Categorizing M2M Devices for Efficient Random Access in LTE

by

Rakib Hasan Chowdhury (ID: 122478)
MD. Rashik Shadman (ID: 122475)
Shakil Hossain (ID: 122481)

A Thesis Submitted to the Academic Faculty in Partial Fulfillment of the Requirements for the Degree of

BACHELOR OF SCIENCE IN ELECTRICAL AND ELECTRONIC ENGINEERING

Department of Electrical and Electronic Engineering
Islamic University of Technology (IUT)
Gazipur, Bangladesh
November 2016

A Dissertation on

Categorizing M2M Devices for Efficient Random Access in LTE

Submitted by

Rakib Hasan Chowdhury **MD. Rashik Shadman** **Shakil Hossain**

Approved by

Prof. Dr. Md. Ashraful Hoque
Head of the Department
Department of Electrical and Electronic Engineering
Islamic University of Technology (IUT)
Gazipur-1704, Bangladesh.

Supervised by

Dr. Mohammad Tawhid Kawser
Assistant Professor
Department of Electrical and Electronic Engineering
Islamic University of Technology (IUT).

Table of Contents

List of Figures

List of Acronyms

LTE	Long Term Evolution
RA	Random Access
M2M	Machine-to-Machine
MTC	Machine-type Communication
3G	Third Generation
RACH	Random Access Channel
PRACH	Physical Random Access Channel
LPWA	Low-Power Wide Area
ITU	International Telecommunications Union
ETSI	European Telecommunications Standards Institute
3GPP	Third Generation Partnership Project
LTE-A	LTE- Advanced
H2H	Human-to-Human
HTC	Human Type Communication
QoS	Quality of Service
RAN	Radio Access Network
EIRP	Effective Isotropically Radiated Power
eNB	eNodeB
IoT	Internet of Things
BI	Backoff Indicator
D2D	Device to Device
PUCCH	Physical Uplink Control Channel
PUSCH	Physical Uplink Shared Channel
SIB -2	System Information Block 2
RA-RNTI	Random Access Radio Network Temporary Identifier
C-RNTI	Temporary Cell Radio Network Temporary Identifier
RAR	Random Access Response

HARQ	Hybrid Automatic Retransmission Request
GSM	Global System for Mobile Communication
IEEE	Institute of Electrical and Electronics Engineers
KPI	Key Performance Indicator
FDD	Frequency Division Duplex
PDSCH	Physical Downlink Shared Channel
MAC	Medium Access Control
ACB	Access Class Barring
EAB	Extended Access Barring
SOOC	Self-Optimizing Overload Control

Acknowledgements

First of all, We would like to thank Almighty Allah for giving us the strength to do our work efficiently. Secondly We would like to thank our honourable supervisor Dr. Mohammad Tawhid Kawser for his guidance and for sharing his knowledge with us. We would also like to thank our honourable Head of the Department Prof. Dr. Md. Ashraful Hoque for providing us the opportunity to perform our thesis work properly. We would like to thank our friends for their support and motivation. Finally, We would like to thank our parents for their constant support.

Abstract

In this paper, we propose a dynamic allocation scheme of random access channel (RACH) slots for categorized machine-to- machine (M2M) devices, in which delay-sensitive services coexist with delay-tolerant ones. Our scheme aims at alleviating the overload and congestion of machine-type (MTC) communication as well as satisfying both delay-sensitive and delay-tolerant services. In order to make full use of RA slots, our proposed scheme adopts a categorized machine-to-machine (M2M) networking, where the M2M devices categorized into several types based on their delay sensitivity and priority. The most delay sensitive and high priority devices will be included in Type 1. Other devices will be categorized in Type 2, Type 3 etc. based on their delay sensitivity and priority. Type 1 will be allocated the maximum number of RA slots. Other types will be provided less number of RA slots. Simulation result shows that our proposed scheme realizes the remarkable increase in improving the number of packets arriving successfully and completing RA.

Chapter 1

INTRODUCTION

Current market penetration and recent predictions confirm that machine-to-machine (M2M) system deployments are increasing exponentially.This is driven by the needs of industries to automate their real-time monitoring and control processes, and the increasing popularity of smart applications to improve human well being. Examples are the automotive industry, which utilizes sensors to monitor the status of critical car components; the smart grid industry, which monitors critical points in the power transportation and distribution networks; and the smart city market,which provides innovative services to citizens by using real-time sensory data from the streets.

Capitalizing on the potential and applicability of M2M, various industrial standardization bodies have commenced embedding the unique needs of M2M systems into legacy wireless communication systems. This is the case of the IEEE and the Internet Engineering Task Force(IETF), which are defining standards suitable for M2M applications based on short- and midrange technologies such as IEEE 802.11 or the IEEE 802.15.4x, with different amendments for low-power industrial and smart applications(e.g., smart cities and smart grids). However,such networks suffer from some limitations, such as the use of shared exempt frequency channels(prone to interference) and limited radio coverage, which compromises the mass deployment of M2M services, but are easily overcome by cellular networks. Indeed, cellular networks constitute a very interesting alternative to providingM2M coverage. The fact that the infrastructure is already installed enables a fast and low-cost deployment of M2M services in a very short time, also providing a simple network topology based on one-hop communications. Unfortunately, current third generation (3G) and Long Term Evolution (LTE) cellular networks were not designed for M2M traffic.

Instead, they were mainly designed to support human-based services such as voice and web browsing, and bandwidth demanding services such as video streaming. Therefore, the network architecture needs to be improved to accommodate new M2M service requirements without sacrificing the quality of human-based services. For example, among others, M2M data traffic is mainly uplink, while current data traffic in human-based applications is mainly downlink; the amount of data per device is very small (e.g., only a few bits of information per transaction); the number of envisioned devices in the network is very high, orders of magnitude above that of humans; also, ultra-high energy efficiency is necessary to ensure the longlife-time of networks once deployed to make sure that M2M deployments can be autonomous and do not require frequent human intervention. For these reasons, leading standardization bodies, such as the International Telecommunications Union (ITU), the European Telecommunications Standards Institute (ETSI), the Third Generation Partnership Project (3GPP), the Telecommunications Industry Association (TIA), and the Chinese Communications Standard Association (CCSA), and global initiatives such as OneM2M, have commenced work on satisfying these and other constraints while not jeopardizing current cellular system usage for human-based applications. For example, in 2009 ETSI launched an M2M technical committee to actively look into architectural design, while 3GPP is incorporating M2M through its machine type communications (MTC) designs, coexisting with human type communications (HTC). In all cases, one of the main objectives of these organizations is to identify open challenges where efficient solutions can be proposed. Among others, the random access channel (RACH) of LTE (and LTE-Advanced, LTE-A, which is essentially the same at the RACH level) has been identified as a key area where improvements for MTC traffic are needed. The fact that the RACH of LTE is still based on a random access mechanism turns it into a potential bottleneck for the performance of cellular networks if the number of MTC devices grows [1]. For this reason, the RACH of LTE has attracted lots of attention from the research community.

EMERGING M2M TECHNOLOGIES

Whereas Zigbee-like solutions may still find their market niche with simple and static applications, the large IoT market is according to our experience well beyond their reach. We are thus witnessing a shift in M2M connectivity technologies which is being discussed in the remainder of this article.

A. Low-Power WiFi Technology

Over the recent years, WiFi (IEEE 802.11) technology has experienced tremendous growth and has become a de-facto 4 solution for home and corporate connectivity. However, WiFi has mostly been out of reach for M2M communications due to its fairly large energy consumption. This has changed as of late, i.e. when the IEEE 802.11 community started to apply duty cycling and hardware optimization with the result of an extremely energy efficient system. On the downside, support of mobility and roaming in WiFi is currently rather poor. In terms of reliability, there is neither guaranteed QoS support, nor adequate tools to combat severe interference typical for unlicensed bands. To this end, it has soon been recognized that the favorable propagation properties of low-frequency spectrum at sub-1GHz may provide improved communication properties when compared to, e.g., conventional WiFi protocols operating at 2.4 and 5GHz bands.

However, the available spectrum at sub-1GHz license-exempt ISM bands is extremely scarce and hence required careful system design considerations. With this in mind, after outlining the purpose and the technical scope of the novel IEEE 802.11ah project, the standardization work of the corresponding TGah task group has commenced in November 2010. The prospective technology is generally based on a variation of IEEE 802.11ac standard, but down-clocked by a factor of 10. It is currently being developed to enable low-cost long-range (up to 1km) connectivity across massive M2M deployments with high spectral and energy efficiencies. Today, thousands of M2M devices may already be found in dense urban areas, which required providing support for up to 6K machines connecting to a single access point. Fortunately, IEEE 802.11ah technology does not need to maintain backward compatibility with the other representatives of the IEEE 802.11 family. Operating over different frequencies, 802.11ah could thus afford defining novel compact frame formats, as well as offering more efficient mechanisms to support a large population of devices, advanced channel access schemes, as well as important power saving and throughput enhancements [2]. As the result, 802.11ah is believed to significantly enrich the family of 802.11 protocols, which already receive increasing attention from mobile network operators willing to introduce low-cost connectivity in unlicensed bands.

B. Unlicensed Low-Power Wide Area Networks

Given that Zigbee-like solutions have not lived up to their expectations, whereas Low-Power WiFi and Cellular M2M systems have commenced to take shape only recently, a novel class of M2M technologies has emerged lately, termed Low Power Wide Area (LPWA), which operate in unlicensed spectrum. Only low data rates and small daily traffic volumes are however foreseen [3], which limits the application to a subset of M2M services with infrequent small data transmissions. LPWA technology today is proprietary with multiple non compatible alternatives. There are also initiatives to propose LPWA technology concepts into the cellular M2M direction within a recent new study item that has been initiated in the 3GPP GERAN (GSM/EDGE Radio Access Network) specification group. Similar to the standardization targets of LTE evolution in 3GPP Radio Access Network (RAN), the motivation is to enable extended coverage beyond GSM coverage today, low device complexity, and long battery lifetimes. Our experience with LPWA shows that it works successfully for large projects, such as the Moscow Smart City deployment [4], where almost 20K sensors have been connected to a modest number of access points. In the trial deployments, we have seen suburban and rural ranges of over 20km, the typical urban ranges of around 5km, and the "difficult" urban ranges of 1-2km. Mobile network operators may hence become the early adopters of this emerging technology building on their well-developed network infrastructure and strong customer trust. For instance, a possible deployment model for an operator may be to install LPWA systems complementary to existing cellular technology and the cell sites that they already have [5]. Despite the time-to-market benefits of LPWA, there are also clear downsides of using unlicensed spectrum for long-range communication. Typical regulation imposes several restrictions on radio transmitters in unlicensed spectrum [6] in terms of effective radiated power (ERP), allowed duty cycles, and listen-before-talk requirements. For long-range transmissions, the limited ERP causes asymmetric link budgets between uplink and downlink directions. The reason is that the ERP is limiting the radiated power after the antenna gain has been applied. However, antennas have significantly different performance between simple devices using a single antenna with around 0dB antenna gain and a base station with an antenna gain of around 19dB. This means that the uplink signal has an additional antenna gain at the receiver in contrast to the downlink signal. For European regulation, this can be partly compensated by selecting a subband for downlink with 13dB higher output power. But even then a link asymmetry of at least 6dB remains. As a consequence, at least 50% of devices experience only uplink connectivity under non-line-of-sight propagation conditions. This is unreliable in the sense that no acknowledgements for successful uplink data delivery are possible. Further, scalability limitations come from the range covered by a single LPWA base station [6]. Projecting that the total number of connected M2M devices is to become approximately 10 times larger than the number of people, easily millions of devices may appear within the coverage area of a single LPWA base station. Many of those will use other radio technologies that share the spectrum with LPWA, such as low-power WiFi (IEEE 802.11ah), Z-Wave, Zigbee, IEEE 802.15.4g, etc. With its low receiver sensitivity for long-range communication, the LPWA device will perceive all of these other transmissions as interference. We therefore foresee that LPWA will only remain viable at the early stage of IoT development when the number of devices is still moderate. However, LPWA can play an important role to support the early IoT market up-take until standardized cellular M2M solutions enter the market, which can handle the anticipated IoT scale in terms of numbers of devices, but also the variety of M2M services.

C. Cellular M2M

Cellular technologies, and especially 3GPP LTE, are becoming increasingly attractive for supporting large-scale M2M 5 installations due to their wide coverage, relatively low deployment costs, high level of security, access to dedicated spectrum, and simplicity of management. However, LTE networks have been neither historically designed with link budget requirements of M2M devices, nor optimized for M2M traffic patterns. Therefore, several improvements targeting M2M solutions have been initiated in 3GPP aiming at augmenting LTE to become more suitable for M2M applications. Given that the numbers of connected machines are expected to grow dramatically, LTE technology requires respective mechanisms to handle a very large number of devices [7]. Correspondingly, an overload control scheme named Enhanced Access Barring has been introduced as part of LTE Rel-11 to avoid overload in RAN, whenever there is a surge in near simultaneous network entry attempts. Further, accounting for the fact that typical M2M data transmissions are infrequent and small, simplified signaling procedures for radio-bearer establishment are necessary to offer energy consumption savings for such M2M devices. In connection to lightweight signaling for small data, M2M device energy consumption can be reduced significantly for infrequent traffic by allowing for longer cycles of discontinuous reception (DRX). In the following sections, we review some of these important improvements in more detail. We intentionally focus our description on LTE which we believe will become the major technology for M2M connectivity even though M2M-centric improvements are being discussed for other 3GPP technologies as well.

M2M PERFORMANCE IMPROVEMENTS BY 3GPP

A. Handling Very Large Numbers of Devices

Our research indicates that smart grid is one of the key M2M use cases incorporating a large number of metering devices that autonomously report their information to grid infrastructure. The motivating smart metering use case therefore serves as a valuable reference "massive M2M" scenario covering many characterizing M2M features (see Fig. 1.1.A). Correspondingly, the involved M2M devices may be divided into several classes according to the priority of their information, e.g. high-priority (alarm messages) and low-priority (measurement data). Potentially, alarm messages constitute a bigger challenge for the network to handle, as they are typically highly synchronized and in addition may require certain latency guarantees [8]. Currently, 3GPP LTE system defines a number of communication channels to deliver uplink transmissions from M2M devices to the network. In particular, the Physical Random Access Channel (PRACH) is employed by a device for its initial network entry, as well as to demand system resources if it does not already have a dedicated resource allocation. In case of many M2M devices connecting to the network near simultaneously, we expect that the use of PRACH would be preferred, but may result in congestion due to its insufficient capacity. More specifically, the PRACH procedure features two distinct stages (see Fig. 1.A). The former is the uplink timing synchronization stage (known as Msg1/Msg2), where the power ramping technique may be used to adjust the transmit power

of a random-access preamble to particular channel conditions. Further, Msg3 is used to transmit a meaningful uplink message to the base station (termed eNodeB or eNB) and Msg4 is utilized for subsequent contention resolution. To understand the impact of a large population of M2M devices on their network access, we construct an event-driven protocol-level PRACH simulator and thoroughly calibrate it against the reference 3GPP methodology documents [9]. Our simulation yields important conclusions on overloaded PRACH performance, when numerous connected-mode M2M devices of different priorities send their information into the network (see Fig. 1.B). In particular, we learn that around 40% of high-priority M2M devices, added to the original (typical) population of 30K low-priority devices, produce a sharp degradation in network access success probability. Interestingly, PRACH preambles selected by the M2M devices randomly may be regarded as non-interfering code based "channels" (see Fig. 1.1.C), where the case when two or more devices select an identical preamble (channel) would correspond to a conventional "packet" collision. This opens door to assessing the contention-based M2M behavior by relying on past knowledge of multichannel random-access protocols. First, careful custom-made approximations can be forged for particular given ranges of PRACH parameters (see Fig. 1.1.B), such as the number of available preambles (M) and contending devices (U), backoff window size, etc. However, these may not be counted as adequate universal solutions and another alternative is straightforward numerical analysis of contention behavior, which would only remain feasible for moderate numbers of users/channels due to high computational complexity. More recently, we have demonstrated the feasibility of applying powerful fluid approximation techniques to rigorously characterize M2M performance, as well as the stability regions of a multichannel random-access system. As our target scenario, we have chosen an industrial automation application (see Fig. 1.1.D), which may require certain data access latency and reliability guarantees (e.g., for supporting priority or critical alarm messages). Along these lines, Fig. 1.1.E indicates analytical PRACH latencies as evaluated with our method, which allows optimizing channel access by properly selecting the Msg1 retransmission probability value for arbitrary numbers of devices and channels. More specifically, in the figure we compare our optimized latency against the values produced with the use of existing PRACH backoff indicator parameters. Our solution thus helps the base station regulate PRACH access by having system wide knowledge across all connected M2M devices. However, if such knowledge is not available, simpler heuristic access control procedures (such as when the retransmission probability is chosen as M over U) may be employed by the M2M devices locally, which sometimes results in close to optimal performance. These results allow for tighter control of important performance indicators, such as data access latency, which may benefit LTE in supporting constrained automation scenarios on the way to the Industrial Internet.

B. Energy Efficiency and Small Data Transmission

In tight connection with access latency and success rate of M2M transmissions goes their energy efficiency, which is accentuated by the fact that M2M devices are typically smallscale and battery-powered. We continue studying the scenario when an IoT application requires a large number of M2M devices to perform a particular action near-simultaneously (e.g., smart meter data readings), or when an unexpected surge, outage, or failure occurs (massive power outage or restoration of power, network failure, etc.) causing multiple devices to (re)connect to the network within a short period of time. In this case, the transmitting devices would still be using the PRACH contention-based random access procedure to obtain uplink synchronization for initial network access or respective data transmission.

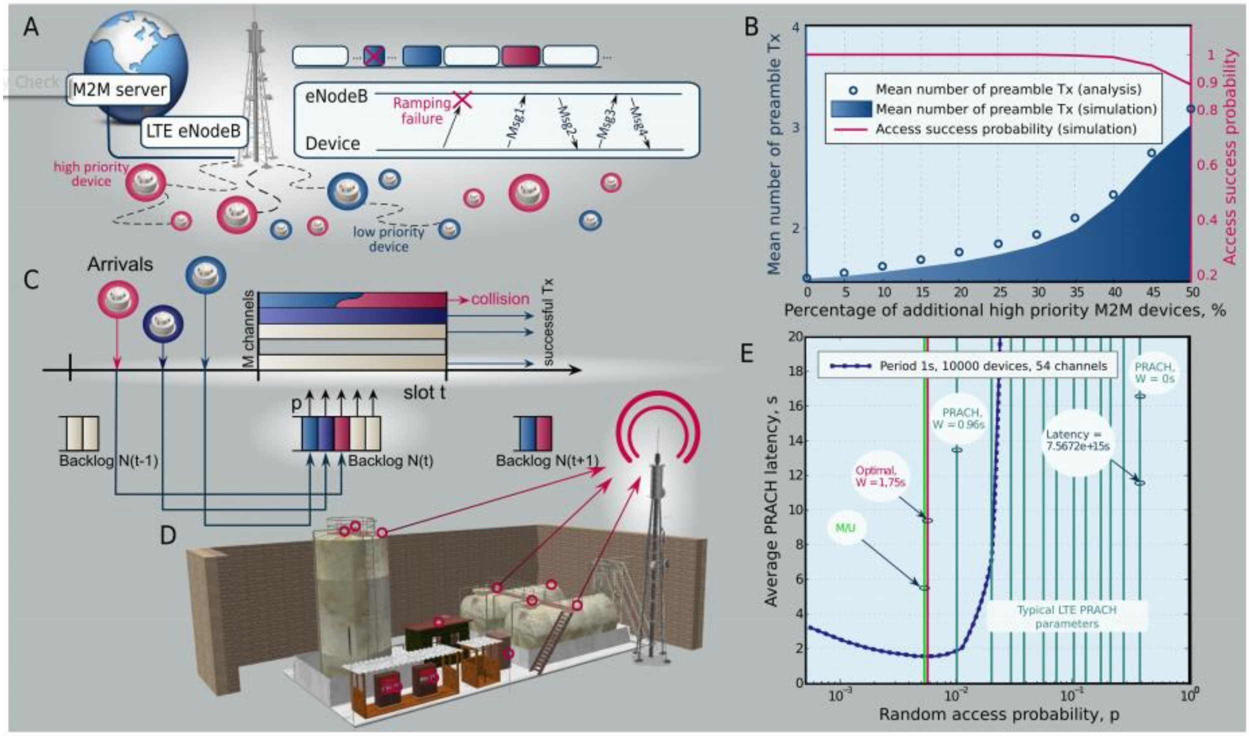


Fig. 1.1: Performance results: handling very large numbers of devices. A: motivating M2M scenario; B: connected-mode performance of different device priorities; C: proposed multi-channel M2M contention model; D: characteristic industrial automation application; E: analytical random-access latencies.

Along these lines, Fig. 1.2.A illustrates the simulated initialnetwork entry performance of 30K M2M devices with respectto their power consumption, collision probability, and access success probability across different PRACH backoff indicator (BI) values. These results are produced for characteristic beta distributed M2M device activation patterns (traffic type 2: beta distribution over 10 seconds), as suggested by 3GPP evaluation methodology (see Table 6.1.1 in [9]), since uniformly distributed activations (traffic type 1: uniform distribution over 60 seconds) do not cause actual network overloads. Our evaluation framework accounts for the main M2M device power consumption levels (inactive, idle, Rx, and Tx) at all states of PRACH signaling procedure (see Fig. 1.2.B) and sheds light on the feasibility of candidate network overload control solutions. In particular, Fig. 1.2.A suggests that a combination of M2M-specific backoff (larger non-standard BI values) and initial backoff (pre-backoff) may successfully alleviate congestion caused by highly-correlated beta-distributed M2M device activation patterns. Further, the focus of our investigation shifts to the dedicated power consumption aspects of M2M devices. Currently, short paging cycles in 3GPP LTE may be highly sub-optimal for M2M devices, especially given lengthy M2M traffic interarrival times and delay tolerant nature of many M2M applications. Hence, extending paging cycle durations in the idle state may help delay-tolerant devices sleep for longer periods of time thus extending their battery lifetimes. The corresponding studies require an appropriate power consumption model (see Fig. 1.2.C), which is capable of capturing typical M2M traffic patterns. Correspondingly, our results in [10] indicate that increasing the current maximum DRX (discontinuous reception) and paging cycle lengths would indeed lead to significant gains in the energy consumption (over 20x) of M2M devices (see Fig. 1.2.D). If additional delay can be tolerated by M2M devices, D2D (device-to-device) communication techniques may further reduce the consumption of power. For instance, one M2M device may act as an aggregation point and relay data from other proximate M2M devices with poor communication link (to avoid excessive

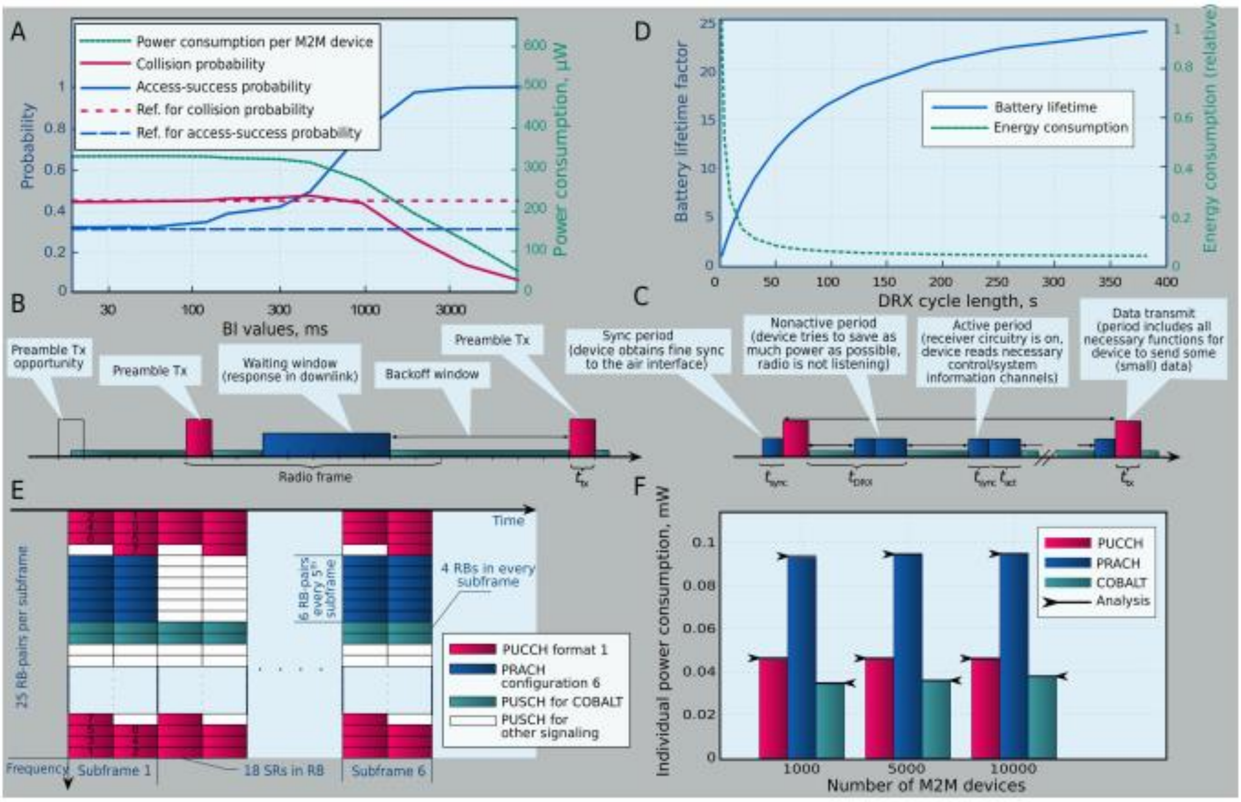


Fig. 1.2: Performance results: energy efficiency and small data. A: initial network entry performance; B: M2M device power consumption levels; C: proposed power consumption model; D: assessment of energy efficiency M2M improvements; E: basic LTE frame structure; F: benefits of COBALT mechanism

Retransmissions. D2D-based “client relay” mechanisms may dramatically reduce energy expenditures of cell-edge M2Mdevices, especially when those only send small data packets,and additionally help relieve a surge in uncontrolled near simultaneous M2M transmissions.

However,if adding more delay is not acceptable, such as in critical control applications,further data access improvements would need to be made reducing PRACH latencies to few milliseconds by, potentially,shortening the signaling sequence in Fig. 1.2.B. Other areas fordata access enhancements concern the size of typical M2Mpayloads (on the order of several bytes), as existing coding mechanisms in LTE are not optimal for short data blocks. Small data also creates inefficiency in control and channel estimation procedures causing excessive overheads, as well as in existing frame structures and resource allocation schemes. In more detail, Fig. 1.2.E illustrates the current frame structure of LTE (for 5MHz bandwidth) as a rectangular grid of resource blocks (RBs). The groups of RBs compose different data access channels, including periodic PRACH allocations and continuous Physical Uplink Control Channel (PUCCH) resources to carry the uplink control information. As both PUCCH and PRACH capacities may be very limited to serve small M2M data from numerous sources, we propose to allocate a part of Physical Uplink Shared Channel (PUSCH), which is otherwise employed for actual human-to-human (H2H) data transmissions, for a dedicated M2M use. Our respective scheme, named contention-based LTE transmission (COBALT), takes advantage of fewer LTE signaling messagesnand a simple collision resolution procedure (see detailed description in [11]). It thus yields better utilization of network resources, lower latencies, and, most importantly, significantly reduced power consumption for M2M devices (see Fig. 1.2.F). An M2M device must trigger the access procedure to the base station (hereinafter eNodeB, which is the term used in LTE) in the following five situations [11]:

1) Upon initial access to the network, i.e., in the association process.

2) When receiving or transmitting new data and the M2M device is not synchronized.

3) Upon transmission of new data when no scheduling request resources are configured on the uplink control channel.

4) In the case of handover (change of associated eNodeB), to avoid a session drop.

5) After a radio link failure, in order to re-establish the connection.

In order to handle all these situations, two different forms of Random Access (RA) procedure are defined in LTE:

- **Contention-based**: where devices compete for the channel access. Since collisions can occur, this type of access is reserved for delay-tolerant access requests.
- **Contention-free**: where the eNodeB allocates specific access resources for those access requests that must have high probability of success (delay-constrained access), e.g., handover.

The contention-based operation of the RACH is based on ALOHA-type access, i.e., transmit the request in the first available opportunity. This means that, in the case of the transmission of simultaneous access requests, the system performance may degrade due to a high probability of collision in the transmission of the preambles. Indeed, the 3GPP and organization members have released some studies regarding the capacity limits of the RA in LTE [12], [13]. In these studies, it has been considered that there is an access opportunity every 5ms and 54 out of the 64 available preambles are used for contention-based access, while the remaining 10 preambles are reserved for contention-free access (e.g. reserved for handover). Under these conditions, the system offers 200 access opportunities per second, which corresponds to a capacity of 10,800 preambles per second. Although this number may seem enough for most envisioned M2M applications, this is the absolute maximum capacity that the system tolerates in the absence of collisions. However, due to the use of ALOHA as the access protocol for the transmission of the preambles and the use of random backoffs in the case of failure, the usual system performance is much lower than this upper limit.

There are some scenarios and applications that may be compromised by such performance limits of the RACH of LTE. For example, this is the case of a power outage, after which all systems try to get connected to the network simultaneously. Another use case is the utility meters reading reports; where all the devices transmit with high correlation in reporting times. A third example is a railway bridge vibration monitoring application where, upon transit of a train along the sensor-equipped bridge, all the sensors react simultaneously and try to transmit through the network. Of course, the last case can be generalized into event-driven applications, such as fire-detection, fault alarm, security threads; this means that devices are idle for long periods of time and become active when an event is triggered. If the number of devices is known, then it is possible to design an optimum scheduling algorithm; however, if the number of devices is unknown, then random access procedures need to be used. These examples will not necessarily be infrequent. Therefore, there is a considerable amount of M2M applications that are characterized by bulk arrivals and require very high energy efficiency to ensure the long lifetime of the network. The 3GPP is fully aware of these limitations of the RACH of LTE and is actively working in improvements to overcome congestion and overloading of the RACH when used for M2M applications [13], [14]. An extensive list of key issues and feasible system improvements are presented in [15]. In addition, different research groups around the globe have been working, and are still progressing, on identifying limitations of the RACH of LTE for M2M communications and proposing alternative solutions to avoid slowing down the penetration of M2M applications into the mass market. In an endeavor to align all the efforts in the same direction, some of the 3GPP organization members have discussed in [13], [14] three key performance indicators to evaluate the performance of the novel proposals being designed today. Additionally, based on the low power consumption requirement issue referred in [15], an energy related indicator should also be considered. This leads to the following four performance indicators:

1) Access Success Probability, defined as the probability to complete the random access procedure in the maximum number of preamble transmissions allowed. This parameter can also be represented by the blocking probability, defined as the probability that a device reaches the maximum number of transmission attempts and is unable to complete an access process.

2) Preamble Collision Rate, defined as the ratio between the number of preamble collisions in the same RA slot and the total number of preambles transmitted on that slot. An equivalent metric consists in measuring the average number of preamble retransmissions required to have a successful access request.

3) Access Delay, defined as the time elapsed between the transmission of the first preamble and the reception of Message 4 by the M2M device.

4) Device Energy Consumption, defined as the total energy spent in transmission and reception tasks, from the first RA attempt until the successful access to the network has been granted. As specified by the 3GPP in [16], each application has its own requirements, and thus different performance indicators need to be considered. It may happen that a specific access technique improves the performance under very specific network conditions and traffic loads, but it performs worse when tested in different conditions. Among the wide range of degrees of freedom to define requirements, the 3GPP has already classified some key applications as: with very low mobility (or static), time-restricted applications, time tolerant applications, infrequent transmission application, and small data transmission applications.

There are many conventional methods of improving RA collision:

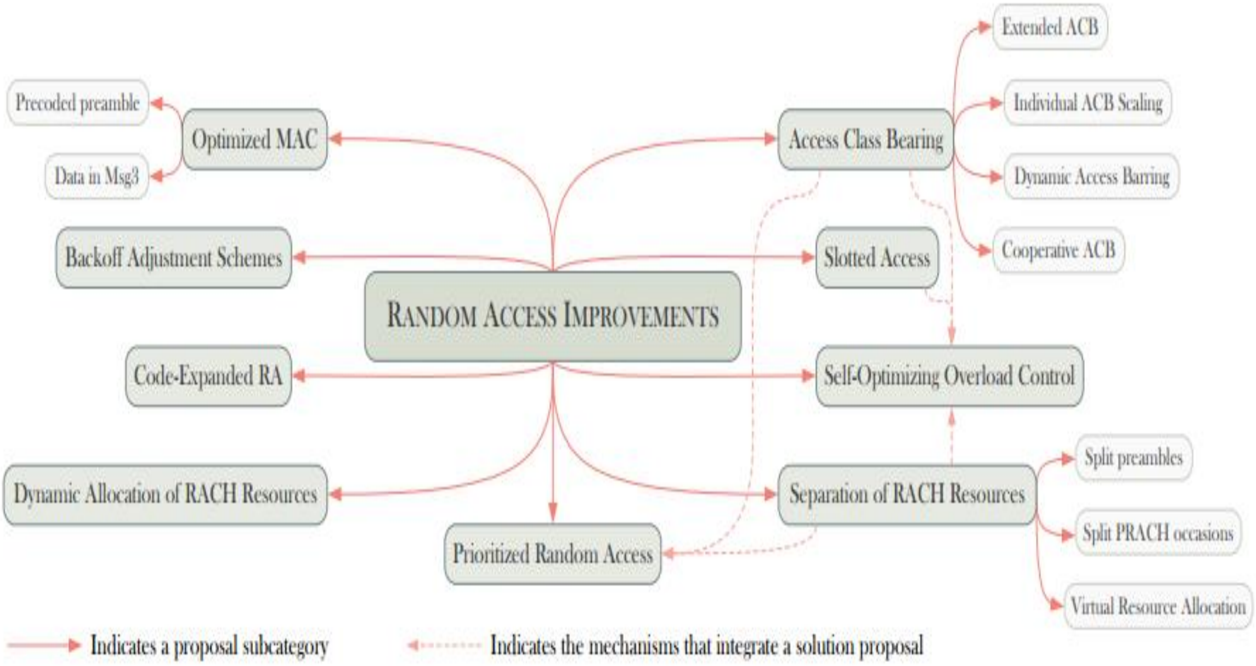


Fig. 1.3. Diagram of proposals to overcome random access overload.

In our scheme , we propose that the M2M devices are categorized based on their delay sensitivity and priority.The most delay sensitive and high priority devices are included in Type 1 . Other devices are categorized as Type 2 , Type 3 , Type 4 etc. according to their delay sensitivity and priority.The types are allocated RA slots based on their delay sensitivity and priority i.e. Type 1 will be able to send preambles in maximum number of RA slots . Other types will get less number of RA slots .

Chapter 2

OVERVIEW OF RANDOM ACCESS PROCEDURE

2.1 INTRODUCTION

For MTC devices, we consider the typical random access procedures defined in LTE protocols [17], where all MTC devices compete for the available wireless resources. It is worth noting that in M2M networks, the contention-based random access procedure usually applies for delay-tolerant access requests. However, contention-free opportunistic access is also supported by pre-allocating particular resources, which are for delay-constrained requests. Because the amount of devices sending request with strict delay demands is quite small compared to delay-tolerant ones and the total number of such MTC devices is often huge, here we focus on the typical contention-based random access procedure, which consists of four steps [17] between MTC devices and the eNB. Moreover, the resource pre-allocation approach is very inflexible, and thus, not able to provide the requests with fine-grained delay requirements. Next, we elaborate on the detailed random access procedures as follows.

When an MTC device attempts to access the network, it needs to send out an access request over the random access channel (RACH), which is comprised of several random access slots, which are used for the transmission of access requests. The length of the RA slot depends on the value of the configuration index. Relying on the existing protocols, the configuration index is valued as six, which means that in the RACH, there is an access opportunity every five milliseconds. In other words, the RACH finishes configuration every five milliseconds. Furthermore, there are in total 64 orthogonal available preambles. Only 54 of them are available for contention-based access, while the remaining 10 preambles are reserved for contention-free access. An access request is completed only if the four steps [18] are successfully finished.

2.2 RANDOM ACCESS PREAMBLE

Random access preambles are the orthogonal bits sequences, called digital signature, used by UEs to initiate RA attempt. RA preambles are generated by cyclically shifting a root sequence, such that every preamble is orthogonal to each other. There are total 64 preambles which are initially divided into two groups, i.e. contention-free RA preambles and contention based RA preambles. The eNB reserves some preambles, say *Ncf*, for contention-free RA, and assigns distinct preambles to different UEs. Rest of the preambles (64 - *Ncf*) are used for contention-based RA, where each UE randomly generates one preamble [19].

2.3 RANDOM ACCESS SLOT

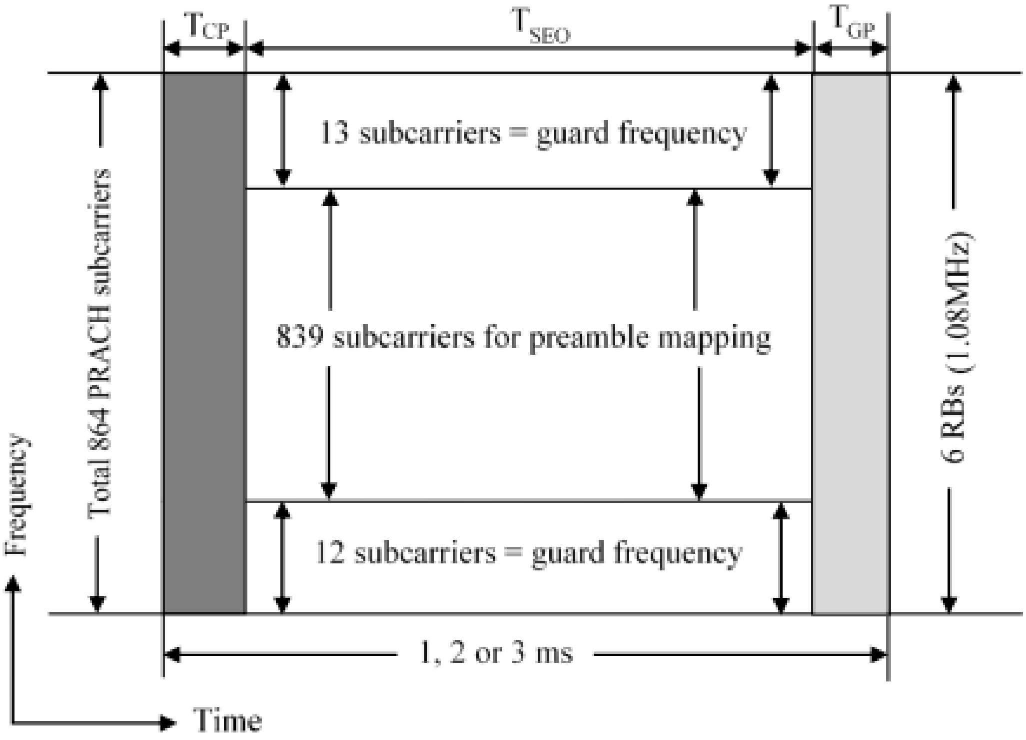


Fig 2.1 : FDD-based RA slot in time-frequency resources

A random access slot (RA slot) refers to the LTE physical radio resources, called Physical Random Access Channel (PRACH), in which RA preambles are mapped and transmitted to the eNB. In Frequency Division Duplex (FDD) operation, an RA slot consists of 6 physical Resource Blocks (RBs) in frequency domain, while the time duration of each RA slot can be 1, 2, or 3 subframe(s) depending on the preamble format [19]. There are a total of 864 subcarriers in one RA slot which are equally distant at 1:25 KHz. All 64 preambles are mapped into 839 RACH subcarriers, while the remaining 25 subcarriers are used as guard frequency subcarriers[19] .

In FDD operation, four different preamble formats are available based on preamble cyclic prefix duration (*TCP*), and preamble sequence duration (*TSEQ*) [19]. A UE can select an appropriate preamble under a specific format depending on the distance from eNB, maximum delay spread, amount of transmission resource needed to transmit RRC request, etc. On the other hand, the number of RA slots in each radio frame is defined by the preamble configuration index. For each preamble format 16 different indices are available, where the eNB allocates radio resources as PRACHs. Depending on the system bandwidth, some LTE systems may not be able to use some preamble configuration indices. However, systems using 20 MHz bandwidth are able to use all of the indices [19]. The eNB periodically broadcasts the preamble information as a part of System Information Block 2 (SIB2) message.

2.4 STEPS OF RANDOM ACCESS

The contention-based RA procedure consists of a four message handshake between the M2M device and the eNodeB, which is described in the next subsections. An access request is completed if the four messages are successfully exchanged.

A. Message 1, Preamble Transmission

Whenever an M2M device requires access to the channel, it selects the next available RA slot of the RACH to transmit an access requests. This consists of a preamble, i.e., a digital signature that the device transmits in an RA slot. There are 64 orthogonal pseudo-random preambles available

for RA and the eNodeB periodically broadcast information in the downlink control channel on which preambles may be used [20]. However, the eNodeB reserves some of them for contention-free access. If two or more devices transmit the same preamble in the same RA slot, a collision occurs. Otherwise, the different preambles can be detected by the eNodeB thanks to their orthogonality. The duration of a preamble depends on the size of the cell, and can vary from1 to 3 ms. The larger the cell-size, the longer the duration of the preamble in order to improve the reliability of reception at the cell edge. The selection of the preamble to transmit for each request is done at random (among those available for contention-based access). Exactly 3 subframes after the transmission of the preamble [21], the M2M device waits for a time window to receive a response from the eNodeB, i.e., Message 2 of the handshake. The duration of this waiting window is broadcast by the eNodeB and is defined between 2 and 10 subframes [22].

B. Message 2, Random Access Response (RAR)

For each successfully decoded preamble, the eNodeB computes an identifier , referred to as the Random Access Radio Network Temporary Identifier (RA-RNTI), which is calculated based on the RA slot where each preamble was sent [21]. Then, the eNodeB transmits a RAR through the Physical Downlink Shared Channel (PDSCH) with the following information:

- Identification of the detected preamble.
- Timing alignment instructions to synchronize uplink transmissions.
- Uplink resource allocation that will be used by the device to transmit the third message of the handshake.
- Assigned Temporary Cell Radio Network Temporary Identifier (C-RNTI).
- In the case of failure, an optional Backoff Indicator (BI) to request the devices to wait for a random period of time before retrying access [21]. This random backoff is used to reduce the probability of preamble collision, dispersing the access attempts along time .

The RAR is addressed to a specific RA-RNTI, i.e., to all the devices that transmitted a preamble on a specific RA slot. The RAR contains different subheaders associated to each detected

preamble. If a device receives a RAR message addressed to the RA-RNTI associated to the RA slot where the preamble was transmitted, but it does not not contain the identifier of the used preamble, it performs a random backoff time (according to the BI parameter attached to the RAR) before scheduling another preamble transmission attempt (Message 1) [21].

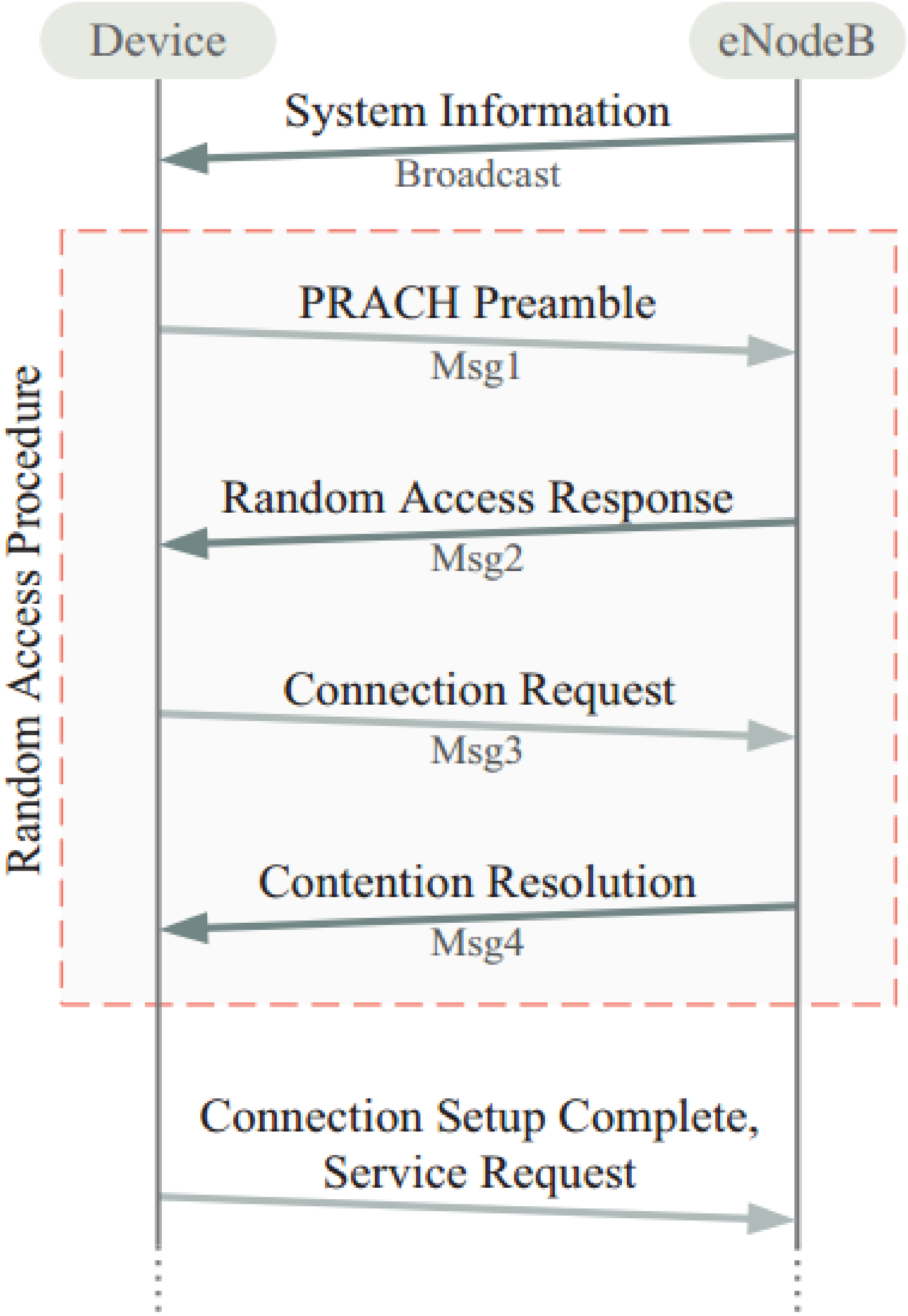


Fig 2.2 : Contention-Based Random Access (RA) Procedure

If multiple devices selected the same preamble and the same RA slot, a collision will occur. The eNodeB may be able to detect the collision based on the different time of arrival. In such case, it will not provide information related to that specific preamble in the next RAR. However, if the devices are at the same distance from the eNodeB, and the two preambles have been received constructively, the collision may be undetected by the eNodeB and the same RAR information will be sent to all the devices that transmitted the same preamble in the same RA slot. This will cause a collision again in Message 3.

C. Message 3, Connection Request

The M2M device transmits a Connection Request message to the eNodeB in the resources granted in the Message 2 associated to the preamble transmitted in the selected RA slot. Message 3 is transmitted with Hybrid Automatic Retransmission Request (HARQ). For the initial access, this message conveys the device identifier (C-RNTI) and the reason for the access request. In the case of an undetected preamble collision by the eNodeB, more than one device will use the same uplink resources to transmit Message 3 and a collision will occur at the eNodeB. Therefore, no acknowledgment will be transmitted by the eNodeB and each device will retransmit Message 3 for the maximum number of retransmissions allowed before declaring access failure and scheduling a new access attempt.

D. Message 4, Contention Resolution

Upon reception of a Connection Request, the eNodeB transmits a Contention Resolution message as an answer to Message 3. A device which does not receive Message 4 declares a failure in the contention resolution and schedules a new access attempt, i.e., a new preamble transmission, starting the process over again. Each device keeps a preamble transmission counter that is increased after each unsuccessful attempt. When the counter reaches the maximum allowed value (informed as system information by the eNodeB), the network is declared unavailable by the device and a random access problem is indicated to upper layers.

Chapter 3

LIMITATIONS OF LTE RANDOM ACCESS

The contention-based operation of the RACH is based on ALOHA-type access, i.e., transmit the request in the first available opportunity. This means that, in the case of the transmission of simultaneous access requests, the system performance may degrade due to a high probability of collision in the transmission of the preambles. Indeed, the 3GPP and organization members have released some studies regarding the capacity limits of the RA in LTE [12], [13]. In these studies, it has been considered that there is an access opportunity every 5ms and 54 out of the 64 available preambles are used for contention-based access, while the remaining 10 preambles are reserved for contention-free access (e.g. reserved for handover). Under these conditions, the system offers 200 access opportunities per second, which corresponds to a capacity of10,800 preambles per second. Although this number may seem enough for most envisioned M2M applications, this is the absolute maximum capacity that the system tolerates in the absence of collisions. However, due to the use of ALOHA as the access protocol for the transmission of the preambles and the use of random backoffs in the case of failure, the usual system performance is much lower than this upper limit.

There are some scenarios and applications that may be compromised by such performance limits of the RACH of LTE. For example, this is the case of a power outage, after which all systems try to get connected to the network simultaneously. Another use case is the utility meters reading reports; where all the devices transmit with high correlation in reporting times. A third example is a railway bridge vibration monitoring application where, upon transit of a train along the sensor-equipped bridge, all the sensors react simultaneously and try to transmit through the network. Of course, the last case can be generalized into event-driven applications, such as fire-detection, fault alarm, security threads; this means that devices are idle for long periods of time and become active when an event is triggered. If the number of devices is known,then it is possible to design an optimum scheduling algorithm; however, if the number of devices is unknown, then random access procedures need to be used. These examples will not necessarily be infrequent. Therefore, there is a considerable amount of M2M applications that are characterized by bulk arrivals and require very high energy efficiency to ensure the long lifetime of the network.

The 3GPP is fully aware of these limitations of the RACH of LTE and is actively working in improvements to overcome congestion and overloading of the RACH when used for M2M applications [13], [14]. An extensive list of key issues and feasible system improvements are presented in [15]. In addition, different research groups around the globe have been working, and are still progressing, on identifying limitations of the RACH of LTE for M2M communications and proposing alternative solutions to avoid slowing down the penetration of M2M applications into the mass market.

In each RA slot, let us consider that 54 preambles are utilized for contention-based random access, and each radio frame contains 2 RA slots. Thus, the maximum number of RA opportunities per second are 10800 (= 54 × 2 × 100), while simultaneous RA opportunities (preambles per RA slot) are still bounded by 108. Also, if 30% contention-based preambles are initially allocated for low data rate MTCDs, the maximum RA opportunity for low data rate MTCDs per second are 3240. In addition, since LTE MAC protocol is slotted-Aloha based, the average RA success rate is around 37%. On the other hand, for massive MTC applications, a single event can drive several thousands of MTCDs to access the network almost simultaneously, and consequently, huge preamble collisions are anticipated.

An example scenario:

Consider an earthquake monitoring scenario in a densely populated urban area. Assume that MTCDs are deployed in a cell of radius 2 km with a density of 60 MTCDs per square kilometer. Thus, the density of MTCDs per cell is 754 (≈ π × 22 × 60). Also, consider that the speed of seismic surface wave is 10 km per second, which will result in 754 access attempts by MTCDs in 200 ms (2×10 1000). In this case, the probability of preamble collision is around 30% (≈ 1 - e- 10800 754 ×:2) with 10800 RA opportunities per second. However, if 30% of the contention-based preambles are dedicated for low data rate MTCDs, then the probability of collision will be 69% (≈ 1 - e- 3240 754 ×:2). Since the collision rate of slotted-Aloha system increases exponentially with increasing rate of RA attempts, the random access in LTE networks is likely to be unstable for massive MTC applications.

Chapter 4

EXISTING SOLUTIONS TO IMPROVE LTE RANDOM ACCESS

A. Optimized MAC

Aiming at applications where M2M devices transmit very small amounts of data, the authors in [23] suggest removing the need to connect to the network to transmit data. The key idea is to transmit data embedded into the access process by attaching data to either the preambles, i.e., Message 1, or into Message 3 of the RA process. While the solution based on the preambles is not very scalable due to the limited amount of available preambles, the transmission of data into Message 3 seems a very interesting and straightforward idea. These options may reduce significantly the amount of control information exchanged between M2M devices and the eNodeB, but at the expense of impeding any mobility or paging capabilities. Furthermore, the authors in [23] only provide the ideas, without performing any thorough mathematical or computational analysis to provide results, and there are no subsequent publication indicating the continuation of this effort.

B. Access Class Barring (ACB)

ACB is actually specified as a mechanism to control the access to the air interface in LTE and LTE-A [22]. 16 different classes are defined and some of them are reserved for high priority special uses, such as emergency services, security services and public utilities. In the case of network overload, the eNodeB transmits a set of parameters related to ACB as part of the system information; this includes a probability factor and a barring timer for the different classes; additionally, it transmits a set of barring bits for the high-priority cases.

Devices which attempt to access the network will draw a random number; if this number is lower than the probability factor, the device is able to attempt an access. Otherwise, the access is barred and the device performs a random backoff time (according to the barring timer value broadcast by the eNodeB) before scheduling the preamble transmission. For high-priority access classes, a string of bits indicates whether the access is being barred or not.

The throughput of the RACH can be improved with this mechanism. However, in the case of serious congestion, the probability factor might be set to a very restrictive value, i.e., dispersing access attempts over time and therefore, increasing the access delay. Several authors

have worked into this idea to adapt it for M2M scenarios. As specified by the 3GPP in [12], a different number of required classes could be defined, depending on the granularity of the control needed among the M2M devices. The main contributions can be summarized as:
1) Individual ACB Scaling: in order to achieve more control granularity, the network shall signal how individual devices or groups of devices will scale the barring parameters. This method is proposed in [12], but there is no indication of any further study regarding this solution.

2) Extended Access Barring (EAB): the basic idea is that devices that belong to delay-tolerant applications are not permitted to access the network in the case of congestion, leaving the contention for devices that are delay-constrained; this method has been proposed by the 3GPP as the most feasible baseline solution and is adopted for radio access overload control [12]. Simulation results have been provided in [24], [25], where different barring factors are given to M2M devices. This scheme slightly improves the access success probability, but access delay is severely increased.

3) Dynamic Access Barring: in this method, the eNodeB continuously monitors the loading state of the network in order to control the number of preamble transmissions on each RA slot. In the case of high traffic load, new arrivals from M2M devices are delayed until the conditions improve. This scheme has not been evaluated as a standalone solution. Instead, it corresponds to an integral part of the Prioritized Random Access proposal [25]. This solution is not compatible with delay-constrained applications, such as critical alarms, because their access cannot be postponed.

4) Cooperative ACB: this solution takes advantage of the high probability that M2M devices are in the coverage area of more than one cell (overlapping macro, pico, or femto-cells). In order to optimize the overall performance of the network, all the ACB parameters from every eNodeB are mutually optimized based on congestions levels at each eNodeB [26]. The impact on the air interface is minimal, as it only uses the probability factor parameters. This approach substantially reduces the delay, achieving 30% of improvement in comparison to the basic ACB scheme. However, this mechanism is only valid when M2M devices are located in the coverage area of more than one cell.

The main drawback of ACB mechanisms is the increased delay that some devices may experience. In addition, these schemes are not well-suited for event-driven applications where congestion can arise in a very short period of time. Even though the 3GPP considers EAB as the solution for overload control [12], other studies found in the literature coincide in suggesting that ACB mechanisms should not be considered as stand-alone solutions to overcome network congestion problems in M2M networks [27], [28].

C. Separation of RA Resources

This set of improvements can be also referred to as *Virtual Resource Allocation* [25]. The separation of resources can be achieved either by splitting the available preambles into Human-to-Human (H2H) and M2M subsets or by allocating different RA slots to H2H and M2M devices [12], [24]. Some studies have considered that H2H devices should be able to use all the resources and only the M2M devices will be restricted to the pre-defined subsets [29]. The separation of resources might help reducing the negative impact on non-M2M devices. Nevertheless, these solutions alone provide limited benefits, because the available resources are severely reduced for M2M devices and the performance tends to be worse under high M2M traffic load.

D. Other solutions

There can be found in the literature other solutions that cannot be classified into any specific group. Either because they combine some of the previous mechanisms, or because they propose some techniques that hold nothing in common with other proposals. These other proposals are:

1) Dynamic Allocation of RACH Resources: in this scheme, the network can allocate additional RA slots to M2M devices in the case of congestion, in order to cope with the additional load. Simulations results presented by the 3GPP in [24] show that this additional allocation can solve most of the cases of access congestion, providing high efficiency to the system. Therefore, the study concludes that allocating additional RA slots for M2M devices should be considered as the basic solution to solve the access overload. However, it is important to bear in mind that this allocation will occupy resources originally intended for data transmission. Therefore, it is not an effective improvement for high traffic load cases, as there is a tradeoff between the amount of access opportunities and the amount of resources available for data transmission.

2) Backoff Adjustment Schemes: for this improvement, different backoff timers are used to delay access attempts, assigning specific values to M2M devices. Although these schemes can provide some improvements for low congestion cases [24], [30], they are not sufficient to cope with peak congestion levels. The main reason for this is the fact that the average access delay will be severely degraded without substantially improving the access probability.

3) Slotted Access: in this scheme, dedicated RA slots are defined for each M2M device to access the network. M2M devices calculate their corresponding RA slot based on their identity and a parameter called RA cycle and broadcast by the eNodeB, which indicates the allocated RA slot periodicity [12]. The main drawback of this approach is the fact that in order to allocate a dedicated RA slot per device in the case of access overload, it is necessary to assign large RA cycles, leading to delays that many M2M applications will not tolerate. Nevertheless, this solution has been considered as an integral part of the Self-Optimizing Overload Control mechanism [31], which will be later explained.

4) Prioritized Random Access: in this proposal, the solution is based on the integration of two mechanisms: Virtual Resource Allocation and Dynamic Access Barring [25]. Virtual Resource Allocation is used to separate the RA resources in five different classes, namely: *i)* H2H, *ii)* low priority, *iii)* high priority, *iv)* scheduled, and *v)* emergency calls. M2M devices can only use the subset of resources according to their class. Dynamic Access Barring is used to bar new arrivals from M2M devices in the case of high traffic load. Simulation-based performance results are presented in [25] for different applications, i.e., voice calls, fleet management, hospital care, smart meters and seismic alarms. These results are compared with the EAB scheme, concluding that this solution is able to achieve better performance in comparison to other EAB methods in terms of both success probability and average access delay.

5) Self-Optimizing Overload Control (SOOC): the work presented in [31] proposes a self-optimizing mechanism that can configure the RA resources according to the load condition. The scheme is comprised of an adaptive integration of other solutions, including Separation of RACH Resources, ACB schemes, and slotted-access scheme. Two classes are added to the LTE-A ACB scheme for M2M devices, i.e., low priority and high priority. If a device is not able to get an access grant on the first attempt, it enters in overloaded control mode; this means that for the next attempt it will perform an ACB scheme before transmitting the next preamble. An important feature of SOOC is that it implements a mechanism to collect information for overload monitoring and adjusts RA resources. When a device receives a RAR, it sends the number of retransmitted preambles to the eNodeB within Message 3. With this information, the eNodeB can determine the congestion level of the RACH. Based on the congestion level, the eNodeB varies the RA slot provisioning. If the number of RA slots reaches a maximum available limit, then the eNodeB temporarily restricts the access to the lowest priority M2M class, until the overload conditions improve. This solution might be capable of handling high traffic loads. Unfortunately, the work presented in [31] only presents a theoretical analysis and no further results have been provided by means of either simulation or real implementation.

6) Code-Expanded RA: in [32], it is proposed a mechanism by which RA slots are assembled in groups referred to as virtual frames and the access is performed over these virtual groups. The mechanism consists in transmitting codewords instead of preambles. A codeword is created when an M2M device transmits one preamble on each of the RA slots that composes the virtual frame. This allows expanding the number of contention resources and, therefore, reducing the collisions. The performance of this proposal has been evaluated through computerbased simulations and the results show that it is especially suited for high traffic loads. The only noticeable drawback of this proposal is its associated energy consumption. Note that, for each attempt, the device must perform more than one preamble transmission per each access attempt.

E. Distributed Queuing

All the previously presented proposals are aimed to enhance the RACH performance considering the possible massive access situations that M2M communications may bring about. However, to some extent, all of them are finally based on ALOHA-like mechanisms. This fact generates a certain level of instability, inefficiency, and uncertainty in the access outcome. There exist other approaches that can tackle these issues, in a more efficient manner. In particular, Campbell and Xu [33] proposed a MAC protocol whose high performance is completely independent of the number of nodes/users sharing a common channel. This is specially fitted to M2M communications, where high density of uncoordinated devices may put a really tight challenge into the access to the system. Since the very first proposal [33], several studies have analyzed the performance of the protocol for a wide set of study case scenarios [34], [35]. All of them demonstrate the stability of its performance and the near optimum behavior in terms of channel utilization, access delay, and energy consumption for all system layouts. Furthermore, several extensions and adaptations for different wireless systems have been also proposed in the last years such as for 3G networks [36], WLAN [37], mobile ad-hoc networks [38] and BANs [39]. In all these cases, the protocol has shown its great performance for any mixture of traffic patterns, loads, and Quality of Service (QoS) requirements.

The key element of the protocol is the so-called Distributed Queuing (DQ) paradigm. In a nutshell, DQ is based on the combination of a m-ary tree splitting algorithm with a smart set of simple rules that allow organizing every device in one out of two virtual queues. These queues actually do not exist physically, but they are logically distributed queues maintained by all the devices in the network. These queues have a partial representation at each device using only four integer counters to represent the total amount of devices in each queue, and the current position of the device in each queue, respectively. The appropriate update of the values of these counters, performed in a distributed manner and autonomously, allows each device to know the exact state of the queues, including their own position within them. In this way, the devices know when their turn to transmit has arrived, indirectly acquiring the access grant for transmission while completely avoiding collisions.

The smart distributed scheduling of the queues permits having almost full utilization of the channel regardless of its capacity, the number of the transmitting nodes, and the traffic pattern. Due to the rules of DQ, it behaves as a random access method for low traffic loads, and it switches smoothly and seamlessly to a reservation access method as the traffic load increases. These dynamics of DQ makes it an ideal candidate to be considered for the RACH of LTE and LTE-A under the presence of a high number of competing devices. These features perfectly match the requirements of M2M communications, especially for massive access when a high number of devices must share the same channel resources. Some ongoing research efforts are being carried out by the research groups led by Luis Alonso at [40] and Jesus Alonso Zarate at [41] in order to propose different ways of applying DQ ideas within LTE and LTE-A systems.

Chapter 5

OBJECTIVE

According to the comprehensive research [42], in the real scene where delay-sensitive and delay-tolerant devices coexist, although the occurrence of delay-sensitive devices, such as the hospital e-care, is very rare, the delay-sensitive devices are strict with delay, and they need instant processing. For the application in hospital e-care, the maximum tolerant delay is five milliseconds, since in a real system, the data become useless after that. On the other hand, delay-tolerant devices, such as smart grids [43], can tolerate several seconds or even minutes, and a great majority of devices is subject to this category.

In our proposal we have categorized the M2M devices based on their delay sensitivity and priority i.e the most delay sensitive and the most prioritized ones will be on Type 1 and further types will be determined according to their delay sensitivity and priority following the descending order. The devices will get to know in which Type they belong from their built-in properties. Type 1 devices will get maximum number of RA slots in a radio frame, Type 2 devices will get less number of slots than that of Type 1 and other types will also get RA slots in this manner resulting that the most delay tolerant devices will get least number of RA slots. This allocation of RA slots among the types will be broadcasted by the eNB through the SIB Type-2 message where PRACH config index will change according to the traffic.

In the existing methods, where both delay sensitive and delay tolerant devices coexist, the delay sensitive ones get useless due to unwanted delays caused by the RAN overload and this scenario sometimes get critical. In our proposed scheme we have given more advantage to the delay sensitive ones over the delay tolerant ones which will have great impact on RAN overload control as the delay sensitive devices will get more chance of random access than the delay tolerant devices.

Chapter 6

SIMULATION AND RESULT

All the related simulation are done in MATLAB version 2015.We have considered both H2H and M2M devices coexist.

Parameters:

Packet arrival rate of M2M devices, λ_m = 0.005 to 0.05(increment with 0.05 step)
Packet arrival rate of H2H devices, λ_h = 0.9
Number of H2H devices, N_h = 1000
Number of M2M devices, N_m = [1000 2000 3000 4000 5000 6000 7000 8000 9000 10000]
Slot time, T_c = 1ms

We have assumed that the packet arrival follows Poisson process and the inter-arrival times are distributed exponentially.

The probability of packet arrival in t_1 duration is $\int_0^{t_1} \lambda e^{-\lambda t} dt = 1 - e^{-\lambda t_1}$. Thus, the probability of no packet arrival in t_1 duration is

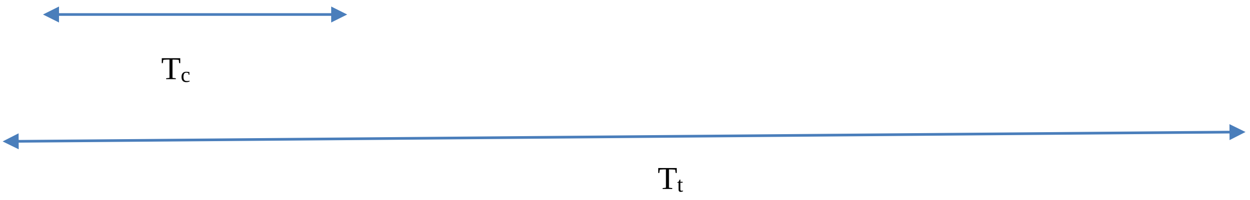
$$1 - \int_0^{t_1} \lambda e^{-\lambda t} dt = e^{-\lambda t_1}$$

We have separately calculated the probabilities of packet arrival for each type in each Tc. We denote the probabilities as Pij where i indicates type no. and j indicates slot no.Such as : P_{11} , P_{12} , P_{13} , P_{14} , P_{21} etc.

While calculating Pij for any type in any slot we have considered the probabilities of no packet arrival for other types of M2M devices and also for H2H devices.

We have assumed the following configuration of time slot

Type 1	Type 1	Type 1	Type 1
Type 2	Type 2	Type 4	Type 2
	Type 3		Type 3



Here Type 1 devices get all the RA slots as they are the most delay sensitive and prioritized one. Then Type 2 devices get 3 RA slots out of 4,then Type 3 devices get 2 RA slots and Type 4 devices get only 1 RA slot.

The probability of one packet arrival in the first RA slot for Type 1 device,

$$P_{11} = (e^{-\lambda_m T_c})^{Nm/4} * (e^{-\lambda_h T_c})^{Nh}$$

The probability of one packet arrival in the second RA slot for Type 1 device,

$$P_{12} = (e^{-\lambda_m T_c})^{Nm/4} * (e^{-2\lambda_m T_c})^{Nm/4} * (e^{-\lambda_h T_c})^{Nh}$$

The probability of one packet arrival in the third RA slot for Type 1 device,

$$P_{13} = (e^{-\lambda_m T_c})^{Nm/4} * (e^{-\lambda_h T_c})^{Nh}$$

The probability of one packet arrival in the fourth RA slot for Type 1 device,

$$P_{14} = (e^{-2\lambda_m T_c})^{Nm/4} * (e^{-2\lambda_m T_c})^{Nm/4} * (e^{-\lambda_h T_c})^{Nh}$$

The probability of one packet arrival in the first RA slot for Type 2 device,

$$P_{21} = (e^{-\lambda_m T_c})^{Nm/4} * (e^{-\lambda_h T_c})^{Nh}$$

The probability of one packet arrival in the second RA slot for Type 2 device,

$$P_{22} = (e^{-\lambda_m T_c})^{Nm/4} * (e^{-2\lambda_m T_c})^{Nm/4} * (e^{-\lambda_h T_c})^{Nh}$$

The probability of one packet arrival in the fourth RA slot for Type 2 device,

$$P_{24} = (e^{-\lambda_m T_c})^{Nm/4} * (e^{-2\lambda_m T_c})^{Nm/4} * (e^{-\lambda_h T_c})^{Nh}$$

The probability of one packet arrival in the second RA slot for Type 3 device,

$$P_{32} = (e^{-\lambda_m T_c})^{Nm/4} * (e^{-\lambda_m T_c})^{Nm/4} * (e^{-\lambda_h T_c})^{Nh}$$

The probability of one packet arrival in the fourth RA slot for Type 3 device,

$$P_{34} = (e^{-\lambda_m T_c})^{Nm/4} * (e^{-2\lambda_m T_c})^{Nm/4} * (e^{-\lambda_h T_c})^{Nh}$$

The probability of one packet arrival in the fourth RA slot for Type 4 device,

$$P_4 = (e^{-\lambda_m T_c})^{Nm/4} * (e^{-\lambda_h T_c})^{Nh}$$

The probability of one packet arrival in time $T_{t,}$

P$_0$ = $e^{-\lambda_m T_t}$

The probability of two packet arrival in time $T_{t,}$

P$_1$ = $\left(e^{-\lambda_m T_t}\right) * (\lambda_m T_t)$

The probability of three packet arrival in time $T_{t,}$

P$_2$ = $\left(e^{-\lambda_m T_t}\right) * (\lambda_m T_t)^2 / 2$

The probability of three packet arrival in time $T_{t,}$

P$_3$ = $\left(e^{-\lambda_m T_t}\right) * (\lambda_m T_t)^3 / 3!$

The probability of four packet arrival in time $T_{t,}$

P$_4$ = $\left(e^{-\lambda_m T_t}\right) * (\lambda_m T_t)^4 / 4!$

Then we have calculated number of packets successfully arrived in a RA slot and completed random access successfully for a large number of RA slots(say 10000)

In case of Type 1 devices,

Number of packets successfully arrived in a RA slot and completed random access successfully for 10000 RA slots,

$(P_1 * P_{11} + P_2 * (P_{11} + P_{12}) + P_3 * (P_{11} + P_{12} + P_{13}) + P_4 * (P_{11} + P_{12} + P_{13} + P_{14})) * 10000$

In case of Type 2 devices,

Number of packets successfully arrived in a RA slot and completed random access successfully for 10000 RA slots,

$(P_1 * P_{21} + P_2 * (P_{21} + P_{22}) + P_3 * (P_{21} + P_{22} + P_{24})) * 10000$

In case of Type 3 devices,

Number of packets successfully arrived in a RA slot and completed random access successfully for 10000 RA slots,

$(P_1*P_{32} + P_2*(P_{32} + P_{34}))*10000$

In case of Type 4 devices,

Number of packets successfully arrived in a RA slot and completed random access successfully for 10000 RA slots,

$P_1*P_4*10000$

We will also calculate the following three performance indicators:

1) Access Success Probability, defined as the probability to complete the random access procedure in the maximum number of preamble transmissions allowed. This parameter can also be represented by the blocking probability, defined as the probability that a device reaches the maximum number of transmission attempts and is unable to complete an access process.

2) Preamble Collision Rate, defined as the ratio between the number of preamble collisions in the same RA slot and the total number of preambles transmitted on that slot. An equivalent metric consists in measuring the average number of preamble retransmissions required to have a successful access request.

3) Access Delay, defined as the time elapsed between the transmission of the first preamble and the reception of Message 4 by the M2M device.

RESULT:

We have calculated the probable number of successful packets that have arrived successfully and completed random access for different types. In our analysis it is assumed that when a packet from a particular type is competing for a RA slot, other types of devices won’t compete and we have also set that assumption for H2H devices. We have compared our result with the existing scenario where a device competing for a RA slot will suffer from collision if any other device with same preamble also compete for that slot.

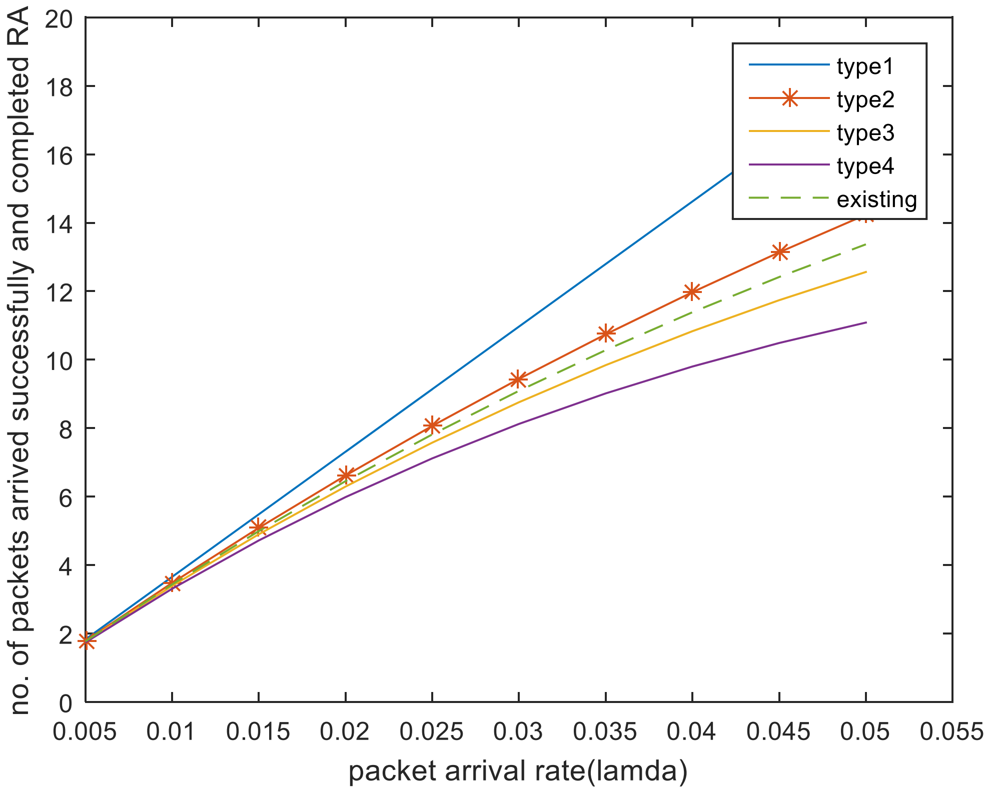


Fig 6.1: no. of packets arrived successfully and completed RA

From the result,we see that Type 1 and Type 2 devices give better performance in case of number of packets arriving successfully and completing RA than existing methods.Type 1 and Type 2 devices include the most delay sensitive devices which get more chance of RA in our proposed method and also the number of packets successfully undergone RA is also higher than the existing methods.Type 3 and Type 4 devices get less chance of RA and the number of packets successfully undergone RA is less than existing method but it won’t be a problem as these are delay tolerant devices.

Chapter 7

CONCLUSION

Machine-to-machine service is one of the important services to be supported in next generation networks. The bursty nature of access attempts generated by M2M devices may congest the RACH of the cellular networks and thus, severely degrade the service quality of the M2M and H2H traffic. To solve the RAN overload problem in LTE and LTE-A networks, we proposed a new mechanism which is applicable to the scenarios where delay-sensitive and delay-tolerant services coexist. The key of our scheme is to categorize the M2M devices based on their delay sensitivity and priority.

In existing method the delay sensitive and high priority M2M devices do not get enough chance for successful random access procedure due to RACH overload. These devices have to compete with the delay tolerant ones to access the network. The delay tolerant and low priority M2M devices can function properly with less opportunity to access the network. But the delay sensitive and high priority devices do not function properly if these devices do not get necessary opportunity to access the network. .After categorizing the M2M devices we see that the delay sensitive and high priority devices get better chance for successful random access procedure

We theoretically discussed the performance analysis. In addition, computer simulations demonstrated the performance superiority of our proposed scheme over the existing scheme in terms of the number of packets arriving successfully and completing RA.

In future we will work to calculate the access success probability, preamble collision rate and access delay of our proposed scheme to prove the effectiveness of our proposal.

References

[1] 3GPP GP-100892, “RACH Capacity Evaluation for MTC,” TSG GERAN # 46, 2010.

[2] IEEE, ”TGah Functional Requirements and Evaluation Methodology”.

[3] ETSI GS LTN 001,”Low Throughput Networks (LTN): Use Cases, Functional Architecture and Protocols”.

[4] I.Vilajosana, J. Llosa, B. Martinez, M. Domingo-Prieto, A. Angles, and X.Vilajosana, “Bootstrapping smart cities through a self-sustainable model based on big data flows,” *IEEE Communications Magazine*, vol. 51, no. 6, pp. 128–134, 2013.

[5] Machina Research, The need for low cost, high reach, wide area connectivity for the Internet of Things, 2014.

[6] ETSI EN 300 200-1, ”Electromagnetic compatibility and Radio spectrum Matters (ERM); Short Range Devices (SRD); Radio equipment to be used in the 25 MHz to 1 000 MHz frequency range with power levels ranging up to 500 mW; Part 1: Technical characteristics and test methods”.

[7] K. Zheng, S. Ou, J. Alonso-Zarate, M. Dohler, F. Liu, and H. Zhu, “Challenges of massive access in highly dense LTE-advanced networks with M2M communications,”*IEEE Wireles Communications*, vol. 21, no. 3, pp. 12–18, 2014.

[8] M. Condoluci, M. Dohler, G. Araniti, A. Molinaro, and J. Sachs, “Enhanced radio access and data transmission procedures facilitating industry-compliant machine-type communications over LTE-based 5G networks,” in *IEEE Wireless Communications*, 2015.

[9] Study on RAN improvements for machine-type communications. 3GPP TR 37.868, 2011.

[10] T.Tirronen, A.Larmo, J.Sachs, B.Lindoff, and N.Wiberg,“Machineto-machine communication with long-term evolution with reduced device energy consumption,” Transactions on Emerging Telecommunications Technologies, vol. 24, no. 4, pp. 413–426, 2013.

[11] S.Andreev, A.Larmo, M.Gerasimenko, V.Petrov, O.Galinina T. Tirronen, J. Torsner, and Y. Koucheryavy, “Efficient small data access for machine-type communications in LTE,” in Proc. of the IEEE International Conference on Communications (ICC), pp. 3569–3574, 2013.

[12] 3GPP TR 37.868 V11.0.0, “Study on RAN Improvements for Machine Type Communications,” Sep. 2011.

[13] 3GPP TSG RAN WG2 #70bis R2-103742, “RACH overload solutions,” ZTE, Stockholm, Sweden, 28th Jun. 2010.

[14] 3GPP TSG RAN WG2 #71 R2-104663, “[70bis#11] LTE: MTC LTE simulations ,” ZTE, Madrid, Spain, 23rd Aug. 2010.

[15] 3GPP TR 23.888 V11.0.0, “System improvements for Machine-Type Communications,” Sep. 2012.

[16] 3GPP TS 22.368 V11.2.00, “Service requirements for Machine-Type Communication,” Sep. 2012.

[17] Laya, A.; Alonso, L.; Alonso-Zarate, J. Is the random access channel of LTE and LTE-A suitable for M2M communications? a survey of alternatives. IEEE Commun. Surveys Tuts. 2014, 16, 4–16.

[18] Cheng, J.-J.; Lee, C.-H.; Lin, T.-M. Prioritized random access with dynamic access barring for RAN overload in 3GPP LTE-A networks. In Proceedings of the IEEE Global Communications Conference (GLOBECOM) Workshops, Houston, TX, USA, 5–9 December 2011; pp. 368–372.

[19] 3GPP TS 36.321 V12.7.0, “Evolved Universal Terrestrial Radio Access: Medium Access Control,” France, Sep. 2015.

[20] Sesia, S. and Baker, M. and Toufik, I., LTE - The UMTS Long Term Evolution: From Theory to Practice. Wiley, 2011, pp. 421–456.

[21] 3GPP TS 36.321 V9.3.0, “Evolved Universal Terrestrial Radio Access (E-UTRA); Medium Access Control (MAC),” Jun. 2010.

[22] 3GPP TS 36.331 V10.5.0, “Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC),” Mar. 2012.

[23] Y. Chen and W. Wang, “Machine-to-Machine Communication in LTEA,” in Vehicular Technology Conference Fall (VTC 2010-Fall), 2010 IEEE 72nd, sept. 2010, pp. 1–4.

[24] 3GPP TSG RAN WG2 #71 R2-104662, “MTC simulation results with specific solutions ,” ZTE, Madrid, Spain, 23rd Aug. 2010.

[25] J.-P. Cheng, C. han Lee, and T.-M. Lin, “Prioritized Random Access with dynamic access barring for RAN overload in 3GPP LTE-A networks,” in *GLOBECOM Workshops (GC Wkshps), 2011 IEEE,* dec. 2011, pp. 368–372.

[26] S.-Y. Lien, T.-H. Liau, C.-Y. Kao, and K.-C. Chen, “Cooperative Access Class Barring for Machine-to-Machine Communications,”Wireless Communications, IEEE Transactions on, vol. 11, no. 1, pp. 27 –32, January 2012.

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

[28] M.-Y. Cheng, G.-Y. Lin, H.-Y. Wei, and A.-C. Hsu, “Overload control for Machine-Type-Communications in LTE-Advanced system,” Communications Magazine, IEEE, vol. 50, no. 6, pp. 38–45, june 2012.

[29] K.-D. Lee, S. Kim, and B. Yi, “Throughput Comparison of Random Access Methods for M2M Service over LTE Networks,” in GLOBECOM Workshops (GC Wkshps), 2011 IEEE, dec. 2011, pp. 373 –377.

[30] X. Yang, A. Fapojuwo, and E. Egbogah, “Performance Analysis and Parameter Optimization of Random Access Backoff Algorithm in LTE,” in Vehicular Technology Conference (VTC Fall), 2012 IEEE, 2012, pp. 1–5.

[31] A. Lo, Y. Law, M. Jacobsson, and M. Kucharzak, “Enhanced LTE-Advanced Random-Access Mechanism for Massive Machine-toMachine (M2M) Communications,” in Proceedings of the 27th Meeting of Wireless World Research Form (WWRF), Oct. 2011.

[32] N. K. Pratas, H. Thomsen, C. Stefanovic, and P. Popovski, “CodeExpanded Random Access for Machine-Type Communications,” CoRR, vol. abs/1207.0362, 2012.

[33] W. Xu and G. Campbell, “A Near Perfect Stable Random Access Protocol for a Broadcast Channle,” in IEEE Proc. ICC92, vol. 1, 1992, p. 370374.

[34] X. W. and G. Campbell, “DQRAP, A Distributed Queueing Random Access Protocol for a Broadcast Channel,” in Computer Commun. Mag., vol. 23, no. 4, Oct. 1993, p. 370374.

[35] Luis Alonso, Ramon Agusti, Oriol Sallent, “A Near-Optimum MAC Protocol Based on the Distributed Queueing Random Access Protocol (DQRAP) for a CDMA Mobile Communication System,” in IEEE Journal on Selected Areas in Communications, vol. 18, no. 9, Sep. 2000, pp. 1701–1718.

[36] Begona Otal, Luis Alonso, Ramon Agusti, “Design and analysis of a cellular mobile communications system based on DQRAP/CDMA MAC protocol,” in IEEE Electronic Letters, vol. 38, no. 3, Jan. 2002, pp. 138–139.

[37] E. Kartsakli, J.Alonso-Zarate, C. Verikoukis, L. Alonso, “Cross-Layer ´ Enhancement for WLAN Systems with Heterogeneous Traffic based on DQCA,” in IEEE Communications Magazine, vol. 46, no. 6, Jun 2008, pp. 60–66.

[38] Jesus Alonso-Zarate, E. Kartsakli, L. Alonso and Christos Verikoukis, ´“Performance Analysis of a Cluster-Based MAC Protocol for Wireless Ad Hoc Networks,” in EURASIP Journal on Wireless Communications and Networking, vol. 2010, 2010.

[39] Begona Otal, Christos Verikoukis, Luis Alonso, “Highly Reliable Energy-saving MAC for Wireless Body Sensor Networks in Healthcare Systems,” in IEEE Journal on Selected Areas in Communications, Special Issue on Wireless and Pervasive Communications in Healthcare, vol. 27, no. 4, May 2009, pp. 553–565.

[40] Wireless Communications and Technologies Group, Universitat Politecnica de Catalunya (UPC), Barcelona. [Online]. Available: `http://wicomtec.upc.edu/

[41] Centre Tecnologic de Telecomunicacions de Catalunya (CTTC), ` Barcelona. [Online]. Available: http://www.cttc.es/

[42] Afrin, N.; Brown, J.; Khan, J.Y. Performance analysis of an enhanced delay sensitive LTE uplink scheduler for M2M traffic. In Proceedings of the Telecommunication Networks and Applications Conference (ATNAC), Christchurch, New Zealand, 20–22 November 2013; pp. 154–159.

[43] Lien, S.-Y.; Chen, K.-C.; Lin, Y. Toward ubiquitous massive accesses in 3GPP machine-to-machine communications. IEEE Commun. Mag. **2011**, 49, 66–74.