens as a result of H2H re-transmissions in the M2M-frame. Also the probability of idle slots caused by the M2M back-off strategy is eliminated since M2M use no back-off. The performance of the FB-QL-

RACH scheme is observed using different BFZ values and a significant improvement in RACH-throughput has been realised.

In addition the interaction between Poisson (typical for H2H) and Periodic (typical for some M2M applications) traffic distributions is considered. The results show that the M2M traffic can be controlled using the QL-RACH scheme with both sync and u-sync Periodic traffic distributions.

The chapter also introduced a new scheme that enables the eNB to automatically update the BFZ value for the H2H B-frame based on a threshold of probability of RACH access blocking. This makes the scheme more practical and has improved the performance especially at lower traffic loads where fewer collisions occur and require a lower BFZ.

Finally, the chapter also presented an analytical model that predicts the RACH-throughput performance of the FB-QL-RACH scheme. The performance of the model is compared with the simulation model and the results shown an agreement between the two models.

Chapter 8. Frame ALOHA and QL-RACH for combined RACH access

8.1 Introduction

The FB-QL-RACH presented in chapter 7 was able to reduce probability of collision between H2H and M2M user groups and significantly improve the RACH-throughput performance. However collisions between the H2H and M2M user groups still occur since they share a frame during the first access attempt.

This chapter introduces a simpler scheme that enables collision free RACH access between the H2H and M2M users. Separate frames (H2H-frame and M2M-frame) are proposed to be used by H2H and M2M during their first attempt and retransmissions (for H2H) to realise Frame-ALOHA for QL-RACH (FA-QL-RACH). Design and operation of the FA-QL-RACH scheme is described in section 8.2. Section 8.3 presents the discussion of the FA-QL-RACH scheme performance and its comparison with other schemes presented in the previous chapters. Finally the chapter is summarised in section 8.4.

8.2 Frame ALOHA for QL-RACH (FA-QL-RACH)

FA-QL-RACH is a scheme introduced to enable collision free combined RACH access between H2H and M2M user groups. A separate frame (H2H-frame) is assigned for H2H user access during both first attempt and for retransmissions. Also M2M users maintain their transmissions in the M2M-frame implemented by the QL-RACH scheme described in the previous chapters. Therefore, as shown in figure 8.1, another global-frame similar to that in

chapter 7 is designed, however there is no common access frame between the H2H and M2M user groups.

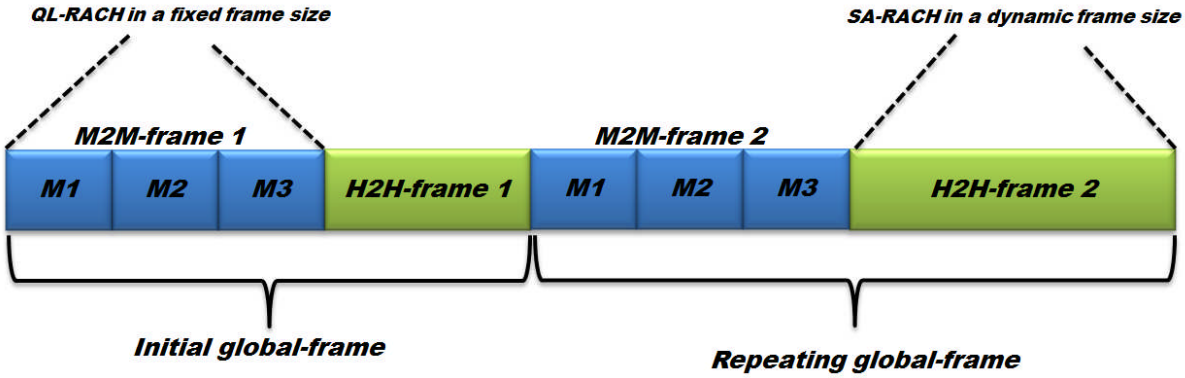

Figure 8.1 FA-QL-RACH scheme frame structure

Figure 8.1 presents the operation of the FA-QL-RACH scheme in which a scenario of combined RACH access is presented. Three M2M users in steady state (having acquired dedicated slots) are considered using the M2M-frame which is fixed to the number of M2M users. After the M2M-frame comes the H2H-frame which is used for H2H traffic and controlled by the SA-RACH scheme. Then both frames repeat as a global-frame within the LTE main PRACH resource. Hence using this configuration in the combined RACH access provides better access procedure than the FB-QL-RACH by eliminating the probability of collision (between H2H and M2M user groups) in (7.1). Therefore, the FA-QL-RACH scheme enables zero collision between the H2H and M2M user groups since each group is allocated a different frame.

Unlike the M2M-frame size which is fixed, the H2H-frame size is dynamic and the size is controlled by the eNB in a similar process described in chapter 7. i.e. each H2H user sends its cumulative number of blocking and successful transmission via message 3 of the RACH request procedure to the eNB after every defined window (frame time) period. This information (from different H2H users) is used by the eNB to obtain the H2H system

blocking probability within the window, compare it with the threshold and determine an appropriate H2H-frame size. For details of the dynamic frame size implementation see chapter 7.

8.3 FA-QL-RACH Results and Discussions

The simulation scenario considered here is similar to that of chapters 6 and 7 where the LTE standard is considered by assuming that both H2H and M2M user groups are deployed within a cell area with a single preamble. Both users have equal right of access with each user group having different load ratios and RACH access scheme. The simulation parameters presented in Table 6.1 are used here.

A single preamble is also used here to examine the RACH-throughput performance of the FA-QL-RACH scheme. The effect of the FA-QL-RACH scheme is shown by comparing its RACH-throughput and delay performances with that of the other schemes presented in the previous chapters. In addition to the legend interpretations of the results presented in chapters 6 and 7, the legend bearing FA-QL-RACH represents the results of the new scheme introduced in this chapter.

Figure 8.2 presents the RACH-throughput performance of the FA-QL-RACH scheme at the upper limit compared with all the schemes presented in the previous chapters.

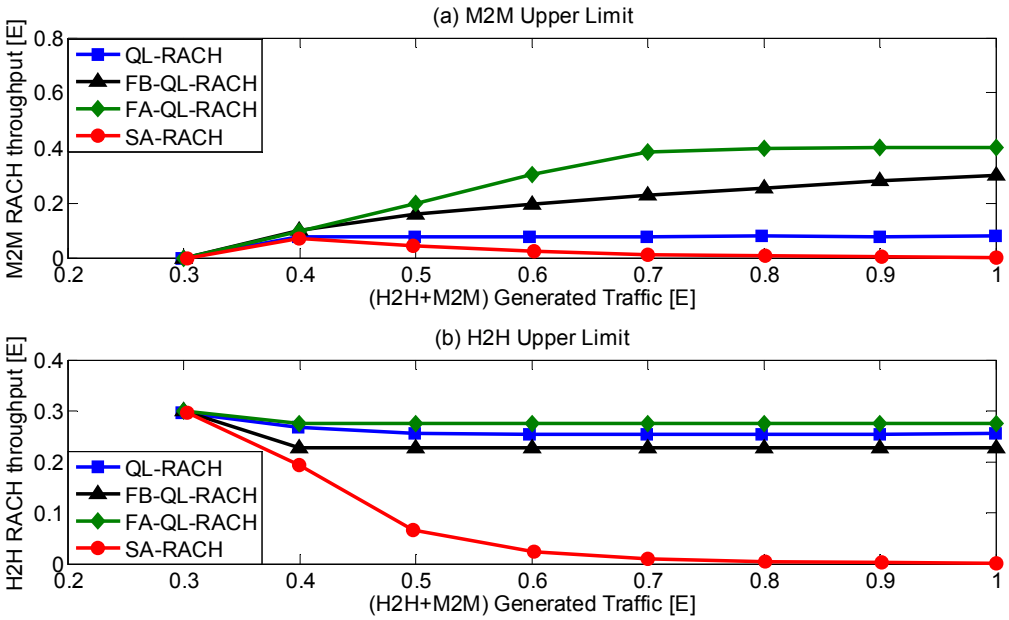

Figure 8.2 Upper limit RACH-throughput performances of FA-QL-RACH, FB-QL-RACH, QL-RACH and SA-RACH schemes

Figure 8.2(a) illustrates the M2M RACH-throughput performance and as shown the result representing the FA-QL-RACH scheme offers better performance than the remaining two proposed schemes (FB-QL-RACH and QL-RACH). The performance increases with increases in the generated traffic and levels off at 0.4E. The best throughput performance provided by the scheme shows the effect of using separate frames for each of the two user groups where collision free RACH access between H2H and M2M user groups is realised. The next better scheme is FB-QL-RACH (note that the dynamic BFZ is used here) which offers a RACH-throughput of about 0.3E. The FB-QL-RACH scheme is lower than FA-QL-RACH because the two user groups still share the access frame during the first attempt, i.e. the frames are only separated during retransmissions as described in chapter 7. The low performance of the QL-RACH scheme is as a result of the random effect of the uncontrolled H2H traffic as explained in chapter 6.

Figure 8.2(b) presents the results of the H2H RACH-throughput performance in a similar arrangement to the results presented in 8.2(a). Also here still the FA-QL-RACH performs better where 0.28E is achieved, followed by QL-RACH and FB-QL-RACH scheme.

In conclusion comparison between the upper limit M2M RACH-throughput performances of the FA-QL-RACH in figure 8.2 and FB-QL-RACH presented in figure 7.4 shows that, the FB-QL-RACH scheme with fixed BFZ value of 50 has higher throughput performance than the FA-QL-RACH scheme. This is due the reason described in chapter 7 that M2M performs better at upper limit with BFZ value of 50 but affects the performance of H2H. This will be clearer if an H2H performance in figure 7.4(b) is compared with that in figure 8.2(b) where the former dropped to about 0.1E (due to fixed BFZ of 50) and the latter has 0.28E. Therefore even though M2M performance of FA-QL-RACH is lower than that of FB-QL-RACH with fixed BFZ of 50, FA-QL-RACH scheme is still better than FB-QL-RACH when looking at the H2H performance.

Figure 8.3 is the lower limit version of the results presented in figure 8.2, where also the FA-QL-RACH performance is compared with the other proposed schemes.

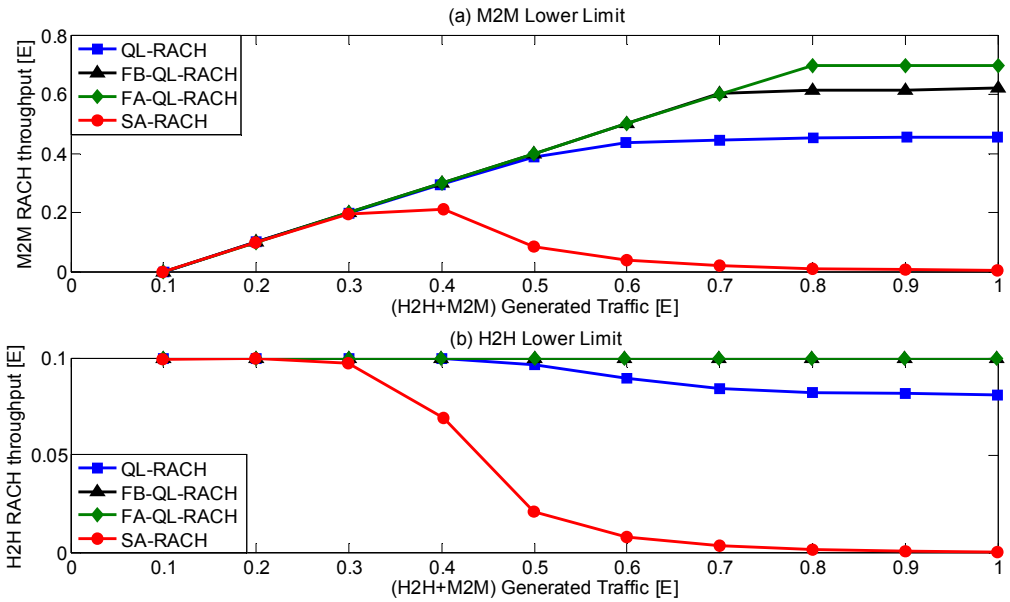

Figure 8.3 Lower limit RACH-throughput performances of FA-QL-RACH, FB-QL-RACH, QL-RACH and SA-RACH schemes

Figure 8.3(a) shows the results of the M2M RACH-throughput performance representing FA-QL-RACH, FB-QL-RACH, QL-RACH and SA-RACH schemes. The results exhibit similar trends to the results presented in figure 8.2(a) with better performance. As shown, the performance of all the schemes are better here since the H2H load is at lower limit (far away from s-ALOHA capacity) as explained in the previous chapter. It can be seen that the performance of the FA-QL-RACH scheme also increases with increases in the generated traffic and levels at 0.7E. On the other hand, figure 8.3(b) presents the H2H RACH-throughput performance with similar behaviour to the results presented in figure 8.2(b).

In general, notice the performance of the FA-QL-RACH scheme, where 0.1E is achieved irrespective of the M2M generated traffic. This result indicates a full protection for the H2H user group which is a better effect to the upper limit performance shown in figure 8.2(a) where a H2H RACH-throughput of about 0.28E (slightly lower than the generated 0.3E) is obtained. Therefore this indicates the scheme has no operational limit in terms of H2H

performance at the lower limit, and drops just 0.02E from the H2H generated traffic at the upper limit.

Figures 8.4 shows the comparison of the M2M and H2H performances of average end-to-end delay of FA-QL-RACH, FB-QL-RACH, QL-RACH and SA-RACH schemes at the upper limit of H2H traffic load.

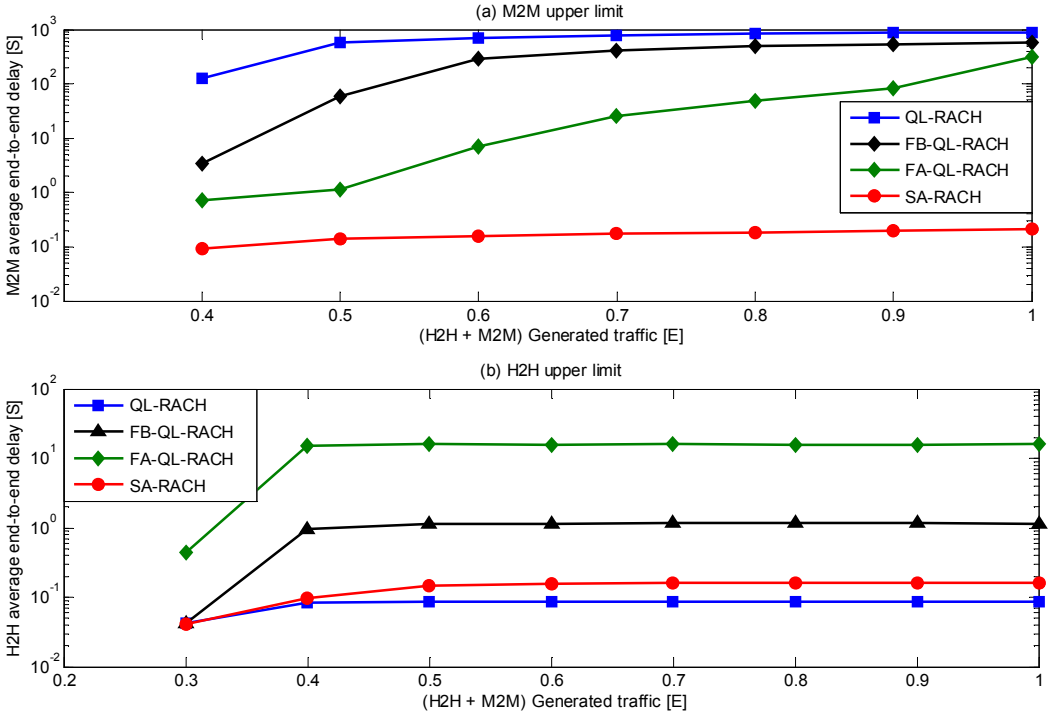

Figure 8.4 Upper limit average end-to-end delay performances of FA-QL-RACH, FB-QL-RACH, QL-RACH and SA-RACH schemes

Figure 8.4(a) presents the M2M average delay and as shown the SA-RACH scheme performance is better than the performances of the other three schemes. This is because the SA-RACH (as the existing scheme) has no transmission restriction i.e. the transmission is not is not controlled which minimises the queue building. On the other hand, by comparing the M2M delay performances of the three schemes, it can be observed that the delay of the FA-QL-RACH scheme has a better performance compared to the remaining other two schemes. The better M2M delay performance achieved by the FA-QL-RACH scheme is because there

is no collision between H2H and M2M user groups (since they don't share frame) where every M2M transmission attempt is successful. This will be clearer by observing M2M delay performance of the QL-RACH scheme which is the worst of the three proposed schemes since here collisions between the two user groups occur and M2M used BEB for retransmissions as explained in chapter 6.

The H2H average delay performances of the schemes are compared in figure 8.4(b) where SA-RACH performance is similar to the M2M since they both used the same scheme. However the delay performance of H2H with SA-RACH scheme is not better here, it can be seen that the QL-RACH scheme has the best delay performance. This is because the M2M traffic is controlled using the QL-RACH scheme and the BEB method used by the M2M gives the H2H users more access rate which helps the delay performance. However, the H2H delays of the other two proposed schemes are worse than the performances of the QL-RACH and SA-RACH schemes. This is because the H2H transmission is restricted here to frame-based with the performance of the FB-QL-RACH scheme better since at least the H2H initial access is allowed in the M2M-frame as described in chapter 7.

Figure 8.5 presents the lower limit M2M and H2H delay by comparing the performances of the FA-QL-RACH, FB-QL-RACH, QL-RACH and SA-RACH schemes.

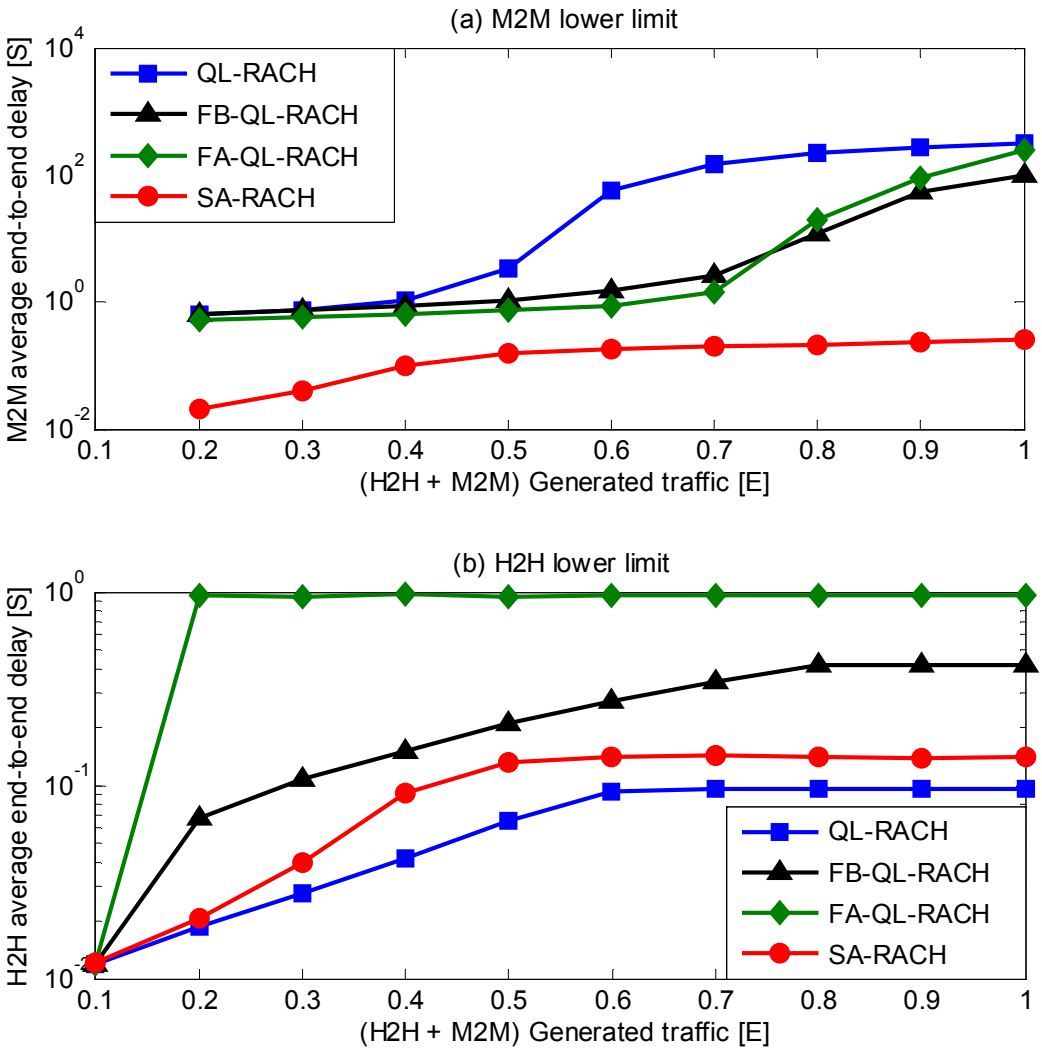

Figure 8.5 Lower limit average end-to-end delay performances of FA-QL-RACH, FB-QL-RACH, QL-RACH and SA-RACH schemes

Figure 8.5(a) shows the M2M delay performance at the lower limit of H2H traffic. Due to the same reason given above, it can be seen that the results exhibit similar behaviour to the performances presented in figure 8.4(a). i.e. the SA-RACH scheme performs better followed by the FA-QL-RACH and FB-QL-RACH schemes and lastly with the QL-RACH scheme having the worst performance. In general a better delay performance is achieved at the lower limit. This is due to the low traffic level of the H2H which reduce the collisions amongst H2H user group and between H2H and M2M user groups in such a way that the system requires low frame size (both H2H-frame and H2H B-frame) for both the FB-QL-RACH and FA-QL-RACH schemes. On the hand, for the same reasons described above the delay results

presented in figure 8.5(b) have similar performance to those in figure 8.4(b) with QL-RACH having better performance and FA-QL-RACH having the worst performance. Also the H2H delay performance at the lower limit is better than that at the upper limit due to the same reason discussed above.

Finally, it is clear from the delay performances of M2M at both upper and lower limit that, the FA-QL-RACH scheme which offers best RACH-throughput performance also presents better delay. This is followed by the FB-QL-RACH and then QL-RACH schemes.

8.4 Summary

This chapter introduced the FA-QL-RACH scheme which is the extension of the FB-QL-RACH scheme (presented in chapter 7) that enables collision free RACH access between the H2H and M2M user groups. Separate frames designed and allocated to the two user groups which enables them to access the RACH without sharing a frame.

In general the FA-QL-RACH scheme has simpler implementation than the FB-QL-RACH scheme since the H2H and M2M user groups do not share frame. In addition to the implementation simplicity, the FA-QL-RACH scheme also provides significant RACH-throughput improvement with better delay performance. The performances indicate that the scheme has eliminated collisions between the H2H and M2M users groups. The simulation results show that the upper limit H2H and M2M performances (both RACH-throughput and delay) are better than that of all the prosed schemes (QL-RACH and FB-QL-RACH) presented in the previous chapters of this thesis. Also the lower limit provides similar performance with better protection to the H2H user group where 100% H2H RACH-throughput is obtained (irrespective of the M2M generated traffic).

Chapter 9. Further Work

The area of M2M (cellular) communication is still in its infancy and from all indications, cellular is the potential network for long range M2M communication. However the new cellular M2M systems will face challenges, one of which has been considered in this work. Different schemes that enable effective interaction between H2H and M2M users have been proposed here. The following sections present potential further work related to the research reported in this thesis.

9.1 The effect of M2M device mobility on the QL-RACH scheme

Q-learning has been used to control M2M users to acquire dedicated slots for collision free RACH access to realise the QL-RACH scheme in this thesis. The scheme is the basis for all proposed M2M RACH access schemes in this work. The QL-RACH scheme is implemented by designing an M2M-frame of a fixed size equals to the number of active M2M users in the system in order to achieve maximum performance as described in chapter 6. The assumption of fixing the number of M2M users in the system is reasonable considering that most of M2M applications are characterised by non-mobile devices. However, some applications that necessitate device mobility may exist and that will require a potential technique to improve the learning process to adapt to the dynamic conditions. A dynamic M2M-frame size that would cater for the M2M users moving in/out of the RACH access contest need to be considered here.

9.2 Priority Based QL-RACH

Looking at the performance of all the proposed schemes in this thesis, it is clear that, they significantly improve the RACH-throughput performance of both H2H and M2M users especially compared with the SA-RACH scheme. However as described in the previous chapters each scheme has its operational limitation(s) in terms of protecting the performance of the H2H as the existing user group.

The performances of the FB-QL-RACH and FA-QL-RACH schemes presented in chapters 7 and 8 are directly controlled by the BFZ and H2H-frame size respectively. For example, the upper limit performance of the FB-QL-RACH with fixed BFZ presented in chapter 7 shows that all the different values of the BFZ used provide a H2H RACH-throughput lower than the desired 0.3E. Also the lower limit performance shows that the performance of H2H using fixed BFZ value of 30 decreases at higher generated traffic. Similar performance is observed for the FB-QL-RACH scheme with dynamic frame size. On the other hand, the FA-QL-RACH scheme at the upper limit is also shown to provide RACH-throughput of 0.28E which is slightly lower than the desired 0.3E.

Therefore, based on the above descriptions, priority could be given to the H2H user group (by adjusting the frame size) to protect its performance from the impact of the M2M traffic. Increasing the frame size delays the M2M access and gives more room for the H2H traffic. An alternative approach is by controlling M2M traffic generation which could reduce the level of its interaction with H2H and also offers the priority to H2H transmissions. For example, looking at the performance at a BFZ value of 30 presented in figure 7.5 where the performance of H2H is fully protected up to 0.5E of M2M generated traffic. Therefore the H2H performance can be protected by limiting the M2M generated traffic to 0.5E or by periodically switching off the M2M traffic generation using a determined duty cycle.

Conclusively, introducing a priority based scheme to protect H2H performance against the impact of the M2M traffic will enhance the efficiency of the FB-QL-RACH and FA-QL-RACH schemes.

9.3 Slot Sharing strategy in the QL-RACH

The implementation of the novel QL-RACH scheme presented in chapter 6 required an M2M-frame of size (number of slots) equal to the number of M2M users. In this way a maximum M2M RACH-throughput performance will be achieved. This is based on the assumption that all the M2M users (devices) in the network have the same/similar traffic behaviour and that the mean RACH request rate is less than the M2M-frame time. i.e. at 100% M2M RACH request rate each M2M user has a request to send in an M2M-frame. However, the fast trends in the current technological development foresee various applications of M2M that may have different traffic patterns. Some of these applications may have devices that generate RACH requests within a time interval higher than that of the time of an M2M-frame. Therefore based on that, some devices (depending on their RACH request inter-arrival time) may be wasting their acquired slots in the M2M-frame and this definitely will reduce M2M-RACH throughput performance. Introducing a slot time sharing technique could enhance the QL-RACH scheme by grouping a number of M2M devices with different RACH request inter-arrival times. A scenario of smart home application of the proposed scheme is presented in Figure 9.1. The figure presents three smart homes that have different M2M applications using number of devices. As shown each home acquired a dedicated slot to be shared by devices that are assumed to have different RACH request inter-arrival times. Therefore the slots could be shared without collision amongst the grouped M2M devices and also maximise the usage of the M2M-frame.

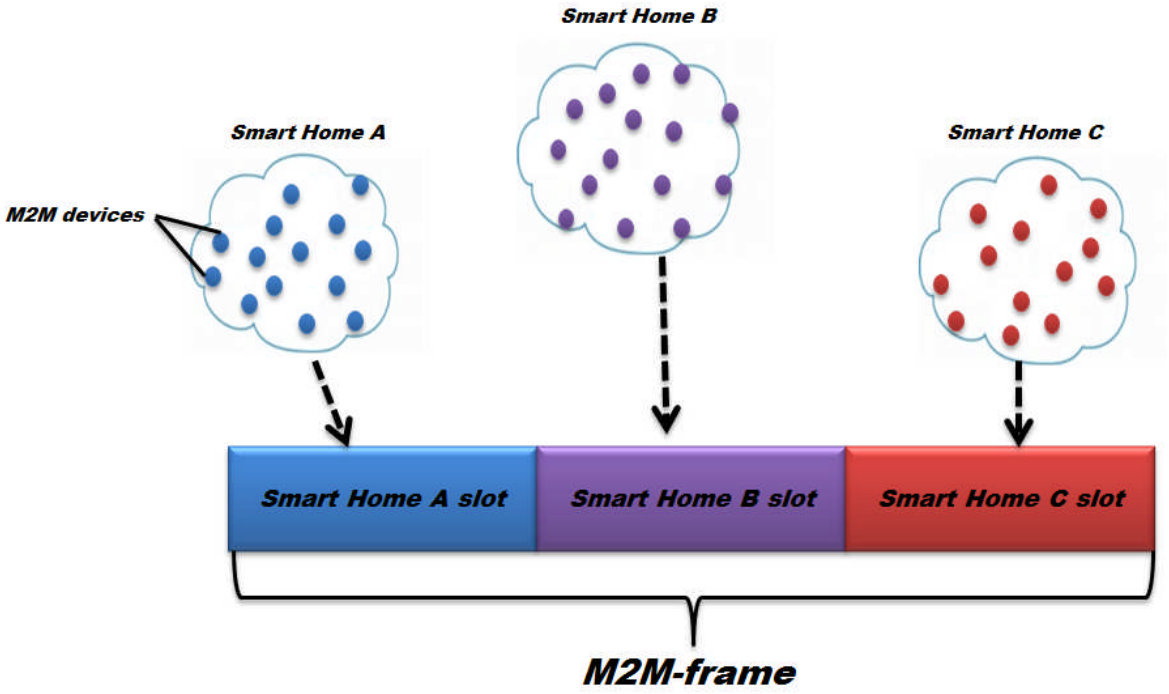

Figure 9. 1 Representation of slot sharing strategy

9.4 Energy Aware QL-RACH scheme

The QL-RACH scheme introduced in this work has not looked at the energy consumption of the M2M devices. On the other hand, energy considerations are becoming necessary in the design of wireless networks [104], which are characterised as a critical issue in M2M communication by the 3GPP standard in one of its technical specifications [105]. Even though the schemes introduced in this works are energy efficient since they have the capability to reduce collisions, however research on how to improve the energy efficiency of M2M RACH access is as equally important as the design of an efficient access scheme. Authors of [106] considered energy saving in M2M RACH access using the slotted ALOHA scheme. The energy efficiency can be improved by adopting the use of message 3 (in LTE network) for the transmission of M2M information data. This is possible since some of the applications of M2M deal with small payload. Another way is the slot sharing strategy above could be implemented by grouping the M2M devices based on power consumption.

Chapter 10. Summary and Conclusions

This work considers the interaction of H2H and M2M traffic sharing the RACH of an existing cellular network. Also Q-learning has been applied to implement a novel RACH access scheme with the introduction of intelligent access and back-off strategies to improve the learning effect.

Performance of the RACH channel using the existing scheme (slotted ALOHA) when additional M2M traffic is supported has been examined where it was realised that a better scheme is required to control the new system. Q-learning is applied to control the M2M traffic while the H2H traffic maintains its existing scheme. It has been shown that the introduced method is capable of controlling the interaction of the H2H and M2M users in sharing the RACH with improved overall system performance. A general summary and conclusion of the work in this thesis is presented here by highlighting the main contributions of the research.

The first chapter provides a general introduction of the thesis and motivation of this work. Chapter 2 presented an overview regarding the concept and motivation of M2M communication. Types of access networks have been used as the basis for M2M classifications as short range, long range and hybrid. In addition some of the possible application scenarios of M2M with examples are presented. Also the concept of Reinforcement Learning with its different solution methods has been discussed. A justification for choosing Q-learning algorithm of the temporal difference method over the other algorithm/methods has been explained.

Brief background information on different types of cellular systems and their resource management is provided in chapter 3. An evolutionary trend of different generations of

cellular network has been presented to highlight the difference and similarities in their technical architecture. This has been coupled with a description of various multiple access techniques used in coordinating cellular network resources. The chapter also provides information about channel classifications, the structure and functions of cellular networks with an emphasis on the GSM and LTE standards. This is to provide sufficient background information to design an efficient MAC layer protocol to access an initial uplink signalling channel of the network. The RACH is the initial uplink signalling channel and relevant in this thesis, therefore details of its structure and access protocol have been provided.

Chapter 4 describes the modelling techniques and methods used to evaluate the system performance as well as validation methods. Three evaluation methods are discussed and the importance of using simulation in the realisation of communications system is described. Simulation is emphasised because the results in this thesis are obtained through simulation using Matlab. To validate the results in this thesis, analytical models have also been developed to match the simulation results.

Chapter 5 introduces and motivates the proposed scheme presented in chapters 6 to 8. The research in this thesis is based on s-ALOHA, therefore the ALOHA scheme is first introduced and the performances of s-ALOHA and p-ALOHA are analysed and compared. The performance of SA-RACH (using s-ALOHA for RACH access) has been analysed and the system is pushed beyond the ALOHA capacity to establish the cause of RACH instability. This is to illustrate the impact of additional M2M traffic if the system is controlled by the SA-RACH scheme. The throughput, average end-to-end delay and blocking probability are used to evaluate the performance of the SA-RACH scheme using: the retransmission limit and back-off interval window, as the two variable parameters that control the retransmission strategy. From the results, it was realised that using a retransmission limit of 7 and back-off interval window of 14 RACH slots provide a better throughput performance. This is also

similar to the GSM standard and therefore the technique is adopted in this work. An analytical model that predicts the RACH-throughput performance is also presented, where new equations are developed using the SA-RACH scheme with retransmissions.

A new RACH access scheme is proposed in chapter 6, to reduce the impact of the additional M2M traffic on the existing H2H users in the cellular network. A learning based RACH access (QL-RACH) scheme is introduced where Q-learning has been used by M2M users to self-organised themselves to avoid collisions by learning a dedicated slot within the M2M-frame. The scheme is generally applicable to all cellular network standards since they all have a similar RACH structure and access protocol. The QL-RACH scheme has shown the ability to enable H2H users to coexist with M2M users by providing channel stability with RACH-throughput performance improvement.

The QL-RACH scheme provides channel stability and improves RACH-throughput; however its performance is affected by the random effect of the uncontrolled H2H traffic (especially at the higher H2H traffic loads). Chapter 7 introduces an enhanced back-off scheme called FB-QL-RACH to be used with the QL-RACH. A separate frame for H2H user retransmission is applied and no back-off for M2M is introduced.

In addition, the chapter also considers the interaction between Poisson (typical for H2H) and periodic (typical for M2M) traffic distributions. It has been shown that the periodic traffic can also be controlled by the QL-RACH scheme.

Chapter 8 introduced a collision free scheme between H2H and M2M users in the RACH contest. This has been achieved by extending the previous proposed schemes where separate frames are designed and allocated to the individual user groups (H2H and M2M) and enable them to access the RACH without sharing a frame. FA-QL-RACH is the realised scheme and has shown better performance than all the other proposed schemes.

10.1 Novel Contributions

As far as we know, this thesis is the first to implement and examines the interaction between H2H and M2M traffic with both user groups having different RACH access scheme. Also this work is the first to consider applying learning to access the RACH channel of a cellular network in the presence of the existing non learning H2H users. The major contributions are summarised as follows:

10.1.1 Intelligent RACH access strategy using Q-learning

- A new RACH access scheme using Q-learning (QL-RACH) to control M2M traffic which has the potential to improve the RACH-throughput performance over the existing s-ALOHA scheme has been developed. The process enables M2M users to learn to avoid each other during the RACH contention period without involvement of any central entity. Combining the QL-RACH scheme with the existing RACH scheme (used by the H2H users) reduces the overall number of collisions since there will be no collisions among the M2M users after convergence. As described in chapter 6, the learning was implemented by designing a virtual frame (from the main cellular frame) of RACH or RA slots on which the M2M users acquire their unique dedicated slots.
- The performance of the QL-RACH scheme has been evaluated by predicting H2H traffic load as the existing traffic in the cellular network. The s-ALOHA capacity limit has been used as the base line for the prediction where 0.3E and 0.1E were chosen as upper and lower H2H load limits respectively. For the upper limit, the QL-RACH scheme only provides RACH stability without improving the RACH-throughput. This is as a result of the random effect of the uncontrolled H2H traffic that dominates the QL-RACH performance. On the other hand, at the lower limit, a significant improvement of the RACH-throughput is obtained and also the RACH channel is

stable. This is because even though the H2H traffic is not controlled, the random effect has no impact on the learning scheme. Also it was observed that at both the upper and lower limit, the implementation of the QL-RACH scheme does not negatively affect the end-to-end delay performance of the H2H users.

- To protect the M2M users from losing their dedicated slots after convergence (especially at the upper limit), chapter 6 also introduced a novel punishment method of the Q-learning. A variable punishment method has been adopted instead of the conventional fixed one.

The above contributions have been published by the IEEE and presented at the *20th European Wireless Conference (EW '14) in Barcelona, Spain.* Also a patent has been filed form the above contributions.

10.1.2 Enhanced H2H back-off for the Q-learning RACH access

- Frame-based Back-off with QL-RACH (FB-QL-RACH) is a novel scheme introduced in chapter 7 as an enhanced back-off scheme. This was implemented using a separate frame for H2H back-off which is used to redirect the H2H traffic in retransmissions where the H2H users are restricted from retransmitting in an a M2M-frame. M2M users maintained the QL-RACH scheme without applying back-off and are not allowed to use the H2H back-off frame. This process is used to reduce the probability of collision between the H2H and M2M users, since the collisions occur only during the first attempt. In general, with the M2M traffic controlled by QL-RACH the overall system collision probability is significantly reduced by the FB-QL-RACH scheme. Also the scheme eliminated the probability of idle slots caused by the M2M back-off. The introduced FB-QL-RACH scheme shown to have better RACH-throughput performance depending of the H2H back-off frame size used.

- ❖ Application of Q-learning has been tested based on the interaction between the conventional Poisson traffic distributions and periodic where it was concluding that the periodic traffic can also be controlled by the QL-RACH scheme.

These contributions have been published by IEEE and presented at *Telecommunication Network and Applications Conference (ATNAC), 2014 Melbourne Australia.*

10.1.3 Dynamic frame size

A new scheme that enables an eNB to control the H2H back-off frame size is also introduced in chapter 7. The scheme changes the nature of the frame size from the fixed to a dynamic using system H2H blocking probability. This is used to maximise the performance of the FB-QL-RACH scheme with a fixed frame size so that the exact required frame is used at a given time. The dynamic frame size scheme has been shown to have better performance compared to the fixed one, since the performance directly depends on the frame size.

10.1.4 New throughput equations for the intelligent RACH access scheme

Analytical models are developed to predict the throughput capability of the SA-RACH scheme, QL-RACH scheme and FB-QL-RACH scheme considering a combined RACH access scenario. New equations for the probability of success and collisions of H2H and M2M are determined and used to define new throughput equations. The model is first developed without considering user retransmission and is then extended to consider the retransmissions using the retransmission strategy adopted by the different schemes as described in chapters 5 and 6. Each of the analytical models developed are compared with a simulation models and the results obtained shown an agreement between the two models. This contribution and that of the dynamic frame size presented in 10.1.3 have been submitted to *Transactions on Emerging Telecommunications Technologies (ETT) journal.*

10.1.5 Collision free H2H and M2M interactions for RACH access

Another novel scheme introduced to enable collision free combined RACH access between H2H and M2M user group is presented in chapter 8. This scheme assigns a separate frame for the H2H user access during both the first attempt and the subsequent retransmissions while the M2M users maintain the M2M-frame. Therefore here there is no sharing of frame between the two groups which makes the scheme simpler than the FB-QL-RACH scheme of chapter 7. In addition to its simplicity the scheme is able to eliminate collisions between H2H and M2M user groups and therefore improve the RACH-throughput with better performance than the FB-QL-RACH scheme. This contribution has been accepted for publication and is to be presented in the *9th International Conference on Next Generation Mobile Applications, Services and Technologies (NGMAST'15) Cambridge, UK.*

Glossary

ACB	Access Class Barring
AGCH	Access Grant Channel
AMPS	Advanced Mobile Phone System
BCH	Broadcast Channel
BCCH	Broadcast Control Channel
BEB	Binary Exponential Back-off
BFZ	Back-off Frame size
BP_{thr}	Blocking probability threshold
BS	Base Station
BSC	Base Station Controller
BTS	Base Transceiver Station
CBCH	Cell Broadcast Channel
CCCH	Common Control Channel
CDF	Cumulative Distribution Function
CDMA	Code Division Multiple Access
CP	Cyclic Prefix
DCCH	Dedicated Control Channel

DECT	Digital Enhanced Cordless Telecommunication
DES	Discrete Event Simulation
DL	Down-link
DwPTS	Downlink part
eNB	Evolved Node B
E-health	Electronic health
ETSI	European Telecommunications Standards Institute
FACCH	Fast Association Control Channel
F-ALOHA	Frame-based ALOHA
FA-QL-RACH	Frame-ALOHA for QL-RACH
FB-QL-RACH	Frame-based Back-off for QL-RACH
FCH	Frequency Correction Channel
FDD	Frequency Division Duplex
FDMA	Frequency Division Multiple Access
FN	Frequency Number
G_{H2H}	H2H generated traffic load
G_{M2M}	M2M generated traffic load
GOS	Grade of Service
GP	Guard Period

GPRS	General Packet Radio Service
GSM	Global System for Mobile
G_{total}	Total System's generated traffic load
$G_{T_{total}}$	Total system's aggregated traffic from both H2H and M2M groups
$G_{T_{BF_{H2H}}}$	H2H-frame aggregated traffic using FB-QL-RACH scheme
$G_{T_{BF_{M2M}}}$	M2M-frame aggregated traffic using FB-QL-RACH scheme
$G_{T_{BF_{total}}}$	Total system's aggregated traffic using FB-QL-RACH scheme
GW	Gate Way
H2H B-frame	H2H Back-off frame
$H2H_{BP}$	H2H blocking probability
H2H	Human-to-Human
H2H-frame	Frame used for H2H transmissions
IEEE	Institute of Electrical and Electronics Engineers
IP	Internet Protocol
IoT	Internet of Things
ISO-OSI	International Standard Organisation-Open Systems Interconnection
LTE	Long Term Evolution
M2M	Machine-to-Machine

M2M-frame	Frame used for M2M transmissions
MAC	Medium Access Control
Matlab	Matrix Laboratory
MIMO	Multiple Input Multiple Output
MS	Mobile Station
MSC	Mobile Switching Centre
N_C	Number of collision
NFC	Near-Field Communication
OFDMA	Orthogonal Frequency Division Multiple Access
p-ALOHA	Pure ALOHA
PAPR	Peak-to-Average Power Ratio
P_B	Probability of Blocking
PCH	Paging Channel
$P_{C\text{-}p\text{-}ALOHA}$	Probability of collision using p-ALOHA protocol
$P_{C\text{-}s\text{-}ALOHA}$	Probability of collision using s-ALOHA protocol
P_{idle_slot}	Probability of idle slots
$P_{M_{succ}}$	Probability of M2M successful transmission in the combined access
$P_{no_{coll_{H2H}}}$	Probability of no collision from H2H
PRACH	Physical Random Access Channel

$P_{Success}$	Probability of success
$P_{H_{succ}Q_{Aloha}}$	Probability of H2H successful transmission in the combined access
Q-Frame	Q-learning Frame
QL-RACH	Q-learning based-RACH access
Q-value	Q-learning acquired weight value during the learning process
RA	Radom Access
RACH	Random Access Channel
RAO	RA Opportunity
RAR	RA Response
RB	Resource Blocks
R_b	Number of blocked RACH requests
RFID	Radio-Frequency Identification
Rs	Number of successful RACH request
r_{total}	Total traffic in retransmissions from both H2H and M2M groups
SACCH	Slow Associated Control Channel
SA-RACH	slotted ALOHA based RACH access
s-ALOHA	slotted ALOHA
SC-FDMA	Single Carrier OFDM
SCH	Synchronisation Channel

SDCCH	Stand-alone Dedicated Control Channel
S(E)	Throughput in Erlang
$S_{H_{SA-RACH_{retrans}}}$	H2H throughput using SA-RACH protocol with retransmissions
$S_{H_{Q_{Aloha}}}$	H2H throughput of the combined SA-RACH and QL-RACH scheme
$S_{H_{Q_{Aloha_{FB}}}}$	H2H throughput of the combined SA-RACH and FB-QL-RACH scheme
$S_{H_{Q_{Aloha_r}}}$	H2H throughput of the combined SA-RACH and QL-RACH schem
SIM	Subscriber Identity Module
$S_{M_{SA-RACH_{retrans}}}$	M2M throughput using SA-RACH protocol with retransmissions
$S_{M_{Q_{Aloha}}}$	M2M throughput of the combined SA-RACH and QL-RACH scheme
$S_{M_{Q_{ALOHA_{FB}}}}$	M2M throughput of the combined SA-RACH and FB-QL-RACH
$S_{M_{Q_{Aloha_r}}}$	M2M throughput of the combined SA-RACH and QL-RACH scheme with retransmission
$S_{p\text{-}ALOHA}$	throughput using p-ALOHA protocol
$S_{S\text{-}ALOHA}$	throughput using s-ALOHA protocol
TB	Tail Bits
TCH	Traffic Chanel
TDD	Time Division Duplex

TDMA	Time Division Multiple Access
t_gen	RACH request generation time
t_inter	inter arrival time
T_L	Generated traffic load
TS	Time Slot
T_{trx}	Transmission time
TV	Television
UWB	Ultra-wideband
Wi-Fi	Wireless Fidelity
WIMAX	Worldwide Interoperability for Microwave Access
WLAN	Wireless Local Area Network
WPAN	Wireless Personal Area Network
WSN	Wireless Sensor Network
UE	User Equipment
UL	Up-link
UpPTS	Uplink part
λ_{total}	Total new RACH request arrivals from both H2H and M2M groups
τ_{ia}	Mean inter-arrival time
1G	First Generation

2G	Second Generation
3G	Third Generation
4G	Fourth Generation
5G	Fifth Generation
3GPP	3rd Generation Partnership Project

References

[1] M. Chen, J. Wan, and F. Li, "Machine-to-Machine Communications: Architectures, Standards and Applications," *KSII Transactions on Internet and Information Systems,* vol. 6, pp. 480-497, 2012.

[2] A. Osseiran, V. Braun, T. Hidekazu, P. Marsch, H. Schotten, H. Tullberg, M. A. Uusitalo, and M. Schellman, "The Foundation of the Mobile and Wireless Communications System for 2020 and Beyond: Challenges, Enablers and Technology Solutions," in *77th IEEE Vehicular Technology Conference (VTC Spring),* pp. 1-5, 2013.

[3] S. Lien, K. Chen, and Y. Lin, "Toward Ubiquitous Massive Accesses in 3GPP Machine-to-Machine Communications," *IEEE Communications Magazine,* vol. 49, pp. 66-74, April 2011.

[4] B. Bertenyi, "3GPP system standards heading into the 5G era," *Eurescom message. The magazine for telecom insiders,* pp. 9-11, 2014.

[5] W.-J. Jo, C.-Y. Oh, Y. Kim, and T.-J. Lee, "Novel Pico-cell Range Expansion with Adaptive RACH Resource Allocation for Random Access of M2M Devices," in *the tenth International Conference on Wireless and Mobile Communications (ICWMC)*, pp. 186-191, 2014.

[6] A. Ksentini, Y. Hadjadj-Aoul, and T. Taleb, "Cellular-based machine-to-machine: overload control," *Network, IEEE,* vol. 26, pp. 54-60, 2012.

[7] T. Taleb and A. Kunz, "Machine type communications in 3GPP networks: potential, challenges, and solutions," *Communications Magazine, IEEE,* vol. 50, pp. 178-184, 2012.

[8] C. Kim, A. Soong, M. Tseng, and X. Zhixian, "Global Wireless Machine-to-Machine Standardization," *Internet Computing, IEEE,* vol. 15, pp. 64-69, 2011.

[9] G. 37.868, "RAN improvements for machine-type communications," Sept, 2011.

[10] C. Luders and R. Haferbeck, "The performance of the GSM random access procedure," in *44th IEEE Vehicular Technology Conference,* 1994, vol.2 pp. 1165-1169, *1994*.

[11] I. Vukovic and I. Filipovich, "Throughput analysis of TDD LTE Random Access Channel," in *IEEE 22nd International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC),* pp. 1652-1656. 2011

[12] Richard S. Sutton and A. G. Barto, *Reinforcement learning: An introduction, Cambridge* MA: MIT Press, 1998.

[13] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey," *arXiv preprint cs/9605103,* 1996.

[14] B. M. Kuboye, B. K. Alese, O. Fajuyigbe, and O. S. Adewale, "Development of Models for Managing Network Congestion on Global System for Mobile Communication (GSM) in Nigeria," *Wireless Networking and Communications,* vol. 1, pp. 8-15, 2011.

[15] K. Sakakibara, H. Muta, and Y. Yuba, "The effect of limiting the number of retransmission trials on the stability of slotted ALOHA systems," *Vehicular Technology, IEEE Transactions on,* vol. 49, pp. 1449-1453, 2000.

[16] M. J. Ferguson, "On the control, stability, and waiting time in a slotted ALOHA random-access system," *Communications, IEEE Transactions on,* vol. 23, pp. 1306-1311, 1975.

[17] V. Galetic, I. Bojic, M. Kusec, G. Jezic, S. Desic, and D. Huljenic, "Basic principles of Machine-to-Machine communication and its impact on telecommunications

industry," in *MIPRO, Proceedings of the 34th International Convention*, pp. 380-385, 2011.

[18] K.-R. Jung, A. Park, and S. Lee, "Machine-Type-Communication (MTC) Device Grouping Algorithm for Congestion Avoidance of MTC Oriented LTE Network Security-Enriched Urban Computing and Smart Grid.", T.-h. Kim, A. Stoica, and R.-S. Chang, ed: Springer Berlin Heidelberg, vol. 78, pp. 167-178, 2010.

[19] Ericsson, ""More than 50 Billion Connected Devices white paper", feb. 2011 http://www.ericsson.com/res/docs/whitepapers/wp-50-billion.pdf." accessed October 2013.

[20] H. Tian, L. Xu, Y. Pei, Z. Liu, and Y. Yang, "Power ramping schemes for M2M and H2H Co-existing scenario," *Communications, China,* vol. 10, pp. 100-113, 2013.

[21] C. Tang, L. Song, J. Balasubramani, S. Wu, S. Biaz, Q. Yang, and H. Wang, "Comparative Investigation on CSMA/CA-Based Opportunistic Random Access for Internet of Things," *IEEE Internet of Things Journal,* vol. 21, pp. 33-41, 2014.

[22] H. Kopetz, "Internet of things," in *Real-time systems*, ed: Springer, pp. 307-323, 2011.

[23] D. S. Watson, M. A. Piette, O. Sezgen, N. Motegi, and L. Hope, "Machine to Machine (M2M) Technology in Demand Responsive Commercial Buildings," in *ACEEE Summer Study on Energy Efficiency in Buildings*, Pacific Grove, CA. Washington D.C, pp. 1-14, 2004.

[24] B. Emerson. (2010). *M2M: the internet of 50 billion devices*. Available: www.huawei.com accessed June 2012.

[25] S. K. Tan, M. Sooriyabandara, and Z. Fan, "M2M communications in the smart grid: Applications, standards, enabling technologies, and research challenges," *International Journal of Digital Multimedia Broadcasting,* 2011.

[26] M. Roberti. (2011). *RFID Enables M2M and M2O*. Available: www.rfidjournal.com accessed September 2014.

[27] M. Sokele, V. Hudek, and A. I. Mincu, "Opportunities for implementation machine-to machine services via 3G mobile networks," in *Proceedings of the 7th International Conference on Telecommunications (ConTEL),* pp. 91-95, 2003.

[28] Z. M. Fadlullah, M. M. Fouda, N. Kato, A. Takeuchi, N. Iwasaki, and Y. Nozaki, "Toward intelligent machine-to-machine communications in smart grid," *Communications Magazine, IEEE,* vol. 49, pp. 60-65, 2011.

[29] M. Starsinic, "System architecture challenges in the home M2M network," in L*ong Island Systems Applications and Technology Conference (LISAT),* pp. 1-7, 2010.

[30] K. Doppler, C. B. Ribeiro, and J. Kneckt, "Advances in D2D communications: Energy efficient service and device discovery radio," in *2nd International Conference on Wireless Communication, Vehicular Technology, Information Theory and Aerospace & Electronic Systems Technology (Wireless VITAE),* pp. 1-6, 2011.

[31] J. M. Peha, "Sharing Spectrum Through Spectrum Policy Reform and Cognitive Radio," *Proceedings of the IEEE,* vol. 97, pp. 708-719, 2009.

[32] C. Ghosh, S. Roy, and D. Cavalcanti, "Coexistence challenges for heterogeneous cognitive wireless networks in TV white spaces," *Wireless Communications, IEEE,* vol. 18, pp. 22-31, 2011.

[33] G. P. Villardi, Y. D. Alemseged, S. Chen, S. Chin-Sean, N. Tran Ha, T. Baykas, and H. Harada, "Enabling coexistence of multiple cognitive networks in TV white space," *Wireless Communications, IEEE,* vol. 18, pp. 32-40, 2011.

[34] A. Laya, L. Alonso, and J. Alonso-Zarate, "Is the Random Access Channel of LTE and LTE-A Suitable for M2M Communications? A Survey of Alternatives," *Communications Surveys & Tutorials, IEEE,* vol. 16, pp. 4-16, 2014.

[35] C. Yu and Y. Yuli, "Cellular Based Machine to Machine Communication with Un-Peer2Peer Protocol Stack," in *70th IEEE Vehicular Technology Conference Fall (VTC 2009-Fall),* pp. 1-5, 2009.

[36] D. Cavalcanti, D. Agrawal, C. Cordeiro, X. Bin, and A. Kumar, "Issues in integrating cellular networks WLANs, AND MANETs: a futuristic heterogeneous wireless network," *Wireless Communications, IEEE,* vol. 12, pp. 30-41, 2005.

[37] A. Lo, Y. W. Law, M. Jacobsson, and M. Kucharzak, "Enhanced LTE-advanced random-access mechanism for massive machine-to-machine (M2M) communications," in *27th World Wireless Research Forum (WWF) Meeting*, 2011

[38] C. Min, W. Jiafu, S. Gonzalez, L. Xiaofei, and V. C. M. Leung, "A Survey of Recent Developments in Home M2M Networks," *Communications Surveys & Tutorials, IEEE,* vol. 16, pp. 98-114, 2014.

[39] M. Gerasimenko, V. Petrov, O. Galinina, S. Andreev, and Y. Koucheryavy, "Impact of machine-type communications on energy and delay performance of random access channel in LTE-advanced," *Transactions on Emerging Telecommunications Technologies,* vol. 24, pp. 366-377, 2013.

[40] Ofcom. *Machine to Machine Communications* Available: http://consumers.ofcom.org.uk/news/machine-to-machine-communications/ accessed September 2014.

[41] M. Chen, J. Wan, S. González, X. Liao, and V. Leung, "A survey of recent developments in home M2M networks," 2013.

[42] L. R. Izquierdo and S. S. Izquierdo, "Dynamics of the Bush-Mosteller learning algorithm in 2x2 games," 2008.

[43] L. R. Marcus, A. T. Michael, F. M. Robert, and A. R. Richard, "Evolution of the air interface of cellular communications systems toward 4G realization," *Communications Surveys & Tutorials, IEEE,* vol. 8, pp. 2-23, 2006.

[44] A. Kumar, D. Y. Liu, and J. Sengupta, "“Evolution of Mobile Wireless Communication Networks 1G to 4G”," *International Journal of Electronics & Communication Technology, IJECT,* vol. 1, 2010.

[45] V. Valenta, Mars, x030C, x, R. lek, G. Baudoin, M. Villegas, M. Suarez, and F. Robert, "Survey on spectrum utilization in Europe: Measurements, analyses and observations," in *Proceedings of the Fifth International Conference on Cognitive Radio Oriented Wireless Networks & Communications (CROWNCOM),* pp. 1-5, *2010*.

[46] P. Singh, R. R. Kumar, and T. Lamba, "A novel multiple access scheme for mobile communications systems," *INDIAN JOURNAL OF RADIO AND SPACE PHYSICS,* vol. 36, p. 430, 2007.

[47] H. Harada and R. Prasad, *Simulation and Software Radio for Mobile Communications*. Boston, London: Artech House, 2002.

[48] Z. Han and K. R. Liu, *Resource allocation for wireless networks*: Cambridge university press, 2008.

[49] T. S. Rappaport, *Wireless Communications Principles and Practice*, Second ed. Upper Saddle River: Prentice HAll, 2009.

[50] A. F. Molisch, *Wireless communications* vol. 15: John Wiley & Sons, 2010.

[51] H. Yin and S. Alamouti, "OFDMA: A broadband wireless access technology," in *IEEE Sarnoff Symposium,* pp. 1-4, 2006.

[52] H. Holma and A. Toskala, *LTE for UMTS-OFDMA and SC-FDMA based radio access*: John Wiley & Sons, 2009.

[53] L. Merakos and D. Kazakos, "On retransmission control policies in multiple-access communication networks," *Automatic Control, IEEE Transactions on,* vol. 30, pp. 109-117, 1985.

[54] N. Abramson, "THE ALOHA SYSTEM: another alternative for computer communications," in *Proceedings of the, fall joint computer conference*, pp. 281-285, *November 17-19, 1970*.

[55] L. G. Roberts, "ALOHA packet system with and without slots and capture," *ACM SIGCOMM Computer Communication Review,* vol. 5, pp. 28-42, 1975.

[56] M. Rahnema, "Overview of the GSM system and protocol architecture," *Communications Magazine, IEEE,* vol. 31, pp. 92-100, 1993.

[57] C. Bettstetter, H. J. Vogel, and J. Eberspacher, "GSM phase 2+; general packet radio service GPRS: Architecture, protocols, and air interface," *IEEE Communications Surveys & Tutorials,* vol. 2, pp. 2-14, 1999.

[58] C. Scholefield, "Evolving GSM data services," in *6th IEEE International Conference on Universal Personal Communications Conference Record,* vol.2 pp. 888-892, *1997*.

[59] C. Peersman, S. Cvetkovic, P. Griffiths, and H. Spear, "The Global System for Mobile Communications Short Message Service," *Personal Communications, IEEE,* vol. 7, pp. 15-23, 2000.

[60] G. Peersman, P. Griffiths, H. Spear, S. Cvetkovic, and C. Smythe, "A tutorial overview of the short message service within GSM," *Computing & Control Engineering Journal,* vol. 11, pp. 79-89, 2000.

[61] G. Heine, *GSM Networks: Protocols, Terminology, and Implementation*. London: Artech House, 1998.

[62] B. H. Walke, *Mobile Radio Networks Networking and Protocols*. West Sussex, England: John Wiley & Sons Ltd, 1998.

[63] A. Larmo, M. Lindstrom, M. Meyer, G. Pelletier, J. Torsner, and H. Wiemann, "The LTE link-layer design," *IEEE Communications Magazine,* vol. 47, pp. 52-59, 2009.

[64] N. Okubo, A. Umesh, M. Iwamura, and H. Atarashi, "Overview of LTE Radio Interface and Radio Network Architecture for High Speed, High Capacity and Low Latency," *NTT DOCOMO Technical journal,* vol. 13, pp. 10-19, Jun. 2011.

[65] "LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and modulation," ed: ETSI, 2009-10.

[66] 3GPP. *LTE Resource Guide* [http://www.anritsu.com/en-GB/Media-Room/Newsletters/files/anritsu_lte_guide.pdf]. accessed April 2012.

[67] B. T. SCHEME, "LTE: the evolution of mobile broadband," *IEEE Communications Magazine,* p. 45, 2009.

[68] H. G. Myung, "Technical overview of 3GPP LTE," *Polytechnic University of New York,* 2008.

[69] (10th, December). *Understanding LTE* [http://www.sharetechnote.com/Docs/anritsu_understanding_lte6.pdf]. accessed September 2014.

[70] J. Zyren and W. McCoy, "Overview of the 3GPP long term evolution physical layer," *Freescale Semiconductor, Inc., white paper,* 2007.

[71] L. Kleinrock and S. Lam, "Packet Switching in a Multiaccess Broadcast Channel: Performance Evaluation," *Communications, IEEE Transactions on,* vol. 23, pp. 410-423, 1975.

[72] J. H. Sarker and S. J. Halme, "Optimizing the use of random access channels in GSM-GPRS," *Wireless Personal Communications,* vol. 22, pp. 387-408, 2002.

[73] G. S. M. (3GPP), "Technical Specification Group GSM/EDGE Radio Access Network," in *Physical layer on the radio path* vol. 3GPP TS 05.01, ed, 2004-2011.

[74] M. E. Rivero-Angeles, D. Lara-Rodriguez, and F. A. Cruz-Perez, "Differentiated Backoff Strategies for Prioritized Random Access Delay in Multiservice Cellular Networks," *IEEE Transactions on Vehicular Technology,* vol. 58, pp. 381-397, 2009.

[75] R. R. Tyagi, F. Aurzada, K.-D. Lee, and M. Reisslein, "Impact of Retransmission Limit on Throughput and Delay of Preamble Contention in LTE-Advanced Random Access," *Engine,* 2012.

[76] "LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Medium Access Control (MAC) protocol specification " vol. V9.0, 2009-10.

[77] L. Sanguinetti, M. Morelli, and L. Marchetti, "A random access algorithm for LTE systems," *Transactions on Emerging Telecommunications Technologies,* vol. 24, pp. 49-58, 2013.

[78] M. Amirijoo, P. Frenger, F. Gunnarsson, J. Moe, and K. Zetterberg, "On self-optimization of the random access procedure in 3G long term evolution," in *IEEE International Symposium on Integrated Network Management-Workshops, 2009. (IM'09),* , pp. 177-184, 2009.

[79] H. Holma and A. Toskala, *WCDMA for UMTS Radio Access For Third Generation Mobile Communications*. West Sussex: John Wiley & Sons, 2004.

[80] x, C. beda, S. Pedraza, M. Regueira, and J. Romero, "LTE FDD Physical Random Access Channel Dimensioning and Planning," in *IEEE Vehicular Technology Conference (VTC Fall),* pp. 1-5, 2012.

[81] R. Patachaianand and K. Sandrasegaran, "System-Level Modeling and Simulation of Uplink WCDMA," in *Fifth International Conference on Information Technology: New Generations (ITNG),* pp. 1071-1076, *2008*.

[82] J. S. Carson, "Introduction to modeling and simulation," presented at the Proceedings of the 36th conference on Winter simulation, Washington, D.C., 2004.

[83] A. Maria, "Introduction to modeling and simulation," presented at the Proceedings of the 29th conference on Winter simulation, Atlanta, Georgia, United States, 1997.

[84] V. S. Frost and B. Melamed, "Traffic modeling for telecommunications networks," *Communications Magazine, IEEE,* vol. 32, pp. 70-81, 1994.

[85] R. Scheaffer, M. Mulekar, and J. McClave, *Probability and statistics for engineers*: Cengage Learning, 2010.

[86] MATHWORKS, *MATLAB the Language of Technical Computing*: The MATH Works inc,, 1999.

[87] J. D. Little and S. C. Graves, "Little's law," in *Building Intuition*, ed: Springer, pp. 81-100, 2008.

[88] W. Stalling, *Wireless Communications and Networks*, Second edition ed.: Pearson Prentice Hall, 2005.

[89] D. Akerberg and F. Brouwer, "On channel definitions and rules for continuous dynamic channel selection in coexistence etiquettes for radio systems," in *IEEE 44th Conference on Vehicular Technology,* vol.2 , pp. 809-813, 1994.

[90] R. L. Freeman, *Telecommunication System Engineering*, Fourth ed. Hoboken, New Jersey: John Wiley and Sons. inc., 2004.

[91] R. L. Freeman, *Fundamentals of Telecommunications*, Second ed.: John Wiley & Sons, INC., 2005.

[92] B. Chandrasekeran. *Survey of network Traffic Models*. Available: http://www.cse.wustl.edu/~jain/cse567-06/ftp/traffic_models3/index.html accessed April 2012.

[93] H. Khedher, F. Valois, and S. Tabbane, "Traffic characterization for mobile networks," in *IEEE 56th Vehicular Technology Conference, Proceedings (VTC 2002-Fall.)*, vol.3, pp. 1485-1489, *2002*.

[94] L. Klein, *Queueing Systems* vol. 1. Canada: John Wiley & Sons Inc., 1975.

[95] A. S. Tanembaum, *Computer Networks*. Boston, U.S.A.: Pearson Education Inc., 2011.

[96] R. Prasad, *CDMA for Wireless Personal Communications*. Norwood: Artech House, Inc., 1996.

[97] J. Sarker and S. Halme, "An Optimum Retransmission Cut-Off Scheme for Slotted ALOHA," *Wireless Personal Communications,* vol. 13, pp. 185-202, 2000.

[98] J. H. Sarker and S. J. Halme, "The prudence transmission method I (PTM I): a retransmission cut-off method for contention based multiple-access communication systems," in *47th IEEE Vehicular Technology Conference,* vol.1, pp. 397-401, *1997*.

[99] A. Mendez and D. Covarrubias, "Stability and optimal retransmission control of S-Aloha as a RACH channel on wireless networks," in *IEEE 54th Vehicular Technology Conference (VTC),* vol.3, pp. 1368-1372, *2001*.

[100] J. E. Wieselthier, A. Ephremides, and L. A. Michaels, "An exact analysis and performance evaluation of framed ALOHA with capture," *Communications, IEEE Transactions on,* vol. 37, pp. 125-137, 1989.

[101] C. Floerkemeier and M. Wille, "Comparison of transmission schemes for framed ALOHA based RFID protocols," in *International Symposium on Applications and the Internet Workshops,* pp. 4 -97, 2006.

[102] C. Yi, P. D. Mitchell, and D. Grace, "ALOHA and Q-Learning based medium access control for Wireless Sensor Networks," in *International Symposium on Wireless Communication Systems (ISWCS),* pp. 511-515, 2012.

[103] K. Byung-Jae, S. Nah-Oak, and M. E. Miller, "Performance analysis of exponential backoff," *Networking, IEEE/ACM Transactions on,* vol. 13, pp. 343-355, 2005.

[104] F. Daquan, J. Chenzi, L. Gubong, L. J. Cimini, Jr., F. Gang, and G. Y. Li, "A survey of energy-efficient wireless communications," *Communications Surveys & Tutorials, IEEE,* vol. 15, pp. 167-178, 2013.

[105] 3GPP, "System improvements for Machine-type communication s (MTC)," 2012.

[106] H. Wang and C. Woo, "Energy-Efficient Random Access for Machine-to-Machine (M2M) Communications," *International Journal of Software Engineering & Its Applications,* vol. 7, 2013.