2-1-2015

Statistical priority-based uplink scheduling for M2M communications

Ahmed Elhamy Mostafa

The American University in Cairo (AUC)

The School of Sciences and Engineering

The Department of Electronics and Communications Engineering

Statistical Priority-based Uplink Scheduling for M2M Communications

By

Ahmed Elhamy Ahmed Mostafa

(ID #: 900-07-2109)

A thesis submitted to

The Department of Electronics and Communications Engineering

In partial fulfillment of the requirements for the degree of

Master of Science (M.Sc.)

under the supervision of Dr. Yasser Gadallah

July 2015

Approval sheet goes here

Dedication

To my mother

Acknowledgment

I would like to thank my thesis supervisor, Dr. Yasser Gadallah, for his guidance and endless support throughout the research process. The meetings and discussions with him were fruitful and useful for the quality of my research.

I would like to thank all the professors who taught me thorough my B.Sc. and M.Sc. degrees courses; Dr. Sherif Abdel Azeem, Dr. Ahmed Abou-Auf, Dr. Hassanein Amer, Dr. Mohab Anis, Dr. Ali Darwish, Dr. Ramez Daoud, Dr. Ayman Elezabi, Dr. Ayman Hassan and Dr. Karim Seddik. In addition, I would like to thank the professors whom I worked with as a teaching assistant; Dr. Hanna Kirolous. I would like also to acknowledge my examiner; Dr. Magdi Fikri.

I would like to thank the department of Electronics and Communications Engineering (ECNG) and the office of graduate studies for their support and assistance.

I would like also to thank all my friends for their moral support to me through my M.Sc. work. Special thanks for my M.Sc. colleagues and dear friends Shady El-Bassiouny and Hassan Halawa.

I would like also to thank the MEMIC team that I enjoyed working with at Intel labs. This team helped me to learn a lot about the field of my research and the team leaders, Dr. Hani El-Gebaly and Dr. Ahmed Ibrahim, had a great understanding of M.Sc. work duties.

Finally, I would like to express my gratitude to my family for their patience, sacrifice and support throughout the years of working on my M.Sc. degree.

Table of Contents

List of Tables

List of Figures

List of Acronyms

1G	1st Generation
2G	2nd Generation
3G	3rd Generation
3GPP	3rd Generation Partnership Project
4G	4th Generation
5G	5th Generation
AGTI	Access Grant Time Interval
AMPS	Analogue Mobile Phone System
AWGN	Additive White Gaussian Noise
BAT	Balanced Alternating Technique
BPSK	Binary Phase Shift Keying
BQA	Bandwidth and QoS Aware
BSR	Buffer Status Report
CBDP	Class-Based Dynamic Priority
CCDF	Complimentary Cumulative Distribution Function
CDMA	Code Division Multiple Access
CoMP	Co-ordinated Multi-point
CP	Cyclic Prefix
CQI	Channel Quality Indicator
CS/CB	Coordinated Scheduling/ Coordinated Beamforming
D2D	Device-to-Device
DFT	Discrete Fourier Transform
DM-RS	Demodulation Reference Signals
DwPTS	Downlink Pilot Time Slot
eICIC	Enhanced ICIC
eNB	Evolved Node B
FCFS	First-Come-First-Served
FDD	Frequency Division Duplexing
FDMA	Frequency Division Multiple Access
FDPS	Frequency Domain Packet Scheduling
GP	Guard Period
GPRS	General Packet Radio Service
GSM	Global System for Mobile communications
H2H	Human-to-Human
HARQ-ACK	Hybrid Automatic Repeat Request-Acknowledgment
HeNB	Home eNB
HSPA	High Speed Packet Access
ICIC	Inter-Cell Interference Cancelation
IDFT	Inverse Discrete Fourier Transform
IFDMA	Interleaved FDMA
IoT	Internet of Things
IP	Internet Protocol
ISI	Inter-Symbol Interference
JT	Joint Transmission
LFDMA	Localized FDMA
LTE	Long Term Evolution
LTE-A	LTE-Advanced
M2M	Machine-to-Machine
MBMS	Multicast Broadcast Multimedia Services
MCS	Modulation and Coding Scheme
MGBR	Minimum Guaranteed Bit Rate
MIMO	Multiple Input Multiple Output

mmWave	Millimeter Wave
MTC	Machine Type Communications
MTCD	Machine Type Communications Device
MTCG	Machine Type Communications Gateway
MU-MIMO	Multi User MIMO
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
PAPR	Peak-to-Average Power Ratio
PMI	Precoding Matrix Indicator
PRACH	Physical Random Access Channel
PRB	Physical Resource Block
PUCCH	Physical Uplink Control Channel
PUSCH	Physical Uplink Shared Channel
QCI	QoS Class Indicator
QoE	Quality of Experience
QoS	Quality of Service
RAT	Radio Access Technology
RE	Resource Element
RI	Rank Indicator
RME	Recursive Maximum Expansion
RTTS	Remaining Time To Serve
SC-FDMA	Single Carrier FDMA
SNR	Signal-to-Noise Ratio
SP	Statistical Priority
SPR	Statistical Priority Report
SR	Scheduling Request
SU-MIMO	Single User MIMO
TDD	Time Division Duplexing
TDMA	Time Division Multiple Access
TDPS	Time Domain Packet Scheduling
TTI	Transmission Time Interval
UDFS	Unique Depth First Search
UE	User Equipment
UMTS	Universal Mobile Telecommunications System
UpPTS	Uplink Pilot Time Slot
WCDMA	Wide-band CDMA
WSN	Wireless Sensor Network

Abstract

Currently, the worldwide network is witnessing major efforts to transform it from being the Internet of humans only to becoming the Internet of Things (IoT). It is expected that Machine Type Communication Devices (MTCDs) will overwhelm the cellular networks with huge traffic of data that they collect from their environments to be sent to other remote MTCDs for processing thus forming what is known as Machine-to-Machine (M2M) communications.

Long Term Evolution (LTE) and LTE-Advanced (LTE-A) appear as the best technology to support M2M communications due to their native IP support. LTE can provide high capacity, flexible radio resource allocation and scalability, which are the required pillars for supporting the expected large numbers of deployed MTCDs.

Supporting M2M communications over LTE faces many challenges. These challenges include medium access control and the allocation of radio resources among MTCDs. The problem of radio resources allocation, or scheduling, originates from the nature of M2M traffic. This traffic consists of a large number of small data packets, with specific deadlines, generated by a potentially massive number of MTCDs. M2M traffic is therefore mostly in the uplink direction, i.e. from MTCDs to the base station (known as eNB in LTE terminology). These characteristics impose some design requirements on M2M scheduling techniques such as the need to use insufficient radio resources to transmit a huge amount of traffic within certain deadlines. This presents the main motivation behind this thesis work.

In this thesis, we introduce a novel M2M scheduling scheme that utilizes what we term the "statistical priority" in determining the importance of information carried by data packets. Statistical priority is calculated based on the statistical features of the data such as value similarity, trend similarity and auto-correlation. These calculations are made and then reported by the MTCDs to the serving eNBs along with other reports such as channel state. Statistical priority is then used to assign priorities to data packets so that the scarce radio resources are allocated to the MTCDs that are sending statistically important information. This would help avoid exploiting limited radio resources to carry redundant or repetitive data which is a common situation in M2M communications.

In order to validate our technique, we perform a simulation-based comparison among the main scheduling techniques and our proposed statistical priority-based scheduling technique. This comparison was conducted in a network that includes different types of MTCDs, such as environmental monitoring sensors, surveillance cameras and alarms. The results show that our proposed statistical priority-based scheduler outperforms the other schedulers in terms of having the least losses of alarm data packets and the highest rate in sending critical data packets that carry non-redundant information for both environmental monitoring and video traffic. This indicates that the proposed technique is the most efficient in the utilization of limited radio resources as compared to the other techniques.

Chapter 1: Introduction

The Internet of Things (IoT) is considered the network of the forseen future [1]. It is the network through which all objects (things) with communication capabilities are connected, without human intervention in general, to achieve certain tasks and goals. Most of the communications through the IoT will therefore be conducted via what is called Machine-to-Machine (M2M) communications [2].

1.1 M2M Communications

M2M communications or Machine Type Communications (MTC) is the communications among devices or machine-type nodes that are usually called Machine Type Communications Devices (MTCDs) in the context of cellular networks. MTCD is the device used to collect information from the environment (like sensing, surveillance and counting). Unlike Human-to-Human (H2H) communications which are carried out among several User Equipment (UE) units, M2M communications are generally characterized with massive deployments, and hence massive access, combined with small data payloads. For example, MTCDs' data can be generated by event triggering (e.g. alarms) or in the form of periodic reports (e.g. environmental monitoring applications). M2M communications are used in a wide variety of applications including periodic measurement reporting (e.g. temperature, pressure, etc.), surveillance (e.g. security cameras), alarm systems (e.g. fire alarms), statistical survey and counting systems (e.g. people and vehicle counting), intelligent transportation systems, healthcare, farming and industrial production lines.

M2M communications is expected to be the dominant traffic source in cellular networks in 5^{th} generation (5G) time frame and beyond. The number of MTCDs is expected to reach 2.1 billion by 2020 [3]. The number of connected devices (UEs and MTCDs) started to exceed the earth's population in 2009 [4]. Due to the decline in the manufacturing cost and the growth of interest by consumer electronics manufacturers, the adoption of M2M solutions in the different fields of business shows strong prospects [3].

The first potential enabler of M2M communications used to be the General Packet Radio Service (GPRS) since it is a well-established protocol with Internet Protocol (IP) native connectivity. The main disadvantage of GPRS is its low capacity (100-150 kbps/cell/MHz) which limits the number of devices that can be connected within a GPRS cell i.e. it limits the scalability of M2M networks [5]. With the proliferation of the IoT and M2M communications of potentially a massive number of devices, Long Term Evolution (LTE) is now seen as the best technology to support MTC due to its Internet compatibility, high capacity, flexibility in radio resources management and scalability.

1.2 Problem Statement and Research Objectives

One of the main challenges in adopting LTE for M2M communications is the problem of radio resource management or *scheduling*. M2M communications have different characteristics when compared to H2H communications. M2M traffic consists of mainly small bursty payloads that exist mostly in the uplink direction (i.e. the direction from a device to the serving base station). In addition, using H2H LTE uplink scheduling algorithms [6] that focus mainly on throughput maximization, radio resources allocation efficiency and preserving the contiguity of radio resources that are assigned to one device is not sufficient. MTCDs have different requirements that include power consumption reduction and adhering to deadlines (time before which data must be transmitted for an emergency alert or to avoid obsolescence of information carried by these data). MTCDs also operate with a wide variety of applications. Each application has its own requirements in terms of deadlines and required Quality of Service (QoS).

Another aspect of the scheduling problem for M2M communications over LTE is the scarcity of radio resources when compared to the expected radio resources needs of MTCDs. It is also worth noting that M2M communications may be sharing the same radio resources of H2H UEs. This highlights the challenges related to the efficiency of dividing the radio resources between H2H and M2M data flows while preserving the Quality of Experience (QoE) promised by LTE for H2H UEs.

The challenge that we focus on in this research is to design an M2M uplink scheduler that is based on the statistical attributes of the data. Since radio resources are limited for massive M2M communications, the scheduling algorithm should consider, as a main factor, the importance of information carried by data traffic of the different MTCDs. The data reported by many monitoring devices are repetitive which means that they do not necessarily carry high-value information all the time. Therefore, the aspect of data similarity needs to be considered while performing the scheduling in case of limited radio resources.

The problem that this research addresses can therefore be stated as "the design of a statistical priority-based uplink scheduling technique for massive M2M deployments over LTE".

In order to address this problem, there is a need to find a method to define which data traffic is most valuable to send i.e. it has unique information of high value. The approach followed in this research is to consider the statistical features of the data sent by MTCDs, such as the time autocorrelation of data as well as the trend of data points. After determining the statistical characteristics of data, we quantify the value of information carried by data. The following step is to utilize this quantified value as a scheduling metric and prove its effectiveness for M2M communications in case of radio resources limitation.

The objectives of this research can therefore be summarized in the following:

- Investigating statistical features of interest in M2M data traffic to identify the value of information carried by the data points.
- Quantifying the statistical-based value of information using a set of utility functions.
- Using the statistical-based value of information as the core of a novel scheduling technique that we introduce for massive M2M deployments.

1.3 Thesis Contributions and Structure

The Contributions of this research can be summarized as follows:

- A novel classification of M2M uplink scheduling techniques (Chapter 3).
- A Balanced Alternating Technique (BAT) to combine both channel-state and system deadlines for M2M uplink scheduling (Chapter 3).
- A novel statistical priority-based scheduling metric for uplink scheduling of massive M2M deployments over LTE (Chapter 4).

- A novel flexible statistical priority-based scheduling algorithm for M2M uplink scheduling (Chapter 5).

This thesis is organized as follows. In Chapter 2, the necessary background of LTE, M2M communications and uplink radio resource scheduling is discussed. In this chapter, a review of LTE basics, LTE frame structure and LTE signals and channels is provided. In addition, an overview of the LTE uplink scheduling process as well as scheduling algorithms classification is introduced, with focus on M2M communications. Chapter 3 includes a literature review of LTE uplink M2M-specific scheduling algorithms. It includes the description and the critique of the M2M scheduling algorithms proposed in literature. This chapter is concluded with a comparison of the surveyed algorithms based on their characteristics and scheduling metrics. In Chapter 4, we introduce our novel statistical priority metric for M2M uplink scheduling over LTE. The new metric is explained with an overview of data statistical features of interest in environmental monitoring data and video classes of traffic. Then, validation results for different types of traffic are shown. In chapter 5, we present a novel flexible dynamic M2M uplink scheduling algorithm that is based on the statistical priority metric that we introduced in Chapter 4. Chapter 6 concludes the thesis.

This research has resulted in a conference paper [7], with another paper being currently prepared.

Chapter 2: Background

2.1 Introduction

In this chapter, we present the necessary background for understanding our research objectives and contributions. This background focuses on two aspects, namely, LTE basics and M2M communications. An overview of LTE scheduling process is provided with clarifying the effect of M2M networks requirements on this process.

2.2 Long Term Evolution (LTE)

LTE is a wireless cellular communications technology that was presented by the 3rd Generation Partnership Project (3GPP) for first time in 2000. It was proposed as a candidate technology that satisfies the specifications set by the IEEE for the 4th generation (4G) technology.

2.2.1 LTE History

The 1st generation (1G) of wireless cellular communications was analog and based on Frequency Division Multiple Access (FDMA) as the multiple access scheme. An example of the 1G technology is the Analogue Mobile Phone System (AMPS) that was commercially deployed in the 1980s in North America [8]. It is no longer supported in cellular networks. The 2nd generation (2G) was digital and its most famous technology, which is still operating till now, is Global System for Mobile communications (GSM). GSM was developed in Europe to be a digital Time Division Multiple Access (TDMA) based cellular scheme [9]. In the USA, US-TDMA was introduced then followed by a Code Division Multiple Access (CDMA) based system which is the IS-95 [9]. Many technologies followed the 2G to represent the 3rd generation (3G) technologies like Universal Mobile Telecommunications System (UMTS) that is based on Wide-band CDMA (WCDMA) and what is called 3.5G and 3.75G technologies like High Speed Packet Access (HSPA) and HSPA+. LTE is Orthogonal Frequency Division Multiple Access (OFDMA) based in the downlink direction and Single Carrier FDMA (SC-FDMA) based in the uplink direction which enables flexible bandwidth and time allocation for mobile terminals. The main goal of the development from generation to another is to increase user data rate and reduce latency and access time.

The first release of LTE, which is known as Release 8, was introduced in 2008 in order to satisfy requirements of reduced latency, reduced access time, rational power consumption for mobile terminals, higher data rates, better cell-edge performance, higher spectral efficiency, supporting mobility at high speed and maintaining fully IP-based network structure. LTE (Release 8) can support up to 300 Mbps downlink data rate (for stationary users given 4×4 spatial multiplexing transmission scheme and 20 MHz system bandwidth) and 75 Mbps uplink data rate [10]. The LTE was further developed resulting in LTE-Advanced (LTE-A) that is developed through releases from 9 onwards (Release 10 is usually referred to as LTE-A). LTE-A is usually denoted as 5th generation (5G) technology.

Table 2.1 summarizes the main added features for each of LTE releases starting from Release 9 to Release 12 [9].

Table 2.1 Main features of LTE releases

Release	Features
Release 9	Home eNB (HeNB) – Supporting Multicast Broadcast Multimedia Services (MBMS)
Release 10 (LTE-A)	Carrier Aggregation – Enhanced Transmission Modes (MIMO Modes) – Relaying – Inter-Cell Interference Cancelation (ICIC)
Release 11	Co-ordinated Multi-point (CoMP) – Enhanced ICIC (eICIC)
Release 12	Small Cells – Multi-Radio Access Technology (Multi-RAT) – Massive MIMO – Device-to-Device (D2D) Communications

2.2.2 LTE Basic Technologies

The great performance results of LTE are mainly based on two main cornerstones; OFDMA as a multiple access scheme and the MIMO capability. OFDMA is a multi-carrier multiple access scheme that is based on dividing the spectrum into narrow-band subcarriers separated in frequency by the inverse of the symbol period. The main advantage of OFDMA is combating frequency selective multipath fading and the channel can be assumed constant over its narrow-band (Bandwidth < Coherence Bandwidth). Figure 2.1 [12] indicates this advantage. In addition, OFDMA enables efficient and flexible power allocation per subcarrier in which more power could be used for transmission over subcarriers with better channel gains. Finally, OFDMA enables flexible radio resource allocation especially in the downlink direction which maximizes the overall data rate. In Orthogonal Frequency Division Multiplexing (OFDM), the basis of OFDMA, symbols could be generated using Inverse Discrete Fourier Transform (IDFT) as shown in Figure 2.2 [12]. Cyclic Prefix (CP) is added at the beginning of the OFDM symbols to combat Inter-Symbol Interference (ISI) and as an enabler for using IDFT in generation [12].

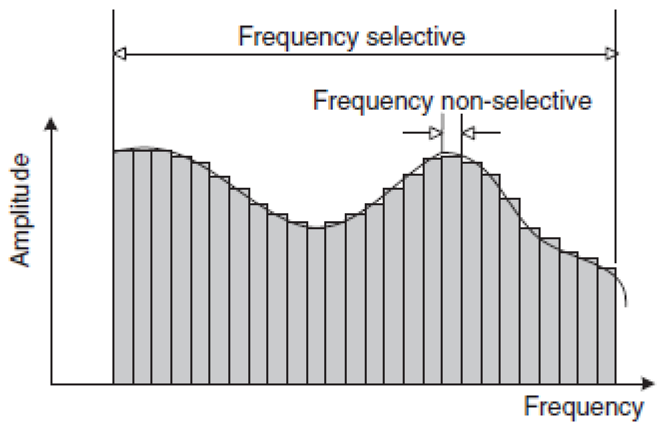

Figure 2.1 Multi-carrier vs. Single-carrier transmission [12]

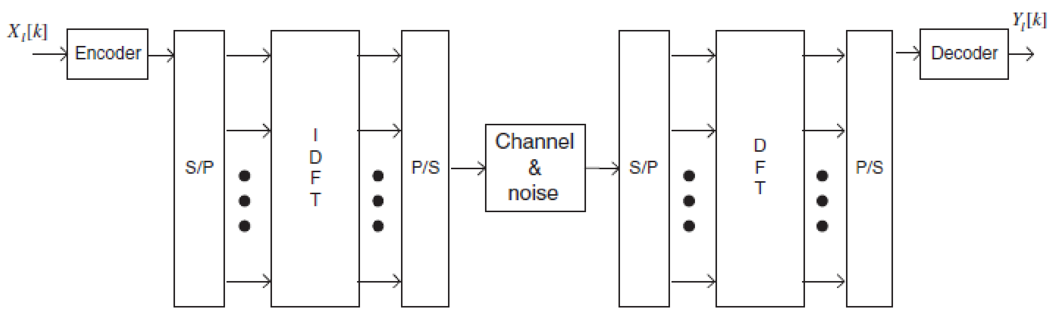

Figure 2.2 OFDM transmission & reception using IDFT and DFT [12]

The OFDM has the great aforementioned advantages but it has a drawback in terms of power consumption. The OFDM suffers a high Peak-to-Average Power Ratio (PAPR) which leads to inefficient power usage for battery-powered devices (e.g. UEs and MTCDs). Hence, in the uplink communication direction, i.e. from UE/MTCD to eNB, SC-FDMA is used by applying DFT spreading before IDFT stage in OFDM transmitter and doing the inverse at the receiver [11]. SC-FDMA (Interleaved FDMA (IFDMA) and Localized FDMA (LFDMA) are two variants of SC-FDMA) shows better performance in terms of PAPR and, as a result, the power consumption at user terminals, at different modulation schemes as shown in Figure 2.3 [11].

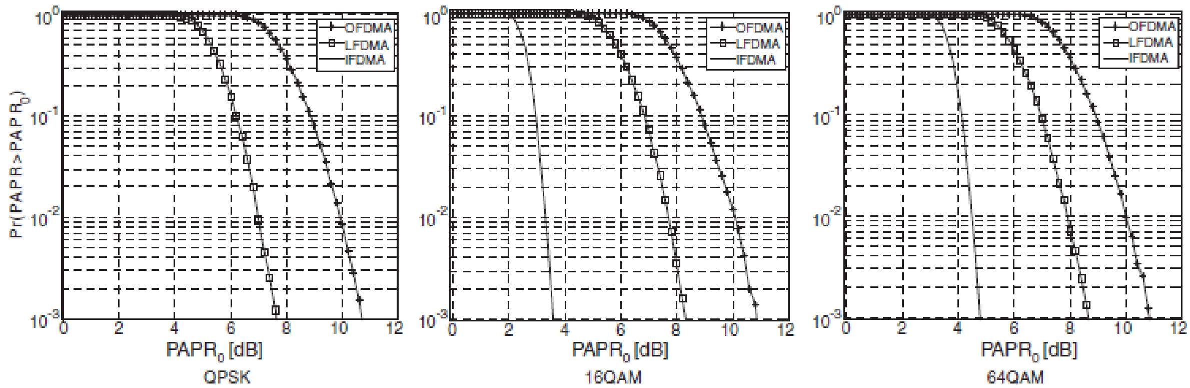

Figure 2.3 PAPR performance comparison [11]

On the other hand, MIMO transmission modes supported in LTE [20], like spatial multiplexing, transmit diversity and receive diversity boost the received Signal-to-Noise Ratio (SNR), enhance diversity (the rate of decay of bit error rate with SNR increase) and reduce bit error probability. LTE-A can support single transmit antenna, Single User MIMO (SU-MIMO), Multi User MIMO (MU-MIMO), transmit diversity and 2/4/8-layer spatial multiplexing.

2.2.3 LTE Future

Organizations like 3GPP, technology companies, governments and researchers have different views in terms of how the LTE-A should be developed to cope with 5G requirements of increasing network capacity considerably, better coverage and reduced latency considerably. Efforts are exerted on three fronts, namely, spectrum efficient usage, spectrum extension and network densification to achieve these goals as depicted in Figure 2.4 [13]. In order to achieve these goals, technologies like Relaying [14] and CoMP [15] are well investigated. In addition, many technologies are adopted, such as D2D communications [16], Small Cells [17], Millimeter Wave (mmWave) [18] and Massive MIMO [19]. Moreover, different communication characteristics of M2M communications [2] enforce new challenges related to dealing with bursty low-rate traffic generated by a massive number of nodes.

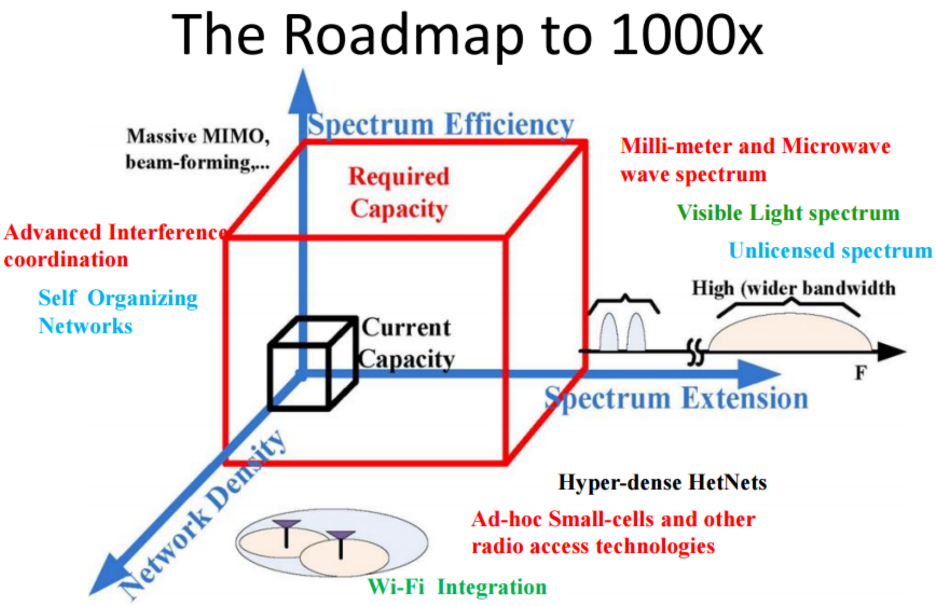

Figure 2.4 5G roadmap [13]

- Relaying: It is the communication between eNB and UE via a relay node that can enhance system performance by extending network coverage, filling coverage holes and increasing communication diversity by amplifying/decoding message and reforwarding it to destination.
- CoMP: It is the cooperation of multiple eNBs to serve a certain UE (usually at cell edge) by Joint Transmission (JT) or Coordinated Scheduling/Beamforming (CS/CB) to boost cell-edge UE data rate.
- D2D Communications: It is the concept of initiating an LTE-based ad-hoc-like connection between 2 or more UEs in case of being in close proximity thus bypassing the communication via the eNB. This helps in freeing the eNB radio resources and which results in providing better QoS to other UEs.
- Small Cells and Heterogeneous Networks: One of the ideas to boost system capacity is to densify the network with low-power base stations which are known as small cells. These small cells are deployed closer to UEs (e.g. femtocells in homes, picocells in hotspots like airport and shopping malls). This results in enhancing the received Signal-to-Noise Ratio (SNR) and the per user throughput.
- mmWave Communications: It is the concept of communicating over the mmWave frequency band where huge bandwidth is available but with severe decaying characteristics. It is a promising technology for indoor communications with challenges in channel characteristics and mobility handling.
- Massive MIMO: It is the concept of increasing the number of antenna elements at the eNB to a high order (64 or 100 elements) to allow for more accurate beamforming, higher diversity and better throughput. The main obstacles for massive MIMO are problems like pilot contamination and channel estimation.

2.3 LTE Fundamentals

2.3.1 Frame Structure

LTE supports communications in Frequency Division Duplexing (FDD) and Time Division Duplexing (TDD) modes between the uplink and the downlink directions of communications, via a flexible configurable frame structure. In the time dimension, the length of the LTE frame is 10ms and it consists of 10 subframes, 1ms each, called Transmission Time Interval (TTI). Each subframe is divided into 2 slots and the slot carries 7 symbols in the normal configuration and 6 symbols in the extended configuration (longer Cyclic Prefix (CP)). Each symbol is sent along with its own CP. Figure 2.5 [8] indicates the time dimension partitioning of the frame.

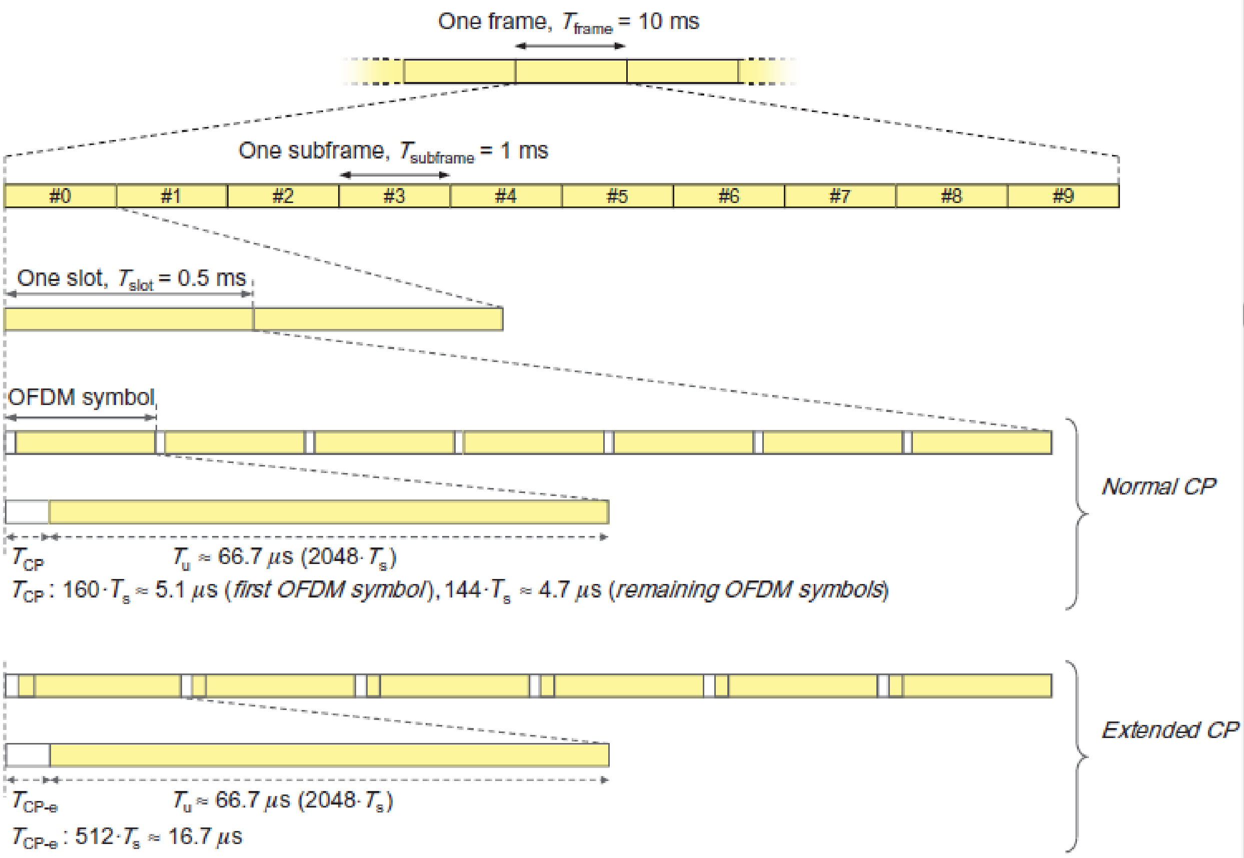

Figure 2.5 Time-dimension structure in FDD configuration [8]

In TDD mode, subframes 1 and 6 (the first subframe is subframe 0) are special subframes. These special subframes consist of special slots, namely, Downlink Pilot Time Slot (DwPTS), Uplink Pilot Time Slot (UpPTS) and a Guard Period (GP) between them. These slots are used as a separation between the subframes used for uplink transmission and the subframes used for downlink transmission [9]. The TDD mode has multiple configurations [9].

In the frequency dimension, the bandwidth is divided into subcarriers with spacing of 15KHz. A time slot for a given subcarrier is referred to as the Resource Element (RE). When an LTE user (UE or MTCD) is allocated radio resources, it is allocated a set of subcarriers to be used within certain time slots. The minimum allocatable unit for one UE/MTCD is called the Physical Resource Block (PRB) [20]. The PRB is a resource grid that consists of 12 subcarriers in one time slot (i.e. it consists of 84 REs for normal CP and 72 REs for extended CP). Figure 2.6 [21] shows the uplink resource grid for 1 time slot.

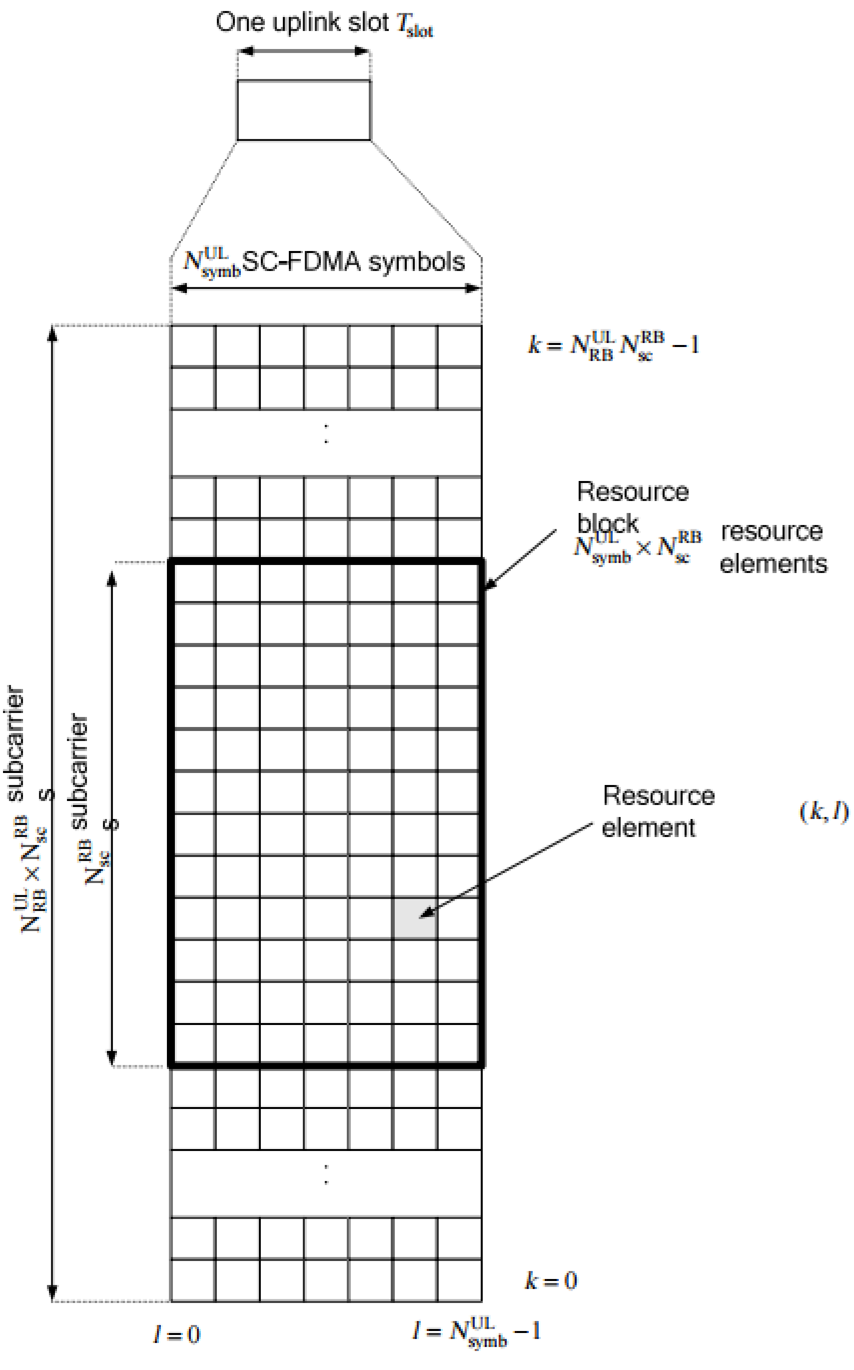

Figure 2.6 Uplink resource grid [21]

2.3.2 Signaling and Channels

In this section, we focus on the signaling and channels in the uplink direction since it is the main traffic direction in M2M communications. There are 3 uplink physical layer channels in LTE [20], [8] which are:

- Physical Uplink Shared Channel (PUSCH): This is the channel over which data is transmitted in the uplink direction. In this channel, PRBs allocated for one UE/MTCD have to be contiguous (It is worth noting that recent standard releases allow allocating two separate chunks of contiguous PRBs per one UE/MTCD [20]). In LTE-A, PUSCH is used also to send Demodulation Reference Signal (DM-RS) from UE/MTCD for channel estimation by the eNB.

- Physical Random Access Channel (PRACH): This is the channel used for contention-based random access. It consists of 64 preambles (Some preambles are contention-free) for random access.
- Physical Uplink Control Channel (PUCCH): This is the channel used to carry control information from the UE/MTCD to the eNB. This control information includes:
    - Scheduling Request (SR): After admitting UEs/MTCDs as associated to eNB (i.e. user association), the UE/MTCD sends an SR to request radio resource allocation to send data through.
    - Channel Quality Indicator (CQI): An index reported by UE/MTCD to the eNB to determine the channel quality. This index helps the eNB to determine the most efficient Modulation and Coding Scheme (MCS) to be used to send data in downlink. That is better channel quality allows for higher modulation order with satisfactory bit error rate.
    - Precoding Matrix Indicator (PMI): An index used in case of spatial multiplexing or Multi-user MIMO (MU-MIMO) transmission modes [20] to suggest the best precoding matrix from a codebook to counter the channel fading effect.
    - Rank Indicator (RI): An index used in case the UE/MTCD has multiple antenna elements to inform the eNB of the receive diversity order (i.e. how many antenna elements will be used for reception by UE/MTCD).
    - Buffer Status Report (BSR): An index reported by UE/MTCD to the eNB to determine the size of the data in UE/MTCD buffer. This index help eNB to know how much data the UE/MTCD needs to send.
    - Hybrid Automatic Repeat Request-Acknowledgment (HARQ-ACK)

2.4 The Scheduling Process

Scheduling is the process implemented by the eNB to allocate radio resources (i.e. PRBs) according to the requests of UEs/MTCDs in downlink or uplink direction. The scheduling process can be divided into 2 stages as shown in Figure 2.7:

- Time Domain Packet Scheduling (TDPS): In this stage, the eNB selects a terminal (UE or MTCD) or a group of terminals to be assigned PRBs according to a certain criteria (e.g. channel state, QoS, fairness, etc.). It is the stage that answers the question of who should get the PRBs.
- Frequency Domain Packet Scheduling (FDPS): In this stage, the eNB selects the PRBs to assign to the terminal, or group of terminals, selected in the first stage of TDPS. This stage answers the question of which PRBs should be assigned to the selected terminal. The eNB allocates PRBs that the terminal can make the maximum use of, i.e. it usually allocates the PRBs at which the given terminal has the best channel conditions.

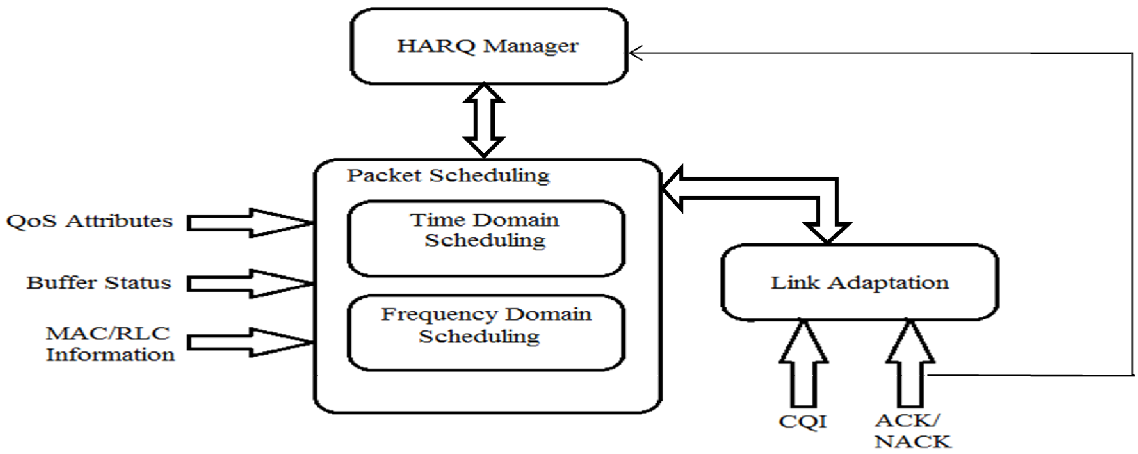

Figure 2.7 Scheduling process

The focus of this research is the uplink scheduling process of M2M communications. The uplink scheduling differs from downlink scheduling in terms of the FDPS stage. In uplink scheduling, opposed to downlink scheduling, the PRBs that are allocated to one UE/MTCD must be contiguous which limits the flexibility of PRB allocation. As an example for uplink scheduling algorithms over LTE, we will consider Recursive Maximum Expansion (RME) [22]. In this scheduling scheme, the value of SNR at every PRB for all UEs/MTCDs is determined. Then, the UE/MTCD with the maximum SNR-PRB pair is allocated this PRB. The UE/MTCD keeps acquiring adjacent PRBs by expanding on the left and right till it is either satisfied (got the needed PRBs) or blocked by PRBs occupied by another UE/MTCD or another UE/MTCD was found to have higher SNR at a certain PRB. The previous UE/MTCD is considered served and the previous step is repeated excluding served UEs/MTCDs and occupied PRBs. The algorithm stops when all PRBs are allocated or all UEs/MTCDs are considered served. In the latter case, there may be non-allocated PRBs which are allocated to one of the UEs/MTCDs occupying the PRBs that are adjacent to the non-allocated PRBs based on maximum SNR at the non-allocated PRBs.

2.5 M2M Communications

2.5.1 M2M Communications General Structure

Several types of devices are involved in M2M communications, such as, MTCDs, Machine Type Communication Gateway (MTCG) and Machine Type Communication Server (MTC Server). The types of transmissions in M2M networks are illustrated in Figure 2.8 [23].

MTCD is the device used to collect information from the environment (like sensing, surveillance and counting). The MTCD sends data to the eNB either directly or via the MTCG. The MTCG acts a cluster head for a group of MTCDs. The MTCG applies some processing on data coming from MTCDs like combining and filtering [24]-[25] (to compress the amount of data to be sent to the eNB). The MTC server is the end-target of the data sent to the eNB. It gets data via the backhaul from the eNB and makes it available for access by human (or automated) users through some application.

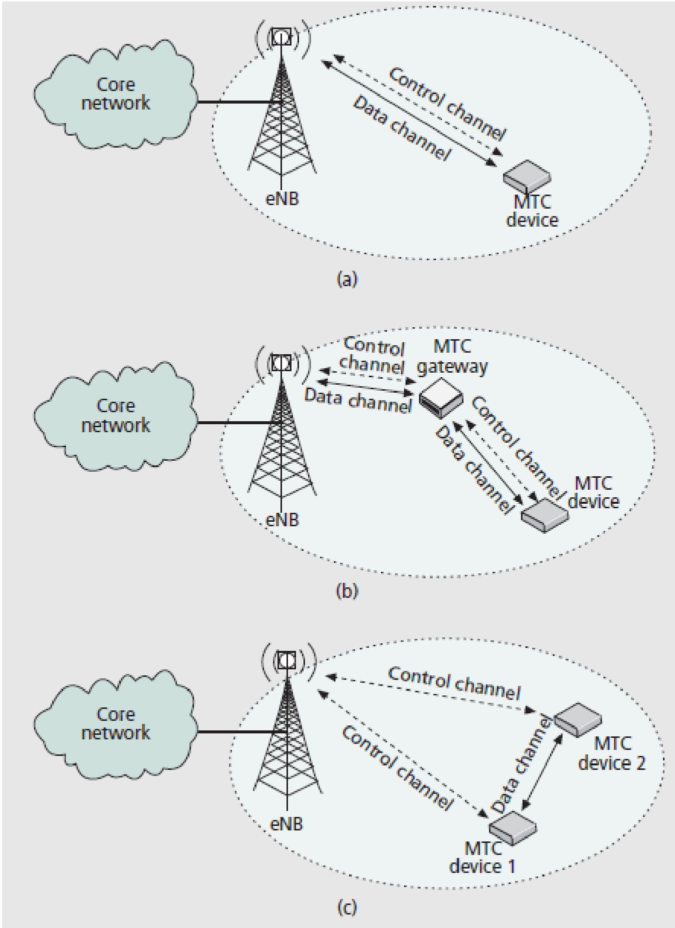

Figure 2.8 M2M communications architecture [23]
(a- Direct Transmission b- Indirect Transmission, c- Peer-to-Peer Transmission)

2.5.2 M2M Communications Characteristics and Requirements

M2M communications characteristics differ from those of H2H communications in several aspects, such as,

- Most of the traffic of M2M communications occurs in the uplink direction from MTCDs to eNB.
- Data traffic is bursty and consists mostly of low-rate small-size packets.
- MTCDs have deadlines that they need to transmit data before them. Abiding by deadlines is necessary to report an alarm for a disaster, maintain a certain data rate or a certain QoS and send data before it becomes useless or obsolete.
- There are various types of MTCDs and they are used in a wide variety of applications. This results in that MTCDs vary widely in terms of requirements of deadlines and needed QoS.

2.6 Data Compression

Data compression is the process of reducing the size of a data stream reported by a device for many purposes such as saving the device's power and reducing data traffic within the network. The main method for data compression is to limit data transmissions by refraining from sending data values that are highly similar to previous sent data points using different methods. In [24], the authors introduce data compression in Wireless Sensor Networks (WSNs) based on two levels of correlation. The first level is the temporal correlation in which the sensor node does not send the measured data value except if it differs from the last sent data value by a difference greater than a certain threshold. The second level is the spatial correlation in which a gateway node (e.g. MTCG) collects data sent by sensor nodes for further compression by comparing the different sensor readings with each other. A sensor reading is sent by the gateway node if differs significantly from the readings of the other sensors. In [25], the authors propose a filtering scheme based on statistical analysis. A gateway node collects measured data values from sensors and sets a distribution of data values based on normal distribution or T-distribution with a certain range. This range is broadcasted to all sensor nodes so that every sensor node can compare the measured data value to the broadcasted range and transmit it only if it is out of range. This technique is proposed as an alternative to using Kalman filtering at sensor nodes.

As we will discuss later in the thesis, we do not rely on application-based compression techniques or group-based decisions to reduce transmitted data. This is due to the fact that imposing such rules on application developers would be restrictive and impractical. In addition, in many cases, network nodes act individually and autonomously with regard to data transmission decisions, which renders group-based decisions inapplicable.

2.7 Chapter Summary

In this chapter, we introduced background about several concepts related to addressing our research problem such as LTE fundamentals, LTE frame structure, physical uplink channels, reference signals, LTE uplink scheduling process, M2M communications and data compression. Finally, we introduce the general structure of M2M communications and M2M traffic characteristics. These concepts have a great effect on the way we address the problem of uplink scheduling for M2M communications.

Chapter 3: Literature Review

3.1 Introduction

Uplink scheduling over LTE is a well-investigated subject since the early releases of LTE. Uplink scheduling algorithms were developed to target close-to-optimal solutions (Maximum Throughput) for the problem of radio resource allocation in a contigous manner per UE or MTCD [6]. Another design aspect that is considered in LTE uplink scheduling is the power efficiency which is crucial for battery-powered devices [2], [26]-[27]. In this chapter, we discuss the M2M-specifc uplink scheduling techniques in literature and provide a comparative summary of these techniques based on our proposed classification. In addition, we propose an M2M uplink scheduling algorithm that strikes the balance between throughput maximization and meeting system deadlines.

The design of uplink scheduling techniques for M2M communications differs from that of conventional uplink schedulers for H2H communications. The first distinguishing factor is that MTCDs in M2M communications may have strict deadline (or class) requirements. MTCDs need to send their data before a given deadline to raise an alarm or to avoid data becoming obsolete. The second factor is the massive number of MTCDs expected to exist in network and the tough fight over limited resources. The third factor is that most techniques try to resolve the contention over radio resources between M2M communications and H2H communications by either dividing resources from the beginning or allowing UEs and MTCDs to compete for the same radio resources in a hybrid manner.

3.2 Uplink Scheduling for M2M Communications

The characteristics of M2M communications needs to be considered upon designing uplink scheduling algorithms. This consideration appears in two main aspects, namely, scheduling metrics and M2M communications resource allocation.

3.2.1 Classification of M2M Scheduling Techniques

M2M uplink schedulers can be classified according to the scheduling metric as follows [7],

- *Channel State-based Schedulers:* In this type, the scheduling algorithm is designed to give higher priority to the MTCDs with the best channel conditions (e.g. highest SNR) aiming at minimizing the bit error rate and maximizing the system throughput. Channel state could be represented in terms of the CQI or the SNR. A mere channel-state based algorithm is not fair since it leads to the starvation of MTCDs with poor channel conditions. In addition, ignoring system deadlines is not suitable in case of the existence of delay-intolerant MTCDs.
- *Delay-based Schedulers:* The scheduling algorithm assigns radio resources to MTCDs with minimum delay tolerance, i.e. least remaining time before the deadline. This a key parameter in M2M communications since data traffic varies from periodic reports of relaxed deadlines to emergency alarms of tough deadlines. The main objective of using delay-tolerance as a scheduling metric is to reduce the deadline-missing ratio. Usually, delay tolerance is combined with other metrics like channel state or buffer size to enhance the scheduling efficiency. There are many ways to estimate delay to be used in scheduling process as in [28]-[29]. A variation of delay-based scheduling is to divide the MTCDs into QoS classes based on their traffic type (video streaming, file transfer, etc.). They are then given priorities

based on traffic type delays and Minimum Guaranteed Bit Rate (MGBR). MTCDs that belong to the highest priority classes are allocated radio resources earlier.

- *Fairness-based Schedulers:* The algorithm is designed to guarantee fair distribution of radio resources. For example, a Round-Robin scheduler divides radio resources equally without any priorities among all devices. It is usually used as a baseline against which scheduling algorithms are compared. It is usually combined with other metrics like channel state or delay tolerance so that no MTCD is harmed due to bad channel conditions (very low probability to get resources if channel only based scheduler is used) or due to being highly delay tolerant. For example, proportional fairness scheduler allocates PRBs according to channel state along with prioritizing MTCDs that were not allocated PRBs previously.
- *Hybrid Schedulers:* The algorithm is designed to combine more than one metric of the aforementioned metrics to reach a balanced outcome in terms of throughput and deadline-missing ratio. These schedulers may also include other metrics such as buffer size to give higher priority to MTCDs having more data in their buffers. Some hybrid scheduling algorithms group MTCDs of similar requirements (e.g. data arrival rate) in clusters to assign them a chunk of PRBs based on their requirements.

3.2.2 M2M Communications Resource Allocation

This aspect concerns how the eNB divides radio resources between M2M traffic and H2H traffic. There are 3 strategies to resolve this contention:

- Strategy 1: To allocate dedicated PRBs for M2M traffic. The amount of dedicated PRBs for M2M traffic varies with the traffic load in M2M traffic flow. In this case, M2M scheduling is isolated from H2H scheduling.
- Strategy 2: To allocate the remaining PRBs after serving H2H traffic (i.e. giving higher priority to H2H traffic). This strategy may lead to MTCDs starvation for PRBs in cases of high H2H traffic or massive M2M deployment. In this strategy, M2M scheduling is separate from H2H scheduling as well.
- Strategy 3: To schedule UEs and MTCDs together and allowing the contention of both traffic flows over the whole available set of PRBs. In this case, the scheduling algorithm is hybrid and it tries to satisfy the requirements of both traffic flows.

3.3 M2M-Specific Uplink Scheduling Algorithms

The authors in [30] proposed 2 scheduling techniques for M2M communications that combine both channel state and MTCDs deadlines as metrics for scheduling decisions.

The first algorithm adapts a channel state-based algorithm to take into account MTCDs deadlines so that it can suit M2M communications requirements. The radio resources are given to MTCDs with highest SNR, i.e. best channel and highest expected throughput, if and only if a deadline-related condition is fulfilled. This condition is that the selected MTCD (highest SNR) must have delay-tolerance less than the average of the delay tolerances of all MTCDs divided by 2. Otherwise, this MTCD is put in sleep mode and the MTCD with the next highest SNR is considered. This condition is used to skip MTCDs with relaxed deadlines even in case of high SNR to reduce deadline-missing ratio. In terms of resolving the contention over PRBs, the algorithm prioritizes the UEs as they get served first and the remaining PRBs are scheduled among MTCDs.

By integrating deadline requirements with channel state in the scheduling decision, the algorithm reduces the deadline-missing ratio for MTCDs compared to a mere channel-based scheduler. In addition, some power efficiency is achieved by putting MTCDs in sleep mode if they have relaxed deadlines. However, with tough deadline M2M networks and with the increase in the number of MTCDs, the percentage of deadline-missing increases exponentially. Moreover, ignoring the amount of data to be sent by an MTCD may lead to putting it to sleep mode due to late deadline but it needs all that time to upload a large bulk of data which may increase the deadline-missing ratio.

The second algorithm proposed in [30] is similar to the first one except that MTCDs deadlines are given higher priority with respect to channel state in the scheduling decisions. The radio resources are given to the MTCD with the least delay tolerance (closest deadline), such that it gets allocated the PRBs at which it has the best channel state. It can be thought of as a delay-based scheduler with intelligent allocation for PRBs per MTCD. Consequently, further reduced deadline-missing for MTCDs could be achieved compared to the first algorithm (but with less throughput). In this algorithm, there is a possibility of resource starvation for delay-tolerant MTCDs since they are not assigned PRBs except if there are no delay-intolerant MTCDs or with their deadlines approaching. This may even lead to missing relaxed deadlines.

In [31], a grouping-based algorithm is proposed. The MTCDs are grouped in clusters that prescribe to a QoS class and hence prioritization in scheduling is done on a cluster basis. Cluster is characterized by the packet arrival rate (γ) and the maximum allowable jitter (δ). The classes with higher γ are considered of higher priority. The cluster is assigned a group of PRBs every ($1/\gamma$) time interval for an Access Grant Time Interval (AGTI) of 1 ms (1 subframe). It is worth noting that if more than one cluster should be assigned the PRBs at the same instant, the clusters with lower packet arrival (lower priority) are postponed to the subsequent subframe. The more the classes, the better the performance which translates into less violation of timing constraints. PRBs are divided equally among cluster members, which may not be optimal. The division of resources between M2M communications and H2H communications is not addressed. Simply, the PRBs which are not used for M2M communications are used by UEs. The algorithm is a simple QoS-based algorithm that can act as a call admission procedure as well. The algorithm has many drawbacks in terms of ignoring deadline requirements, not addressing the resource division between M2M communications and H2H communications and giving PRBs for an AGTI to clusters irrespective of their traffic load, which is a waste of radio resources. In addition, the algorithm requires that every MTCD should have 5 PRBs assuming a wide system bandwidth of 20 MHz (100 PRBs) allocated so the cluster can have up to 20 MTCDs. This may not be suitable for limited bandwidth networks, since the number of cluster members will be limited to 10, 5, 3, 1 MTCDs for 10, 5, 3, 1.4 MHz bandwidth, respectively, which will result in many service denials in massive M2M deployments.

In [32], a grouping-based algorithm is proposed as an improvement to the algorithm in [31]. The MTCDs are grouped in clusters that prescribe to a QoS class characterized by the packet arrival rate (α) and the size of data to be downloaded or uploaded (β). The classes with higher $\alpha\beta$ multiplication are considered of higher priority. The cluster is assigned a group of PRBs every ($1/\alpha\beta$) time interval. The periodicity of resource allocation to a cluster indicates the data traffic load that is uploaded or downloaded by the cluster. The PRBs are shared between M2M communications and H2H communications. The PRBs that are not used by UEs are allocated to MTCDs based on cluster priority. PRBs within the cluster are distributed on a First-Come-First-Served (FCFS) basis, which is not optimal but more efficient. The algorithm takes the data size into account, makes use of unused PRBs by UEs and accounts for M2M massive deployment which is an improvement over the algorithm in [31]. MTCDs deadlines and resource division between M2M communications and H2H communications remain unaddressed.

The authors in [33] introduce the idea of predictive scheduling. It can be described as follows. When a device in an M2M network requests resources i.e. sends an SR, it is most probable that other devices in the vicinity (i.e. neighboring devices) will be requesting to upload data as well. Consequently, when radio resources are granted before an SR is sent by neighbor MTCD, this reduces the delay in resource allocation and can help devices with their deadlines. The algorithm is suitable for specific applications such as cascaded alarm systems and Wireless Sensor Networks (WSNs).

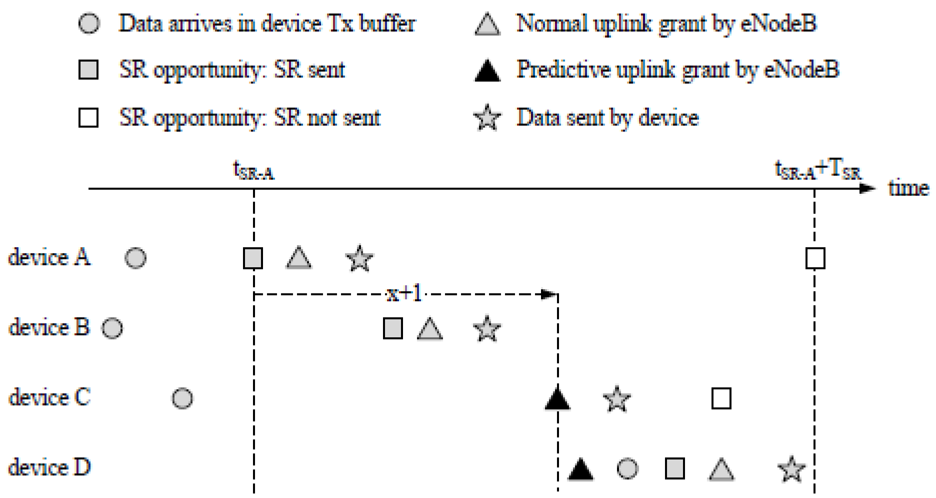

Figure 3.1 The predictive scheduling concept [33]

Figure 3.1 [33] clarifies the concept of predictive scheduling where device A sends an SR and gets granted resources. By checking the neighbor devices we see the following

- Device B has a scheduled SR in less than (x+1) ms so it is not granted resources.
- Device C has a scheduled SR after (x+1) ms so it is predictively granted resources that are utilized to send data (this is a successful prediction and resources are not wasted).
- Device D has a scheduled SR after (x+1) ms so it is predictively granted resources that are not utilized to send data since there is nothing in device's buffer (this is an unsuccessful prediction and resources are wasted).

The prediction success increases with the increase in the periodicity of SRs (T_{SR}) at the expense of the mean uplink latency. However, the algorithm risks wasting radio resources which may need to be adjusted for enhanced efficiency.

The authors in [34] study the problem of the division of radio resources between MTCDs and UEs in a heterogeneous cellular network. The design goal is to avoid MTCDs starvation for PRBs and to prevent the performance degradation of H2H UEs. PRBs are divided between both traffic flows based on traffic levels. For example *p*% of PRBs (10%, 20% are used in algorithm demonstration) are assigned to serve M2M communications. However, the authors did not elaborate on how to deduce the value of p% from the traffic load in M2M and H2H flows. The division of resources helps eNB to apply scheduling separately for every traffic flow. For the M2M traffic flow, the scheduling is QoS-based. Applications are divided into 3 QoS classes of 3 different QoS Class Indicators (QCIs), with priorities defined within M2M traffic since there is no competition over resources with H2H traffic, as suggested by authors according to Table 3.1 [34]. Unused PRBs by M2M flow are reallocated to the H2H flow.

Table 3.1 Proposed QoS classes for M2M communications in [34]

QCI	Type	Priority	Packet Delay	Packet Error Rate	Applications
10	GBR	1	100 ms	10^{-3}	Real-time & Delay Sensitive (Alarms)
11	GBR	2	200 ms	10^{-3}	Real-time (Live Monitoring)
12	nGBR	3	NA	NA	Delay Tolerant (Metering)

In [35], a Class-Based Dynamic Priority (CBDP) scheme is presented. This scheduling technique suggests that a hybrid scheduler for both M2M and H2H traffic simultaneously is the most efficient which opposes the strategy of the previous scheduler in [34]. The authors design their algorithm in a hybrid manner in order to be able to prioritize delay-sensitive M2M traffic with respect to delay-tolerant H2H traffic which is not achieved when dedicated PRBs are just assigned to M2M traffic. For example, this is very suitable for MTCDs that trigger alarms. For every UE or MTCD, a value called Remaining Time To Serve (RTTS) is defined which represents the remaining time till the deadline for the given UE or MTCD. UEs/MTCDs are divided into N classes according to their RTTS. The eNB allocates PRBs to UEs of class 1 first followed by MTCDs of class 1 then the cycle continues as in Figure 3.2 [35]. The CBDP outperforms baseline schedulers like round-robin and proportional fair. However, low traffic density (60 MTCDs) is assumed with large bandwidth (10 MHz) which does not expose scheduler to tough testing under massive M2M deployment conditions. In addition, the deadline-missing ratio would have been a useful evaluation metric to measure.

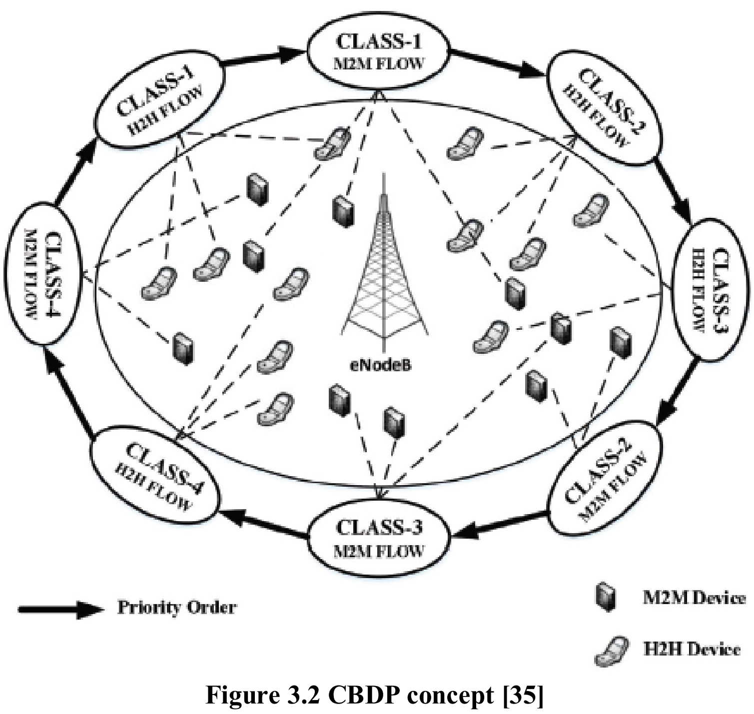

Figure 3.2 CBDP concept [35]

The authors in [36] propose an algorithm that schedules the MTCDs based on the deadline requirements and the buffer size (size of the data to be sent). Both factors are combined to form a new metric called urgency. The devices with the highest urgency are scheduled first. Urgency is calculated as:

$$U_i = \begin{cases} \dfrac{B_i}{\max\{B\}} \cdot \dfrac{T_{SF}}{(D_i - t)} & (D_i - t) > 1ms \\ 1 & (D_i - t) \leq 1ms \end{cases} \quad (3.1)$$

where U_i is urgency metric for request i, B_i = BSR index of request i corresponding to the buffer size as defined in [37], $\max\{B\}$ is the maximum BSR index, i.e. 63, T_{SF} is the LTE subframe duration, i.e. 1ms, t is the current time in ms and D_i is the deadline for request i in ms. The algorithm is delay-based and takes into consideration the amount of data the MTCDs need to send. This achieves the balance between the deadline and the size of uploaded data and increases the probability of meeting deadlines. The algorithm did not address the contention between UEs and MTCDs for PRBs.

In [23], resources are distributed in a mixed environment of H2H and M2M users in order to maximize the sum of utility functions for all UEs and all MTCDs. The utility itself is a function of QoS and the strictness of deadlines. Classes of service where divided into 4 classes:

- Class 1 (Elastic Applications): Applications with tolerance in delay requirements like file transfer for UEs, or downloading a file from MTC server for MTCDs.
- Class 2 (Hard Real-time Applications): Applications where delay requirements exist like telephony for UEs and tracking of vehicles and assets for MTCDs.
- Class 3 (Soft Real-time Applications): Delay sensitive or delay bounded applications (more delay leads to less utility i.e. worse QoE) like audio/video services for UEs or e-Health monitoring in M2M.
- Class 4 (Rate-Adaptive Applications): Applications adjusting rate to available resources which is assumed by authors to be most of M2M applications.

Each class has a utility function (how much is it useful to send data at a given data rate) which is clarified in Figure 3.3 [23]. The main shortcomings of the study lie in the fact that not all classes were considered in testing (only 1 and 4) and the scheduling was designed for downlink while most of M2M traffic is uplink.

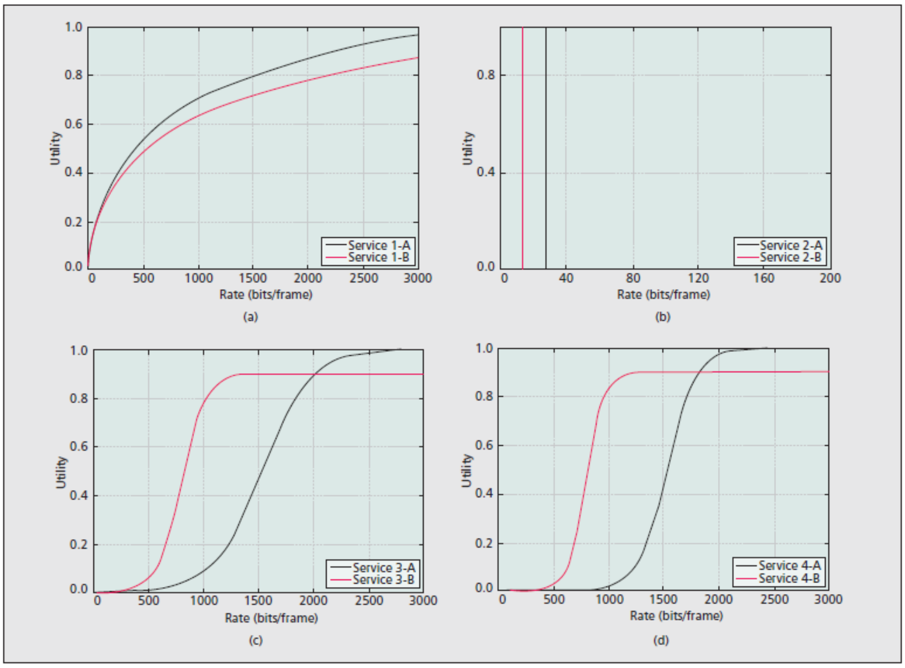

Figure 3.3 Utility functions of application classes in [23]

In [38], the authors propose a multi-step scheduler that combines many scheduling metrics for scheduling decisions. In the first step, MTCDs are divided into groups where each group is assigned a portion of resources based on required QoS and buffer status. QoS is defined by application type where emergency messaging is prioritized over video streaming and monitoring in order. QoS weighting factor is calculated as a function of end-to-end delay budget, required bit rate and delay threshold (time after which application gets even more QoS weight). The second step is TDPS. In this step, a group of MTCDs is chosen based on channel, fairness and required QoS. The third step is FDPS in which search-tree-based resource chunk distribution is used. Unique Depth First Search (UDFS) is used to eliminate choices that contradict with the contiguity constraint or the maximum number of resource chunks per MTCD. The algorithm combines many scheduling metrics like channel-state, fairness and QoS in a hierarchical scheme that may help in reducing algorithm complexity. The authors do not elaborate on the performance of the scheduler in an M2M network that has emergency alerts or the mixed traffic environments of H2H and M2M communications.

The authors in [39] propose a hybrid scheduling algorithm for a heterogeneous network that has both H2H and M2M communications. The traffic is divided into 2 queues and each queue is scheduled separately:

- The first queue includes all H2H users (UEs) and delay-sensitive MTCDs.
- The second queue includes all remaining (delay-tolerant) MTCDs.

In the first queue, delay-sensitive MTCDs of high priority are considered with H2H traffic. UEs and MTCDs are scheduled based on a combination of metrics that include buffer waiting time, proportional fairness and delay thresholds. While, the MTCDs in the second queue are scheduled based on a combination of channel-state-based and round-robin-based schedulers. The channel-state based scheduler is used to maximize throughput while, the round-robin-based scheduler is used to avoid the starvation of MTCDs with bad channel conditions and to reduce deadline-missing for these MTCDs. Considering round-robin in the scheduling of delay-tolerant

MTCDs acts as a guard band against missing deadlines or unfairness due to bad channel conditions. It is worth noting that there are dedicated PRBs for each queue. The first queue is prioritized over the second queue. Results show that the scheduler outperforms a single queue scheduler. Prioritizing H2H traffic over M2M traffic may be not suitable in case of large number of UEs. For example, 100 UEs in the network cause 1000 MTCDs to miss deadlines 50% of the time.

3.4 Comparison Summary

In this section, we summarize the literature survey of M2M uplink schedulers over LTE by comparing them in terms of utilized scheduling metrics and how they dealt with M2M and H2H resources contention. Table 3.2 includes a summary of the scheduling metrics used by every scheduler surveyed in this chapter. It is worth noting that class-based scheduling means dividing MTCDs into groups. All schedulers are general purpose except for the predictive scheduler [33] which is a special case since it is specifically designed for cascaded alarm systems and WSNs.

Table 3.2 Summary comparison of M2M uplink schedulers based on the scheduling metric

Scheduler	SNR (Channel)	Delay (Deadline)	Fairness	Packet Size	Class	General Purpose
Basic Schedulers						
Channel-state-based	√	X	X	X	X	√
Delay-based	X	√	X	X	X	√
Class-based	X	√	X	X	√	√
Round-Robin	X	X	√	X	X	√
Proportional Fair	√	X	√	X	X	√
Literature Survey Schedulers						
Channel Based Delay Aware [30]	√	√	X	X	X	√
Delay Based Channel Aware [30]	√	√	X	X	X	√
Grouping (Packet Arrival, Jitter) [31]	X	X	X	X	√	√
Grouping (Packet Arrival, Data Size) [32]	X	X	X	√	√	√
Predictive Scheduling [33]	X	X	X	X	X	X
QoE Preserving Hybrid Scheduler [34]	X	√	X	X	√	√
Class Based Dynamic Priority (CBDP) [35]	X	√	X	X	√	√
Enhanced Delay Sensitive [36]	X	√	X	√	X	√
Utility-based [23]	√	√	X	X	√	√
Bandwidth and QoS Aware (BQA) [38]	√	√	√	X	√	√
M2M Aware Scheduling Algorithm (M2MA-SA) [39]	√	√	√	X	√	√

In Table 3.3, the algorithms are summarized in terms of how they resolved the contention over the PRBs in the system between M2M traffic and H2H traffic. Recall the following strategy definitions,

- Dedicated: A portion of PRBs is dedicated for serving M2M communications traffic separately.
- Remainder: MTCDs are assigned the PRBs remaining after H2H traffic is served.
- Shared: The pool of PRBs is available for UEs and MTCDs to compete for according to the scheduling scheme.
- Not Applicable (NA): The scheduler did not address the contention problem.

Table 3.3 Summary comparison of M2M uplink schedulers based on M2M resource contention handling strategy

Scheduler	Strategy
Channel Based Delay Aware [30]	Remainder
Delay Based Channel Aware [30]	Remainder
Grouping (Packet Arrival, Jitter) [31]	NA
Grouping (Packet Arrival, Data Size) [32]	Shared
Predictive Scheduling [33]	NA
QoE Preserving Hybrid Scheduler [34]	Dedicated
Class Based Dynamic Priority (CBDP) [35]	Shared
Enhanced Delay Sensitive [36]	NA
Utility-based [23]	Shared
Bandwidth and QoS Aware (BQA) [38]	Shared
M2M Aware Scheduling Algorithm (M2MA-SA) [39]	Shared*

* The scheduler in [39] allocates PRBs for delay-tolerant MTCDs remaining from serving H2H traffic, but delay-sensitive MTCDs are served with H2H traffic.

3.5 Balanced Alternating Technique (BAT) for M2M Uplink Scheduling

We propose a simple scheduling technique that alternates between SNR and MTCDs deadlines to make uplink scheduling decisions. This technique is called Balanced Alternating Technique (BAT) [7]. The main goal of this technique is to strike the balance between throughput maximization and meeting system deadlines. The M2M traffic flow is assigned dedicated PRBs for serving MTCDs based on their density with respect to UEs in H2H traffic flow. This balance is achieved dynamically according to the characteristics of MTCDs traffic. The algorithm is implemented as follows

Step 1: Assign $p\%$ of PRBs for serving MTCDs. For example, $p\%$ can be calculated as follows

$$p\% = \min(\frac{W \times M}{U}, 1) \times 100\% \tag{3.2}$$

where W is a weighting factor much less than 1 that is determined according the traffic load of MTCDs and UEs, M is the number of active MTCDs and U is the number of UEs.

Step 2: Consider a set of N PRBs and a set of M active MTCDs (active MTCDs are the ones that request data transmission and have not missed deadline yet). Every MTCD m ($m = 1 \ldots M$), has a deadline D_m and an SNR value at the n^{th} PRB of S_{mn}.

Step 3: Divide time into equal intervals (x TTIs per interval). Each interval is divided into 2 parts; the channel-state scheduler part (*Step 4a*) for $q\%$ of the time interval and the deadline-based scheduler part (*Step 4b*) for the remaining time.

Step 4a: Channel-based scheduler part

- Sort S_{mn} in a descending order to select the MTCD(m)-PRB(n) pair of maximum SNR, i.e. best channel state.
- Assign PRB *n* to active MTCD *m*.
- Allocate PRBs on the right and the left of PRB *n* to MTCD *m* and keep expanding in both direction till any of the following conditions applies [22]:
    - o MTCD *m* acquires enough PRBs to send its data.
    - o Expanding in both directions is blocked by PRBs allocated to other MTCDs.
    - o Expanding in both directions is blocked by PRBs at which other MTCDs have higher SNR.
- Consider MTCD *m* as served and remove the allocated PRBs from the set of available PRBs.
- Repeat *Step 4a* till all PRBs are allocated.
- If some PRBs remain after considering all MTCDs are served, these PRBs are allocated to one of the MTCDs acquiring the adjacent PRBs.

Step 4b: Deadline-based scheduler part

- Sort D_m in an ascending order to select an active MTCD *m* with closest deadline.
- Assign PRB *n* (at which the selected MTCD *m* has maximum SNR S_{mn}) to MTCD *m*.
- Allocate PRBs on the right and the left of PRB *n* to MTCD *m* and keep expanding in both direction till any of the following conditions applies [22]:
    - o MTCD *m* acquires enough PRBs to send its data.
    - o Expanding in both directions is blocked by PRBs allocated to other MTCDs.
- Consider MTCD *m* as served and remove the allocated PRBs from the set of available PRBs.
- Repeat *Step 4b* till all PRBs are allocated.

By increasing *q%*, the channel-state based scheduler is used for a longer time. Hence, higher throughput could be achieved by favoring MTCDs having better channel conditions. This is suitable for networks that have relaxed deadline constraints or low density of MTCDs. While decreasing *q%* prolongs the time in which deadline-based scheduler is used. Hence, better performance in terms of less deadline-missing ratio by favoring MTCDs with more strict deadlines.

The simulation parameters and MTCD configurations are shown in Tables 3.4 and 3.5 respectively. In our research, a tailored MATLAB-based scheduling simulation tool is created. The scheduling simulator simulates a single cell in which MTCDs send data in the uplink direction according to the configuration of the traffic, the scheduling technique and the MTCDs setup. This tool is used to collect statistics, such as, throughput and deadline-missing ratio.

Table 3.4 BAT simulation parameters

Parameter	Value
Number of MTCDs (M)	100 – 160 – 200
Requested RBs / Available RBs	0.720 – 1.152 – 1.440
Number of Base Stations	1
Average SNR Range	Uniform (4dB,10dB)
Number of Subframes	1000

Parameter	Value
Number of Runs	30 Independent Runs
Bandwidth (MHz)	5
Number of PRBs	25
Channel Model	Additive White Gaussian Noise (AWGN)
Scheduling Algorithms	• Round-Robin • Channel Based • Enhanced Delay Sensitive [36] • BAT
q%	50% - 60% - 70% - 90%

Table 3.5 BAT MTCD configuration

#	Application	Delay Tolerance (ms)	Packet Size (Bytes)	Packet Rate (sec^{-1})	%
1	Emergency Alarms	Constant (10)	32	Sudden*	25%
2	Surveillance Camera	Uniform (100,200)	512	2	25%
3	Regular Monitoring	Uniform (800,1000)	128	1	50%

* By sudden packet rate in Table 3.5 we mean 5 packets within a short period of 200ms once in the simulation time.

By increasing *q%*, we increase the time in which SNR-based scheduling is used. This increases the throughput at the expense of a higher deadline-missing ratio. On the other hand, decreasing *q%*, delay-based scheduling is used for a longer time leading to meeting more system deadlines with decreased throughput. We look at another metric which is the deadline-missing ratio for emergency alarms (i.e. alarm deadline misses / total number of deadline misses) due to their importance. SNR-based algorithms (like BAT with *q%* = 90%) miss deadlines of alarms with higher percentage. Table 3.6 shows the performance of the BAT algorithm with changing *q%* for 30 independent runs after applying 95% confidence analysis on results.

Table 3.6 BAT simulation results

q%	Throughput (Kbps) (±Δ)	% Deadline-Missing (±Δ)	% Alarm Deadline-Missing (±Δ)
Number of MTCDs = 100			
50%	6.15 (±0.28)	15.67% (±0.86%)	15.55% (±2.54%)
60%	6.16 (±0.25)	15.93% (±0.80%)	17.18% (±2.70%)
70%	6.22 (±0.26)	16.39% (±0.88%)	19.14% (±3.19%)
90%	6.41 (±0.24)	20.86% (±1.19%)	38.19% (±3.27%)
Number of MTCDs = 160			
50%	5.16 (±0.26)	21.27% (±0.72%)	20.99% (±1.99%)

60%	5.20 (±0.24)	21.46% (±0.74%)	22.88% (±2.22%)
70%	5.25 (±0.25)	22.27% (±0.93%)	26.47% (±2.69%)
90%	5.43 (±0.24)	27.75% (±1.26%)	42.77% (±2.75%)
Number of MTCDs = 200			
50%	4.72 (±0.22)	23.18% (±0.72%)	22.21% (±2.13%)
60%	4.77 (±0.21)	23.38% (±0.80%)	23.92% (±2.29%)
70%	4.84 (±0.21)	24.38% (±0.83%)	27.99% (±2.32%)
90%	5.02 (±0.20)	31.11% (±1.23%)	45.69% (±2.22%)

Furthermore, BAT with *q%* = 50% and 90% is compared against basic schedulers like the fully SNR-based scheduler and round-robin scheduler and schedulers from literature like enhanced delay sensitive scheduler [36]. It was found that BAT-50% approaches the performance of the enhanced delay sensitive scheduler in terms of deadline-missing ratio. In addition, BAT-90% approaches the performance of the mere SNR-based scheduler in terms of MTCD throughput. Hence, BAT strikes the balance between SNR-based and Delay-based schedulers and *q%* could be adjusted according to the status of the M2M network and traffic characterization. The value of *q%* could be decreased when the number of MTCDs with tough deadline requirements is high. However, if the network is mainly composed of delay-tolerant MTCDs, *q%* could be increased to maximize system throughput. It is worth noting that SNR-based scheduler and BAT-90% have high deadline-missing ratio for emergency alarms. Delay-based scheduler and BAT-50% give more consideration to deadlines and less emergency alarm packets miss deadlines as a result. Figures 3.4, 3.5 and 3.6 summarize the performance evaluation results for BAT and other algorithms.

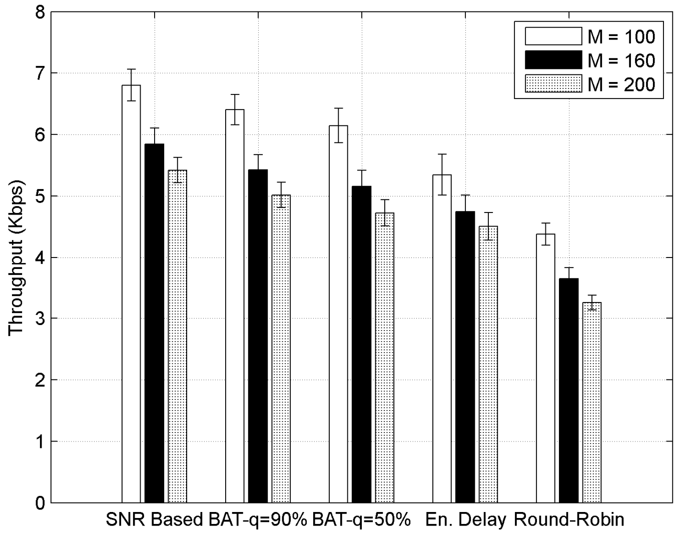

Figure 3.4 BAT throughput results [7]

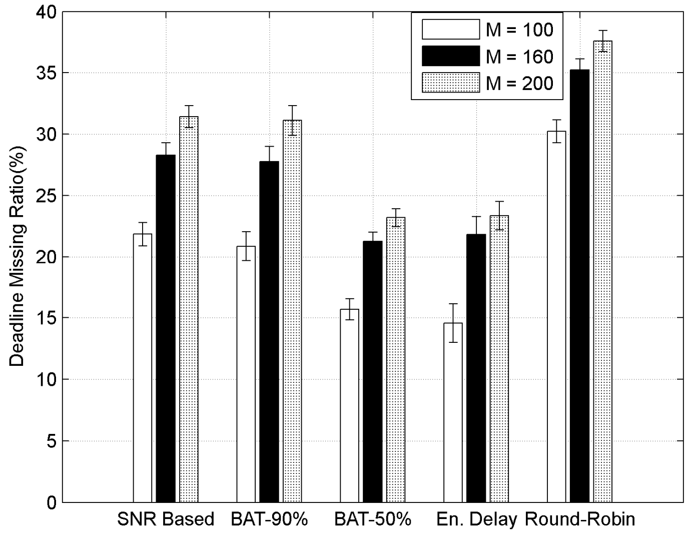

Figure 3.5 BAT deadline-missing ratio results [7]

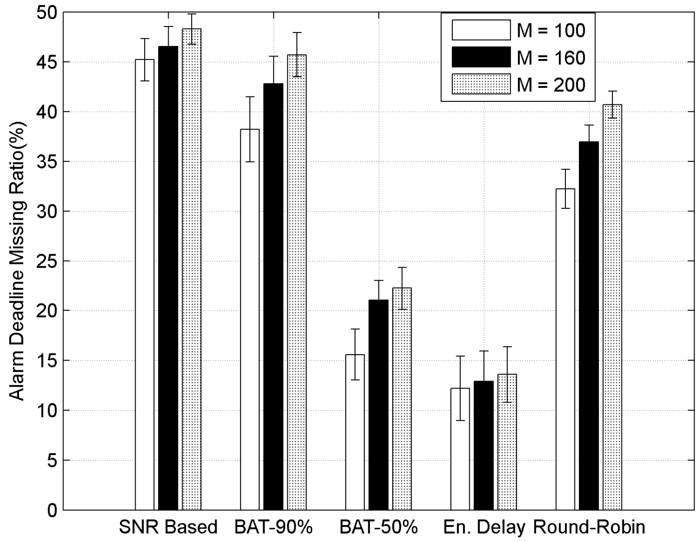

Figure 3.6 BAT alarm deadline-missing ratio results [7]

3.6 Chapter Summary

In this chapter, we classified the uplink scheduling techniques in M2M communications based on scheduling metrics and the strategy of resolving contention over radio resources between M2M and H2H communications. We presented the M2M-specific uplink scheduling techniques that are proposed in literature. We then compared them based on the classification that we proposed in this chapter. Finally, we presented BAT for M2M uplink scheduling to strike the balance between throughput and deadline requirements. When *q%* is increased, BAT approaches the performance of SNR-based scheduler in terms of MTCD throughput at the expense of missing more MTCDs deadlines. While decreasing *q%* results in reducing the deadline-missing ratio for MTCDs and approaching the performance of a delay-based scheduler.

Chapter 4: Statistical Priority-based Scheduling Assessment

4.1 Introduction

It is expected with the proliferation of the IoT that the network should be able to handle access of a massive number of MTCDs. Radio reseources are scarce and satisfying the requests of all MTCDs may not always be possible. Consequently, assigning priorities to different types of traffic in M2M communications is a solution to schedule radio resources. The scheduling techniques that we discussed in Chapter 3 are based on MTCD-related criteria like channel state and delay tolerance. In this chapter, we propose a novel method to assign priority to MTCDs based on the awareness of the data that they need to report. In M2M communications, A lot of the data values collected by deployed devices are ptentially redundant. This offers an opportunity for allocating scarce resources on the basis of data uniqueness. The data to be sent by any MTCD is assigned a priority based on the value of information entailed in data. For example, it is more valuable to assign resources to a temperature sensor whose current reading is more different from its previous reported reading compared to another sensor whose reading suffered less change. Another example is that, reporting abrupt data changes from normal average is more beneficial than reporting a steady-state data stream. We call this type of priority the *Statistical Priority* (SP).

In order to be able to quantify the value of information carried by a certain piece of data reported by an MTCD, there is a need to classify the types of data obtained from M2M networks. Each class of data has a different way to define its SP. For example, in video data, a great interest is given to the correlation between successive frames. On the other hand, the data point reported by a sensor monitoring temperature will gain importance if it exceeds a safety threshold. After data classification and defining its important statistics, a score based on SP should be calculated and incorporated into the scheduling decisions in massive M2M networks, which is what will be discussed in the following sections.

4.2 M2M Data Classification

M2M data can be classified into three classes [34], namely environmental monitoring data, video data and alarm data.

4.2.1 Environmental Monitoring Data

Environmental monitoring data are the data produced by sensors that monitor a given phenomenon such as, temperature, humidity, pressure, light intensity, gases levels for gas leakage detection, liquid level in containers, etc. These data may be collected for recording and archiving. For example, environmental phenomena could be monitored to keep records for the temperature of a city throughout the year and to help predictions and meteorological studies. In addition, data may be collected for safety reasons related to industrial processes and monitoring warehouses. For example, checking temperature of furnaces and liquid leakage may enhance safety measures in factories and warehouses. The environmental monitoring data are generally characterized by small packet sizes and low data rates. There are several statistical parameters that can help define the value of information reported by MTCDs performing environmental monitoring, namely, "threshold", "value similarity" and "trend similarity".

4.2.1.1 Threshold

Data reported by an MTCD that exceeds a certain upper/lower threshold which may not be normal or safe and may need an action to handle it. The data points also become even more important when the difference between them and the threshold increases.

If an MTCD reports a data value at instant i which is x_i, where we have an upper threshold of interest Th_{upper} and a lower threshold of interest Th_{lower}, x_i is said to have a highly valuable information if:

$$x_i > Th_{upper} \text{ or } x_i < Th_{lower} \tag{4.1}$$

4.2.1.2 Value Similarity:

Let the data point reported by an MTCD at a given instance be different from the last reported data point by Δ. If the absolute value of Δ exceeds a certain threshold, this means that the reported data point is not redundant and its information has a unique value. For example, let us assume a temperature sensor that reports the data points of values (27.1, 27.1, 27.15 and 27.4 degrees Celsius). If the first data point is reported, the second data point is redundant since it is the same like the previous one (especially when the time difference between them is small). The third data point is not considered redundant but it has less valuable information when compared to the fourth data point. Hence, setting Δ to a value of 0.1 to define the data points of high-value information is reasonable. The data point becomes more important when the change level threshold is exceeded by a higher difference since it presents a higher change in the measurement of the monitored phenomenon.

If an MTCD reports a data point x_i at instant i and last reported data point is x_p. x_i is said to have a high information value if:

$$|x_i - x_p| > \Delta \tag{4.2}$$

4.2.1.3 Trend Similarity

When a sequence of data points reported by an MTCD maintain a constant trend (increasing or decreasing) for a series of points, this may be more valuable to report than points oscillating around an average value with small variations. For example, if an MTCD needs to report humidity data points of (40%, 41%, 39% and 40%). This can be seen as minor fluctuations around an average value. On the other hand, if it reports data points of humidity of (39%, 40%, 41% and 42%), it is easily concluded that there is an increasing trend in humidity data. This may be of more interest to know and report than the modestly fluctuating case.

If an MTCD to report a data point x_i, at instant I, x_i is said to have a high information value if [40]:

$$(x_i - x_{i-1}) \times (x_{i-1} - x_{i-2}) > 0 \tag{4.3}$$

4.2.2 Video Data

Video data are of the common types of data in M2M networks. Their sources are mainly cameras that are used in many applications e.g. surveillance, people counting and object counting. These cameras have higher data rates (for seamless transition between consecutive frames, cameras operates at a minimum of 30 frames/second) and harder delay tolerance requirements compared to environmental monitoring data. Throughout this research, the video is represented in RGB format, where any pixel at a given position (*x, y*) and time instant (*t*) has an 8-bit represented level (Maximum Level = 255) for each of the R, G and B components. Video encoding, compression and special frames are within of our focus in this research. The interest in video data lies behind giving higher value of information for frames carrying more changes. For example, a surveillance camera recording a stable status with no changes should be given lower priority in data transmission in case of scarcity of radio resources. The following sub-sections list some important statistics that are useful to define the value of information entailed in a given video frame.

4.2.2.1 2-D Frame Correlation

The best statistic to measure the value of information carried by a given video frame is the correlation with previous frames. A frame that is very highly correlated (correlation ≈ 1) with previously transmitted frame has less importance and value if compared to frames carrying new events and many changes and hence less correlated with last reported frame. For example, a surveillance camera recording at late night or early morning will barely carry useful information to tell since the frames are almost constant (except at the time of threats).

Consider an MTCD that reports a video frame at instant *t* represented by the vector $\boldsymbol{F(x, y, t)}$ = $[F_R(x, y, t)\ F_G(x, y, t)\ F_B(x, y, t)]$, where *x* and *y* are the coordinates of the pixel in the frame, while F_R, F_G and F_B represent the level of RGB components at this pixel. The 2-D correlation between the current frame and a previous frame at instant t_o can be defined by the vector $\boldsymbol{C}$ as follows

$$\boldsymbol{C} = \begin{bmatrix} C_R \\ C_G \\ C_B \end{bmatrix} = \begin{bmatrix} Corr(F_R(x,y,t), F_R(x,y,t_o)) \\ Corr(F_G(x,y,t), F_G(x,y,t_o)) \\ Corr(F_B(x,y,t), F_B(x,y,t_o)) \end{bmatrix} \tag{4.4}$$

4.2.2.2 Binary Change Percentage

Another statistic to measure is the change in a given video frame compared to another one, and as a result its value of information. The goal is to calculate the percentage of pixels that had a change in each of the R, G and B components of the video frame. With high percentage of changed pixels (10% or more), the value of information carried by the frame becomes more important and unique.

Consider an MTCD that reports a video frame of width *X* and length *Y*, i.e. a size of *XY*, at instant *t* represented by the vector $\boldsymbol{F(x, y, t)}$. The percentage of change in the R, G and B components of the video frame with respect to another video frame at instant t_o could be defined by the vector $\boldsymbol{Ch}$ as follows

$$\boldsymbol{Ch} = \begin{bmatrix} Ch_R \\ Ch_G \\ Ch_B \end{bmatrix} = \begin{bmatrix} \sum(|\ F_R(x,y,t) - F_R(x,y,t_o)| > 0)/XY \\ \sum(|F_G(x,y,t) - F_G(x,y,t_o)| > 0)/XY \\ \sum(|F_B(x,y,t) - F_B(x,y,t_o)| > 0)/XY \end{bmatrix} \times 100\% \tag{4.5}$$

where ($|F_R(x,y,t) - F_R(x,y,t_o)| > 0$) and the similar ones in G and B components are logical vectors that have only zeros and ones.

4.2.3 Alarm Data

Alarm data are the data that report the occurrence of abnormal conditions, in general, and they are mainly event-triggered. The alarms may be used as an alert for fires or non-authorized building entry. They can be modeled as a sequence of very small payload data packets with highly strict deadline requirements.

4.3 Statistical Priority (SP)

Statistical Priority (SP) is a quantification of the value of information entailed in a data unit based on its data stream statistics (i.e. relationship with previous data within the same stream) and reporting periodicity.

The main goals of SP can be stated as follows

- Giving more priority to data packets carrying non-redundant information of higher value or importance in a manner that is adaptive to the various M2M applications.
- Guaranteeing a minimum rate of data transmission by an MTCD. In other words, to make sure that every MTCD reports at least one comprehensive unit of information every defined period (T).

4.3.1 Reporting SP in LTE

In LTE, there are 3 physical uplink channels that were introduced in Chapter 2. PUSCH is used to carry uplink data. PRACH is used for random access requests. PUCCH is used to send uplink control information from UEs/MTCDs to the eNB; such as, CQI, PMI, RI, SR and BSR.

We propose to report SP via a Statistical Priority Report (SPR) that could be sent by the MTCD through the PUCCH. It can follow similar structure to that of BSR. This means that there should be an SPR value for every radio bearer (logical channel). The SPR value can be reported using 6 bits (64 levels) like BSR. Although, this approach overload the PUCCH with more control traffic but this will greatly help reduce the amount of data traffic to be sent over LTE M2M networks.

4.3.2 SP Mathematical Model

SP is a bounded output function whose output takes a value between 0 and SP_{max} (a positive real value) and can be modeled as follows,

$$SP = f(A, Y, t, T) = \max(a_o f_o(t) \times f_r(t, T), \sum_{i=1}^{n} a_i f_i(y_i)) \qquad (4.6)$$

where SP is the statistical priority. Y is an array of statistical functions that are used to calculate SP. T is the period over which at least one reading should be reported from an MTCD. A is an array of weights ($a_o, a_1, \ldots, a_n$) that are associated with functions determining SP. These weights can take values according to what follows,

$$0 \le a_o \le SP_{\max}$$
$$0 \le \sum_{i=1}^{n} a_i \le SP_{\max} \tag{4.7}$$

It is worth noting that, the second constraint in (4.7) applies only to non-mutually exclusive statistics (i.e. it is not applicable for exceeding upper/lower threshold). If we assume that the data stream sent from an MTCD is *x(t)*, then *y* is any function that can be applied to the data stream like the average of last group of reported readings of an MTCD, the difference between the last 2 readings or the difference between the current reading and a certain threshold. *Y* is an array of functions applied to data stream ($y_1, \ldots, y_n$), *t* is time and *T* is the period over which at least one reading should be reported from an MTCD. For example, a sensor may provide a reading every 0.1*T* and at least one reading should be reported every *T* seconds (i.e. at least one transmission every 10 readings in this case). The functions $f_o(t)$ and $f_i(y_i)$ are bounded functions while f_r can take only one of 2 values as follows,

$$0 \le f_o(t) \le 1$$
$$0 \le f_i(y_i) \le 1 \tag{4.8}$$
$$f_r \in \{0,1\}$$

where $f_o(t)$ is a periodic (period = *T*) bounded function that represents a periodic priority of data packet at time *t*. This function satisfies the first goal of SP of guaranteeing a minimum rate of data transmission by an MTCD. For example, $f_o(t)$ could be assumed to be a discrete impulse train of period *T*

$$f_o(t) = \sum_N \delta(t - NT), N = 0,1,2,..,a_o = SP_{\max} \tag{4.9}$$

This means that, $SP = SP_{max}$ once every period *T* which makes the data highly prioritized for transmission and guarantees transmitting at least one packet every *T*. The function $f_o(t)$ can be any periodic function bounded between 0 and 1. Actually, an impulse train is not a good choice since it enforces transmission trial for only one time in the period *T* which is risky in case of a competition between multiple MTCDs with $SP = SP_{max}$. In addition, it may be better for the network that the MTCD reports the data at another instant within the period *T* not at instants of integer multiples of *T*. Hence, $f_o(t)$ could be written, for example, as a sinusoidal function as follows

$$f_o(t) = 0.5(\cos(\frac{2\pi t}{T}) + 1), a_o = SP_{\max} \tag{4.10}$$

In this case, the MTCD will try to report the reading during the whole period of *T* with varying the priority that reaches a maximum once per period. For example, if the resources are available, the MTCD can report its reading even with low priority as opposed to zero priority in case of discrete impulse train. In addition, the MTCD may be able to send data packets more than once during the period *T*.

It is worth noting that periodic data reporting is irrelevant to the value of the information carried by the data. Hence, $f_r(t,T)$ is used optionally (could be neutralized by making it equal to 1) to enforce only one packet transmission for a certain MTCD by setting it to 0 if any data packet was sent during the current interval *T*. This will help in cases of scarcity of the radio resources and insuring only one measurement of the phenomenon in case there is nothing interesting in measurements like abrupt changes. Therefore, $f_r(t,T)$ could be defined as a function that is equal to 1 if no data packet is sent during the current period and is equal to 0 when a packet is sent during the current interval T, that is only one packet is sent in period T, and in the next period it is reinitialized to 1 again. Hence, it can modeled as follows

$$f_r(t,T) = \sum_{\tau=\max(NT);\tau<t}^{t} s(\tau) < 1$$
$$s(t) = \begin{cases} 1 & packet\_sent \\ 0 & no\_packet\_sent \end{cases} \tag{4.11}$$

where $s(t)$ is set when no packet is sent at a given instant t and reset when a packet is sent at t. So, if the summation of this function over the current period is 0 (no data sent) then $f_r(t,T)$ is 1 and the MTCD has the chance to send its packet. On the other hand, if the summation is greater than or equal to 1, a data packet is sent during this interval and, as a result, no need to send a data packet except if it is prioritized due to the value of the information it carries and $f_r(t,T)$ is sent to 0.

Till now, data is sent periodically for one time or more in an adjustable manner. Consequently, the question remains as to how the SP is affected by the value of information carried by data to fulfill the second goal. The answer lies in the remaining part of (4.6).

As stated before, y is $f(x(t))$ (function of the data stream). Let's assume that in (4.6) $n = 1$ and $a_1 = SP_{max}.f_1(y_1)$ is a bounded function as in (4.8) and can be modeled as a sigmoid (logistical or s-curve) function

$$f_1(y_1) = sigmoid(b,c) = \frac{1}{1+e^{-b(y_1-c)}} \tag{4.12}$$

where c represents the inflection point of the function between producing an output of 0 or 1 such that $f_1(y_1 = c)$ is 0.5, b controls how steep the function drops or rises from 0 to 1. If $b < 0$, the sigmoid function is reflected around the vertical axis. Figure 4.1 shows different configurations of the sigmoid function at $c = 0$.

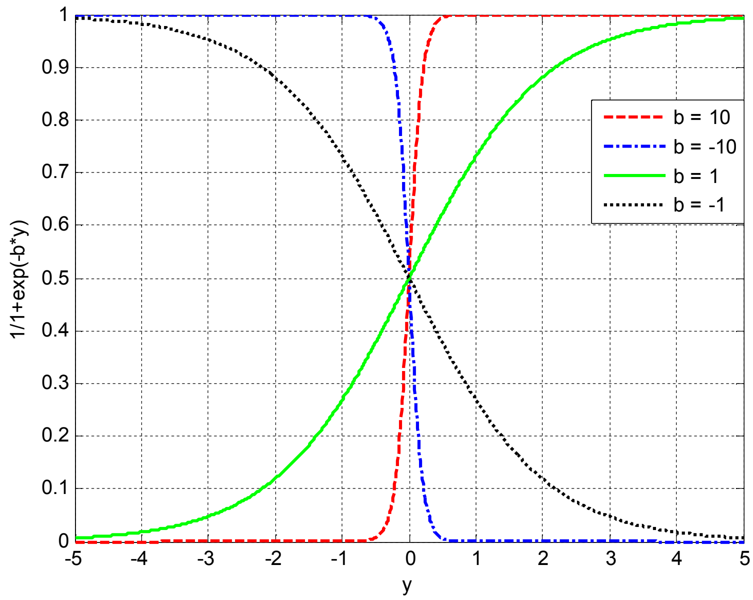

Figure 4.1 Sigmoid function

We consider 4 cases where the information carried in the data is valuable to report.

Case 1: if $x(t)$ > *Threshold* (*Th*) where this threshold is some kind of a safety margin above which problems may occur. In this case, $x(t)$ may act as an alarm that should be sent with high priority. Then, the function of interest y_1 can be represented as

$$y_1 = x(t) - Th \tag{4.13}$$

When $x(t)$ exceeds threshold (high-value information), y_1 becomes positive. When threshold is not exceeded, y_1 is negative. Hence, $f_1(y_1)$ could be represented as *sigmoid(100,0)* (where b can be any large number for steep transition).

A similar case that can be considered is when safety margin is reversed i.e. there is a danger when the reported data goes below a certain threshold. For example, when liquid level approaches a sensor at a given height and the distance between liquid and sensor decreases to a predefined threshold, an alarm is triggered. In this case, we can either use define a $y_{1'} = -y_1$ or making b for sigmoid equals to –100 instead of 100.

Case 2: if the data are considered valuable when the difference between $x(t)$ and last reported measurement x' by the MTCD exceeds a certain threshold. For example, if the temperature reported by sensor is 16.0, a following value of 16.1 has higher value of information compared to 16.01. Consequently, y_1 can be represented as,

$$y_1 = |x(t) - x'| - \Delta \tag{4.14}$$

Hence, $f_1(y_1)$ could be represented as *sigmoid(b,0)* (where b can be selected according to how much we are interested in valuing or devaluing the relative values of $x(t)$ as compared to x').

Case 3: Sometimes we are interested in the trend of the data reported by MTCDs (increasing or decreasing). If data show a trend, it is more valuable to send and deserves higher priority. Then, the function of interest y_1 can be represented as,

$$y_1 = \prod_{i=0}^{2d-1} (x(t-i) - x(t-i-1)), d = 1,2,... \tag{4.15}$$

where d represents the depth of the trend function (trend in the last 2 measurements, the last 4 measurements, etc.). If the measurements follow an increasing or decreasing trend, their multiplied differences will be positive (for an even number of measurements), otherwise y_1 will be negative. Hence, $f_1(y_1)$ could be represented as *sigmoid(100,0)* (where b can be any large number for steep transition). On the other hand a small value of b can be used to give higher priority to trends with bigger differences between data values.

Case 4: Another important statistical feature is the time autocorrelation of the data stream sent by the MTCD. Autocorrelation can be utilized to reduce the amount of frames sent by a surveillance camera by not sending frames that are highly similar (highly correlated) to the last sent frame. Let's consider a video stream $\boldsymbol{F}(x,y,t)$ where x and y are the horizontal and vertical indices of a pixel, respectively, and t represents time or frame number. The 2-D correlation can be calculated for $\boldsymbol{F}(x,y,t)$ and $\boldsymbol{F}(x,y,t_o)$ where $\boldsymbol{F}(x,y,t_o)$ is the last sent frame by the surveillance camera. The correlation for the three components of the video stream as well as the general correlation is calculated as follows

$$Corr\_red = \sum_x \sum_y F_R(x,y,t) \times F_R(x,y,t_o)$$

$$Corr\_green = \sum_x \sum_y F_G(x,y,t) \times F_G(x,y,t_o) \quad (4.16)$$

$$Corr\_blue = \sum_x \sum_y F_B(x,y,t) \times F_B(x,y,t_o)$$

$$Corr = \min(|Corr\_red|, |Corr\_Green|, |Corr\_blue|)$$

So the function of interest y_1 can be represented as,

$$y_1 = Corr - Corr\_Th \quad (4.17)$$

where *Corr_Th* is the maximum correlation value that represents an effective change between 2 video frames. In this case, negative y_1 indicates that the current frame x is a valuable frame so it carries new information as inferred from the fact that it is not highly correlated with last transmitted frame. Hence, $f_1(y_1)$ could be represented as *sigmoid(-1000,(1-Thresold)/2)* (where b is a very large negative number since practically a frame that is correlated by up to ±0.995 carries different information and c is a mid-point between threshold and maximum correlation of 1).

A special case exists for alarms. The data from alarms are undoubtedly important and urgent. Hence, *SP* is directly set to be SP_{max}.

Choosing the type of statistic that indicates the value of information entailed in data is dependent on the application. Moreover, multiple statistics may be important to varying degrees. So, *SP* can be a function of Y as an array of statistical functions of interest that have different weights specified in the array A that can sum up to SP_{max} as can be concluded from (4.7). The mathematical model for *SP* is very flexible which makes it suitable for different M2M applications, different data statistics of interest and periodic data transmission.

4.4 Statistical Priority Validation

In this section, SP is validated for 2 different types of data, namely, environmental monitoring data and video data. The goal of the validation is to demonstrate the following.

- The use of SP helps reduce the amount of data that needs to be sent from an MTCD by giving less priority to repeated data points with low value of information.
- Sending a reduced set of data points is sufficient to represent the full set of data with its changes and characteristics.

4.4.1 Statistical Priority Validation for Environmental Monitoring Data

In [41], the authors introduce a data set for temperature and humidity measurements that are used for anomaly detection. The data were collected using 2 Tiny-OS based TelosB motes [42]. The data points were collected from the 2 sensors in 2 environments, an outdoor and an indoor environment, for 6 hours at a rate of 1 data point every 5 seconds. Some data anomalies were introduced by applying water vapour resulting from a water boiler to one of the sensors. The following table summarizes the data sets in [41].

Table 4.1 Data sets in [41]

#	Mote	Environment	Phenomenon	Number of Data Points
1	1	Indoor	Temperature	4417
2	2	Indoor	Temperature	4417
3	1	Outdoor	Temperature	5039
4	2	Outdoor	Temperature	5041
5	1	Indoor	Humidity	4417
6	2	Indoor	Humidity	4417
7	1	Outdoor	Humidity	5039
8	2	Outdoor	Humidity	5041

For validating the effectiveness of SP, we need to show the value that the information in data points have when considering variable data statistics and how combining different statistical features in the evaluation of the value of information can help represent the data sufficiently with less number of points. Recall that SP is calculated as per equation (4.6).

Let $f_o(t) = 0$, hence SP is a fucntion of the statistical features only. We will consider 4 statistical features ($n = 4$). That is, $Y = [y_1; y_2; y_3; y_4]$, where y_1 is the difference between the data point and the upper threshold. Y_2 is the difference between the data point and the lower threshold. Y_3 represents the difference between the current data point and the last reproted data in relation to a certain threshold. $Y_4 = 1$ if there is a constant trend (increasing or decreasing) in last three data points (including the current data point) and $y_4 = 0$ otherwise. Therefore,

$$\begin{aligned} y_1 &= x(t) - Upper\_Threshold, \\ y_2 &= x(t) - Lower\_Threshold, \\ y_3 &= |x(t) - x'| - \Delta, \\ y_4 &= \prod_{i=0}^{2d-1} (x(t-i) - x(t-i-1)), d = 1 \end{aligned} \tag{4.18}$$

Recall that x' is the last reported value by the MTCD and Δ is the threshold of the level of difference between the current data point and last reported data point that would be of interest (has high-value information). The thresholds for temperature and humidity data are set as follows in Table 4.2. While, Table 4.3 indicates the sigmoid function paramteres for the different statistical functions.

Table 4.2 Thresholds for temperature and humidity data

Threshold	Temperature	Humidity
Upper Threshold	28° C	48%
Lower Threshold	27° C	42%
Difference Threshold (Δ)	0.1° C	0.1%

Table 4.3 Sigmoid function parameters for different statistical features

Statistical Feature	b_i	c_i
y_1	100	0
y_2	–100	0
y_3	5	0
y_4	100	0

The weights are selected to sum up to SP_{max} = 10. Different validation scenarios for the 8 data sets given in Table 4.1 are examined using the following methodology. The reason for validation assessment is to demonstrate the capability of representing a large data set using a reduced set of data points when these points are selected based on statistical priority without loss of data characteristics.

- First, by investigating every statistical function separately and checking how many data points carry information of interest for a given statistical function, i.e. trying to answer these questions:
  - How many data points exceed the upper threshold?
  - How many data points go below the lower threshold?
  - How many data points are reported if the MTCD reports only data points differing from last reported data point by Δ?
  - How many data points show a constant trend along with its two previous data points?
- Second, by considering a combination of the statistical functions where A = [4; 4; 3; 3]. It is worth noting that these weights are selected based on trial and to give all factors almost equal weight with some priority to exceeding upper or lower thresholds. The generated *SP* values are evaluated against 30 independent uniformly distributed random sets of *SP* values to check how many data points will be reported by the MTCD at all cases. This case roughly simulates an MTCD that tries to send its data in a cellular network by contending over PRBs with many other MTCDs that have a maximum SP that is uniformly distributed over SP range. It is worth noting that, a data point is considered for reporting if its SP is greater than or equal to the counter random *SP* value from the random *SP* sequence.

4.4.1.1 First Step

Table 4.4 Environmental monitoring data validation (step 1)

#	Phenomenon	Number of Data Points	$y_1 > 0$	$y_2 < 0$	$y_3 > 0$	$y_4 > 0$
1	Temperature	4417	1791	343	139	1109
			40.55%	7.77%	3.15%	25.11%
2	Temperature	4417	990	530	68	974
			22.41%	12.00%	1.54%	22.05%
3	Temperature	5039	1835	2573	203	1973
			36.42%	51.06%	4.03%	39.15%
4	Temperature	5041	2105	2466	364	2252
			41.76%	48.92%	7.22%	44.67%
5	Humidity	4417	118	157	495	1141
			2.67%	3.55%	11.21%	25.83%
6	Humidity	4417	66	2	565	1197
			1.49%	0.05%	12.79%	27.10%
7	Humidity	5039	1549	1063	1154	2304
			30.74%	21.10%	22.90%	45.72%
8	Humidity	5041	1562	808	1190	2376
			30.99%	16.03%	23.61%	47.13%

The numbers in Table 4.4 show that a large portion of the data points carries non-valuable information. For example, the number of the data points that exceed the upper value threshold does not exceed 42% in all sets. Another example is that the number of the data points that represent a considerable change in value compared to the last reported value, i.e. change level is greater than minimum change threshold, is less than 24%.

4.4.1.2 Second Step

In this step, an SP-based score is given to all data points according to the aforementioned statistical functions. The weights are chosen to be A = [4; 4; 3; 3]. It is worth noting that sum of weights for non-mutually exclusive functions is SP_{max}. Figure 4.2 shows a sample of temperature data of data set 1. Figure 4.3 shows sample of the SP values in the range at which data spike occurs which shows that not all data points have high-value information even when MTCD reports a drastic change in measurements that it reports. It could be easily concluded that data points at stable time, i.e. no considerable change in measurements like time instants away from spike, have a much lower value of information.

The Complimentary Cumulative Distribution Function (CCDF) (P (X > x)) of the score is evaluated as a representation of the distribution of SP values. Figure 4.4 depicts the SP score distribution for the combined statistical function for the 8 data sets (temperature data sets and humidity data sets respectively). The figures show that the data points of outdoor sets have higher SP-based scores due to the commonality of the temperature/humidity increasing trend and exceeding upper thresholds and this is expected. In addition, some data sets like data set 1 has a low number of data points carrying high-value information since only about 20% of data points have an SP score greater than 5 (50% of SP_{max}).

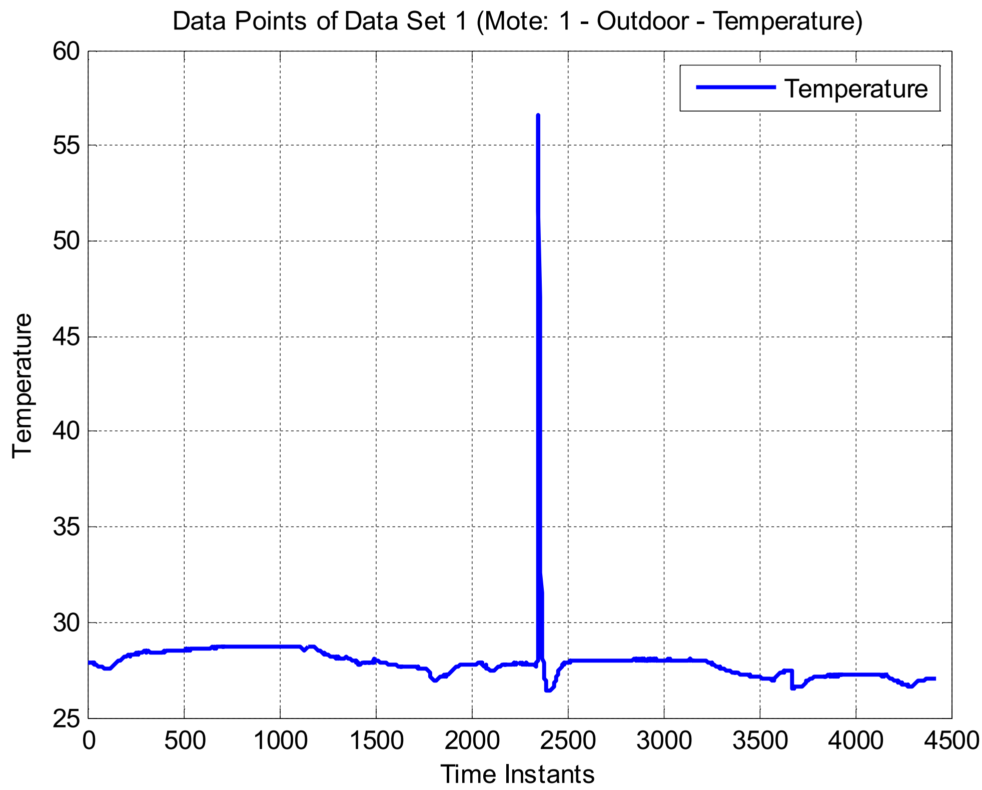

Figure 4.2 Data points of data set 1

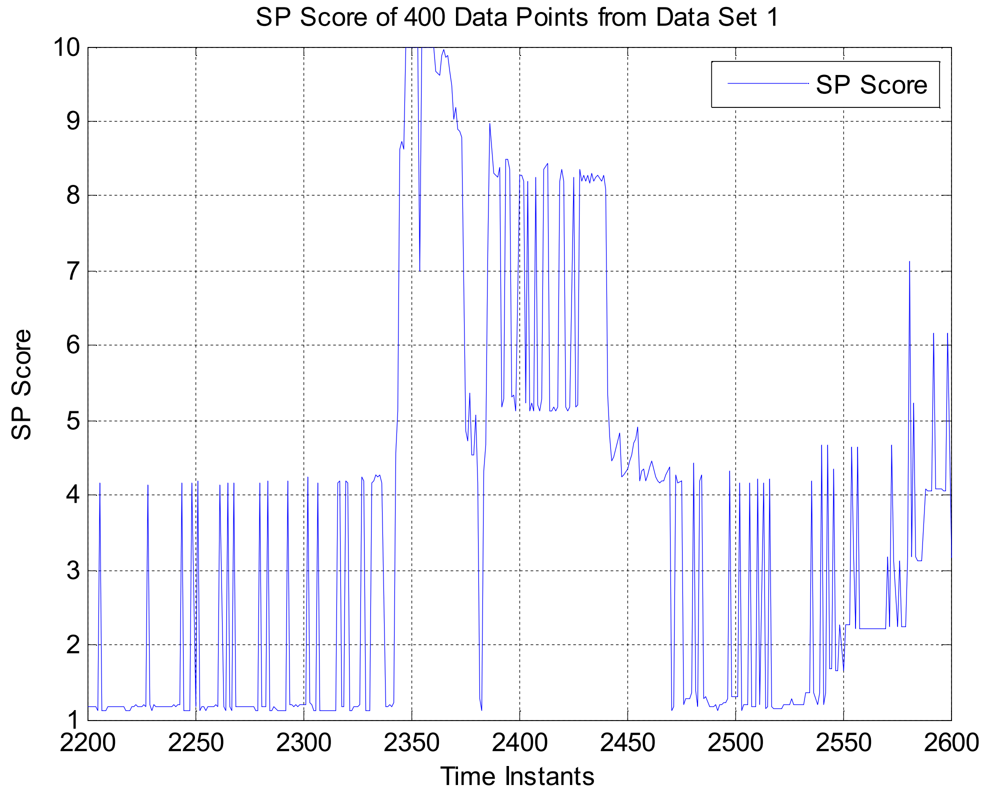

Figure 4.3 Sample SP score of 400 data points in data set 1

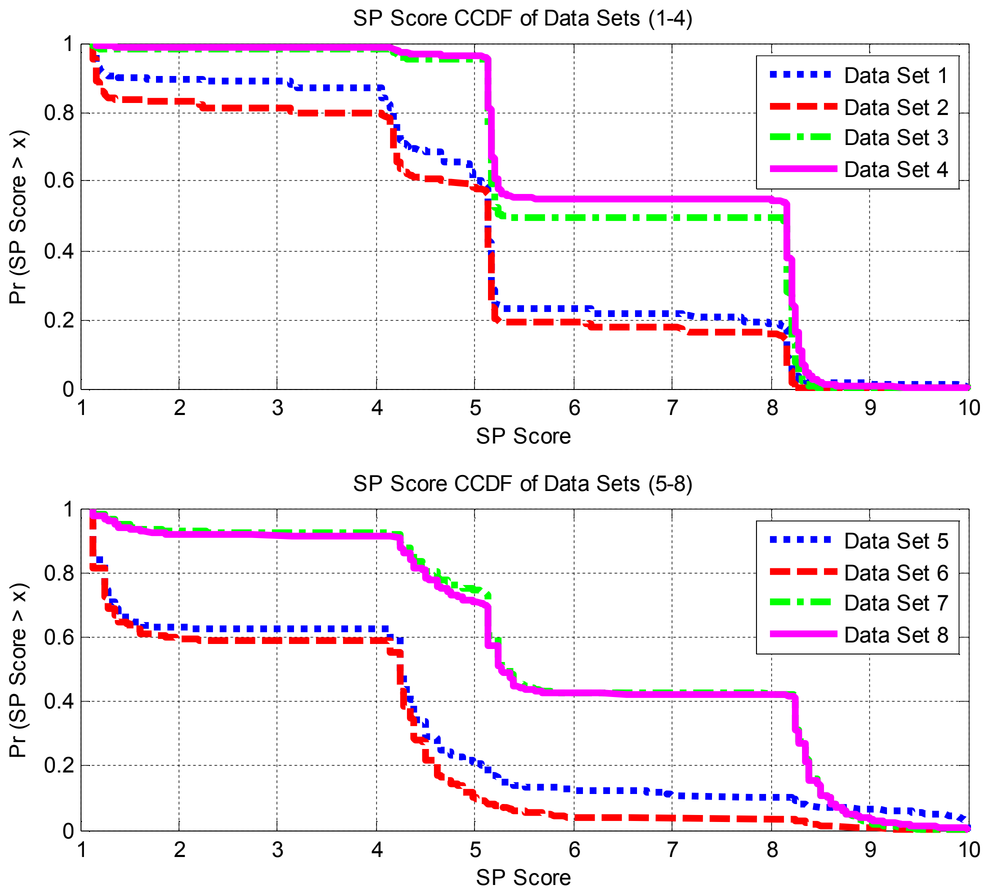

Figure 4.4 CCDF of SP score of data sets (1-8)

In this step, every data set with its SP score distribution is compared against 30 independent random uniformly distributed (from 0 to SP_{max}) SP score sequences. These random sequences model the SP score of another MTCD that competes over radio resources with the MTCD of concern and the data from each of the 8 data sets should have an SP score that is greater than the virtual maximum SP score to be allowed to report its data. Over 30 independent runs, we check how much resource savings have been made (How many data points are reported with respect to the total number of data points?) and whether the reduced number of data points affects the representation of information stored in data. Table 4.5 summarizes the answer to the first question related to radio resource savings. The results in Table 4.5 are averaged for the 30 runs and 95% confidence analysis is applied. The set of graphs in Figure 4.6 shows qualitatively (for one sample run of one sample of a random SP sequence) that the information in data is preserved when reporting less number of data points for all data sets. It is worth noting that that the reduced set data points are interpolated for clarification purposes.

Table 4.5 shows a reduction in the number of data points reported for all data sets. The MTCD needs to only report (22% – 62%) of the data points while preserving the same features or information entailed in data as shown in the graphs of Figure 4.5.

Table 4.5 Environmental monitoring data validation (step 2)

#	Phenomenon	Number of Data Points (Total)	Number of Data Points ($SP \geq$ Random SP)
1	Temperature	4417	1715 (±10)
			38.83%
2	Temperature	4417	1449 (±7)
			32.80%
3	Temperature	5039	2945 (±12)
			58.44%
4	Temperature	5041	3115 (±11)
			61.80%
5	Humidity	4417	1038 (±9)
			23.47%
6	Humidity	4417	967 (±9)
			21.89%
7	Humidity	5039	2434 (±12)
			48.31%
8	Humidity	5041	2364 (±11)
			46.90%

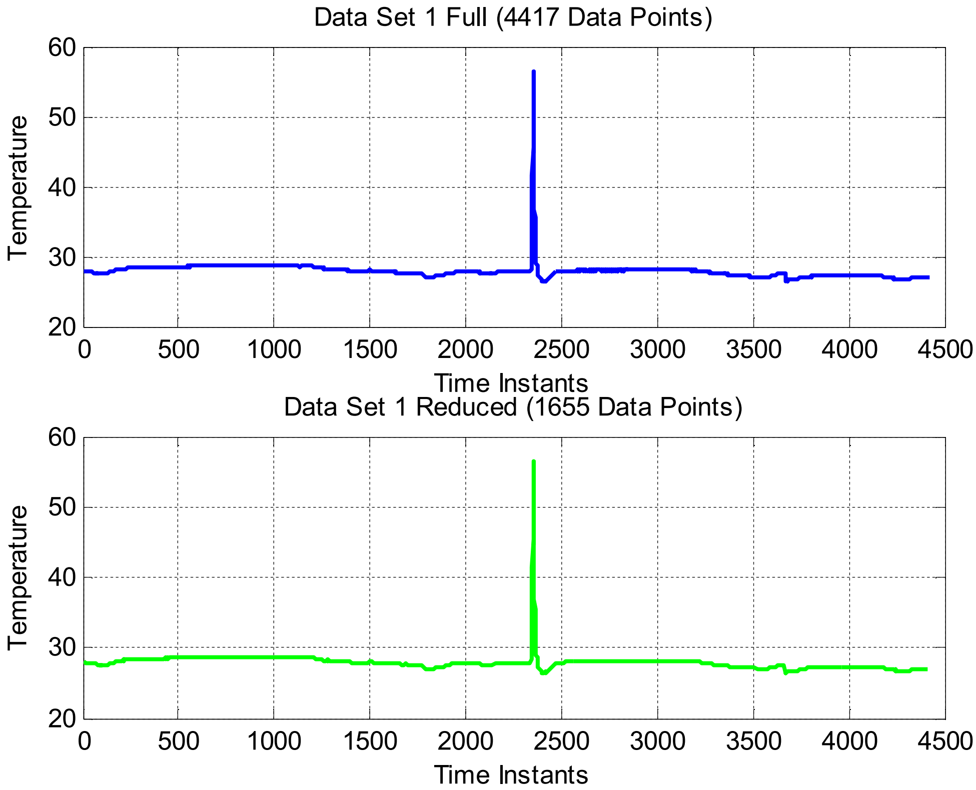

Figure 4.5-a Environmental monitoring data validation (step 2 – data set 1)

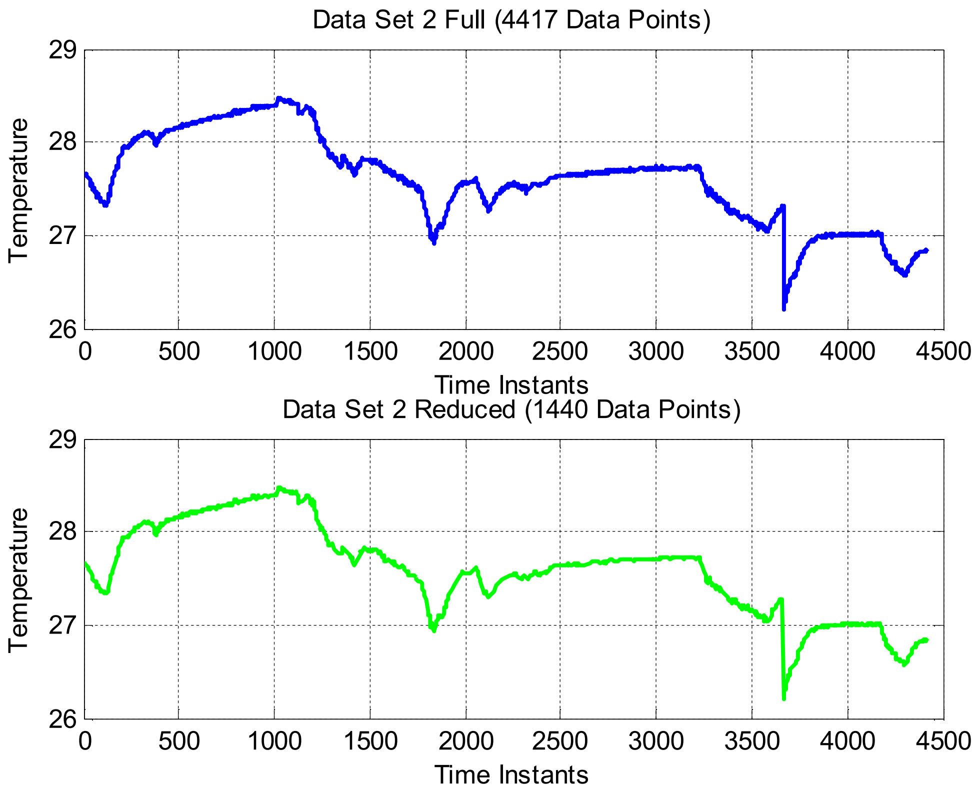

Figure 4.5-b Environmental monitoring data validation (step 2 – data set 2)

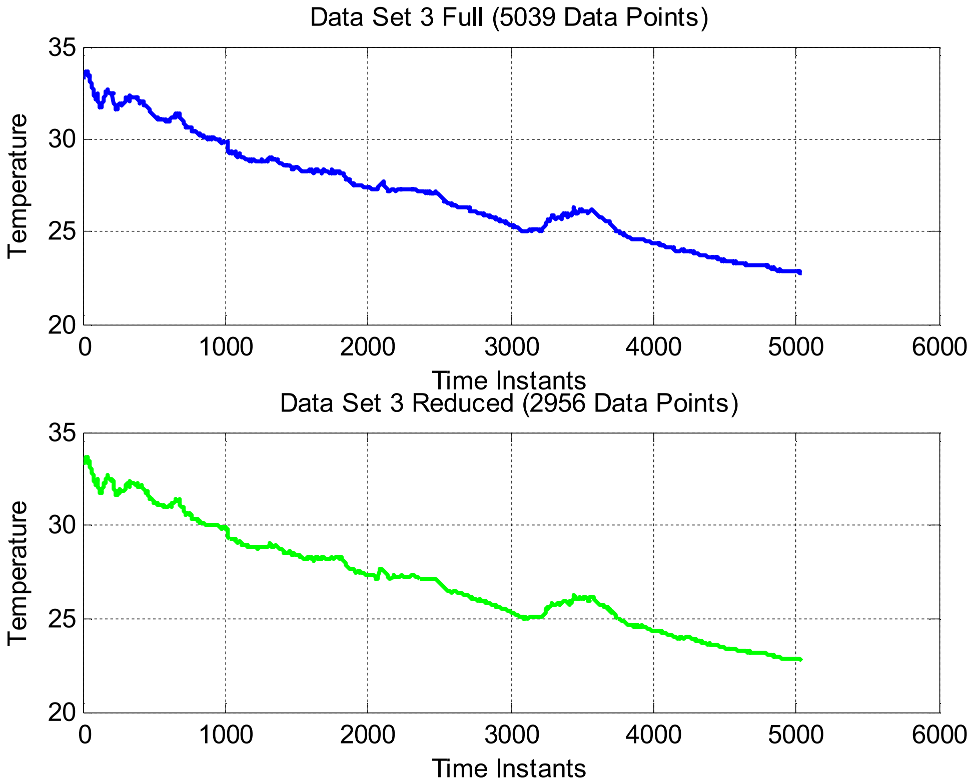

Figure 4.5-c Environmental monitoring data validation (step 2 – data set 3)

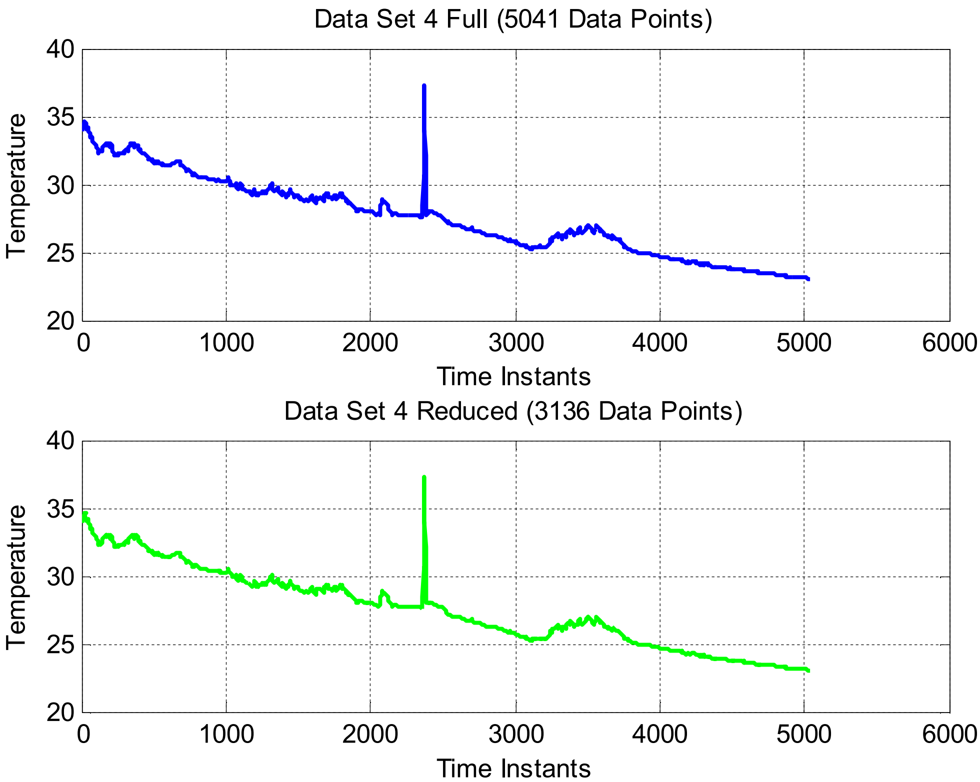

Figure 4.5-d Environmental monitoring data validation (step 2 – data set 4)

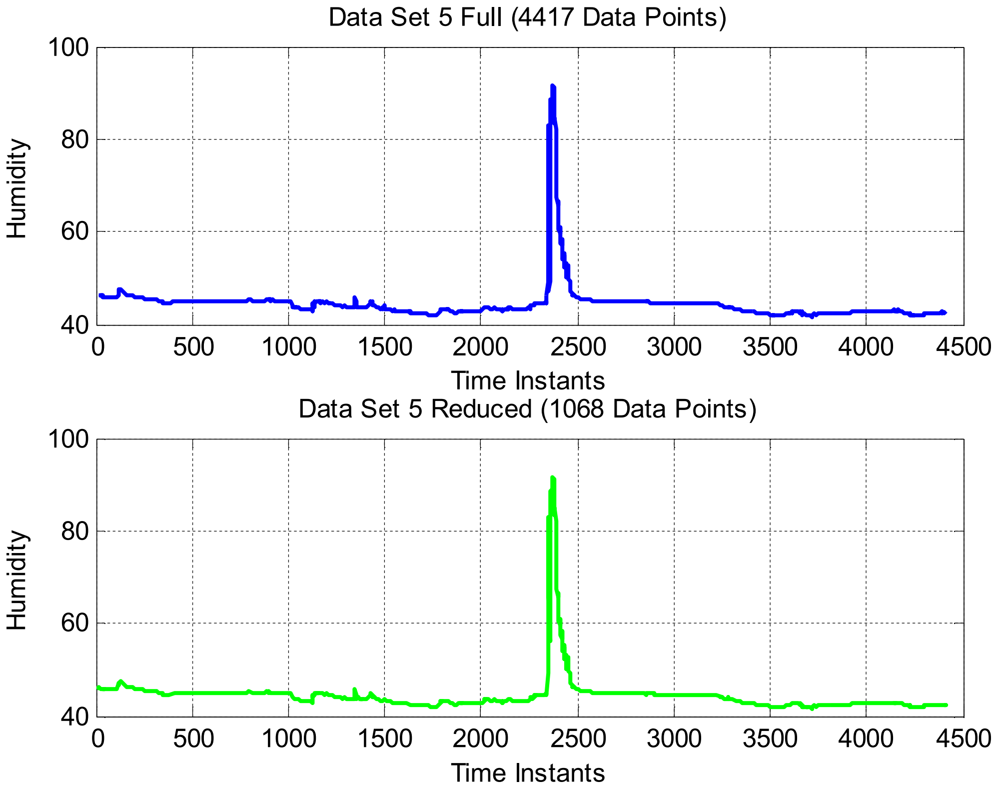

Figure 4.5-e Environmental monitoring data validation (step 2 – data set 5)

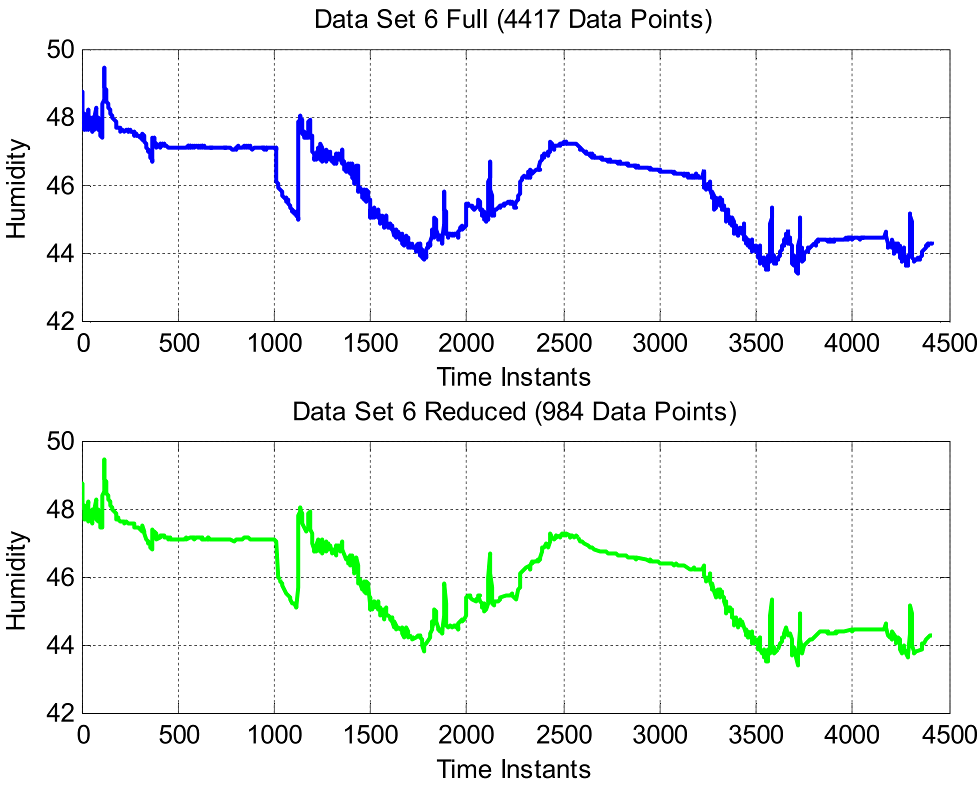

Figure 4.5-f Environmental monitoring data validation (step 2 – data set 6)

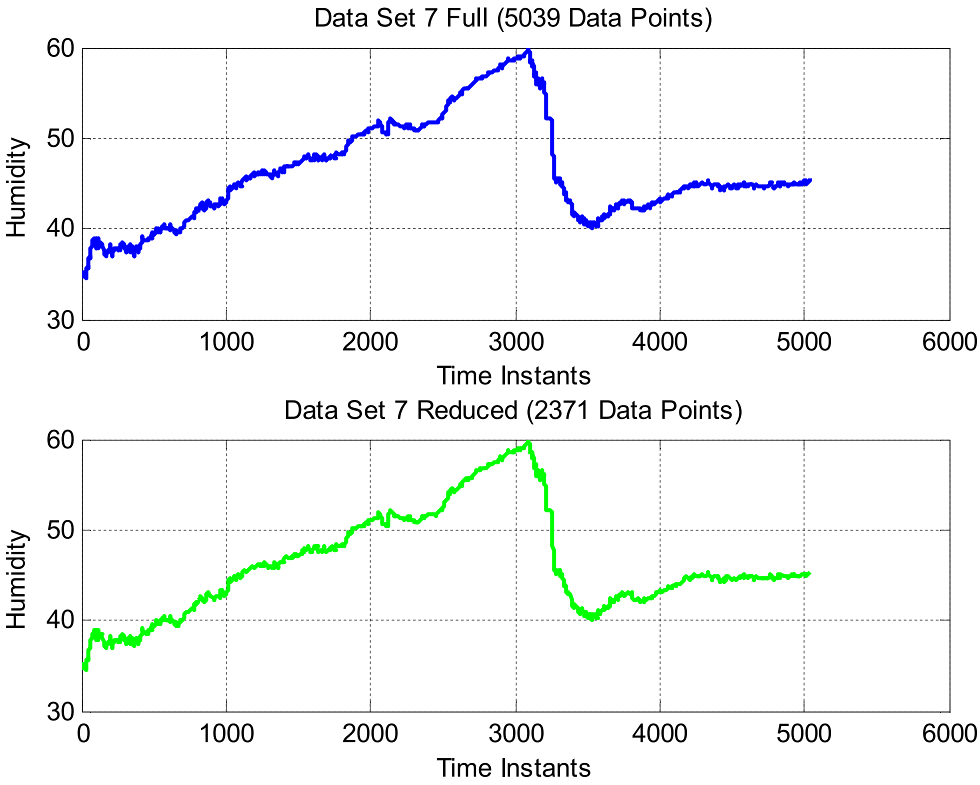

Figure 4.5-g Environmental monitoring data validation (step 2 – data set 7)

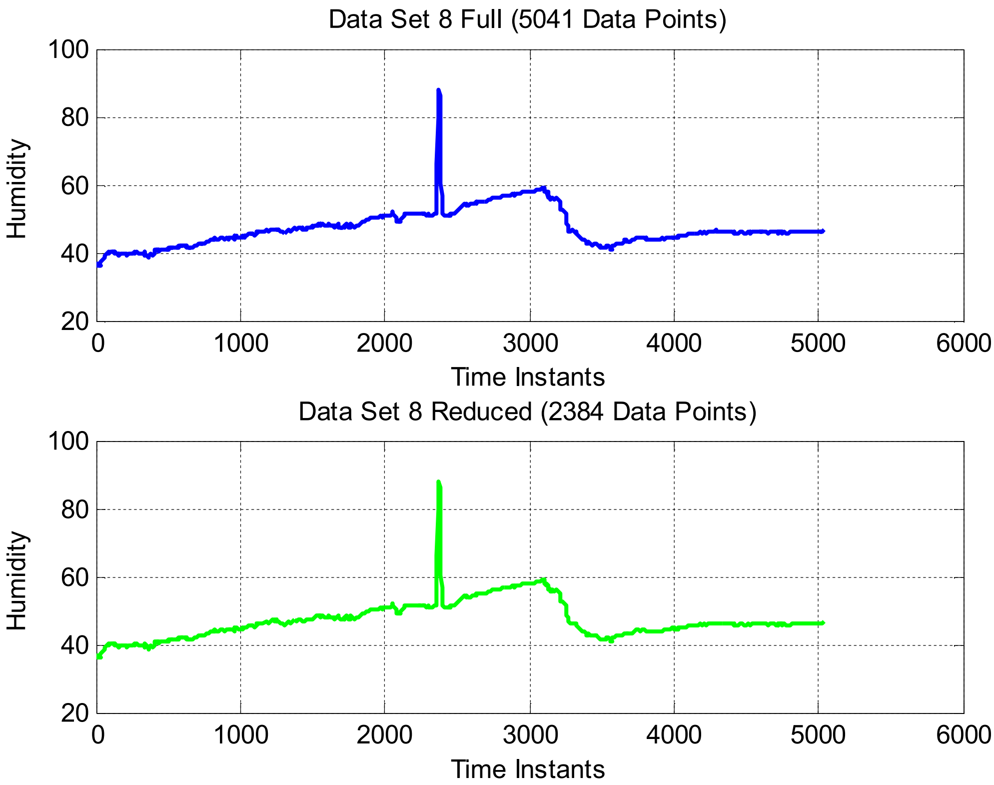

Figure 4.5-h Environmental monitoring data validation (step 2 – data set 8)

4.4.2 Statistical Priority Validation for Video Data

In this sub-section, we prove the validity of SP for video data over M2M communications. The video data comes from a generic surveillance camera sample video that monitors a street with vehicles moving [43]. The video consists of 3903 frames so we chop it into 10 shorter video clips for faster software processing (Each clip consists of 400 frames except the last one has 303 frames only). The video format is RGB24 which means that every pixel has 3 values for red, green and blue components of 8 bits, each. The following table summarizes the specifications of the used video [43].

Table 4.6 Video data specifications

Parameter	Value
Width	640
Height	360
Frame Rate	30 Frame/Second
Duration	130.1 Second
Number of Frames	3903
Video Format	RGB24
Bits per Pixel	24 Bits

For validating the effectiveness of SP, we need to calculate the value of information carried by each frame in the video. Recall that SP is calculated as per equation (4.6).

Let $f_o(t) = 0$, hence SP is fucntion only in the statistical features. We will consider 2 statistical features separately (n = 1), where, $Y = [y_1]$. Here, y_1 represents in one case the 2-D correlation between the current video frame and the last frame sent by the camera MTCD. In the other case, y_1 represents the percentage of pixels that did not change their values as shown in (4.5). The 2 metrics are not combined since they quantify the same value which the change between 2 video frames.

Two validation scenarios for the 10 short video clips are examined using the following methodology

- First, by investigating every statistical function separately and checking how many frames carry information of interest to a given statistical function, i.e. trying to answer these questions:
  - How many frames have correlation with the last reported frame by less than a certain threshold?
  - How many frames have a percentage of pixels that did not change compared to the last reported frame by less than a certain threshold?

The thresholds for the 2-D correlation and the percentage of non-changing pixels are set as shown in Table 4.7. If the 2-D correlation or the percentage of non-changing pixels is less than the specified maximum, the frame is considered to have a high-value information and it is reported as result.

Table 4.7 Thresholds for temperature and humidity data

Threshold	Value
Maximum Absolute 2D Correlation	0.995
Maximum Percentage of Non-Changing Pixels	90%

- Second, by calculating the *SP* value where $A = [10]$. We consider each feature separately and hence it is given the full weight of 10. The sigmoid function parameters for every statistical feature is are shown in Table 4.8. Then, the SP values are evaluated against 30 independent uniformly distributed random sets of *SP* values. This is done in order to determine how many frames will be reported by the MTCD in both cases. Also recall that $f_i(y_i)$ is defined as per equation (4.12).

Table 4.8 Sigmoid function parameters for video statistical features

Statistical Feature	b_i	c_i
2D Correlation	–1000	0.0025
Percentage of Non-Changing Pixels	–20	0.92

4.4.2.1 First Step

The goal of this step is to estimate the number of video frames that carry valuable information, i.e. the frame differs significantly from last transmitted frame, and to realize how considering statistical features of video could be used to minimize video traffic without loss of essential information. The results in Table 4.9 show that a portion of the frames carries non-valuable information and consequently can be ignored. The portion of frames with high-value information ranges between 42% and 83% for 2-D correlation method and between 39% and 80% for the percentage of non-changing pixels method. It is worth noting that the video information is preserved but with less frame rate since only frames with important actions are considered and the Rx side codec needs to compensate for the ignored frames (e.g. reducing frame rate when playing video or interpolating frames). Another observation is that these methods are useful for night shift surveillance cameras where there is little or no motion in the surrounding environment. We assume that special type frames [44] are reported but this is not the focus of our work. Another assumption is that a full video with all frames is recorded at the camera while the reduced video version is just used for wireless transmission over the cellular network for off-site monitoring.

Table 4.9 Video data validation (step 1)

#	Number of Frames	2D Correlation < 0.95	Percentage of Non-Changing Pixels < 0.9
1	400	199	223
		49.75%	55.75%
2	400	175	196
		43.75%	49.00%
3	400	240	195
		60.00%	48.75%
4	400	298	253
		41.75%	63.25%

5	400	331	312
		82.75%	78.00%
6	400	252	157
		63.00%	39.25%
7	400	276	275
		69.00%	68.75%
8	400	303	315
		75.75%	78.75%
9	400	319	320
		79.75%	80.00%
10	303	155	154
		51.16%	50.83%

4.4.2.2 Second Step

In this step, an SP-based score is given to all frames according to the aforementioned statistical functions and their sigmoid parameters. Every video with its SP scores is compared against 30 independent random uniformly distributed (from 0 to SP_{max}) SP score sequences. These random sets model the SP score of another MTCD and the data from each of the 10 video clips should have a greater SP score to be able to report a certain frame. Over 30 independent runs, we examine how much resource savings have been made (How many frames are reported with respect to the total number of frames?). Table 4.10 summarizes the results of the second step of this validation. The results are averaged for the 30 runs and 95% confidence analysis is applied. The portion of frames that were transmitted successfully ranges between 67% and 87% for 2-D correlation method and between 54% and 70% for the percentage of non-changing pixels method. 2-D correlation shows better performance in terms of success to transmit video frames more than or equal to the minimum needed for preserving video information (minimum needed frames could be deduced from the first step).

Table 4.10 Video data validation (step 2)

#	Number of Frames	2D Correlation	Percentage of Non-Changing Pixels
1	400	280 (±2)	245 (±3)
		70.07%	61.21%
2	400	269 (±2)	235 (±3)
		67.36%	58.68%
3	400	321 (±2)	240 (±3)
		80.13%	60.04%
4	400	347 (±1)	245 (±3)
		86.70%	61.32%
5	400	343 (±1)	279 (±3)
		85.80%	69.87%
6	400	336 (±1)	218 (±3)
		83.95%	54.38%
7	400	332 (±2)	261 (±3)
		82.90%	65.27%
8	400	332 (±1)	267 (±3)
		83.03%	66.68%
9	400	338 (±1)	266 (±3)
		84.39%	66.55%
10	303	225 (±2)	179 (±3)
		74.30%	59.10%

4.5 Chapter Summary

In this chapter, we introduced a detailed description of the statistical priority which is our proposed metric to be used for M2M scheduling. We discussed how to calculate SP for different types of MTCDs such as environmental monitoring sensors, cameras and alarms. We presented demonstrations of statistical priority calculations on real data sets (sensor data points and video clips) and proved that it can be used to represent a full data with a reduced subset by considering the data with higher statistical priority only for reporting without loss of data features. Using SP is promising as a scheduling metric since it gives higher priority to data points with high-value information allowing the chance to send less data packets, i.e. less radio resources, while preserving data stream features.

Chapter 5: Statistical Priority-based Scheduler for Massive M2M Deployments

5.1 Introduction

In this chapter, we present a statistical priority-based scheduler that utilizes the Statistical Priority (SP) metric introduced in Chapter 4 as its scheduling metric. We state the design goals of the scheduling technique. We then carry out the performance evaluation of the our proposed scheduling algorithm through MATLAB simulation experiments. Finally, experimental results are analyzed to show the advantages of the statistical priority-based scheduling scheme.

5.2 Statistical Priority-based Scheduling Scheme for M2M Communications

The design of the proposed scheduling algorithm is subject to the requirements and the characteristics of M2M communications that we discussed in Chapter 2. We present a statistical priority-based uplink scheduling algorithm for M2M communications that is flexible since it can adapt to different types of MTCDs as the main contribution of the thesis.

5.2.1 Design Goals

Based on understanding the M2M challenges and characteristics and previous contributions in the literature, we determine the design goals of our proposed scheduling algorithm and its motivation as follows

- *Radio Resource Efficiency*: In M2M communications, massive deployments of MTCDs are expected. The radio resources may be insufficient to serve that huge number of MTCDs in addition to UEs that share the same bandwidth. M2M uplink scheduling algorithms should be designed to operate under scarcity of radio resources and to use them efficiently.
- *M2M Data Characteristics Adaptation*: The data reported by MTCDs in M2M networks have high redundancy. For example, the data reported by a temperature sensor within an hour do not reflect much change (refer to data analysis of temperature sensor data in Chapter 4). Another example is the data reported by surveillance cameras in case no activity in the secured area. This means that an M2M uplink scheduling algorithm should be able to consider the value of information carried by data reported by MTCDs in order to prioritize MTCDs with data of higher importance to make efficient use of the scarce radio resources.
- *Flexibility with respect to Data Type*: The MTCDs perform different tasks in a wide variety of fields. Consequently, M2M uplink scheduling algorithm should be able to deal with data from different sources. The algorithm should consider priority of data based on the MTCD type and the importance of the information carried by these data. In addition, the method of evaluating the importance of the information carried by MTCDs differs from one MTCD type to another, as discussed earlier in Chapter 4.

5.2.2 Statistical Priority-based Algorithm Design

The algorithm is designed to meet the aforementioned goals and it can be divided into three steps that can be described as follows.

Step 1: Assign *p%* of PRBs for serving MTCDs. For example, *p%* can be calculated as follows

$$p\% = \min(\frac{W \times M}{U}, 1) \times 100\% \tag{5.1}$$

where W is a weighting factor much less than 1, M is the number of active MTCDs and U is the number of UEs. (For example if $M = 1000$, $U = 10$, $W = 0.001$ (e.g. the average packet size for a UE is 1000 times the average packet size for an MTCD), then $p\% = 10\%$).

Step 2: Consider a set of N PRBs and a set of M active MTCDs (active MTCDs are the ones that request data transmission and have not missed deadline yet). Every MTCD m ($m = 1 \dots M$), has a deadline D_m, an SP value of SP_m and an SNR value at the n^{th} PRB of S_{mn}.

Step 3:

- Sort SP_m in a descending order to select an active MTCD m that carries data with maximum statistical priority score (i.e. most valuable information).
- Assign PRB n (at which the selected MTCD m has maximum SNR S_{mn}) to MTCD m.
- Allocate PRBs on the right and the left of PRB n to MTCD m and keep expanding in both direction till any of the following conditions applies [22]:
  - MTCD m acquires enough PRBs to send its data.
  - Expanding in both directions is blocked by PRBs allocated to other MTCDs.
- Consider MTCD m as served and remove the allocated PRBs from the set of available PRBs.
- Repeat *Step 3* till all PRBs are allocated.

The following points are worth noting:

- In *Step 1*, the radio resources are divided between M2M and H2H traffic flows, so that each traffic flow gets its dedicated PRBs (i.e. M2M traffic resource allocation is done separate from H2H traffic resource allocation). The division of PRBs is based on network conditions. After this step, the algorithm focuses solely on the M2M uplink scheduling process.
- In *Step 2*, SNR could be replaced by any channel-state metric like SINR or CQI
- Any MTCD that misses deadline is deactivated, except there are data packets in its buffer, till it requests to transmit data.

5.3 Statistical Priority-based Scheduling Performance Evaluation

The main goal of this sub-section is to answer the following questions.

- Is statistical priority-based scheduling resource efficient and suitable for M2M communications?
- Would it be essential to use statistical priority-based scheduling rather than SNR-based scheduling or Deadline-based scheduling?

The following subsections provide answers to these questions.

5.3.1 Idea Validation

In M2M communications, using a deadline-based scheduler is the most efficient in terms of meeting MTCDs deadlines and reducing the deadline-missing ratio. However, this occurs only when radio resources are sufficient to serve all requests from MTCDs to send data in the uplink direction. When radio resources are not sufficient, it can be shown that serving MTCDs based on any metric rather than deadline, e.g. statistical priority, is more efficient in reducing deadline-missing ratio. This can be illustrated through the following 2 examples.

5.3.1.1 A Simple Example

Scenario 1: Assume that there are 2 MTCDs (M_1, M_2) that request 4 PRBs each for data transmission. There is 1 PRB available every 1ms. There deadlines are D_1 = 5ms and D_2 = 8ms. There 8 PRBs available for in 8ms which is sufficient to satisfy MTCDs requests.

Case 1-1: Deadline-based scheduler is used

1ms	2ms	3ms	4ms	5ms	6ms	7ms	8ms
M_1	M_1	M_1	M_1	M_2	M_2	M_2	M_2
				D_1			D_2

Average deadline-missing ratio = 0%

Case 1-2: Metric-based scheduler is used (i.e. MTCDs are sorted according a scheduling metric like SP or SNR)

Options:

- (M_1, M_2)

1ms	2ms	3ms	4ms	5ms	6ms	7ms	8ms
M_1	M_1	M_1	M_1	M_2	M_2	M_2	M_2
				D_1			D_2

- (M_2, M_1)

1ms	2ms	3ms	4ms	5ms	6ms	7ms	8ms
M_2	M_2	M_2	M_2	M_1			
				D_1 Missed			D_2

Average deadline-missing ratio = (0.5)(0%) + (0.5)(50%) = 25%

Conclusion 1: Deadline-based scheduling has a deadline-missing ratio that is less than random priority scheduling in case of the sufficiency of radio resources.

Scenario 2: Now, if we add a third MTCD M_3 with deadline D_3 = 6ms with the same request, PRBs will be sufficient to satisfy 2 requests only i.e. insufficient PRBs. Hence, a new conclusion is reached.

Case 2-1: Deadline-based scheduler is used

1ms	2ms	3ms	4ms	5ms	6ms	7ms	8ms
M_1	M_1	M_1	M_1	M_3	M_3	M_2	M_2
				D_1	D_3 Missed		D_2 Missed

Average deadline-missing ratio = 66.67%

Case 2-2: Metric-based scheduler is used

Options:

- (M_1, M_2, M_3)

1ms	2ms	3ms	4ms	5ms	6ms	7ms	8ms
M_1	M_1	M_1	M_1	M_2	M_2	M_2	M_2
				D_1	D_3 Missed		D_2

- (M_1, M_3, M_2)

1ms	2ms	3ms	4ms	5ms	6ms	7ms	8ms
M_1	M_1	M_1	M_1	M_3	M_3	M_2	M_2
				D_1	D_3 Missed		D_2 Missed

- (M_2, M_1, M_3)

1ms	2ms	3ms	4ms	5ms	6ms	7ms	8ms
M_2	M_2	M_2	M_2	M_1	M_3		
				D_1 Missed	D_3 Missed		D_2

- (M_2, M_3, M_1)

1ms	2ms	3ms	4ms	5ms	6ms	7ms	8ms
M_2	M_2	M_2	M_2	M_3	M_3		
				D_1 Missed	D_3 Missed		D_2

- (M_3, M_1, M_2)

1ms	2ms	3ms	4ms	5ms	6ms	7ms	8ms
M_3	M_3	M_3	M_3	M_1	M_2	M_2	M_2
				D_1 Missed	D_3		D_2 Missed

- (M_3, M_2, M_1)

1ms	2ms	3ms	4ms	5ms	6ms	7ms	8ms
M_3	M_3	M_3	M_3	M_2	M_2	M_2	M_2
				D_1 Missed	D_3		D_2

Average deadline-missing ratio = (1/6)(33.33%) + (1/6)(66.67%) + (1/6)(66.67%) + (1/6)(66.67%) + (1/6)(66.67%) + (1/6)(33.33%) = 55.56%

Conclusion 2: In case of insufficient radio resources, deadline-based scheduling may have a deadline-missing ratio that is higher than the average deadline-missing ratio of the random priority scheduling.

5.3.1.2 An Advanced Example

Moving to a more advanced example, we have 2 scenarios.

Scenario 1: Fixed number of MTCDs with variable service time

A fixed number of MTCDs ($M = 100$) is assumed. MTCDs delay-tolerance (deadline from t = 0) is uniformly distributed over the range (10ms – 950ms). The system bandwidth is 1.4 MHz (6 PRBs per subframe). The service time is varied by varying the number of PRBs required by each MTCD from 6 PRBs (to be served in 1ms) to 112 PRBs (to be served in 19ms). The deadline-missing ratio is calculated for the deadline-based scheduler and the metric-based scheduler. Table 5.1 summarizes the deadline-missing ratio percentage, averaged over 30 independent runs, which are depicted in Figure 5.1. It is worth noting that minimum deadline-missing ratio is calculated based on the assumption of optimal utilization of PRBs i.e. all PRBs are used by MTCDs that did not miss deadlines. Hence,

$$Min\_Miss\_Ratio = \max(1 - \frac{\#Available\_PRBs}{\#Requested\_PRBs}, 0) \quad (5.2)$$

Table 5.1 Deadline-missing ratio (Scenario 1 – Advanced Example)

Requested PRBs per MTCD	% Deadline-Missing Ratio		
	Minimum	Deadline-based Scheduler	Random Priority Scheduler
6	0%	0%	4.67%
12	0%	0%	8.93%
24	0%	0%	18.33%
48	0%	2.67%	32.87%
60	5%	17%	38.6%
72	**20.83%**	**41%**	**43.8%**
96	**40.63%**	**69.53%**	**52.47%**
112	49.11%	79.3%	57.47%

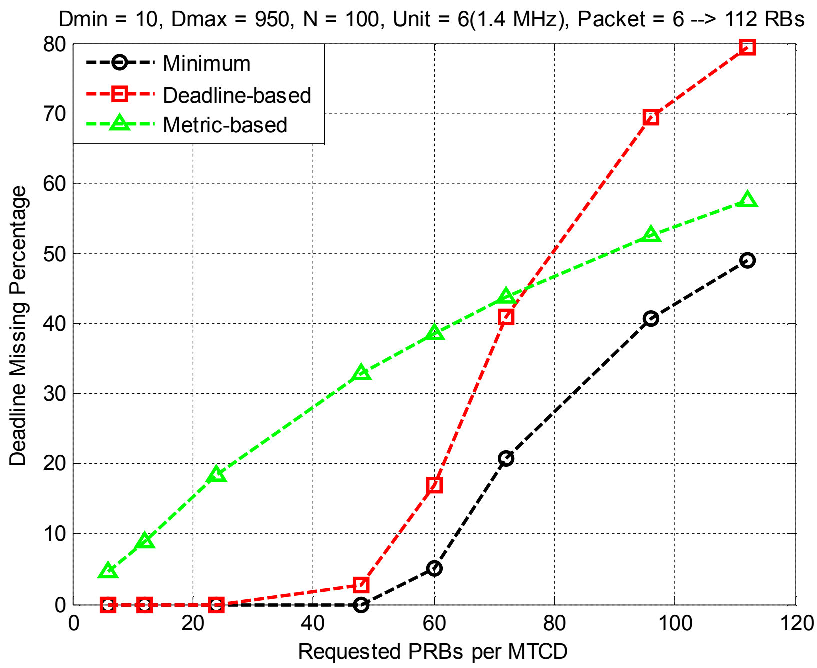

Figure 5.1 Deadline-missing percentage (Scenario 1 – Advanced Example)

The results of this experiment show that scheduling MTCDs on a non-deadline basis is more efficient (makes better usage of PRBs) by reducing the percentage of deadline-missing when PRBs are not sufficient for satisfying MTCDs requests (approximately when minimum deadline-missing is 25% i.e. 75% of request could be satisfied as an upper bound).

Scenario 2: Fixed service time with variable number of MTCDs

This time, we make radio resources insufficient by varying the number of MTCDs in the M2M network. MTCDs delay-tolerance (deadline from t = 0) is uniformly distributed over the range (10ms – 950ms). The system bandwidth is 1.4 MHz (6 PRBs per subframe). The service time is fixed by fixing the number of PRBs required by each MTCD to be 60 PRBs (to be served in 10ms). The number of MTCDs is varied from 50 to 900 MTCDs. Table 5.2 summarizes the deadline-missing ratio percentage, averaged over 30 independent runs, which are depicted in Figure 5.2.

Table 5.2 Deadline-missing ratio (Scenario 2 – Advanced Example)

Number of MTCD	% Deadline-Missing Ratio		
	Minimum	Deadline-based Scheduler	Random Priority Scheduler
50	0%	0%	23.2%
100	**5%**	**17%**	**38.6%**
150	**36.67%**	**61.8%**	**50.42%**
200	52.5%	81.3%	59.1%
300	68.33%	94.03%	70.2%
500	81%	99.13%	81.6%
700	86.43%	99.77%	86.81%
900	89.44%	99.87%	89.9%

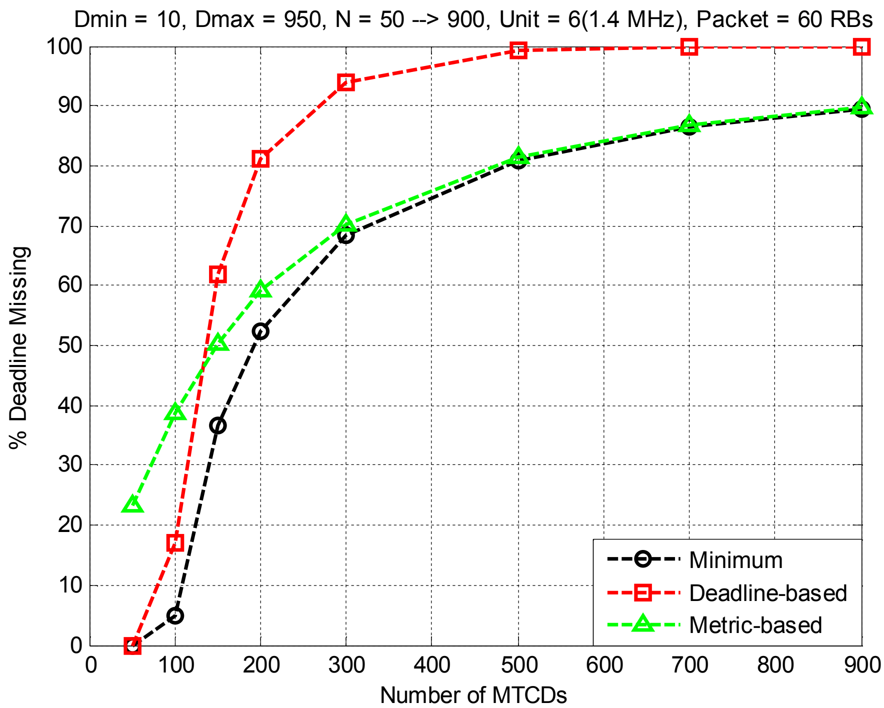

Figure 5.2 Deadline-missing percentage (Scenario 2 – Advanced Example)

The second scenario results in the same conclusion as the first one. There is an inflection point at which scheduling MTCDs on a non-deadline-basis is more efficient in utilizing PRBs. The rate of increase in deadline misses for the deadline-based scheduler with increasing resource scarcity is higher than the non-deadline-based scheduler. In summary, deadline-based scheduler performs better at lower traffic load but at certain point in high traffic load it is not the correct decision to use deadline-based scheduler.

This answers the first question of that we posed in the beginning of Section 5.3. When radio resources are insufficient, we need a technique (that is not delay-based) to select the data of higher importance to be given higher priority in uplink scheduling since available PRBs cannot satisfy all requests. Statistical priority, as discussed in Chapter 4, can be used to determine data importance and when used to send data it preserves data characteristics to a great extent.

5.3.2 Simulation Setup and Scenarios for Statistical Priority-based (SP-based) Scheduling

In order to answer the second question that we posed in the beginning of Section 5.3, we compare statistical priority-based scheduling with delay-based scheduling in more complex M2M deployments where MTCDs have different profiles (i.e. different delay tolerance, different data size, different data type, different traffic structure). In addition, SNR-based scheduler does not address deadline priorities so it can be a good candidate for comparison with the statistical priority-based scheduler. In this comparison, we also introduce novel performance evaluation metrics that focus on measuring the amount of critical data that was successfully transmitted rather than measuring mere throughput that disregards the variation in importance of the different data packets. The simulation parameters, MTCD configuration, MTCD traffic, and performance evaluation metrics are given in Tables 5.3, 5.4, 5.5 and 5.6 respectively.

Table 5.3 Simulation parameters for SP-based scheduling evaluation

Parameter	Value
Number of MTCDs (M)	100 – 200 – 300 – 400
Number of Base Stations	1
Average SNR Range	Uniform (4dB,10dB)
Number of Subframes	200000
Number of Runs	10 Independent Runs
Bandwidth (MHz)	3 – 5
Number of PRBs	15 – 25
Channel Model	Additive White Gaussian Noise (AWGN)
Scheduling Algorithms	• Deadline-based Scheduler • SNR-based (Channel-state) Scheduler • SP-based Scheduler

Table 5.4 MTCD configuration

#	Application	Delay Tolerance (ms)	Packet Size (Bytes)	%
1	Emergency Alarms	Constant (10)	32	10%
2	Surveillance Camera	Uniform (125,250)	512	10%
3	Regular Monitoring (Temperature)	Uniform (800,900)	128	40%
4	Regular Monitoring (Humidity)	Uniform (800,900)	128	40%

Table 5.5 MTCD traffic description

#	Application	Packet Content	Traffic Description
1	Emergency Alarms	Alert	• 5 packets within 200ms
2	Surveillance Camera	Compressed Low-quality Video Frame	• 30 frames per second • 400 frames per MTCD • Random start time • Frames extracted from [43]
3,4	Regular Monitoring (Temperature and Humidity)	Data Point	• 1 data point per second • 200 data points per MTCD • Data points extracted from [41]

Table 5.6 Performance evaluation metrics description

Evaluation Metric	Description
Overall deadline-missing ratio	The ratio between packets that missed deadline to the overall number of packets for all MTCD types
Alarm deadline-missing ratio	The ratio between alarm packets that missed deadline to the overall number of alarm packets
Critical packets success rate (Sensors)	The ratio between successfully sent packets for MTCDs of type (3, 4) to the critical number of packets for every MTCD of type (3, 4). The critical number of packets is the minimum number of packets to claim that summarized data represent the full data set for every MTCD.
Critical packets success rate (Cameras)	The ratio between successfully sent packets for MTCDs of type (2) to the critical number of packets for every MTCD of type (2).

In addition, SP parameters are selected to boost the importance of data points or frames that have high-value information. The high-value information for sensors data points as described in Chapter 4 is found in statistical features like exceeding a threshold, following a trend, experiencing a big change with respect to last reported data. In case of video frames, high-value information exists in a frame if it is fairly uncorrelated with last sent video frame. For alarms, all packets are of maximum importance ($SP = sp\_max$) and prioritized over all other types of traffic. Tables 5.7-a, 5.7-b and 5.7-c summarize SP parameters used in simulation.

Table 5.7-a Statistical features thresholds for MTCDs

Statistical Feature	Temperature Sensor	Humidity Sensor	Camera
Upper Threshold (y_1)	28° C	48%	N/A
Lower Threshold (y_2)	27° C	42%	N/A
Difference Threshold (y_3)	0.1° C	0.1%	N/A
Trend (y_4)	Yes	Yes	N/A
2D Correlation (y_5)	N/A	N/A	0.995

Recall that $f_i(y_i)$ is defined as per equation (5.3)

$$f_i(y_i) = sigmoid(b_i, c_i) = \frac{1}{1+e^{-b_i(y_i - c_i)}} \tag{5.3}$$

Table 5.7-b Statistical features sigmoid parameters for MTCDs

Statistical Feature	Temperature Sensor		Humidity Sensor		Camera	
	b	c	b	c	B	c
Upper Threshold (y_1)	100	0	100	0	N/A	N/A
Lower Threshold (y_2)	-100	0	-100	0	N/A	N/A
Difference Threshold (y_3)	5	0	5	0	N/A	N/A
Trend (y_4)	100	0	100	0	N/A	N/A
2D Correlation (y_5)	N/A	N/A	N/A	N/A	-1000	0.0025

Table 5.7-c Statistical features weights for MTCDs

Statistical Feature	Temperature Sensor	Humidity Sensor	Camera	Alarm
Upper Threshold (a_1)	4	4	0	N/A
Lower Threshold (a_2)	4	4	0	N/A
Difference Threshold (a_3)	3	3	0	N/A
Trend (a_4)	3	3	0	N/A
2D Correlation (a_5)	0	0	10	N/A
Periodic Priority (a_0)	5	5	5	N/A
Default	N/A	N/A	N/A	10

Overall, 8 scenarios (2 values of bandwidth × 4 values of the number of MTCDs) are simulated according to the aforementioned simulation parameters to represent different traffic loads.

5.3.3 Simulation Results and Discussion

The evaluation metrics stated in Table 5.7 are measured for the 8 simulation scenarios. Each simulation scenario is performed using three uplink schedulers, deadline-based, SNR-based, and statistical priority-based, for 10 independent runs. The final results represent the average of these runs with 95% confidence interval analysis. It is worth noting that each of the performance evaluation metrics focus on a certain type of MTCD traffic (i.e. sensor data, video and alerts).

5.3.3.1 Overall Deadline-Missing Ratio

Overall deadline-missing ratio results, that are shown in Figures 5.3-a and 5.3-b, are in line with the previous conclusion that non-deadline-based schedulers have less deadline misses than deadline-based schedulers when radio resources are not sufficient for satisfying all data transmission requests. SNR-based and SP-based schedulers help reduce the ratio of deadline misses for different densities of MTCDs as shown in Figures 5.3-a and 5.3-b. We see from the figure that SNR-based scheduler outperforms our proposed SP-based scheduler due to the fact that the SNR-based scheduler allocates PRBs to the MTCDs with the best channel conditions so they utilize the scarce radio resources efficiently (i.e. without packets drops or need for additional PRBs per packet). However, the disadvantages of SNR-based scheduling will be revealed through the other performance evaluation metrics.

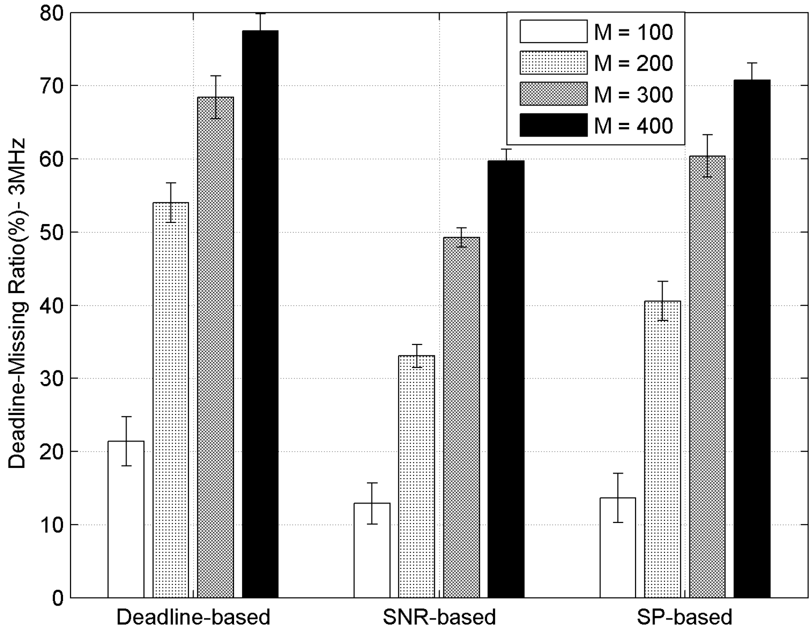

Figure 5.3-a Overall deadline-missing ratio (BW = 3MHz)

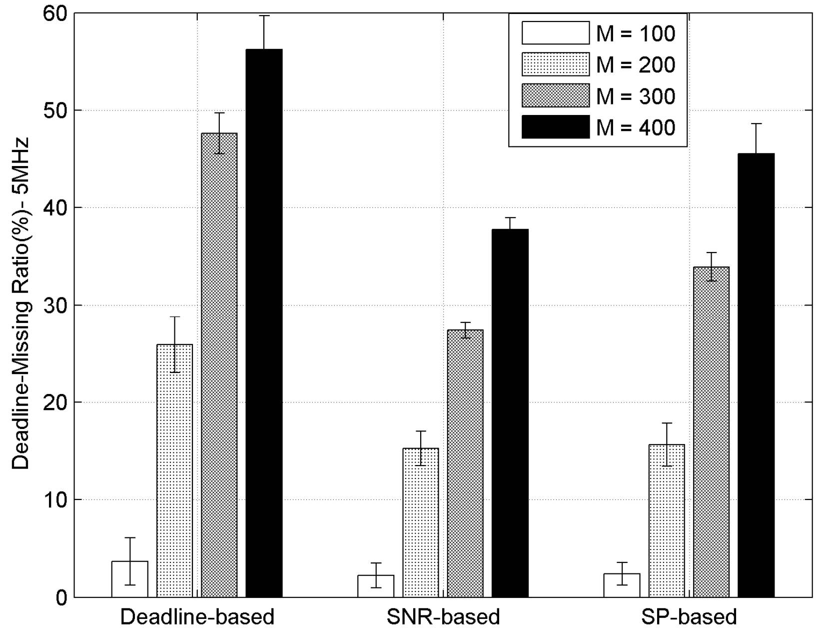

Figure 5.3-b Overall deadline-missing ratio (BW = 5MHz)

5.3.3.2 Alarm Deadline-Missing Ratio

In this performance evaluation metric, we focus on alarm MTCDs due the high importance of the data they send and the strict deadline requirement they have. A perfect scheduler should be able not to miss any alarm packet (or miss as minimum packets as possible if alarm redundancy is assumed). The alarm deadline-missing ratio results in Figures 5.4-a and 5.4-b show that SP-based scheduler outperforms deadline-based scheduler with an alarm deadline-missing ratio that does not exceed 4% for all simulation cases. On the other hand, the SNR-based scheduler deals with alarm MTCDs and other MTCDs on equal basis while alarm packets need to be given higher priority due to the high-value information they carry about abnormal conditions or threats. Hence, it performs significantly worse than the other schedulers from the perspective of this evaluation metric by missing the largest number of alarm packets. The reason behind the superior performance of the statistical priority-based is that it gives higher priority to alarm MTCDs based on data importance or SP score in which alarm MTCDs always surpass other MTCDs. On the other hand, the deadline-based scheduler gives higher priority to alarm MTCDs based on deadline where other non-alarm MTCDs may have closer deadlines.

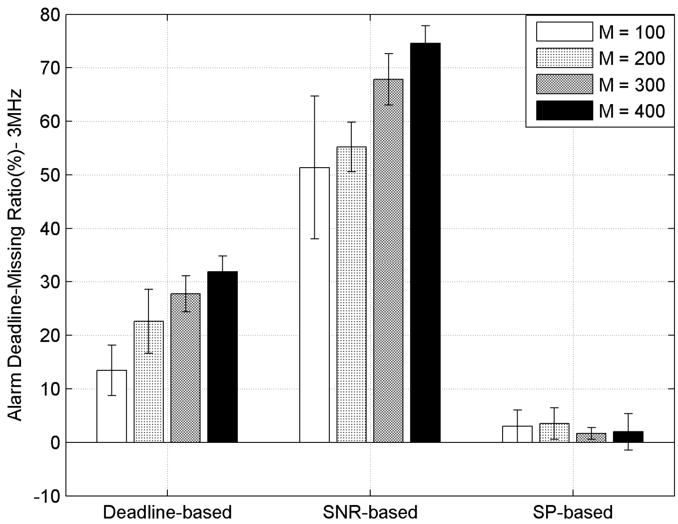

Figure 5.4-a Alarm deadline-missing ratio (BW = 3MHz)

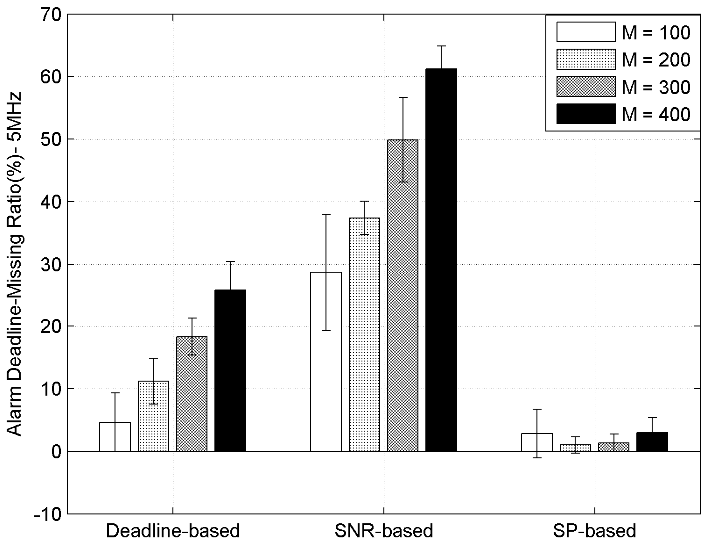

Figure 5.4-b Alarm deadline-missing ratio (BW = 5MHz)

5.3.3.3 Critical Packets Success Rate (Sensors)

When we evaluate the performance of the scheduler for regular monitoring MTCDs (i.e. MTCDs of types 3 and 4), more focus should be given to the success of the scheduling technique in terms of efficient usage of radio resources to send data of high-value information. Regular monitoring devices (i.e. sensors) usually send repetitive data values as shown in Chapter 4 where we concluded that the same data features could be preserved even with sending a reduced number of data values selected on the basis of their importance calculated by statistical priority. Hence, to evaluate the success of a given scheduler with respect to regular monitoring MTCDs, we measure the success rate of sending the critical packets ratio of sent data packets to the number of critical packets that must be sent to fully represent the whole data stream. This success ratio is measured for every regular monitoring MTCD as in (5.4) and the average is calculated. The number of critical packets is obtained from Chapter 4 results (recall Table 4.5 in Chapter 4). With limited radio resources, it is not efficient to strive to send as much data as possible. Rather, the focus should be on ensuring the success of sending critical data packets.

$$Critical\_Packets\_Success\_Rate = \min\left(\frac{\#Sent\_Packets}{\#Critical\_Packets}, 1\right) \times 100\% \qquad (5.4)$$

The results of this performance evaluation metric are shown in Figures 5.5-a and 5.5-b. It is clear that the statistical priority-based scheduler outperforms the other schedulers. The success rate is almost 100% for system bandwidth equal to 5 MHz at any traffic load. When system bandwidth is reduced to 3 MHz, i.e. less number of PRBs, SP-based scheduler suffers the least losses and keeps the success rate greater than 90% when number of MTCDs is 300 or less.

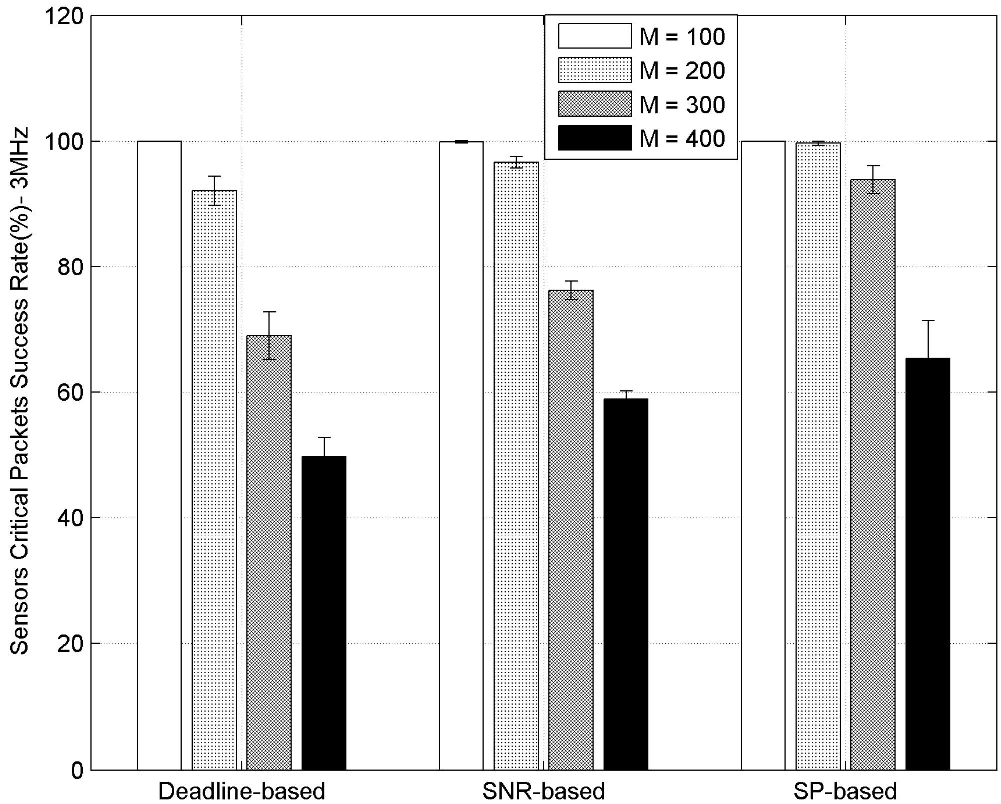

Figure 5.5-a Critical packets success rate (Regular monitoring MTCDs) (BW = 3MHz)

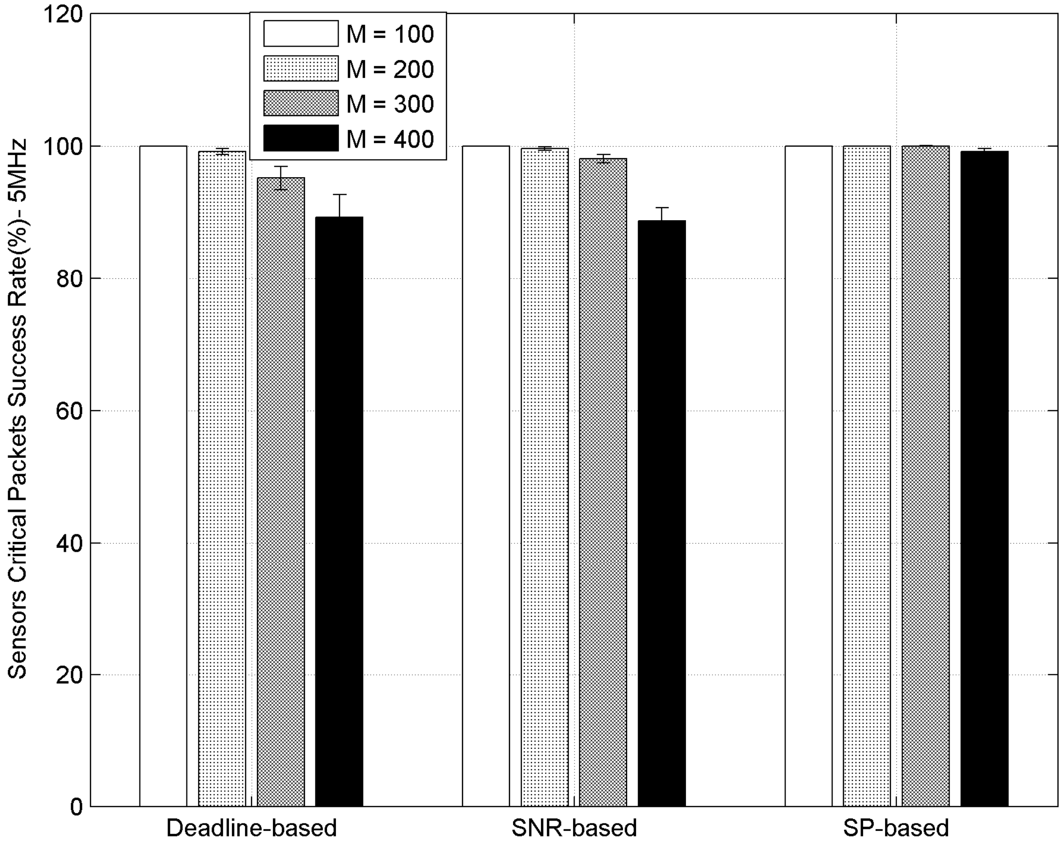

Figure 5.5-b Critical packets success rate (Regular monitoring MTCDs) (BW = 5MHz)

5.3.3.4 Critical Packets Success Rate (Cameras)

The previous critical success rate evaluation metric is used for surveillance camera MTCDs as well. Sending the critical frames is sufficient to represent video information. The success rate for surveillance cameras MTCDs is less than regular monitoring MTCDs due to the larger packet size, denser traffic and stricter deadlines. However, a 70-80% success rate by visual experiment is sufficient to represent the information in video without much negative effect as far as human eye is concerned. The number of critical packets is obtained from Chapter 4 results (recall Table 4.9 in Chapter 4). The success ratio is measured for every surveillance camera MTCD as in (5.4) and the average is calculated. The SP-based scheduler outperforms the other schedulers. The success rate does not drop below 70% except for worst cases of radio resources limitation (i.e. 300 MTCDs and bandwidth of 3 MHz or 400 MTCDs at both values of bandwidth) as shown in Figures 5.6-a and 5.6-b.

It is worth noting that SNR-based scheduler shows the minimum deadline-missing ratio for all MTCDs as shown in Figures 5.3-a and 5.3-b. However, SNR-based scheduler does not consider the importance of data while making scheduling decisions. As a result, it uses the radio resources inefficiently to send data packets from MTCDs with high SNR even if they carry redundant information at expense of MTCDs with high-value information and lower SNR. Hence, SNR-based scheduler has less critical packet success rate when compared to statistical priority-based scheduler that misses more data packets.

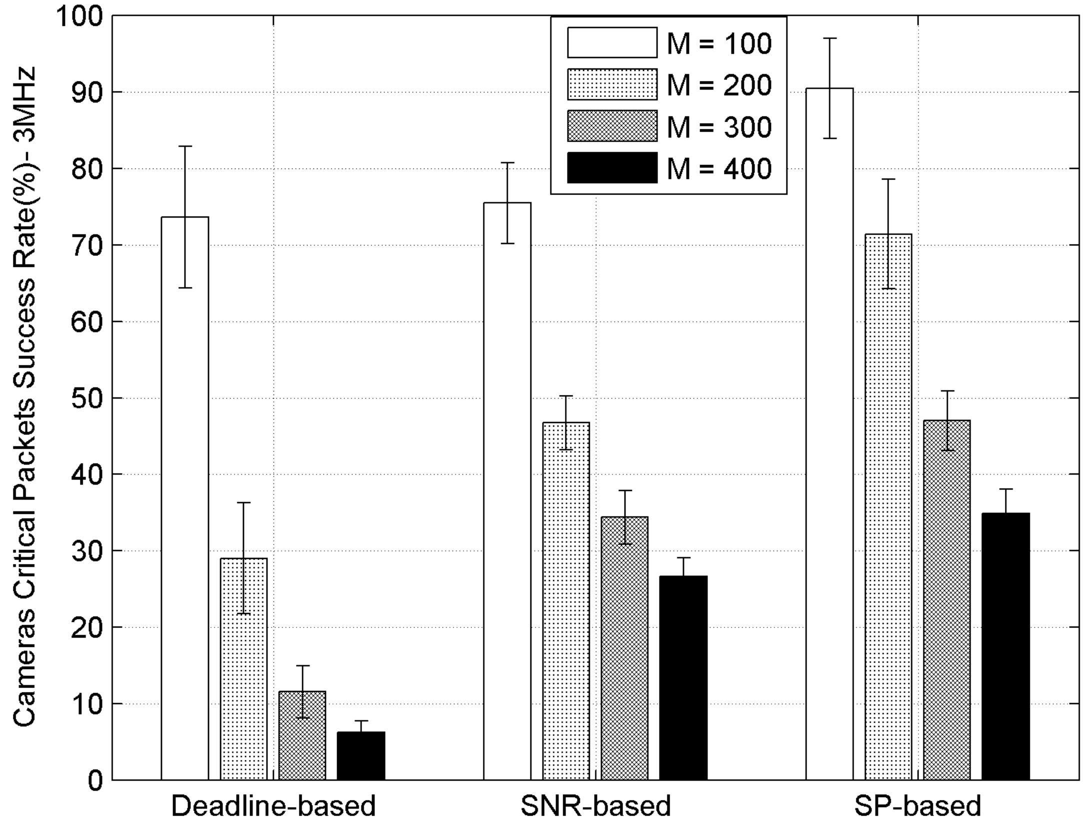

Figure 5.6-a Critical packets success rate (Surveillance cameras MTCDs) (BW = 3MHz)

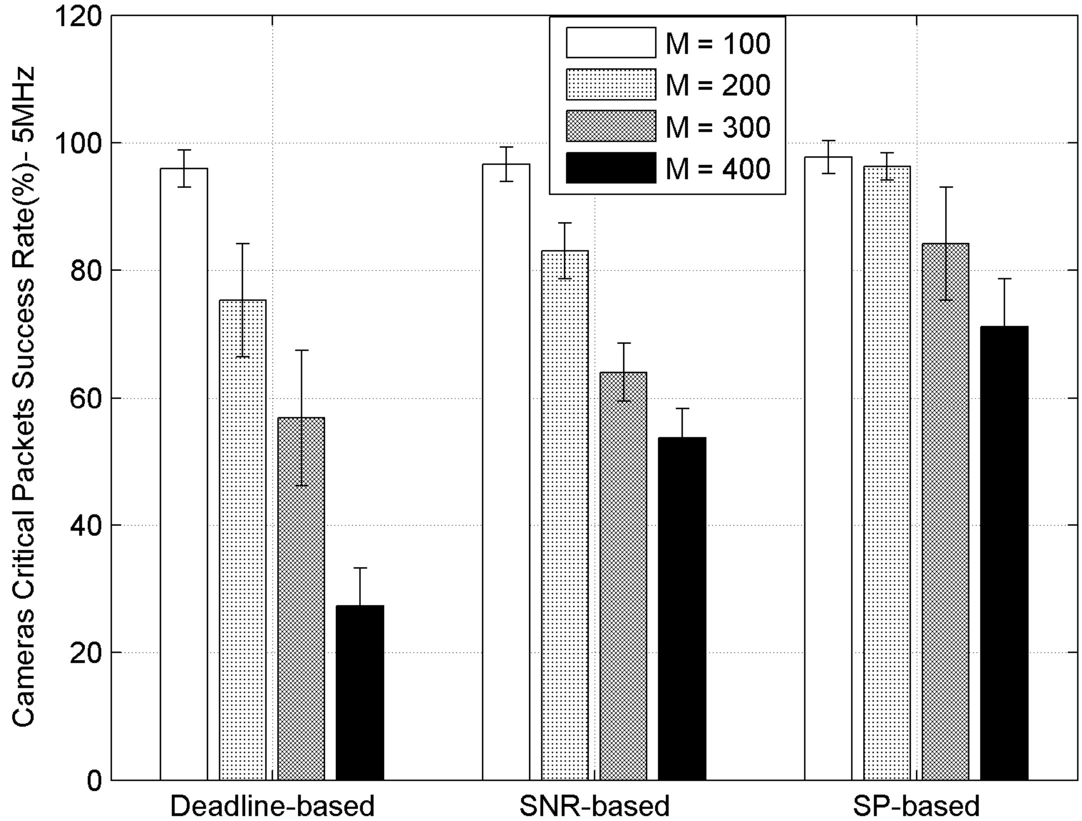

Figure 5.6-b Critical packets success rate (Surveillance cameras MTCDs) (BW = 5MHz)

5.4 Chapter Summary

In this chapter, we introduced a new statistical priority-based scheduling technique for M2M communications. We presented the design goals of our proposed scheduler. Then, we introduced our scheduler design and proposed the usage of statistical priority for M2M scheduling in resource-limited cellular networks. The simulation results show that statistical priority-based scheduler outperforms baseline schedulers which are the deadline-based scheduler and the SNR-based scheduler with the least alarm packet loss and the maximum success rate for transmitting critical packets for sensors and cameras.

Chapter 6: Conclusions and Future Research

6.1 Work Summary and Conclusions

In this thesis, we addressed one of main challenges facing the support of M2M communications over LTE which is the radio resource allocation or scheduling in the uplink direction. We provided a review and a classification of M2M-specific uplink scheduling algorithms in literature. Then, we introduced the concept of a novel scheduling technique based on statistical priority for massive M2M deployments. Statistical priority is calculated by testing different statistical features of the data that could be used to indicate the importance of a certain data value or video frame. The importance of a data value of a monitoring sensor is determined by testing the points against upper and lower thresholds, checking for magnitude similarity with previous data points and checking if a series of data points follow an increasing or decreasing trend. On the other hand, the importance of a video frame is determined through its correlation with previous frames and with lower correlation, the video frame is deemed to carry more changes as compared to the previous ones. Using this concept we can validate that a stream of data values or video frames could be represented sufficiently with a small subset of the stream. This selected subset of the high-value information is based on high statistical priority scores.

We then implemented a scheduler that utilizes this metric for radio resource allocation in M2M networks that has multiple types of MTCDs, such as environmental monitoring sensors, surveillance cameras and alarms. The new technique was compared with channel-based scheduling and deadline-based scheduling. Experiments show that the new technique outperforms the channel-based scheduling and deadline-based scheduling from different perspectives. The statistical priority-based scheduler has the least deadline-missing ratio of data packets generated by alarm MTCDs, which has the highest importance, where the deadline-missing ratio does not exceed 4% for different traffic densities (i.e. number of MTCDs). In addition, statistical priority-based scheduler makes the best use of the scarce radio resources by allocating them to the MTCDs carrying data with highest informative value. This can be concluded from calculating the average ratio of critical packets (i.e. non-redundant information packets that must be sent to preserve data stream features) sent by MTCDs of different types. Environmental monitoring MTCDs succeed to send more than 99% of the critical packets on average in case of using a system bandwidth of 5 MHz and more than 93% of the critical packets on average in case of using a system bandwidth of 3 MHz (except the case of 400 MTCDs where PRBs are not sufficient at all but statistical priority-based still guarantees the best performance). The same superiority is also achieved for surveillance cameras. In this case, the statistical priority-based scheduler succeeds to transmit a minimum of 71% of critical video frames for all traffic densities at a system bandwidth of 5 MHz. The use of a system bandwidth of 3 MHz leads to deterioration in performance especially for high traffic densities (i.e. 300 or 400 MTCDs) for all schedulers. However, the statistical priority-based scheduler still shows superior performance to the other schedulers in this case. Although, the SNR-based scheduler allows transmitting more data packets in total, it exploits the radio resources in sending redundant data from MTCDs with good channel conditions. This drawback is addressed by using statistical priority-based scheduling.

6.2 Future Research

The M2M uplink scheduling can further be investigated from the following perspectives.

- Statistical Priority Optimization: Many optimizations can be done for statistical priority calculation. One of these optimizations is choosing optimal sigmoid function parameters (b and c) to enhance statistical priority calculation. This has the potential of getting better performance for statistical priority-based scheduling.
- Trust Factor: In our proposed scheduling scheme, an assumption is made that MTCDs report a real and honest SPR about the data they want to transmit. This is a backdoor by which MTCDs could always claim that they carry important non-redundant information. Hence, a trust factor could be introduced to reward or penalize MTCDs based on SPR honesty in a way similar to outer-loop link adaptation used in dealing with CQI reports when selecting a MCS for data transmission.
- Spatial Statistical Features: In our research we focused on temporal statistical features which are statistics over time dimension at each MTCD. By spatial statistical features we mean the statistical relations among data streams sent by different MTCDs. MTCDs deployed in close vicinity may send data values that are highly correlated (i.e. redundant information). Hence, there could be a further step for more efficient utilization of radio resources which is to consider statistical priority on the space dimension as well as time dimension.
- Combining SP with Other Metrics: It is an interesting question to research about which is how the scheduling technique will perform if different scheduling metrics were combined together along the guidelines of the generalized scheduling algorithm in 5.2.3. Another interesting question is to define a procedure on how to switch from a scheduling metric to another according to network varying conditions.
- MTCG-based Scheduling: Most M2M uplink scheduling schemes focus on one-hop communications. Performing scheduling of M2M radio resources by allocating them to MTCGs may help resolve the tough contention due to massive access by MTCDs and simplify the scheduling process. The eNB can allocate PRBs to MTCGs and a second-layer resource allocation is performed by the MTCG to assign PRBs to MTCDs. In addition, small cells could be promising to act as MTCGs which is another way of breaking down the problem while using one-hop communication between MTCD and serving small cell eNB.

Appendix 1: M2M-Specific Uplink Scheduling Techniques Detailed Procedures

First Algorithm [30]

There are L PRBs available for K MTCDs. Let the k^{th} MTCD has a deadline (delay tolerance) d_k, needs a number of PRBs for uploading its traffic equal to F_k and its SNR at the l^{th} PRB is $\gamma_{k,l}$.

Step 1: Exclude the PRBs needed for UEs requirements; the remaining set of PRBs is used for M2M communications.

Step 2: Sort active MTCDs in a descending order based on their SNR at different PRBs (based on $\gamma_{k,l}$).

Step 3: For the best $\gamma_{k,l}$,

- If $(d_k < (\sum_{i=1}^{K} d_i)/2K)$ then allocate PRB l to MTCD k and reduce F_k by 1.
- Else, MTCD k is put to sleep for $d_k/2$ period and redo step 3 for next highest $\gamma_{k,l}$.

Step 4: For MTCD k selected in step 3, check if it has the highest SNR in PRBs adjacent to l^{th} PRB.

- If (condition is true) then allocate adjacent PRB to MTCD k and reduce F_k by 1 then repeat step 4 till the condition is dissatisfied or $F_k = 0$.
- Else, Repeat step 3 after considering MTCD k as served and cannot compete for other PRBs.

Step 5: If allocation fails due to delay tolerance condition in step 3 for several times then ignore this condition (Algorithms works as a conventional channel aware scheduler).

Second Algorithm [30]

There are L PRBs available for K MTCDs. Let the k^{th} MTCD has a deadline (delay tolerance) d_k, needs a number of PRBs for uploading its traffic equal to F_k and its SNR at the l^{th} PRB is $\gamma_{k,l}$.

Step 1: Exclude the PRBs needed for UEs requirements; the remaining set of PRBs is used for M2M communications.

Step 2: Sort active MTCDs in an ascending order based on their delay tolerance (based on d_k).

Step 3: For the MTCD with the lowest d_k,

- Allocate PRB l at which MTCD k has best SNR ($\gamma_{k,l}$) (Best SNR for this specific MTCD over all PRBs) and reduce F_k by 1.

Step 4: Allocate adjacent PRBs to MTCD k till $F_k = 0$ or no more adjacent PRBs exist. MTCD k is assumed to be served and step 3 is repeated.

Grouping by Packet Arrival and Jitter [31]

Let an MTCD i with packet arrival rate γ_i and maximum allowable jitter δ_i requests service and there are M formed clusters for MTCDs each characterized by (γ_m, δ_m). The m^{th} cluster is allocated the PRBs for an Access Grant Time Interval (AGTI) of 1 ms (1 subframe) every $1/\gamma_m$. It is worth noting that if more than one cluster should be assigned the AGTI at the same instant, the clusters with lower packet arrival (lower priority) are postponed to the subsequent subframe.

Step 1: Check if there is a cluster with parameters (γ_i, δ_i)

- If condition is true then proceed to step 2a.
- Else proceed to step 2b.

Step 2a: Check if there enough PRBs that can be allocated to a newly added MTCD to the cluster (Each MTCD should get as a minimum of 5 PRBs).

- If condition is true then MTCD i joins the cluster.
- Else the MTCD is denied service.

Step 2b: Calculate δ^* and check if $\delta^* \leq \delta_i$ and $\delta^* \leq \delta_m$ for m = 1 … M (packet arrival rate of all clusters). Where

$$\delta^* = \tau + \sum_{k=1}^{i-1} \left\lceil \frac{\gamma_k}{\gamma_i} \right\rceil, i = 2 \dots M, \tau = AGTI\ Period \quad \text{(A1.1)}$$

- If condition is true then a new cluster is created with parameters (γ_i, δ_i). The reason for that is to guarantee that other MTCDs in other clusters can meet their jitter requirements.
- Else the MTCD is denied service.

Grouping by Packet Arrival and Data Size [32]

Step 1: After determining UEs and MTCDs that are granted access to the eNB [32]. The MTCDs are divided into clusters according to α and β.
Step 2: The m^{th} cluster is allocated the PUSCH/PDSCH PRBs every $1/\alpha_m\beta_m$.
Step 3: PRBs are assigned within the cluster on an FCFS basis.

Predictive Scheduling [33]

Assume a network of MTCDs, where a given device sends a SR. It is worth noting that every MTCD has a chance to send a SR only once in period T_{SR}.

Step 1: eNB grants this device some radio resources and checks if this device has neighbours or not.

- If condition is true then proceed to step 2.
- Else the algorithm is done.

Step 2: eNB checks for every neighbor if it has a scheduled chance for SR after a period of (x+1) ms where ($x \geq 0$).

- If condition is true for a certain neighbor device then this device is granted radio resources after (x+1) ms or more.
- Else if the condition is false for all neighbor devices then the algorithm is done.

QoE Preserving M2M Aware Hybrid Scheduler [34]

Step 1: Assign $p\%$ of system bandwidth to serve MTCDs.
Step 2: Assign PRBs for MTCDs according to their QCI indicated in Table 3.1 till all MTCDs are satisfied or all PRBs are exploited.
Step 3: Check if there are remaining PRBs after serving all MTCDs.

- If condition is true then consider these PRBs for serving UEs.

Step 4: Assign PRBs to UEs based on conventional QCI table set by LTE standard [45].

Class Based Dynamic Priority (CBDP) Scheduler [35]

For every UE or MTCD, a value called Remaining Time To Serve (RTTS) is defined which represents the remaining time till deadline for the given UE or MTCD.
Step 1: UEs are divided into N classes according to their RTTS, where class 1 is used for delay-intolerant UEs, class N is used for delay-tolerant UEs and classes from 2 to $N-1$ are used for the remaining applications.
Step 2: MTCDs are divided into N classes according to their RTTS, where class 1 is used for delay-intolerant MTCDs, class N is used for delay-tolerant MTCDs and classes from 2 to $N-1$ are used for the remaining applications.
Step 3: PRBs are allocated for UEs of Class 1.
Step 4: PRBs are allocated for MTCDs of Class 1.
Step 5: PRBs are allocated for UEs of Classes 2 to $N-1$.
Step 6: PRBs are allocated for MTCDs of Classes 2 to $N-1$.
Step 7: PRBs are allocated for UEs of Class N.
Step 8: PRBs are allocated for MTCDs of Class N.
Step 9: Update RTTS for all unsatisfied or partially satisfied UEs or MTCDs except for those belong to class N in both traffic flows.
It is worth noting that by updating RTTS, terminals could advance to higher service classes as time passes. Figure 3.2 shows the CBDP concept.

For the allocation of PRBs within every class for steps 3 – 8, a heuristic FDPS is proposed to preserve contiguity of PRBs per UE/MTCD with highest efficiency. The steps of the following algorithm are as follows,
Step a: Construct NEED matrix (A 2D matrix, the first dimension is the index of UE or MTCD and the second dimension is the unallocated PRBs). The value or each term in the NEED matrix ($NEED_{ij}$) represents the number of PRBs needed by a given UE/MTCD i if it is allocated a chuck of PRBs starting from PRB j.
Step b: Check for minimum $NEED_{ij}$ in NEED matrix and assign PRBs from PRB j to PRB ($j + NEED_{ij} - 1$) to UE/MTCD i
Step c: Mark the previous set of PRBs as allocated.
Step d: Move to the next UE/MTCD i+1.

We assume that $NEED_{ij}$ is a function of SNR and considers pre-allocated PRBs (For example, if a UE/MTCD needs 10 PRBs starting from PRB j but the last 5 PRBs are pre-allocated, then $NEED_{ij}$ is set to 5).

Enhanced Delay Sensitive Scheduler [36]

Step 1: Calculate urgency for data upload requests.

$$U_i = \begin{cases} \dfrac{B_i}{\max\{B\}} \cdot \dfrac{T_{SF}}{(D_i - t)} & (D_i - t) > 1ms \\ 1 & (D_i - t) \le 1ms \end{cases} \quad \text{(A1.2)}$$

where U_i is urgency metric for request i, B_i = Buffer Status Report (BSR) index of request i corresponding to the buffer size as defined in [36], $\max\{B\}$ is the maximum BSR index, i.e. 63, T_{SF} is the LTE subframe duration, i.e. 1ms, t is the current time in (ms) and D_i is the deadline for request i in(ms).
Step 2: Serve requests with highest urgency first.

Utility Based Scheduler [23]

The PRBs are allocated such that utility sum is maximized.

$$\hat{R} = \arg\max_{R \in S} \{ \sum_{i \in H} U_i^H (R_i^H) + \lambda \sum_{j \in M} U_j^M (R_j^M) \} \quad \text{(A1.3)}$$

Where $R \in S$ represents a possible resource allocation matrix, R_i^H and R_j^M represent the achievable data rate of the i^{th} UE and j^{th} MTCD. U_i^H (R_i^H) and U_j^M (R_j^M) represent utility functions according to the achievable data rates and $\lambda \in [0, 1]$ is the unified weighting factor of M2M communication.

Bandwidth and QoS Aware (BQA) Scheduler [38]

- Step 1: Each group of MTCDs reports its needs to resource estimator so that a portion of PRBs is assigned to every group.
- Step 2: Within each group apply TDPS to pick some MTCDs from the active ones to compete for PRBs allocated for the group based on fairness and priority according to the following scheduling metric (Λ) for MTCD i at a time instant t,

  $$\Lambda_i = \frac{Rinst_i(t, n_i)}{Ravg_i} \cdot \sum_k W_{i,k}(t) \quad \text{(A1.4)}$$

  where *Rinst* is the instantaneous data rate if assigned *n* PRBs, *Ravg* is the average data rate of MTCD *i*. $W_{i,k}$ is the QoS weight of bearer *k* of MTCD *i* at time *t* which is calculated as a function of required bit rate and delay budget per bearer as in [38].
- Step 3: Within the selected MTCDs apply FDPS based on channel-state, contiguity and proportional fairness. Select the resource allocation branch that maximizes overall FDPS metric in UDFS tree. Figure A1.1 [38] shows an example from [38] for FDPS scheduling.

	RC_0	RC_1	RC_2
UE_0	9	6	3
UE_1	11	10	7

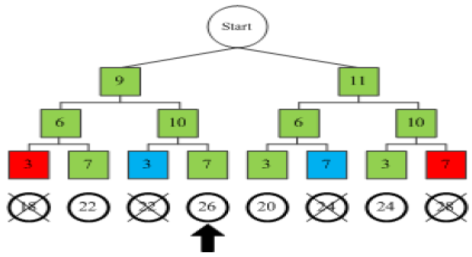

Figure A1.1 FDPS Scheduling Example [38]

M2M Aware Scheduling Algorithm (M2MA-SA) [39]

Algorithm Steps (First Queue)

There are X PRBs available for N UEs (and MTCDs). Let the i^{th} user at a time instant t has an instantaneous data rate r_i, average data rate $\overline{r_i}$ over a time window and target data rate T_i. For the i^{th} user, a satisfaction parameter M_i is defined as the ratio between average data rate $\overline{r_i}$ and target data rate T_i. The set of PRBs assigned to i^{th} user is denoted by F_i. D_i is the waiting delay for data in the buffer of the i^{th} user. TH_i is the service delay threshold of the i^{th} user.

Step 1: Calculate metric λ_{ic} for the i^{th} user at c^{th} PRB as follows,

$$\lambda_{ic}(t) = \frac{r_{ic}(t)}{M_i(t)} \cdot e^{\left(\frac{D_i(t)}{TH_i}\right)} \qquad \text{(A1.5)}$$

Step 2: Select maximum λ_{ic} and start to serve user i by allocating PRB c to it.

Step 3: For user i selected in step 2, check if it has the highest SNR in PRBs adjacent to c^{th} PRB.

- If (condition is true) then allocate adjacent PRBs (right and left) to user i and update F_i and repeat step 3 till the condition is dissatisfied or F_i is sufficient to transmit the data of user i. This is known as Iterative Maximum Expansion (IME).
- Else, Repeat step 2 after considering MTCD i as served and cannot compete for other PRBs.

Step 4: Repeat steps 2 and 3 till all users are served or all PRBs are allocated.

Step 5: If some PRBs remain, allocate them to the users that got the adjacent PRBs if they still need PRBs to fully transmit data.

Algorithm Steps (Second Queue)

There are W PRBs available for Q MTCDs. Let λ_{ic} as the utility for the i^{th} MTCD on PRB c at a time instant t as calculated in the algorithms steps of the first queue. The time is divided in equal sections of length T. Each section is divided into 2 cycles round-robin cycle and maximum utility cycle as shown in Figure A1.2 [39].

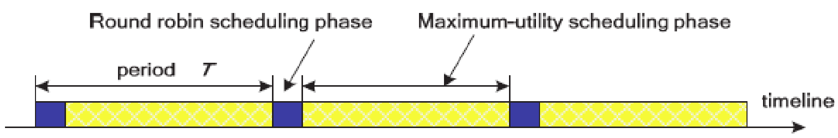

Figure A1.2 M2MA-SA cycles for the second queue [39]

Step 1: Check users about to exceed deadline threshold and put them in a sub-queue called timeout queue.

Step 2: In the round-robin cycle, timeout queue MTCDs are scheduled in a round-robin manner till all queue members are satisfied and the queue is empty.

Step 3: For the remaining period of T period, calculate metric λ_{ic} for the i^{th} user at c^{th} PRB as in (A1.5) and schedule users based on maximum metric.

Step 4: Select maximum λ_{ic} and start to serve MTCD i by allocating PRB c to it.

Step 5: For MTCD i selected in step 4, check if it has the highest SNR in PRBs adjacent to c^{th} PRB.

- If (condition is true) then allocate adjacent PRBs (right and left) to MTCD i and update F_i and repeat step 3 till the condition is dissatisfied or F_i is sufficient to transmit the data of MTCD i.
- Else, Repeat step 4 after considering MTCD i as served and cannot compete for other PRBs.

Step 6: Repeat steps 4 and 5 till all users are served or all PRBs are allocated (or T period ends).

Appendix 2: Generalized M2M Uplinks Scheduling Algorithm Design

The statistical priority-based algorithm can be generalized to combine more M2M scheduling metrics along the guidelines of BAT algorithm proposed in Chapter 3 [7]. There are different scheduling metrics that could be considered in M2M uplink scheduling. It is a design goal to make the algorithm configurable according to network conditions like the availability of radio resources, density of MTCDs deployment and switching/combining different scheduling metrics. The steps of the algorithm can be stated as follows.

Step 1: Assign $p\%$ of PRBs for serving MTCDs. For example, $p\%$ can be calculated as per equation (A2.1).

$$p\% = \min(\frac{W \times M}{U}, 1) \times 100\% \qquad \text{(A2.1)}$$

Step 2: Consider a set of N PRBs and a set of M active MTCDs (active MTCDs are the ones that request data transmission and have not missed deadline yet). Every MTCD m ($m = 1 \ldots M$), has a deadline D_m, a BSR value of B_m, an SP value of SP_m and an SNR value at the n^{th} PRB of S_{mn}.

Step 3: Divide time into equal intervals (x TTIs). Each interval is divided into 2 parts; the channel-state scheduler part (*Step 4a*) for $q\%$ of the time interval and the hybrid scheduler part (*Step 4b*) for the remaining time.

Step 4a: Channel-based scheduler part

- Sort S_{mn} in a descending order to select the MTCD-PRB pair of maximum SNR, i.e. best channel state.
- Assign PRB n to active MTCD m.
- Allocate PRBs on the right and the left of PRB n to MTCD m and keep expanding in both direction till any of the following conditions applies [22]:
  - o MTCD m acquires enough PRBs to send its data.
  - o Expanding in both directions is blocked by PRBs allocated to other MTCDs.
  - o Expanding in both directions is blocked by PRBs at which other MTCDs have higher SNR.
- Consider MTCD m as served and remove the allocated PRBs from the set of available PRBs.
- Repeat *Step 4a* till all PRBs are allocated.
- If some PRBs remain after considering all MTCDs are served, these PRBs are allocated to one of the MTCDs acquiring the adjacent PRBs.

Step 4b: Hybrid scheduler part

- Sort the MTCDs based on the selected scheduling metric in the specified order according to Table A2.1 to select an active MTCD m at the top of the list.
- Assign PRB n (at which the selected MTCD m has maximum SNR S_{mn}) to MTCD m.
- Allocate PRBs on the right and the left of PRB n to MTCD m and keep expanding in both direction till any of the following conditions applies [22]:
  - o MTCD m acquires enough PRBs to send its data.
  - o Expanding in both directions is blocked by PRBs allocated to other MTCDs.

- Consider MTCD *m* as served and remove the allocated PRBs from the set of available PRBs.
- Repeat *Step 4b* till all PRBs are allocated.

To elaborate on hybrid scheduler part, let us assume that A is the event that deadline is enabled as a scheduling metric, B is the event that BSR is enabled as a scheduling metric, C is the event that SP is enabled as a scheduling metric. Hence, there are 8 possible options in *step 4b*.

Table A2.1 Hybrid scheduler options

Case	*A*	*B*	*C*	Scheduling Metric	Sorting Direction
1	0	0	0	None/Default	-
2	0	0	1	Statistical Priority (SP_m)	Descending
3	0	1	0	BSR (B_m)	Descending
4	1	0	0	Deadline (D_m)	Ascending
5	1	0	1	Virtual Deadline (VD_m)	Ascending
6	1	1	0	Urgency (U_m)	Descending
7	0	1	1	Virtual BSR (VB_m)	Descending
8	1	1	1	Virtual Urgency (VU_m)	Descending

We explain some of the cases that are not straightforward in Table A2.1 in the following.

- Case 1: This case may be used to indicate a default setting where there is no scheduling metric and hence all MTCDs have equal priority so selection could be based on random selection, round-robin, or first-come-first-served basis.
- Case 5: In this case, a new metric Virtual Deadline (VD_m) is calculated for every MTCD as follows,

$$VD_m = D_m \times (2 - scaled(SP_m)) \tag{A2.2}$$

where scaled(SP_m) is calculated such that, if SP_m ranges from 0 to sp_max, then scaled(SP_m) ranges from 0 to 2. This scaling can be simply a linear scaling (A2.3) or any non-linear function as per the need,

$$scaled(SP_m) = SP_m \times \frac{2}{sp\_\max} \tag{A2.3}$$

The concept of the new metric is to create a closer virtual deadline when data carried by an MTCD has a high-value information (as explained in Chapter 4) and vice versa for low importance data.

For example, consider 2 MTCDs with deadlines $D_1 = D_2 = 1000$ms and $SP_1 = 7$, $SP_2 = 4$, $sp\_max = 10$. Both MTCDs have same deadline but the first MTCDs carries more important information. By simple calculations we obtain $VD_1 = 600$ms and $VD_2 = 1200$ms and hence the first MTCD is scheduled first.

- Case 6: Urgency (U_m) is calculated as in [36].
- Case 7: In this case, a new metric Virtual BSR (VB_m) is calculated for every MTCD as follows,

$$VB_m = B_m \times scaled(SP_m) \quad \text{(A2.4)}$$

The concept of the new metric is to create a higher virtual BSR value when data carried by an MTCD has a high-value information and vice versa for low importance data. It is worth noting that scaling of *SP* given by equation (A2.3).

- Case 8: In this case, urgency [36] is modified by the scaled value of *SP* to create a Virtual Urgency (VU_m) that is calculated for every MTCD as follows,

$$VU_m = U_m \times scaled(SP_m) \quad \text{(A2.5)}$$

The MTCDs carrying more important data get a higher virtual urgency value and scheduled first.

The following should be noted as we configure the algorithm:

- By increasing the value of *q%* in *Step 3*, the channel-state based scheduler is used for a longer time. Hence, higher throughput could be achieved by favoring MTCDs having better channel conditions. This is suitable for networks that have relaxed deadline constraints or low density of MTCDs. While decreasing *q%* prolongs the time in which hybrid scheduler is used. This results in better performance in terms of less deadline-missing ratio or favoring MTCDs carrying data of higher importance or informative value.
- By setting *q%* = 0%, channel-state scheduler is bypassed.
- In hybrid scheduler part, resource allocation decisions are made based on SP, BSR, deadline requirements, or a combination of them as stated in *Step 4b*.
- Any MTCD that misses deadline is deactivated, except there are data packets in its buffer, till it requests to transmit data.

References

[1] A. Gluhak, S. Krco, M. Nati, D. Pfisterer, N. Mitton and T. Razafindralambo, "A Survey on Facilities for Experimental Internet of Things Research," *IEEE Communications Magazine*, vol. 49, no. 11, pp. 58-67, November 2011

[2] F. Ghavimi and H.-H. Chen, "M2M Communications in 3GPP LTE/LTE-A Networks: Architectures, Service Requirements, Challenges, and Applications," *IEEE Communications Surveys and Tutorials*, vol. 17, no. 2, May 2015

[3] "The M2M Adoption Parometer 2013," Vodafone, 2013

[4] D. Evans, "The Internet of Things, How The Next Evolution of The Internet is Changing Everything," Cisco, White Paper, April 2011

[5] A. G. Gotsis, A. S. Lioumpas and A. Alexiou, "M2M Scheduling over LTE: Challenges and New Perspectives," *IEEE Vehicular Technology Magazine*, vol. 7, no. 3, pp.34-39, September 2012

[6] N. Abu-Ali, A. M. Taha, M. Salah and H. Hassanein, "Uplink Scheduling in LTE and LTE-Advanced: Tutorial, Survey and Evaluation Framework," *IEEE Communications Surveys and Tutorials*, pp. 1-27, September 2013

[7] A. Elhamy and Y. Gadallah, "BAT: A Balanced Alternating Technique for M2M Uplink Scheduling over LTE," *IEEE Vehicular Technology Conference (VTC-Spring)*, May 2015

[8] E. Dahlman, S. Parkvall and J. Sköld, *4G LTE/LTE-Advanced for Mobile Broadband*, 1st Edition. Burlington, MA, Elsevier, 2011

[9] S. Sesia, I. Toufik and M. Baker, *LTE The UMTS Long Term Evolution from Theory to Practice*, 2nd Edition. West Sussex, UK, Wiley, 2011

[10] T. Nakamura, "Proposal for Candidate Radio Interface Technologies for IMT-Advanced Based on LTE Release 10 and Beyond (LTE-Advanced)," 3GPP ITU-R WP5D 3rd Workshop on IMT-Advanced, October 2009

[11] Y. S. Cho, J. Kim, W. Y. Yang and C.-G. Kang, *MIMO-OFDM Wireless Communications with MATLAB*, IEEE Press & Wiley, 2010

[12] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, Cambridge University Press, 2005

[13] "2020: Beyond 4G, Radio Evolution for the Gigabit Experience," Nokia Siemens Networks, White Paper, August 2011

[14] C. Hoymann, W. Chen, J. Montojo, A. Golitschek, C. Koutsimanis and X. Shen, "Relaying Operation in 3GPP LTE: Challenges and Solutions," *IEEE Communications Magazine*, vol. 50, no. 2, pp. 156–162, 2012

[15] D. Lee, H. Seo, B. Clerckx, E. Hardouin, D. Mazzarese, S. Nagata and K. Sayana, "Coordinated Multipoint Transmission and Reception in LTE-advanced: Deployment Scenarios and Operational Challenges," *IEEE Communications Magazine*, vol. 50, no. 2, pp. 148–155, 2012

[16] A. Asadi, Q. Wang and V. Mancuso, "A Survey on Device-to-Device Communication in Cellular Networks," *IEEE Surveys on Communications*, April 2014

[17] M. Z. Shakir, D. L. Perez, M. A. Imran and K. A. Qaraqe, "Small-cells: Capacity, Mobility and Energy Efficiency Perspectives," *International Conference on Communications (ICC) Tutorials*, June 2014

[18] A. Ghosh, T. A. Thomas, M. C. Cudak, R. Ratasuk, P. Moorut, F. W. Vook, T. S. Rappaport, G. R. MacCartney, S. Sun and S. Nie, "Millimeter-Wave Enhanced Local Area Systems: A High-Data-Rate Approach for Future Wireless Networks," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1152–1163, June 2014

[19] E. G. Larsson, O. Edfors, F. Tufvessonm and T. L. Marzetta, "Massive MIMO for Next Generation Wireless Systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–95, February 2014

[20] 3GPP TS136.213 v.10.1.0, "Physical Layer Procedures (Release 10)," April 2011

[21] 3GPP TS136.211 v.11.3.0, "Physical Channels and Modulation (Release 11)," July 2013

[22] L. R. de Temiño, G. Berardinelli, S. Frattasi and P. Mogensen, "Channel-Aware Scheduling Algorithms for SC-FDMA in LTE Uplink," *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, pp. 1-6, September 2008

[23] K. Zheng, F. Hu, W. Wang, W. Xiang and M. Dohler, "Radio Resource Allocation in LTE-Advanced Cellular Networks with M2M Communications," *IEEE Communications Magazine*, vol. 50, no. 7, pp. 184-192, July 2012

[24] M.-H. Li, C.-C. Lin, C.-C. Chuang, and R.-I. Chang, "Error-Bounded Data Compression using Data, Temporal and Spatial Correlations in Wireless Sensor Networks," *IEEE International Conference on Multimedia Information Networking and Security (MINES)*, pp.111-115, November 2010

[25] S. T. Hong and J. W. Chang, "A New Data Filtering Scheme Based on Statistical Data Analysis for Monitoring Systems in Wireless Sensor Networks," *IEEE 13th International Conference on High Performance Computing and Communications (HPCC)*, pp. 635-640, September 2011

[26] F. Ghandour, M. Frikha, and S. Tabbane, "A Fair and Power Saving Uplink Scheduling Scheme for 3GPP LTE Systems," *International Conference on the Network of the Future (NOF), 2011*, pp. 6-9, November 2011

[27] R. Chen, R. Q. Hu, Q. Li and G. Wu, "Energy Efficiency in a Delay Constrained Wireless Network," *IEEE Wireless Communications and Networking Conference (WCNC),* pp. 836-841, April 2013

[28] A. Baid, R. Madan and A. Sampath, "Delay Estimation and Fast Iterative Scheduling Policies for LTE Uplink," *IEEE 10th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt)*, pp. 89-96, May 2012

[29] N. Lee and S. Kwon. 2015. "Uplink QoS Scheduling Algorithm with Delay Estimation for LTE Systems," *Wireless Personal Communications Journal*, vol. 80, no. 3, pp. 1131-1146, February 2015

[30] A. Lioumpas and A. Alexiou, "Uplink Scheduling for Machine-to-Machine Communications in LTE-based Cellular Systems," *IEEE GLOBECOM Workshops*, pp. 353–357, December 2011

[31] S.-Y. Lien, K.-C. Chen and Y. Lin, ''Toward Ubiquitous Massive Accesses in 3GPP Machine-to-Machine Communications,'' *IEEE Communication Magazine*, vol. 49, no. 4, pp. 66–74, April 2011

[32] Y. Hsu, K. Wang and Y. Tseng, "Enhanced Cooperative Access Class Barring and Traffic Adaptive Radio Resource Management for M2M Communications over LTE-A," *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, pp. 1–6, October 2013

[33] J. Brown and J.Y. Khan, "Predictive Resource Allocation in the LTE Uplink for Event Based M2M Applications," *IEEE International Conference on Communications (ICC)*, pp. 95–100, June 2013

[34] I. Abdallah and S. Venkatesan, "A QoE Preserving M2M Aware Hybrid Scheduler for LTE Uplink," *International Conference on Selected Topics in Mobile and Wireless Networking (MoWNeT)*, pp. 127-132, August 2013

[35] M. K. Giluka, N. Rajoria, A. C. Kulkarni, V. Sathya and B. R. Tamma, "Class Based Dynamic Priority Scheduling for Uplink to Support M2M Communications in LTE," *IEEE World Forum on Internet of Things (WF-IoT)*, pp. 313-317, March 2014

[36] N. Afrin, J. Brown and J. Y. Khan, "Performance Analysis of an Enhanced Delay Sensitive LTE Uplink Scheduler for M2M Traffic," *Australasian Telecommunication Networks and Applications Conference (ATNAC)*, pp.154-159, November 2013

[37] 3GPP TS 36.321 V10.9.0 "Medium Access Control (MAC) Protocol Specification (Release 10)," July 2013

[38] S. N. K. Marwat, C. Goerg, T. Weerawardane, A. Timm-Giel and R. Perera, "Impact of Machine-to Machine (M2M) Communications on Disaster Management in Future Mobile Networks," *General Sir John Kotelawala Defence University Annual Symposium*, Sri Lanka, 2012

[39] S. Zhenqi, Y. Haifeng, C. Xuefen and L. Hongxia, "Research on Uplink Scheduling Algorithm of Massive M2H and H2H Services In LTE," *IET International Conference on Information and Communications Technologies (IETICT 2013)*, pp. 365-369, April 2013

[40] S. Yang, Z. Gao, R. Huang and X. Qiu, "A Data Correlation-based Virtual Clustering Algorithm for Wireless Sensor Network," *International Conference on Mobile Ad-hoc and Sensor Networks* (*MSN*), pp. 335-336, December 2011

[41] S. Suthaharan, M. Alzahrani, S. Rajasegarar, C. Leckie and M. Palaniswami, "Labelled Data Collection for Anomaly Detection in Wireless Sensor Networks," *International Conference on*

Intelligent Sensors, Sensor Networks and Information Processing (*ISSNIP*), pp. 269-274, December 2010
[42] "TelosB Mote Platform," Datasheet, Crossbow
[43] https://www.youtube.com/watch?v=aK80efdO7fE
[44] Cisco, "Fundamentals of Digital Video"
[45] 3GPP TS 23.203 V11.6.0, "Policy and Charging Control Architecture (Release 11)," June 2012