



HAL
open science

Machine-to-machine communication congestion mechanism

Ahmad Hani El Fawal

► **To cite this version:**

Ahmad Hani El Fawal. Machine-to-machine communication congestion mechanism. Networking and Internet Architecture [cs.NI]. ENSTA Bretagne - École nationale supérieure de techniques avancées Bretagne, 2018. English. NNT: 2018ENTA0010 . tel-02374816

HAL Id: tel-02374816

<https://theses.hal.science/tel-02374816v1>

Submitted on 21 Nov 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THESE DE DOCTORAT DE

L'ÉCOLE NATIONALE SUPÉRIEURE
DE TECHNIQUES AVANCÉES BRETAGNE
COMME UNIVERSITÉ BRETAGNE LOIRE

ÉCOLE DOCTORALE N°601
*Mathématiques et Sciences et Technologies
de l'Information et de la Communication*
Spécialité : *Télécommunications*
Par

Ahmad Hani EL FAWAL

Machine-to-Machine Communication Congestion Mechanism

Thèse présentée et soutenue à Brest, le 3 Décembre 2018
Unité de recherche : Lab-STICC UMR CNRS 6285

Rapporteurs avant soutenance:

M. Karim ABED MERAIM
Professeur, Université d'Orléans, Rapporteur

M. Vincent VIGNERON
MCF, HDR, Université d'Évry-Val-d'Essonne, Rapporteur

Composition du Jury:

M. Stéphane AZOU
Professeur, École Nationale d'Ingénieurs de Brest, Président du jury

M. Karim ABED MERAIM
Professeur, Université d'Orléans, Rapporteur

M. Vincent VIGNERON
MCF, HDR, Université d'Évry-Val-d'Essonne, Rapporteur

M. Denis HAMAD
Professeur, Université du Littoral-Côte-d'Opale, Examinateur

Mme Roua YOUSSEF
Maître de conférences, Université de Bretagne Occidentale, Examinatrice

M. Ali MANSOUR
Professeur, ENSTA-Bretagne, Directeur de thèse

M. Benoît CLEMENT
Enseignant-Chercheur, HDR, ENSTA Bretagne Brest, Membre invité

**Titre: Mécanisme de congestion en
M2M communication**

Mots clés: LTE-A, LTE-M, NB-IoT, M2M, H2H, CTMC

Résumé : Nos travaux s'orientent vers les problèmes de réseaux sans fil liés à la coexistence des communications machine-to-machine (M2M) et humain-humain (H2H). On souhaite souligner l'impact mutuel entre les trafics M2M et H2H dans un contexte d'Internet des objets (IoT, Internet Of Things) en particulier lors des catastrophes. Les communications M2M, qui devraient connaître une croissance exponentielle dans un avenir proche, constitueront un facteur important pour affecter tous les réseaux mobiles. On prévoit un grand nombre d'appareils M2M qui entraînera inévitablement des problèmes de saturation et aura des impacts remarquables sur les trafics, les services et les applications M2M et H2H. Pour étudier les influences mutuelles M2M et H2H, nous développons un nouveau modèle Markovien à temps continu (CTMC) pour simuler, analyser et mesurer les différentes stratégies d'accès aux réseaux sans fil. Notre modèle nous a permis de contourner certaines limitations des simulateurs professionnels de LTE-A

(Long Term Evolution-Advanced) comme SimuLTE en terme d'un nombre massif d'appareils M2M, une flexibilité de certains paramètres ou pour élaborer plus des outils statistiques. Lors d'un sinistre et suite à un énorme nombre de M2M souhaitant accéder aux réseaux sans-fil, nous avons constaté un épuisement rapide de la bande passante allouée dans les réseaux LTE-M : Long Term Evolution for Machines) ou Narrow Band for IoT (NB-IoT). Pour résoudre ce problème, nous proposons une nouvelle approche appelée Adaptive eNodeB (A-eNB) pour les réseaux LTE-M et NB-IoT. Selon nos simulations, l'A-eNB peut résoudre progressivement le problème de surcharge tout en assurant une satisfaisante qualité de service (QoS) pour le trafic H2H. Avec le concept d'A-eNB, un réseau LTE-M pourra adapter ses ressources pour faire face à une augmentation progressive du nombre de connexions M2M accédant au réseau LTE-M / NB-IoT et en même temps réduire l'impact sur le trafic H2H.

Title: Machine-to-Machine Communication Congestion Mechanism

Keywords : LTE-A, LTE-M, NB-IoT, M2M, H2H, CTMC

Abstract: This Ph.D. work aims to study the Machine-to-Machine (M2M) congestion overload problem and the mutual impact among M2M and Human-to-Human (H2H) traffics in IoT (Internet of Things) environments specifically during disaster events. M2M devices with their expected exponential booming in the near future, will be one of the significant factors to influence all mobile networks. Inevitably, the expected huge number of M2M devices causes saturation problems, and leads to remarkable impacts on both M2M and H2H traffics, services and applications. To study the M2M and H2H mutual influences, we create a new platform model based on Continuous-Time Markov Chain (CTMC) to simulate, analyze and measure radio access strategies due to the limitations of existing Long Term Evolution-Advanced (LTE-A) simulators (i.e, SimuLTE)

in term of massive M2M devices, parameter flexibility and statistical tools. Additionally, during disaster events, a fast bandwidth depletion of the limited bandwidth assigned to M2M devices in Long Term Evolution for Machines (LTE-M) and Narrow Band for IoT (NB-IoT) networks is expected due to the high arrival request of M2M device network access. To address this problem, we propose a new approach named Adaptive eNodeB (A-eNB) for both LTE-M and NB-IoT networks. The A-eNB can solve gradually the overload problem, while keeping the H2H traffic Quality of Service (QoS) not to be affected badly. The network adaptation is provided through a dynamic LTE-M resource reservation aiming to increase the number of M2M connections accessing the LTE-M/NB-IoT network and to decrease the impact on H2H traffic.

Acknowledgments

In the Name of "ALLAH", the Most Gracious, the Most Merciful.

First and foremost I praise and acknowledge "ALLAH", the most beneficent and the most merciful.

Secondly, my humblest gratitude to the Holy Prophet "MOHAMAD" (Peace be upon him) whose way of life has been a continuous guidance for me.

This thesis appears in its current form due to the assistance and guidance of several people. It gives me great pleasure to express my gratitude to all those who supported me and have contributed in making this thesis possible:

- My thesis director Prof. Ali MANSOUR who gave me the opportunity to join the Lab-STICC in ENSTA-Bretagne. His continuous support, motivation and untiring guidance have made this dream come true. His vast knowledge, calm nature and positive criticism motivated me to strive for nice results. I would also express my sincere thanks to his wife "Houwaida", his son "Taha" and his daughter "Mariam" for their hospitality, the warmth they extended to me during different vacations spent in the villages, beaches and mountains. They were my second home in France.
- Dr. Mohamad NAJEM for his valuable suggestions and strong inputs which push my thesis towards its nice endings. I have been extremely lucky to have a friend who cared so much about my work, and who responded to my questions and queries so promptly.
- Prof. Karim ABED MERAÏM, Dr. Vincent VIGNERON for reviewing my thesis and for all their valuable comments that helped me to improve it; they have put my thesis on the right scientific road-map.
- Prof. Stéphane AZOU, Prof. Denis HAMAD, Dr. Roua YOUSSEF as examiners. Great thanks for their comprehensive comments and thoughtful suggestions.

- Dr. Benoit CLEMENT for his useful comments in the CSI (second year) and for attending my defense as an invited member.
- Dr. Denis LE JEUNE, Dr. Frédéric LE ROY, Dr. Jean-Christophe LE LANN, Dr. Joël CHAMPEAU, M. Jean-Yves LE ROUX for their support throughout the Ph.D. years.
- Dr. Mustafa HAMZE for supporting me whenever needed in AUL. Not to forget as well his invaluable embracement in my career.
- AUL ALUMNI president Eng. Abbass SALLOUM who provided me with different assistance and support throughout my journey.
- This moment also reminds me of the influence of my father, my grandfather, my grandmother, my aunt "Raghda" and my father-in-law "Adnan". I wished if they share with me this moment that they and I dreamt of. But, it is God's will that I should respect with full of obedience and homage. May their souls rest in peace forever.
- I would like to acknowledge my mom and my aunt "Fayzeh", my uncle "Zouheir", my uncle "Ghassan" and his wife "Raghda" for their constant support, prayers and best wishes. They uplifted my morale whenever I needed. Thanks goes also to my mother-in-law "Fadia" for her support. She made double efforts to bridge any gap that I left while I was in France.
- I am indebted also to my brother and mentor "Hayssam" for his generous funding, which removed financial concerns from my decision to embark on this long journey. Additionally, he keeps encouraging me to meet tight deadlines during the course of my thesis. I would like also to express my deep appreciation to his wife "Bouchra", my nephew "Karim" and my nieces "Reem", "Leen" and "Sara". They all deserve my warm thanks.
- Deep appreciation to my sister "Hounada", her husband "Mohamad", my nephews "Yaman" and "Rayan" and my niece "Ranim". They have been a never-ending

source of love, encouragement and motivation.

- Last, but not least, my beloved wife "Reem", my daughters "Alaa", "Aya" and "Jana", my son "Mohamad" had to grudgingly learn to accept my separation from them and still gave me nothing but support, day after day, my love and gratitude for them can hardly be expressed in words. Their great patient, love and strong emotional support made my life pleasant even in the hardest times.

Acronyms

3G: 3rd Generation

3GPP: 3rd Generation Partnership Project

4G: 4th Generation

5G: 5th Generation

A-eNB: Adaptive evolved Node B

A-eNB-EMG: Adaptive evolved Node B EMerGency stage

ACB: Access Class Barring

AC: Access Class

APF: Access Probability Factor

BGP: Border Gateway Protocol

BL: Bandwidth Limit

CANAL: Coexistence Analyzer and Network Architecture for Long term evolution

CAPEX: CAPital EXpenditure

Cat: Category

clIoT: clean slate IoT

CMMPP: Coupled Markov Modulated Poisson Processes

CP: Control Plan

CTMC: Continuous-Time Markov Chain

DL: Down-Link

DRX: Discontinuous Reception

DSL: Digital Subscriber Line

DTMC: Discrete Time Markov Chain

EAB: Extended Access Barring

EC-GSM: Extended Coverage Global System for Mobile

EDP: Event-Driven Pattern

eMTC: enhanced Machine Type Communication

eNodeB: evolved Node B

ETSI: European Telecommunications Standards Institute

FDD: Frequency Division Duplex
FIFO: First Input First Output
FTP: File Transfer Protocol
GERAN: GSM/EDGE Radio Access Network
GSM: Global System for Mobile
HARQ: Hybrid Automatic Repeat Request
H2H: Human-to-Human
HP: High Priority
HTC: Human-Type-Communication
IoT: Internet of Things
IP: Internet Protocol
IPP: Interrupted Poisson Process
L3 RN: Layer 3 Relay Node
LENA: LTE Network Simulator
LIFO: Last Input First Output
LoRa: Long Range
LP: Low Priority
LPWAN: Low Power Wide Area Network
LTE: Long Term Evolution
LTE-A: Long Term Evolution-Advanced
LTE-M: Long Term Evolution for Machines
MAE: Mean Absolute Error
MAR: Mobile Autonomous Reporting
MATLAB: MATrix LABoratory
MBB: Mobile Broad-Band
M2M: Machine-to-Machine
M2M-NRT: Machine-to-Machine Non-Real-Time
M2M-RT: Machine-to-Machine Real-Time
MIMO: Multiple Input Multiple Output
MMPP: Markov Modulated Poisson Processes

MTC: Machine Type Communication
NB-IoT: Narrow Band-IoT
OFDMA: Orthogonal Frequency-Division Multiple Access
OMNeT: Objective Modular NeTwork
OPNET: Optimized Network Engineering Tool
OSPF: Open Shortest Path First
PC: Personal Computer
PDCP: Packet Data Convergence Protocol
PPP: Point to Point Protocol
PRACH: Physical Random Access Channel
PRB: Physical Resource Block
PUP: Periodic Update Pattern
QAM: Quadrature Amplitude Modulation
QCU: Queuing Control Unit
QPSK: Quadrature Phase-Shift Keying
QoS: Quality of Service
RA: Random Access
RAC: Resource Allocation Control
RACH: Random Access CHannel
RAN: Radio Access Networks
RAR: Random Access Response
RB: Resource Block
RE: Resource Element
RMSE: Root Mean Squared Error
RN: Relay Node
RRC: Radio Resource Control
SCR: Service Completion Rate
SDT: Small Data Transmission scheme
SENSEI: Making Sense of Human - Human Conversation
SH-GW: Smart Home Gateway

SIM: Subscriber Identity Module
SNR: Signal-to-Noise Ratio
SOOC: Self-Optimizing Overload Control
TCP: Transmission Control Protocol
TCP/IP: Transmission Control Protocol/Internet Protocol
UDP: User Datagram Protocol
UE: User Equipment
UL: Up-Link
VoIP: Voice over Internet Protocol
VNI: Visual Networking Index
WiFi: Wireless Fidelity
WiMAX: Worldwide Interoperability for Microwave Access

Contents

Abstract	i
Acknowledgments	iii
Acronyms	vi
Contents	xi
List of Tables	xvii
List of Figures	xvii
1 General Introduction	1
1.1 Internet of Things (IoT)	2
1.1.1 Background	2
1.1.2 IoT requirements	4
1.1.3 IoT service enablers	4
1.2 Thesis components	6
1.2.1 Research challenges and thesis motivations	6
1.2.2 Thesis objectives	8
1.2.3 Thesis contributions	9
1.2.4 Thesis organization	11
2 Machine-to-Machine (M2M)	13
2.1 Introduction	14
2.2 M2M communication system	15
2.2.1 M2M device domain	16
2.2.2 Network domain	17
2.2.3 Application domain	17

2.3	M2M applications	17
2.4	M2M traffic characteristics and challenges	20
2.5	M2M congestion	21
2.6	M2M state-of-the-art solutions	23
2.7	Conclusion	32
3	RACH Overload Congestion Mechanism for M2M in LTE-A	33
3.1	Introduction	35
3.2	LTE-A overview	36
3.2.1	LTE-A data-rate	36
3.2.2	LTE-A access methods	38
3.3	RACH procedure overload problem	40
3.3.1	RACH procedures	42
3.3.2	Contention-based RACH procedure	43
3.3.3	RACH overload control mechanisms in the literature	44
3.3.4	Comparative analysis of RACH overload control mechanisms	46
3.4	Simulations, results and discussions	48
3.5	Conclusion	51
4	Adaptive Evolved Node B (A-eNB) in LTE-M	53
4.1	Introduction	54
4.2	LTE-M networks	55
4.2.1	LTE-M overview	55
4.2.2	LTE-M Data-rate	57
4.2.3	LTE-M bandwidth limitation (case study)	57
4.3	Literature review	60
4.4	Adaptive eNodeB for LTE-M	61
4.5	Simulation scenarios	65
4.6	Results and discussions	66
4.6.1	Regular eNodeB	66
4.6.2	A-eNB Initial stage	67

4.6.3	A-eNB Emergency stages	69
4.6.4	A-eNB Predictive emergency scenarios	70
4.7	Conclusion	71
5	CTMC modeling for H2H/M2M coexistence in LTE-A/LTE-M	73
5.1	Introduction	74
5.2	Continuous-Time Markov Chain (CTMC)	75
5.3	Existing solutions found in the literature	76
5.4	Coexistence Analyzer and Network Architecture for Long term evolu- tion (CANAL)	79
5.4.1	Queuing Control Unit (QCU)	80
5.4.2	Resource Allocation Control (RAC) unit	83
5.5	CTMC analytical methodology	84
5.5.1	Representing the system as a set of states	85
5.5.2	Generating the equilibrium equations	87
5.5.3	Linear system solution	89
5.5.4	Performance and metrics	90
5.6	Simulations, results and discussions	91
5.6.1	M/M/1 Queuing Model	91
5.6.2	Single traffic simulations and results	92
5.6.3	Basic simulations and results	96
5.6.4	M/M/1 results	98
5.7	Conclusion	99
6	CTMC modeling for M2M/H2H coexistence in an NB-IoT Adaptive eNodeB	101
6.1	Introduction	102
6.2	NB-IoT networks	104
6.2.1	NB-IoT overview	104
6.2.2	NB-IoT Data-rate	105
6.2.3	NB-IoT bandwidth limitations	106

6.3	Existing solutions found in the literature	107
6.4	Adaptive eNodeB for NB-IoT	109
6.5	CTMC analytical model	112
6.5.1	CTMC model for one traffic	113
6.5.2	CTMC model for M2M/H2H traffics	116
6.6	Simulation scenarios and Result discussions	119
6.6.1	Case-study:	119
6.6.2	Model accuracy	119
6.6.3	Simulation parameters:	120
6.6.4	Comparison and result discussions:	121
6.6.5	Worst-case scenario:	123
6.6.6	M2M and H2H coexistence scenarios:	123
6.7	Conclusion	126
7	Heterogeneous traffic modeling for M2M traffic with H2H coexistence	129
7.1	Introduction	130
7.2	Traffic modeling	132
7.3	Markov Modulated Poisson Processes Model	136
7.4	Case-study	139
7.5	Model and metrics	140
7.5.1	CANAL Model	140
7.5.2	Performance and metrics	142
7.6	Simulations	142
7.6.1	Simulator	142
7.6.2	Regular eNodeB scenarios, results and discussions	143
7.6.3	Adaptive eNodeB scenarios, results and discussions	146
7.7	Conclusion	148
8	General conclusion and perspectives	151

I	References	155
	List of Publications	158
	Bibliography	159
II	Appendix	169
A	OMNeT++	171
	A.1 Introduction	171
	A.2 Installation	172
	A.3 OMNeT++ components	173
	A.4 Simulation Launcher	174
	A.5 User interface	174
	A.6 Tic-Toc example	175
B	INET	179
	B.1 Introduction	179
	B.2 Installation	180
	B.3 Practical example in INET	181
C	SimuLTE	187
	C.1 Introduction	187
	C.2 Installation	187
	C.3 Single Cell example	188
	C.4 ENSTAUL model	191
D	CTMC analysis code in MATLAB	205
	D.1 CTMC analytical methodology	205
	D.2 Matlab code	206
E	M/M/1 Queuing Model in SimuLink	215
	E.1 Overview	215
	E.2 Structure	215

E.3 Results and Displays 216

List of Tables

1.1	Comparative analysis for 3GPP Cellular IoT and Non 3GPP technologies	6
2.1	Maximum number of UEs.	25
3.1	Bandwidth notations for LTE-A bandwidth.	38
3.2	Simulation parameters.	49
3.3	LTE-A traffic Models used in two different platforms.	49
3.4	Maximum number of M2M devices.	50
4.1	H2H traffic models.	66
5.1	Description of notations used in CANAL architecture.	81
5.2	Description of notations used in the analytical methodology.	86
5.3	Emergency scenario results.	98
6.1	Case-study M2M storms.	107
6.2	MAE and RMSE for analytical and simulation results	120
7.1	The adaptivity of an A-eNB in normal cases	147
7.2	The adaptivity of an A-eNB in emergency cases	148
C.1	Simulation Parameters	193
C.2	Traffic Models	194

List of Figures

2-1	M2M communication system model	16
2-2	M2M applications.	18
2-3	M2M architecture for wireless connectivity in e-Health scenarios.	19
3-1	RE, RB and PRB in LTE-A time-frequency resources.	37
3-2	RA slot in a LTE-A frame.	41
3-3	Contention-free RACH procedure.	42
3-4	Contention-based RACH procedure.	44
4-1	Bandwidth Limitation for LTE-M carrier within LTE-A carrier.	58
4-2	Flow chart for LTE-M Adaptive eNode-B over normal cycle	62
4-3	Initial Stage.	63
4-4	Adaptive bandwidth during emergency stages	64
4-5	Impact of M2M traffic on H2H traffic in a LTE-A legacy scenario for different traffics.	67
4-6	FTP-DL improvement in LTE-M comparing to LTE-A	68
4-7	Improvement on FTP-DL in a LTE-M	68
4-8	Improvement on M2M traffic in emergency stage1	69
4-9	M2M saturation points comparing to different A-eNB stages	71
5-1	CANAL Architecture.	80
5-2	H2H/M2M HP and LP Queuing.	82
5-3	RAC phases for CANAL architecture.	84
5-4	Generic CTMC Model.	87

5-5	$\pi_{(i)}$ result comparison for $c = 6$	93
5-6	$\pi_{(i)}$ result comparison for $c = 25$ RB	95
5-7	SCR_{H2H} vs SCR_{M2M} in dense area cases.	96
5-8	SCR_{H2H} vs SCR_{M2M} in emergency cases.	98
6-1	NB-IoT deployment.	105
6-2	Bandwidth Limitation (BL) for NB-IoT within LTE-A.	106
6-3	Flow chart for NB-IoT Adaptive eNode-B over a normal cycle.	110
6-4	Adaptive bandwidth during an emergency scenario.	111
6-5	Simple CTMC states.	114
6-6	Generic CTMC states	117
6-7	Overload problem in a regular eNodeB during emergency scenario.	121
6-8	Resolving M2M storms by an A-eNB during an emergency scenario.	122
6-9	$M2M-S_c$ vs $H2H-S_c$ in Normal cases.	124
6-10	$M2M-S_c$ vs $H2H-S_c$ in Dense area cases.	125
6-11	$M2M-S_c$ vs $H2H-S_c$ Emergency cases.	126
7-1	The expected arrival rate over time for non-synchronized and synchronized M2M traffics.	134
7-2	The interpretation as MMPP for a M2M traffic.	137
7-3	Generic MMPP model	137
7-4	Regular eNodeB in a normal scenario.	144
7-5	Regular eNodeB in a emergency scenario.	146
A-1	The main window of OMNET++ user interface.	175
A-2	Tic-Toc with two nodes.	176
A-3	Tic-Toc with six nodes.	177
B-1	Wireless network in INET Framework.	182
B-2	The NED source of a wireless network in INET Framework.	184
C-1	Single cell configuration.	189
C-2	Single cell traffic.	190

C-3 Single cell graphical run-time environment. 190
C-4 ENSTAUL model 192
E-1 M/M/1 Queuing System 217

Chapter 1

General Introduction

Glossary

3GPP: 3rd Generation Partnership Project

5G: 5th Generation

A-eNB: Adaptive evolved Node B

CAPEX: CAPital EXpenditure

CANAL: Coexistence Analyzer and Network Architecture for Long term evolution

CTMC: Continuous-Time Markov Chain

EC-GSM: Extended Coverage Global System for Mobile

eMTC: enhanced Machine Type Communication

eNodeB: evolved Node B

GSM: Global System for Mobile

H2H: Human-to-Human

IoT: Internet of Things

LoRa: Long Range

LPWAN: Low Power Wide Area Network

LTE: Long Term Evolution

LTE-A: Long Term Evolution-Advanced

LTE-M: Long Term Evolution for Machines

M2M: Machine-to-Machine

MBB: Mobile Broad-Band
MMPP: Markov Modulated Poisson Processes
NB-IoT: Narrow Band-IoT
PC: Personal Computer
PRB: Physical Resource Block
QoS: Quality of Service
RACH: Random Access CHannel
WiFi: Wireless Fidelity
WiMAX: Worldwide Interoperability for Microwave Access

Our main objective in this thesis is seeking for a comprehensive resolution for Machine-to-Machine (M2M) congestion that drives to overload problems in an Internet of Things (IoT) world full of sensors and actuators. In the next section, an IoT overview is introduced in order to build an essential background for our thesis. Meanwhile at the end of this chapter, we focus on the thesis itself in term of structure, objectives and contributions.

1.1 Internet of Things (IoT)

1.1.1 Background

In [1], an appropriate definition of IoT that might be a good starting point: "a world where physical objects are seamlessly integrated into the information network and where the physical objects become active participants in business processes". Nowadays, the soar of technology especially in the telecommunication field has resulted in an unprecedented techno-civilization among all day-to-day tasks, duties and missions. It is true that the "*Internet*" has evolved the human lives ultimately, but then again the new adopted technology IoT has promoted dummy machines to smart, autonomous and interactive machines. Thus, new types of applications are on the rise to integrate smart house appliances, electric vehicles, security equipment, energy-saving devices, automation tools, telecommunication components and computers into

a single ecosystem with a shared user interface. This ecosystem aims collecting machines payloads. Then, by analyzing the accumulated data in a smart system, these machines become smart enough to control our surroundings autonomously which expects to result significant improvements in various human lives (*e.g.*, civil, industrial, agriculture, education, health care, *etc.*).

As a complementary step, IoT technology is carried by different available wireless and cellular networks (*e.g.*, Global System for Mobile (GSM), Worldwide Interoperability for Microwave Access (WiMAX), Zigbee, Wireless Fidelity (WiFi), Long Range (LoRa), SigFox, Long Term Evolution-Advanced (LTE-A), Long Term Evolution for Machines (LTE-M), Narrow Band-IoT (NB-IoT), Extended Coverage Global System for Mobile (EC-GSM), *etc.*). The more the network provides ubiquitous coverage, secure connectivity, data-rate efficiency and mobility, the more the network earns competitive advantages to attract IoT services (*e.g.*, smart cities, smart grid, smart transportation and mobility, smart home, smart buildings and infrastructure, smart manufacturing, smart health, food and water tracking and security, *etc.*).

Financially, the new era of communications "IoT" is expected to provide a distinct revenue of 3 Trillion USD in 2025 [2] as an eye-catching benefit for mobile operators to host such substantial opportunity. However, it is not just a case of rising tides lifting all boats. Actually, it becomes more than necessary to take the advantage of this opportunity provided by IoT to accommodate its traffic into existing networks with minimal CAPital EXpenditure (CAPEX) [3]. Dealing with challenges (*e.g.*, supporting a massive number of devices, bandwidth limitation, dealing with emergency cases, Human-to-Human (H2H) and M2M coexistence problems, overload problems, M2M congestion *etc.*) while providing the lowest CAPEX along with an effective and sustainable solutions become hot topics for the academia, researchers and operators. This thesis seeks for comprehensive resolutions and suggestions for most of the aforementioned challenges.

At first let us have a close look on IoT requirements that should be tightly met by any innovative technology to reach their goals, as explained the next section.

1.1.2 IoT requirements

Standardisation is a key-enabler for any wireless or cellular network competing to host IoT services. Also it might help any new-born technology to scale to IoT objectives.

Based on Nokia Networks white paper in [4], the key requirements of a cellular IoT to enable its services are:

- Long battery life: Industries aim to achieve a minimum of 10 years of battery operation for simple daily connectivity with a small amount of data exchanged.
- Low device cost: Industries target a module cost of less than 5 USD for a single communication module.
- Extended coverage: The target for the IoT connectivity link budget is an enhancement of 15-20 dB. This coverage enhancement would typically be equivalent to the signal penetrating a wall or floor, enabling deeper indoor coverage.
- Support for a massive number of devices: Supporting more than 52K M2M devices connected to a single base station [5]

All previous requirements should be respected in IoT proxy networks in order to relay IoT data. The evolutionary wireless communications can do the trick as the best IoT service enablers. The standardised Low Power Wide Area Networks (LPWAN) technologies and 3GPP solutions possess several characteristics that make them particularly attractive for IoT. But, which solution live up to IoT requirements as best IoT service enabler? This open question is discussed in the next section.

1.1.3 IoT service enablers

In the upcoming years, a ubiquitous coverage, a long battery life, a low device cost, a low deployment cost and a massive number of devices support are mandatory to satisfy IoT requirements, as mentioned in section (1.1.2).

To reach this goal, the 3rd Generation Partnership Project (3GPP) adopts its legacy LTE-A technology by proposing an enhanced version called "LTE-A Pro" supporting the narrow-band machine type communications in two featured technologies [6]:

- 1) "enhanced Machine Type Communication" (eMTC) also known as LTE-M, which reduces the bandwidth to 1.4 MHz using only 6 Physical Resource Blocks (PRB¹).
- 2) NB-IoT which reduces the bandwidth to 180 KHz using only 1 PRB.

These two solutions adopt all their specifications based on IoT requirements, trying to seize the maximum portion of the market.

"LTE-A Pro" new technologies have competitive advantages of using the same hardware and a share spectrum by making them compatible with the legacy LTE-A.

Additionally, as mentioned in section (1.1.2), supporting a massive number of devices will be one of the key requirements for the new innovative IoT solutions called "LPWAN solutions" [7] characterized by its low-rate and long-range transmission. On one hand, LPWAN (*e.g.*, LoRa, SigFox, *etc.*) are expected to play a significant role in smart cities especially when it comes to the massive number of connected devices:

- SigFox is able to connect around 1 million devices per base station [8].
- LoRa is able to connect a large number of devices (*e.g.*, 62K devices using a SX1301 gateway [9]).

But, on the other hand, the forthcoming 3GPP cellular IoT technologies (NB-IoT, LTE-M and EC-GSM) are striving to share the market with the legacy LTE network in order to reach the total potential volume of 20 billion of communicating things by 2020 [10].

Reusing LTE for narrowband IoT solutions takes advantage of LPWAN solution because it is possible to reuse the same hardware and share spectrum by making LTE-M and NB-IoT compatible with LTE, without running into coexistence. In 2020, H2H communications will use several Gigabytes of mobile broadband data per day. By contrast, connected things (M2M communications) may use hundreds of kbytes

¹A Physical Resource Block (PRB) represents the minimal unit that can be scheduled for a User Equipment (UE) to send or receive data.

per day on average. Consequently, IoT traffic is expected to consume only a small amount of the mobile broadband data. Furthermore, most of the IoT traffic will not follow the same amount of data consumption as H2H communications. Therefore, deploying LTE-M and NB-IoT is as simple as a software upgrade to enable a full IoT network with significantly better coverage than the legacy LTE network [7]. In Table 1.1, the result of a comparative analysis for 3GPP Cellular IoT and Non 3GPP is shown.

Table 1.1: Comparative analysis for 3GPP Cellular IoT and Non 3GPP technologies [7, 8, 11, 12].

	Spectrum [MHz]	Bandwidth [KHz]	Max. Data Rate [Kbps]	Max. Range [Km]	Capacity [K device]	Deployment / Infrastructure	Module Price [\$]
LTE-M	LTE (in-band)	1400	1000 (DL/UL)	100	80 (4 DL/UL per day)	LTE carrier	[7-10]
NB-IoT	LTE (in-band) LTE (Guard-band) GSM band	180	170 DL & 250 UL	35	5 (4 DL/UL per day)	GSM or LTE carriers	5
LoRa	169/433/915 (USA) 868 (Europe)	[125-250]	50 DL/UL	[3-15]	e.g. SX1301: 62 (1 packet/hour)	Via LoRa Gateways	[2-5]
SigFox	902 (USA) 868 (Europe)	0.1	0.1 (DL/UL)	[3-50]	1000 per BS	Via SigFox infrastructure	2

Finally, LPWAN vs 3GPP solution competition has just started. On one hand, 3GPP cellular IoT technologies are expected to attract a huge amount of connected devices from the LPWAN if it can offer a better IoT platform that allows customers to scale and manage their business requirements more efficiently. But on the other hand, many challenges are expected to face the deployment of 3GPP solution which opens different research topics, as highlighted in the next section.

1.2 Thesis components

1.2.1 Research challenges and thesis motivations

Since the dawn of mobile networks, most of operator services have been focusing on H2H traffic needs. Integrating M2M traffic into the same network implies dealing with new challenges. These new challenges influence the academia with new research topics that should find appropriate resolutions. In our thesis, we are motivated to study and provide solutions to the following challenges:

1. Supporting a massive number of devices:

In order to support IoT services, a key aspect of cellular radio networks is the accommodation of large number of devices per cell. However, existing cellular networks have been adapted to serve H2H Mobile Broad-Band (MBB) services with few concurrent access opportunities. Hence, it is important to evolve and develop capabilities of the evolved Node B (eNodeB²) to handle the traffic and concurrent connection of IoT devices that belong to different services. In recent studies, researchers have focused on the enhancement of wireless architectures and systems to support connected devices along with existing H2H traffic within networks like LTE-A, LTE-A Pro and future 5th Generation (5G). As a part of this effort, we are eager to study the key challenge to accommodate a huge number of devices per cell from cellular system's access control plane perspective.

2. Studying the influence of M2M traffic on the Quality of Service (QoS) of H2H traffic: LTE-A was coined initially to support H2H communications (*e.g.*, web-browsing, phone calls, internet, televisions, *etc.*). M2M devices (*e.g.*, smart cities, mobile health, smart meters, *etc.*), with their sporadic small packet sizes, pour their payloads into the same LTE-A channel but with different aspects and specifications. Due to these differences, it would be an imperfect match for these two traffics to share the same LTE-A network. With different perspectives and a diversity of applications of M2M and H2H traffics, many challenges are expected as result of this coexistence. In this context, we are motivated to stress the network to its maximum limit and then measure the impact of M2M traffic on H2H traffic that shares the same bandwidth.

3. Bandwidth limitation in emergency cases: Although, M2M devices transmit small-sized packages in different time intervals, but due to their specificity and functionality they send their payloads in form of synchronized storms, unlike traditional H2H communications. Therefore, one of the prominent challenges

²An eNodeB is an access device in the LTE-A network which provides connectivity to a User Equipment (UE).

that will face the mobile operators is the fast saturation of the limited bandwidth of LTE-M with the rise of M2M devices especially in many emergency cases in which all these devices are requesting to send their data simultaneously (*e.g.*, terrorist attacks, tsunamis, power outages, *etc.*). The saturation problem leads inevitably to a remarkable impact on both M2M and H2H traffics, services and applications.

To tackle this drawback, we might seek for an adaptive solution that addresses this problem which can result a "Clean LTE Traffic" free from any M2M congestion.

4. Reducing the mutual influences between H2H and M2M traffics: Based on all previous influence expectations among H2H and M2M traffics, we are motivated to propose a model to study H2H and M2M coexistence through a mathematical framework which can help in studying and analyzing the mutual impact between M2M and H2H traffic coexistence.
5. Heterogeneous M2M traffics: Unlike H2H traffic, M2M traffic is highly homogeneous in most of the cases because M2M traffic uses small chunks of data in a low transmission rate, mostly with predictable communication times and durations [13]. But with M2M synchronization behavior along with diverse applications from different domains that have different payloads, timings and data-rates an accumulative traffic is expected to be received from different sources, which forms an overall heterogeneous traffic. In this thesis, we are keen on addressing the heterogeneity of M2M traffic along with H2H traffic and its consequences on both traffics.

1.2.2 Thesis objectives

The main objective of this dissertation is to study the mutual impact among H2H and M2M traffics while dealing with M2M congestion that leads to different overload problems and network saturations as result of integrating the M2M traffic into existing radio networks.

Our work focuses on analyzing various M2M congestion over 3GPP cellular IoT technologies (LTE-A, LTE-M, NB-IoT) via relevant scenarios where the challenges listed in section (1.2.1) may have a drastic impact on the network performance while proposing possible solutions for each type of challenge.

To recap, based on our preceding discussion, the principal research questions addressed in this dissertation are the followings:

1. First question: How far can the existing cellular networks scale to a surge of M2M devices accessing the network simultaneously?
2. Second question: What will happen to the network which faces bursty M2M traffic along with a limited network bandwidth especially in emergency and disaster events?
3. Third question: What are the impacts of IoT on H2H traffic?
4. Fourth question: What is the threshold number of occurrences for M2M traffic that respect a good balance among both H2H and M2M traffic QoS?
5. Fifth question: How would the network behave with a heterogeneous M2M traffic?

Based on these five questions, we start knuckling down to find relevant answers, as we present in the next section.

1.2.3 Thesis contributions

1. First question's answer: in our conference paper entitled "RACH Overload Congestion Mechanism for M2M Communication in LTE-A: Issues and Approaches" [14], we review M2M communication technology from LTE-A perspective and we outline the random access challenges in high dense areas where the LTE-A network is striving to fulfill the massive number of M2M devices. Moreover, we compare the most common mechanisms found in the literature that deal with the RACH (Random Access Channel) procedure issues and challenges by

analyzing the existing solutions and approaches to avoid RACH overload congestion in the M2M communications. To this end, we have developed different M2M scenarios using SimuLTE³ Modeler to investigate the impact of M2M communications on LTE-A networks in emergency events.

2. Second question's answer: we propose a novel scheme "LTE-M Adaptive eNodeB" (A-eNB) in another conference paper entitled "LTE-M Adaptive eNodeB for Emergency Scenarios" [16].

The A-eNB gradually solves overload problems, while keeping H2H traffic QoS not to be affected badly. Moreover, we adaptively manage network resources to allow both traffic to efficiently access the LTE network via SimuLTE open source modeler. Eventually, an evaluation of the mutual impact of M2M and H2H coexistence is also presented.

3. Third question's answer: in our conference paper entitled "CTMC modeling for M2M/H2H coexistence in a NB-IoT Adaptive eNodeB" [17], a Continuous-Time Markov Chain (CTMC) in a NB-IoT network is proposed helping the H2H/M2M coexistence to become more approachable especially during disaster scenarios.
4. Fourth question's answer: we bridge a research gap by extending the CTMC model as a stochastic process tool in a LTE-M network to characterize the H2H/M2M coexistence based on analytical equations. Afterwards, we validate the proposed model through extensive Monte-Carlo simulations. Eventually, it becomes approachable to characterize the impact of H2H/M2M coexistence in one LTE-A/LTE-M radio resource allocation in dense areas and under disaster scenarios. The CTMC model, simulations and results are published in our Journal article entitled "CTMC modeling for H2H/M2M coexistence in LTE-A/LTE-M networks" [18].

³SimuLTE is a simulation tool enabling complex system level performance-evaluation of LTE and LTE Advanced networks. SimuLTE [15] is an open source project that allows researchers simulate and benchmark their solutions on an easy-to-use framework. It borrows the concept of modularity from OMNeT++ thus it is easy to extend.

5. Fifth question's answer: in another Journal article entitled "M2M Heterogeneous traffic modeling with H2H coexistence during disaster events" [19], we propose a Markov Modulated Poisson Processes (MMPP) framework that models processes whose arrival rates vary randomly over time. In this context, MMPP helps in studying the heterogeneity of M2M traffic (with variable arrival rates) effects along with H2H traffics using Markov chains as a stochastic process tool to characterize the H2H/M2M coexistence based on different analytical equations.

1.2.4 Thesis organization

The rest of the thesis is organized as follows:

- Chapter (2) presents a comprehensive explanation on M2M communication while highlighting on its architecture, applications, traffic characteristics, challenges and congestion.
- In chapter (3), the existing approaches found in the literature which address the impact of M2M devices on LTE-A networks are summarized, then we illustrate this impact in different scenarios by comparing the results in two different platforms (Cluster and Personal Computer (PC) platforms) using SimuLTE modeler during emergency events.
- In chapter (4), we propose a novel scheme called Adaptive evolved Node B (A-eNB) for LTE-M to address the bandwidth limitation challenge. Our proposed A-eNB solves the M2M overload congestion gradually, while keeping the QoS of the H2H traffic within the acceptable standards. The network adaptation is provided through a dynamic LTE-M bandwidth re-allocation with the objective of maximizing the number of M2M connections and minimizing the impact on the H2H traffic.
- In chapter (5), we characterize mathematically the key performance of M2M and H2H communications (*e.g.*, completion rate, network utilization, *etc.*), then we

propose a CTMC model with the aim of studying the mutual impact of both M2M and H2H traffics. Additionally, we created a new framework called Coexistence Analyzer and Network Architecture for Long term evolution (CANAL) in order to measure and analyze the M2M congestion and to test our proposed solution (A-eNB).

- In chapter (6), we have tested our new concept called A-eNB as an extension to the classical eNodeB in NB-IoT networks. Our A-eNB keeps broadening the NB-IoT bandwidth gradually while shrinking LTE-A bandwidth till soaking-up all M2M storms. Moreover, this solution provides an essential resolution of presumable M2M storms especially during disaster and emergency events.
- Chapter (7) addresses the heterogeneity of M2M traffic along with H2H traffic and its consequences. Additionally, a MMPP model is used as a framework to model the system behavior and to study the impact on both H2H and M2M traffics along with the network efficiency.

Chapter 2

Machine-to-Machine (M2M)

Glossary

3G: 3rd Generation

3GPP: 3rd Generation Partnership Project

DSL: Digital Subscriber Line

EC-GSM: Extended Coverage Global System for Mobile

eNodeB: Evolved Node B

ETSI: European Telecommunications Standards Institute

FTP: File Transfer Protocol

H2H: Human-to-Human

IoT: Internet of Things

IP: Internet Protocol

L3 RN: Layer 3 Relay Node

LENA: LTE Network Simulator

LoRa: Long Range

LTE: Long Term Evolution

LTE-A: Long Term Evolution-Advanced

LTE-M: Long Term Evolution for Machines

M2M: Machine-to-Machine

MATLAB: MATrix LABoratory

MTCG: Machine Type Communication Gateway
OPNET: Optimized Network Engineering Tool
PDCP: Packet Data Convergence Protocol
PRACH: Physical Random Access Channel
PRB: Physical Resource Block
QoS: Quality of Service
RAN: Radio Access Networks
RA: Random Access
RN: Relay Node
SENSEI: Making Sense of Human - Human Conversation
SH-GW: Smart Home Gateway
SIM: Subscriber Identity Module
UE: User Equipment
VNI: Visual Networking Index
VoIP: Voice over Internet Protocol
WiFi: Wireless Fidelity
WiMAX: Worldwide Interoperability for Microwave Access

2.1 Introduction

Machine-to-Machine (M2M) communications are new technological genre whereby intelligent devices behave as autonomous agents with their environment without any human intervention. M2M communications could be considered as a landmark in the telecommunication field by doing the trick for a "Techno-chauvinism" future life [20].

Now-a-days, competitive prices are provided by the telecommunication operators that use different standards such as 3rd Generation (3G), Long Term Evolution (LTE), *etc.* to support such M2M paradigm. Additionally, the soar use of Internet Protocol (IP)-devices, such as sensors and actuators, has enabled the booming in M2M market in various fields of our daily lives [21]. Furthermore, using autonomous machines is a double edge sword. On one hand, it facilitates our modern civilization, but on

the other hand, it leads to many potential issues such as traffic congestion, overload problems, security threats, heterogeneous data formats, *etc.* [22].

In this chapter, we tend to present a comprehensive explanation on M2M communications covering:

- M2M communication system model.
- M2M current and future applications.
- M2M traffic characteristics and challenges.
- M2M congestion issues and solutions.

In order to delve more into M2M communications, we need to have a general idea about M2M system and how it works, as explained in the next section.

2.2 M2M communication system

In simple terms, M2M is a technology that enables devices to interact with each other over a wired or wireless communication network without any human intervention. M2M technology usually employs sensors, actuators, transceivers, *etc.*, that acquire data from end devices or events which is then transmitted via a communication network to an application software that processes the received data into meaningful information.

A generic end-to-end architecture of M2M system model consists of three domains as proposed by the European Telecommunications Standards Institute (ETSI) and shown in Figure 2-1:

1. M2M device domain.
2. Network domain.
3. Application domain.

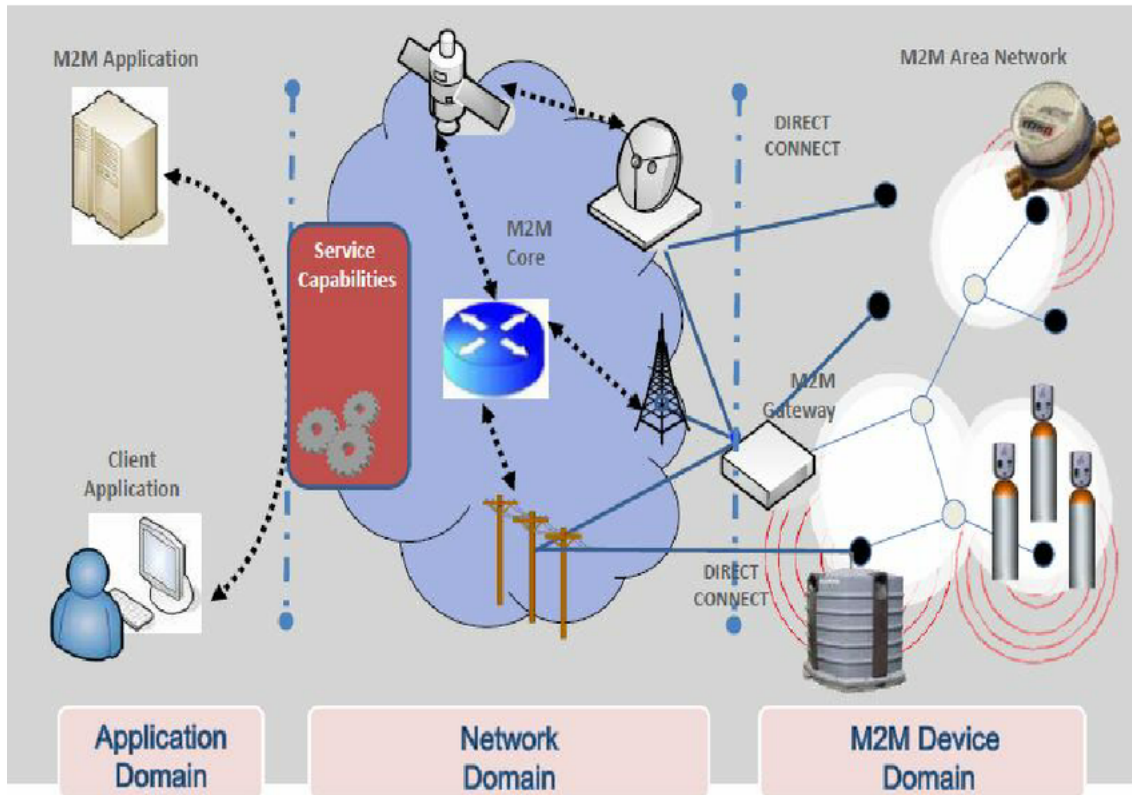


Figure 2-1: M2M communication system model [23].

2.2.1 M2M device domain

The main role of this domain is to submit data from M2M devices¹ to the Network domain.

M2M devices are the system endpoints which can connect to a Network domain either directly using an embedded Subscriber Identity Module (SIM), or via a gateway², which receives, manages and aggregates data (*e.g.*, ZigBee, Bluetooth, *etc.*), then delivers it to the Network domain [21].

¹M2M devices are autonomous devices which submit sensory data as part of an intelligent system that solves a certain problem in real life [23].

²M2M Gateways: Equipment that uses M2M capabilities to ensure M2M devices interconnection to the Network domain [23].

2.2.2 Network domain

It bridges the connection between a M2M device domain and a M2M application domain (*e.g.*, 3G, LTE, Worldwide Interoperability for Microwave Access (WiMAX), Digital Subscriber Line (DSL), *etc.*) by conveying the collected information from the M2M device domain to the M2M application domain [21].

2.2.3 Application domain

M2M applications might be either end users for a specific M2M solution, or application providers who can build more sophisticated M2M solutions and services [23].

Finally, to exemplify the three domains: a M2M device (*e.g.*, sensor, meter, *etc.*) is used to capture an "event" (temperature, inventory level, *etc.*), which is relayed through a network (wireless, wired or hybrid) to an application (software program), that translates the captured event into meaningful information (*i.e.*, items need to be restocked). But, what kind of applications could be served by M2M? and what are the categories that represent most of M2M devices? These questions are discussed in the next section.

2.3 M2M applications

As result of emerging low-power capability along with low-cost M2M devices, a significant demand on M2M applications arises, which enforces many operators to knuckle down to improve our life duties (*e.g.*, clerical work, home tasks, human activities, *etc.*) with various M2M applications that automate these duties without any human intervention [24].

The most famous M2M applications, shown in Figure 2-2, can be categorized as follows:

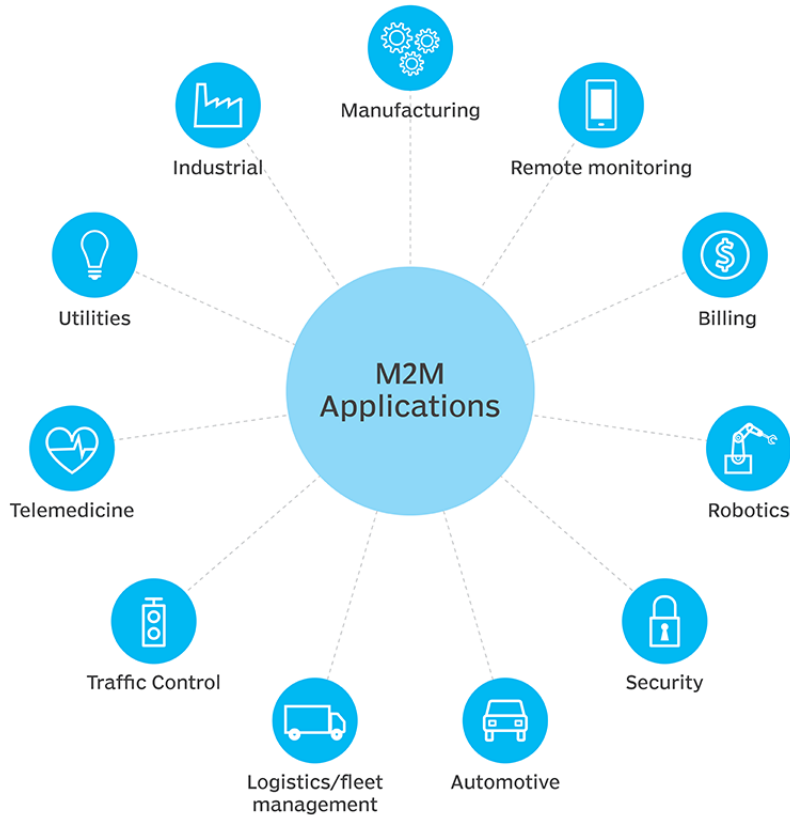


Figure 2-2: M2M applications.

- Security and public safety: Security is one of the most serious concerns for private, commercial and public sectors. Loads of researchers are engaged in security and safety development projects. A great demand for effective security systems makes the M2M communication technology a perfect choice for the simplification and automation of security and public safety monitoring and management. M2M technologies provide cost effective, rapid and flexible deployment for remote surveillance, remote burglar alarms, personal tracking and public infrastructure protection.
- e-Health: Health-care applications might include automatic medical data collection and retrieving, tracking and monitoring of patients and drugs, identification and authentication of patients in hospitals, as shown in Figure 2-3.

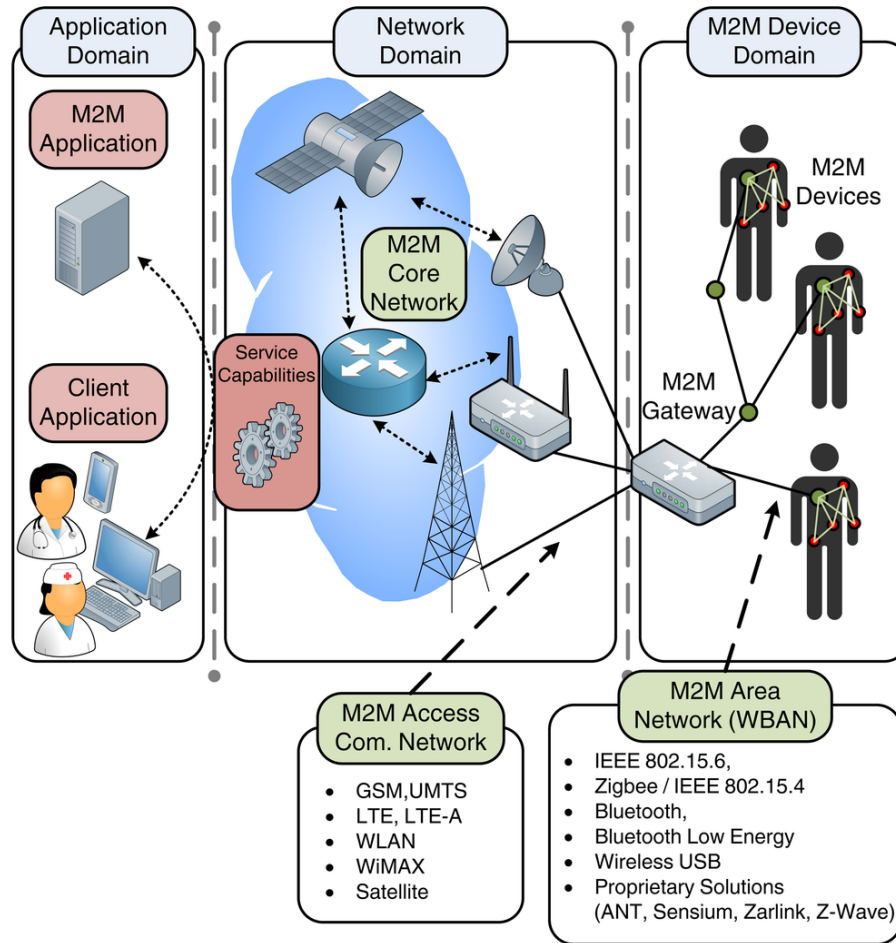


Figure 2-3: M2M architecture for wireless connectivity in e-Health scenarios [25].

Typically, the patient wears one or more M2M sensors which record health indicators (*e.g.*, heart-rate, blood pressure, *etc.*) in order to enable the M2M applications for e-health and to collect the patient's information. Due to many limitations (*i.e.*, battery consumption), it is expected that they require to forward the collected data with some short range technology (*e.g.*, Zigbee, Bluetooth, *etc.*) to a device that can act as a M2M aggregator of the collected information such as a M2M gateway. Then, using one of the network domain operators (*e.g.*, GSM, Long Term Evolution-Advanced (LTE-A), WiMAX, *etc.*), it connects the M2M gateway to the M2M server that stores and possibly reflects to the collected data and subsequently the M2M application user (*i.e.*, health-care remote monitoring).

- Intelligent transportation: Being ahead of the curve in an intelligent transportation, requires vehicles equipped with M2M sensors and actuators that can send valuable information to M2M control centers and transportation companies to route the traffic, monitor the status of the transported goods, track the physical locations of fleet vehicles and deliver updated schedule information to customers (*e.g.*, logistics services, assisted driving, fleet management, e-ticketing, passenger services, smart parking, journey time estimation, *etc.*).
- Smart environment: The ubiquitous use of M2M communications can realize the idea of having a smart environment at homes, offices and every corner of any futuristic smart city. Meanwhile, collecting, managing and manipulating of M2M information can result in making better decisions based on real-time information which reduces significantly living costs, and more efficient utilization of resources in smart homes, offices, shops, smart lighting, agriculture, smart metering, green environment, *etc.*

With these various type of M2M applications, many challenges are expected as result of having different characteristics and data patterns which are explained in the coming section.

2.4 M2M traffic characteristics and challenges

M2M traffic is a new-born traffic which has different characteristics, aims and challenges. In this section, we highlight on some of the important M2M characteristics while envisioning the expected challenges that might be faced and can cause nuanced problems in the Internet of Things (IoT) world.

- a) Access problem: Because of the coordinated behavior of M2M devices, simultaneous transmissions are expected for certain applications. These transmissions imply that a huge number of access requests are sent to the network at the same time causing congestion and overload problems. To solve this problem, enhance-

ments to the channel request procedure is expected to be a real challenge for all operators.

- b) **Bursty traffic:** Most of M2M devices submit their payloads during short periods on a synchronized manner, which forms an accumulative huge traffic in a split second. This bursty traffic hinders data transmission which can cause a fast depletion of available bandwidth throughout certain bottlenecks. For this reason, an adaptive bandwidth and resource allocation becomes a must to address this problem.
- c) **Resource use efficiency:** In many use-cases, M2M devices dispatch small payloads to their related applications. These payloads must be carried by a communication system (*e.g.*, 3G, LTE, WiMAX, *etc.*). Most of these systems are designed to carry huge sizes of data (*e.g.*, phone calls, video-streaming, file exchanges, *etc.*) using big sized resource unit. In the case of M2M traffic, a smaller resource unit should be studied carefully in order to end up with efficient use of resources.
- d) **Traffic priority:** with a diversity of M2M applications, a priority strategy is a must to schedule access permissions especially during emergency situations. To reach this goal, a comprehensive analysis of different traffics should be conducted, and a priority strategy should be studied, tested and implemented to end up with reliable communications regardless of any environment and disastrous events.

Most of the previous challenges are leading to a common symptom: "congestion", which can affect the network performance badly by causing overload problems. Dealing with congestion problems on different levels is the main topic of the next section.

2.5 M2M congestion

A prediction from networking leader Cisco in its annual Visual Networking Index (VNI) Forecast and Methodology (2016-2021) [26] reveals that the number of devices connected to IP networks will be three times as high as the global population in 2021. There will be 3.5 networked devices per capita by 2021, up from 2.3 networked devices

per capita in 2016. Accelerated in part by the increase in devices and the capabilities of those devices, IP traffic per capita will reach 35 GB per capita by 2021, up from 13 GB per capita in 2016.

As result of this technological acceleration, an exponential increase in M2M traffic is more likely to happen [21]. Consequently, a severe data capacity becomes a serious challenge for all mobile operators, which are requested to avoid any expected congestion at all costs.

The new game changer, M2M traffic, should find its way in the market smoothly without any obstacle that hinders its progress and can cause a fluctuation in the telecommunication markets. Shielding the networks from congestion by providing appropriate solutions will facilitates the operators task in addressing the aforementioned growth in M2M traffic.

To recall, as we have explained in section 2.2, a network traffic management has three levels, as shown in Figure 2-2:

- **Access level:** The first level of traffic management deals with Radio Access Networks (RAN). Prior to any data exchange between a device and the network, an access request should be accepted in order to establish a communication channel. Having an excessive number of devices imply sending a huge number of access requests which leads to many overload problems resulting in access attempt failures.
- **Network level:** The second level of traffic management is the core network mechanisms. Knowing that any network working as proxy to deliver M2M payloads to the related application has a limited bandwidth. The enormous number of M2M devices expected to send their payloads simultaneously, requires a fast bandwidth depletion which is one of the main operator's concern to scale any network that aims to evolve M2M devices in its future plan.
- **Application level:** The third level of traffic management might be a way of preventing network congestion by giving network operators the opportunity to prioritize M2M operations. In a perfect application level management solution,

a gateway will facilitate network operators prioritize data traffic from different M2M applications and impose network usage policies ensuring customer satisfaction.

Therefore, combining all these three levels of traffic management could be one of the most effective approaches to address congestion and overload problems. But, what are the proposed solutions found in the literature? How researches dealt with the congestion problem? A detailed answer is presented in the coming section.

2.6 M2M state-of-the-art solutions

The presence of a potentially huge number of M2M devices accompanied with intelligent communication systems, create critical communication issues including scalability, heterogeneity, access problems, *etc.* To implement M2M technology in different applications, these challenges need to be addressed to achieve low power, cost efficient, reliable, ubiquitous communication. In this section, we have highlighted some potential communication challenges and their proposed solutions related to M2M communications.

M2M scalability

The authors of [27] stress the fact that LTE networks were designed to fulfill Human-to-Human (H2H) application needs, which are usually broadband applications. In contrast, M2M applications are narrowband applications because they transmit sequences of small packet sizes (*e.g.*, temperature, humidity, location, *etc.*) as irregular bursts. This difference is a potential future challenge. Consequently, the authors of [27] used Optimized Network Engineering Tool (OPNET) modeler to simulate two different scenarios. The first scenario showed that the maximum number of end stations that could be served, regardless of the amount of transmitted data, by a base station using a LTE network without overloading, was about 250 end stations. They proposed to increase the number of cells or sectors in an evolved Node B (eNodeB) to overcome this limitation. Their second scenario showed that the maximum number

of end stations, which can be allowed to download simultaneously a file sized 100 KB from the eNodeB without overloading, was about 180 end stations. The authors propose to change the time interval to send the data irregularly rather than sending it during the whole hour.

In [28], the capacity limits of 4G mobile networks for emerging M2M services was envisioned. On one hand, the authors use OPNET modeler to simulate two different scenarios and they came up with similar results as the ones found in [27]. On the other hand, they suggest two solutions to handle the problem of the limitations. The first solution is to scale the network capacity by building out more cell towers and base stations of smaller cell sizes (*e.g.*, picocell, femtocell, *etc.*) or upgrading the network to the next generation networks (4G) such as LTE and WiMax. Their second solution is a new promising mechanism called "Mobile Data Offloading", where its main idea is to use a complementary network communication technologies (such as IEEE 802.11, Wireless Fidelity (WiFi)-Direct, Zigbee, Bluetooth, *etc.*) to deliver mobile data traffic originally planned for transmission over cellular networks. However, both new solutions still need additional efforts and expenditures from the mobile operators in order to know where the limits are and what amount of M2M traffic can a mobile network handle with respect to specific circumstances.

In [29], the author studies and examines the overload control problem of M2M devices and the performance of network access in the LTE communication network. A LTE Network Simulator (LENA) is used by the author to simulate the LTE core network architecture and to analyze the M2M throughput. The simulation includes: M2M environment, a connection with a large number of User Equipment (UE)s for an eNodeB in order to examine the performance of eNodeB with a large number of mobile devices. The results show that the maximum number of connection entity in an M2M environment is 320 mobile devices.

The previous three articles can be summarized in the following two questions:

- First question: What is the maximum number of UEs that could connect simultaneously without overloading the eNodeB in LTE Networks?

The answer is detailed in Table 2.1.

Table 2.1: Maximum number of UEs.

Article	Simulator	UE range	Result
[27]	OPNET	series of: 150,200,250,350	250 UEs
[28]	OPNET	series of: 260,360,500	400 UEs
[29]	LENA	series of: 5,10,20,100,200,250,275,320	320 UEs

- Second question: What is the maximum amount of transmitted data that could be downloaded simultaneously without overloading the eNodeB in LTE Networks?

The results show that using OPNET simulator about 200 UEs can connect simultaneously to download a 100 KB file from an FTP server.

M2M and H2H mutual impacts

In [30], the impact of integrating the M2M communication was discovered, with their rather low data rates and small packet sizes, on the LTE data traffic such as video, voice and file transfer. If we know that in LTE networks a limited number of Physical Resource Blocks (PRB)s are available (*i.e.*, 6 PRBs for a bandwidth of 1.4 MHz), and if we added that the minimal allocated number of PRBs for one M2M device is one PRB, we conclude that the spectral efficiency of the LTE system could be declined severely in case of emergency like fire or flooding. Consequently, the authors use OPNET modeler to simulate a scenario considering three types of LTE traffic: video, voice and file transfer. The results show two different conclusions: the video and voices do not reveal a considerable effect of the increasing M2M traffic load within the LTE cell. While, the file transfer users suffered from a degradation of the download performance, which causes a considerable delay.

In [31], an analysis of the mobile network behavior integrated with M2M end stations has been conducted, especially during natural disasters and emergency events. The authors use OPNET modeler to simulate a scenario considering two types of users: Voice over Internet Protocol (VoIP) users and file transfer users. The results

show that the performance of voice users remain unaffected by the additional users during an emergency event, while file upload and M2M traffic experience a significant delay (around 4 times higher delay).

In [32], the authors realize that there are two major differences between M2M and H2H communications:

1. M2M devices are "homogeneous" because they behave similarly while running the same application, while H2H traffic are "heterogeneous".
2. M2M traffic is "coordinated" because of the synchronization behavior, while H2H traffic is "uncoordinated".

Therefore, some changes are required on the human based traffic models in order to be applicable in M2M communications, especially with the increasing number of M2M devices and their severe impact on LTE networks. This could lead us to differentiate between File Transfer Protocol (FTP), voice and video users in order to compare the end-to-end delays when using different types of data on LTE networks. To this end, various M2M traffic performance is investigated by analyzing its impact on LTE network simulating different scenarios that consist of a fixed number of FTP, voice and video users (10 for each) and an increasing number of M2M devices (300, 450 and 600 M2M devices) are deployed in the cell. The results of [32] show that the FTP users suffer from a large delay in the file upload time due to low priority comparing to M2M devices. While voice and video users show no major influence because of M2M traffic, as the priority of voice and video users is more than M2M devices. Finally, the authors of [33] proposed to introduce the aggregated traffic modeling approach in the LTE-A model in order to improve the network performance by using a Relay Node (RN). RNs are one of the significant features of LTE-A that could be utilized as intermediate terminals between M2M devices and eNodeBs for data aggregation.

To sum up, if we consider the results of [30] and [31] shown that the performance of voice users remain unaffected by the additional users, while file upload and M2M traffic experience a significant delay around four times higher, and if we continue to [27, 28, 29] in which the authors spot on the maximum number of UEs that could

overload an eNodeB especially when loads of M2M devices are contending to access the network in dense areas. The answers vary as follows: 250 UEs [27], 320 UEs [29] and 400 [28], which show a kind of contradiction in between the aforementioned results. This contradiction needs additional analysis -which will be performed in the coming chapters- according to the different components and various parameters among several scenarios and platforms.

Overload problem mechanisms

In [34], a key functionality was highlighted to manage and control M2M devices overload problem by rejecting messages using back-off timers in the M2M devices as well as long timers for mobility management and prioritization. What is important in this article is the "Conflict Management" approach during a disaster. Supposing that a conflict occurred between a fire control system in an automated home, which sends a signal to turn on some sprinkles, but on the other hand turning sprinkles can cause flood situation which enforces the flood control system to send a signal to shut off the home's water main, which renders the sprinkles useless. The authors of [34] suggested to use Making Sense of Human - Human Conversation (SENSEI)³ project, which can detect and resolve conflicts between operations in a specific framework.

In [35], it was spotted that M2M and H2H devices must perform Random Access (RA) requests using the Physical Random Access Channel (PRACH) prior to any data transmission. Although M2M payload sizes are small, but when a large number of M2M devices try to communicate over the same channel, all devices contend to access the shared radio channels which causes a network overload problem. To this end, a reinforcement learning-based eNodeB selection algorithm is suggested to allow M2M devices in an overlapping area of multiple eNodeBs can choose the eNodeB that maximizes their Quality of Service (QoS) performances. In this case, M2M devices can observe, learn, and adapt the eNodeB selection decision independently. The simulation is made using MATrix LABoratory (MATLAB). The results show that

³Making Sense of Human - Human Conversation (SENSEI) is a European project which aims to realize smart environment for the real world internet and developed a global framework of common interfaces for sensor and actuator.

when two PRBs are available, the congestion can be alleviated, and the throughput is higher compared to that for one PRB. Also, the access probability can be increased without throughput drop. Moreover, by allocating different numbers of PRBs to the M2M devices, the performance can be also controlled. At the end, the authors opened two research topics:

1. By adopting an efficient eNodeB selection mechanism which guarantees QoS and balances the network load. Because unfortunately, until now most of the approaches proposed in the literature do not consider the Machine Type Communication Gateway (MTCG⁴) selection mechanisms while addressing RA overload control.
2. By developing a group management and addressing of M2M devices which could be vital for M2M communication. But the challenging task is to allocate RA-slots for the whole group while reducing signaling overhead.

Finally, the overload problem mechanisms may open many research topics via discussing relevant solutions such as traffic characterizations, heterogeneity/homogeneity patterns, security threats, network accesses, maximum simultaneous connections *etc.*, as we discuss in different sections later such as in section 3.3.

Data aggregation solutions

In [36], the issue of cell-edge users who suffer from a huge degradation throughput due to poor coverage and low Signal-to-Interference Noise Ratio (SINR) was addressed. This issue could be solved by integrating a RN⁵ into LTE-A networks. As M2M devices usually transmit small data at a particular time, the allocation of an entire PRB to a single M2M device degrades the spectral efficiency. In the context of large scale networks, allocating an entire PRB to each device may not be a feasible idea. A proposed solution is to address this problem by multiplexing the data of M2M users at

⁴A MTCG is responsible for providing a suitable path and facilitating local control for M2M communication.

⁵A Relay Node (RN) is a low power devices used for coverage extension of cells.

the Packet Data Convergence Protocol (PDCP) layer of a RN. As a result, the eNodeB sees requests from RNs as a single user request. Therefore, different M2M devices can share a single PRB at a single time to increase the spectral efficiency. Since RN has not been implemented, the authors simulate the aggregation and multiplexing of data for 180 sensors by increasing data transmission rate of a sensor M times and reduce the number of sensors $1/M$ times. The simulation was made using OPNET modeler, by deploying 10 LTE-A users with up-link FTP application along with 90 e-healthcare and 90 logistics devices with a single sensor in the network scenario. E-healthcare devices are given a higher priority comparing to logistic devices while FTP users have the least priority. The results show that it is clearly shown that the aggregation of M2M traffic improves FTP users performance in term of upload time, comparing to an unaggregated data from M2M devices, where FTP user performance is relatively poor. At the end, the authors of [32] conclude that implementing a RN with aggregation and multiplexing capabilities can improve the LTE-A network performance when huge amount of M2M devices are deployed along with H2H users.

In [37], the capability of a PRB to transmit large data in favorable channel conditions has been spotted. Meanwhile, the LTE-A frequency spectrum is a scarce resource and the vendor pays huge investment capital to obtain it. As the M2M device usually transmit small data at a particular time, the allocation of the entire PRB to single M2M device degrades the spectrum efficiency. Since, the number of machines are growing; the idea of allocating whole PRB to each machine can reduce the overall efficiency. Therefore, to enhance spectral efficiency and to ensure the same QoS provision to M2M and normal traffic, architectural changes in LTE/ LTE-A are required. To this end, the authors studied the performance of the LTE-A network connected with a huge number of M2M devices with the integration of a Layer 3 Relay Node (L3 RN⁶) utilized for aggregation and multiplexing of M2M data traffic. A proposed solution is addressed to solve this problem, where the data of the M2M

⁶In a Layer 3 Relay Node (L3 RN), the received signal on the down-link performs the decoding and goes further to the higher layers to process user data by ciphering, combining and encoding again to forward the signal to User Equipment (UE). Using this relay, best throughput enhancement can be achieved but on the other hand, this relay introduces a large processing delay.

users is multiplexed at the PDCP layer of the RN. Consequently, the RN disguises the requests for PRBs from the eNodeB as a single user requesting for resources. Hence, several M2M devices could share a single PRB. Using this strategy, the spectrum efficiency can be increased significantly as compared to normal M2M communication where one PRB is assigned to one M2M user. The simulation is performed using OPNET modeler with different M2M traffic load in the network. 10 FTP users are deployed in all scenarios. However, the number of trailers is varied in the scenarios to evaluate the network performance. The number of trailers with a single sensor under study is 60, 90, 120 and 150 in these scenarios. The results show that the performance of the network is significantly improved for the aggregated M2M data traffic as compared to the unaggregated traffic. The reason of this improvement is that the aggregation of M2M data make ensures the usage of less number of PRBs for the same amount of data transmission. Hence, it ultimately increases the efficiency of the network. Finally, a proposed solution can be tested by implementing a fully functional RN with aggregation and multiplexing capabilities in a LTE-A model using OPNET modeler, which can help in improving the LTE-A network performance when huge amount of M2M devices are deployed along with LTE-A users.

In [38], the expected large number of small sized M2M messages was highlighted, as a result of the huge number of devices that are contending to access a LTE-A network simultaneously. Existing mobile standards are neither designed to handle small-sized payloads efficiently nor to support simultaneous access of thousands of devices. Consequently, this leads to network congestion. Therefore, the current mobile standards must be extended in order to support devices which send and receive small-sized data. To address this problem, the authors come up with a new proposed data aggregation scheme for uplink M2M traffic in LTE-A networks. For this purpose, a LTE-A layer 3 in-band RN is used to aggregate uplink M2M traffic by sharing the PRBs among various devices. Therefore, in the proposed data aggregation scheme, small data packets are aggregated at the PDCP layer of the RN in order to maximize the multiplexing gain without aggregating the additional headers such as those from the PDCP, Radio Link Control (RLC) and Medium Access Control (MAC).

The simulation is constructed using OPNET modeler where 200 M2M devices are placed in the proximity of the RN. The number of M2M devices is incremented by 200. The simulation results show the efficient utilization of PRBs in uplink with the proposed M2M data aggregation. The results show also that without multiplexing the RN serves nearly 2400 devices with 5 PRBs in uplink. However, in case of multiplexing, the number of devices served by the RN nearly doubles.

In [39], an important question regarding M2M communication was highlighted on how to deal with the differences between H2H and M2M traffic patterns. As M2M devices generate various types of traffic patterns (*e.g.*, periodic, event-driven, multimedia streaming patterns, *etc.*) depending on the application. Traffic is generated by a massive number of nodes and in comparison with H2H communication, M2M applications are expected to generate a diverse range of services, including narrow-band applications which will be transmitting data infrequently. In a worst scenario, when thousands of the M2M devices attempt to send some information at the same time, the overloading of the accessing gateway (*i.e.*, base station in mobile network) occurs. When the various data traffics from a large number of sensors are gathered through a network, an unpredictable pattern is created at the core network. In case the sensors will send data separately to the remote destination node (without any aggregation point), the eNodeB will be overloaded. For that reason, it is necessary to aggregate data from individual sensors inside the home gateway. The data aggregation performed using a new Smart Home Gateway (SH-GW) which brings a significant reduction on network traffic comparing to the situation when the sensors are communicating independently. The M2M data is sent through the mobile network less frequently (sending period depends on the aggregation scheme), but more effectively. The implementation of this approach is feasible due to less-strict requirements on a delay in M2M services. The new SH-GW should be able to provide some type of data aggregation scheme which results in the sending of M2M data through a mobile or fixed network more effectively. The need for aggregation scheme is given due to the differences between H2H and M2M communications. In comparison with H2H communication, the M2M traffic follows some specific patterns: relatively small

amount of data per transmitted packet; large number of transmitted messages. Since current mobile networks are primarily designed to fulfill the H2H type of traffic, the aggregation scheme is used for shaping M2M traffic as close as possible to H2H traffic. The data aggregation technique minimizes the number of accesses to the cellular network and does not overload the base station disproportionately.

2.7 Conclusion

In this chapter, we spot on the general M2M communication system model while explaining its three domains. Then, we list the most famous M2M applications in different sectors that fulfill the significant demand on M2M applications. Moreover, some of the important M2M characteristics are highlighted along with the expected challenges in the future IoT world. Then, the M2M congestion has been explained and the network traffic management proposed on the three levels: access level, network level and application level to address each type of congestion. At last, we review the proposed state-of-the-art solutions found in the literature to address M2M challenges and problems.

To conclude, many challenges are expected to arise as result of diverse M2M data patterns and different behaviors which lead inevitably to network congestion that causes many eNodeB overload problems. Addressing these problems become essential to live up to IoT expectations and objectives. The question now is to find the most appropriate technology(ies) to carry M2M data with minimal drawbacks. Many technologies are competing to be on the top forthcoming communication generation (*e.g.*, Extended Coverage Global System for Mobile (EC-GSM), LTE-A, LTE-M, NB-IoT, LoRa, Sigfox, WiMAX, *etc.*). The winner(s) is/are expected to attract most of the communication market with an unprecedented revenue.

In the next chapters, we will explore the ability of LTE-A, LTE-M and NB-IoT to take balancing acts towards M2M traffic with the coexistence of H2H traffic. Then, we will test the adaptivity of the network to fulfill M2M and H2H requests during disaster events with acceptable limits of performance degradation.

Chapter 3

RACH Overload Congestion Mechanism for M2M in LTE-A

Glossary

3GPP: 3rd Generation Partnership Project

ACB: Access Class Barring

AC: Access Class

APF: Access Probability Factor

BT: Barring Timer

cIoT: clean slate IoT

DL: Down-Link

EAB: Extended Access Barring

EC-GSM: Extended Coverage Global System for Mobile

ETSI: European Telecommunications Standards Institute

eNodeB: evolved Node B

FTP: File Transfer Protocol

H2H: Human-to-Human

HTC: Human-Type-Communication

IoT: Internet of Things

IP: Internet Protocol

LoRa: Long Range
LPWAN: Low Power Wide Area Network
LTE: Long Term Evolution
LTE-A: Long Term Evolution-Advanced
LTE-M: Long Term Evolution for Machines
M2M: Machine-to-Machine
MIMO: Multiple Input Multiple Output
MTC: Machine Type Communication
NB-IoT: Narrow Band-IoT
OFDMA: Orthogonal Frequency-Division Multiple Access
PC: Personal Computer
PRB: Physical Resource Block
QAM: Quadrature Amplitude Modulation
QoS: Quality of Service
RA: Random Access
RACH: Random Access Channel
RAR: Random Access Response
RB: Resource Block
RE: Resource Element
RRC: Radio Resource Control
SOOC: Self-Optimizing Overload Control
UE: User Equipment
UL: Up-Link
VoIP: Voice over Internet Protocol

3.1 Introduction

The dawn of Machine-to-Machine (M2M) or Machine-Type-Communication (MTC), is taking place inevitably in the coming years. During the new technological take-off, a new era of M2M communications is going to rule the new opening business markets (*e.g.*, smart cities, e-health care, logistics, surveillance and security systems, smart metering, in-car satellite navigation systems, *etc.*) [40]. M2M communications will be handled either by the current mobile infrastructure Long Term Evolution-Advanced (LTE-A) networks [35] - in particular with the innovative 3rd Generation Partnership Project (3GPP) cellular Internet of Things (IoT) solutions such as Narrow Band-IoT (NB-IoT), Long Term Evolution for Machines (LTE-M) and Extended Coverage Global System for Mobile (EC-GSM), or by the non 3GPP Low Power Wide Area Network (LPWAN) solutions such as Long Range (LoRa), SigFox, clean slate IoT (cIoT), *etc* [7].

Unfortunately, current LTE-A networks were designed to fulfill the Human-to-Human (H2H) needs (*e.g.*, internet browsing, voice messages, video streaming, *etc.*) in which, huge amount of data and files are downloaded. But on the flip side, M2M communications, with their huge expected number, are mainly found to automate many types of services which require uploading only few bits of information (*e.g.*, temperature, humidity, location, *etc.*) [40].

Moreover, it is highly recommended to develop an innovative M2M/LTE-A approach which supports the network infrastructure in order to accommodate the new M2M application requirements without any sacrifice in the Quality of Service (QoS) of the legacy H2H communications or Human-Type-Communication (HTC) [40].

Needless to say that LTE-A networks should serve the expected massive number of M2M devices contending to access a LTE-A network using RACH (Random Access Channel) procedures [40].

This contention causes a remarkable performance degradation (*e.g.*, huge delay, packet loss, *etc.*) especially when a large number of M2M devices are trying to access the network over the same channel (*e.g.*, alarms triggered by unexpected events, failures

of the power grid, earthquakes, flooding, *etc.*). Consequently, it will lead to a network overload problem [35].

This unavoidable challenge sheds the light on the implementation of RACH procedure in LTE-A networks as a key-point improvement which attracts the research community in order to propose solutions for this potential bottleneck in mobile networks [41]. Moreover, many questions arise trying to investigate the impact of M2M devices communication on LTE-A networks. The maximum number of M2M devices that could be handled by an evolved Node B (eNodeB)¹ is still a challenging point. In this chapter, we summarize the existing approaches found in the literature which address the impact of M2M devices on LTE-A networks, then we illustrate this impact in different scenarios by comparing the results in two different platforms using SimuLTE modeler during emergency events.

3.2 LTE-A overview

3.2.1 LTE-A data-rate

As 3GPP cellular technologies (*e.g.*, LTE-A, NB-IoT, LTE-M, *etc.*) are expected to rule the forthcoming technological world, studying the characteristics, specifications and features of each one becomes a part of any envisioned development for the futuristic telecommunication technologies.

In order to study the LTE-A data-rate², we propose a presentation of the time-frequency resources, as shown in Figure 3-1.

In LTE-A, time-frequency resources are subdivided according to the following structure:

- *Time domain*

In time domain, the largest unit of time is the radio frame (10 ms), which is

¹evolved Node B is an access device in the LTE-A network which provides connectivity to a User Equipment (UE).

²The data-rate indicates the number of bits that can be transmitted in a subframe/TTI (Transmit Time Interval).

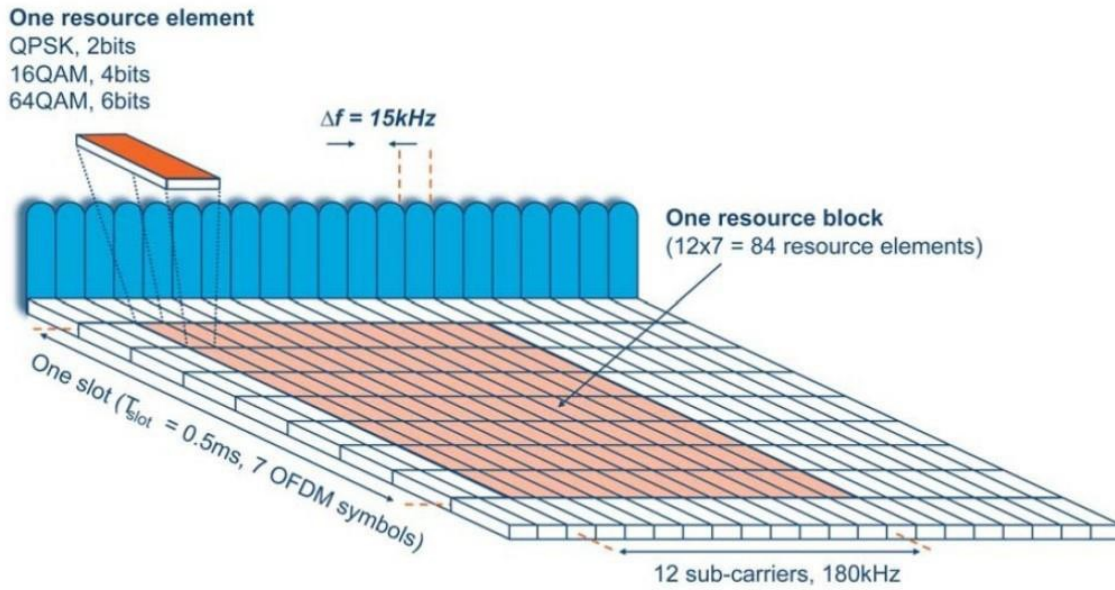


Figure 3-1: RE, RB and PRB in LTE-A time-frequency resources [42].

subdivided into ten subframes (1 ms). Each subframe is split into two slots (0.5 ms). Each slot comprises seven symbols (0.5/7 ms) [43].

- *Frequency domain*

In frequency domain, resources are grouped in units, such that one unit of:

- One sub-carrier³ for a duration of one symbol is termed as a Resource Element (RE) with 15 *KHz* spacing.
- 12 sub-carriers for a duration of one slot is termed a Resource Block (RB) with 180 *KHz* bandwidth.
- 12 sub-carriers for a duration of one sub-frame is termed a Physical Resource Block (PRB) with 180 *KHz* bandwidth.

All notations used in a LTE-A bandwidth are summarized in Table 3.1.

Thus a RB comprises $7 \times 12 = 84$ REs, while a PRB comprises $7 \times 12 \times 2 = 168$ REs.

Moreover, a PRB represents the minimal unit that can be scheduled for a User Equipment (UE) to send or receive data; if we recall that according to [5], it is

³A sub-carrier is a small channel spaced at 15 *KHz* with the adjacent channel.

Table 3.1: Bandwidth notations for LTE-A bandwidth.

Notation	Description
MIMO	Multiple Input Multiple Output
OFDMA	Orthogonal Frequency-Division Multiple Access
PRB	Physical Resource Block
QAM	Quadrature Amplitude Modulation
RB	Resource Block
RE	Resource Element
UE	User Equipment

expected to have more than 52K devices trying to send their payloads simultaneously during a disaster event, that explains the necessity of studying the maximum data rate in classical LTE-A.

In LTE-A, each RE can carry 6 bits in the best modulation scheme 64 Quadrature Amplitude Modulation (QAM). Consequently, the maximum data-rate can be estimated as follows:

$$R_{max} = n \times o \times l \times p \times e \quad (3.1)$$

where; R_{max} : the maximum data rate, n : number of sub-carriers, o : number of Orthogonal Frequency-Division Multiple Access (OFDMA) symbols, l : number of slots, p : number of PRBs, e : number of carried bits per RE.

Assuming 20 MHz of a total bandwidth (with a single antenna) and as one PRB is equivalent to 180KHz, then there are 100 PRBs available with 2 MHz accumulative guard-band⁴ in 20 MHz total bandwidth. So, the maximum data-rate for the 100 PRBs is about 100 kb/ms (100 Mbps approx.). Knowing that with 4x4 Multiple Input Multiple Output (MIMO) antenna the maximum data-rate is about 300 Mbps approx.

3.2.2 LTE-A access methods

In order to gain access to the network resources and according to European Telecommunications Standards Institute (ETSI), there are three different methods of access

⁴For each PRB (180 KHz) we have 20 KHz guard-band which is a narrow frequency range that separates two PRBs.

in LTE-A networks that can M2M devices choose from [41]:

1. *Direct access:*

In this access method, M2M and H2H devices can directly access the LTE-A network via eNodeBs. While this access method is the simplest one (no need for any intermediate device or complicated mechanism), it may lead to an eNodeB overload problem when a huge amount of M2M devices are contending in a limited amount of access resources network [41]. Therefore, a new RACH overload control mechanism is required for the random access procedure in M2M communications to solve this problem [35].

2. *Gateway access:*

A M2M gateway is a dedicated device, added to the network infrastructure, used to provide a suitable path and to facilitate local control for M2M communication [35]. In dense areas, M2M gateways are used to manage the huge volume of M2M devices by relaying data between the eNodeB and its connected M2M devices [41]. However, providing access to M2M devices via gateways worth studying it, especially when researchers are interested in satisfying the QoS requirements for both M2M and H2H devices. Consequently, an efficient M2M gateway selection mechanism is needed while addressing the RACH overload control [35].

3. *Coordinator access:*

In this method, adjacent M2M devices can be grouped before transmission which reduces the redundant signaling and avoids congestions. A M2M member group can be chosen in order to play a role of a temporary M2M gateway which has to collect the data from all members in the same group and delivers it to the eNodeB [41]. In this method, although there is no need for an additional equipment to be added to the network infrastructure but a more complicated mechanism should be designed to select the group coordinator and to manage group members' requests. Moreover, the challenging task is to develop an

adaptive algorithm for preamble allocation which improves the overall network performance (*e.g.*, "Clustering Techniques" presented in [44, 45]).

Regardless of the access method used to request an access to the network, consequently the device used to provide this access method (M2M device, M2M gateway or coordinator M2M device), any device will be able to transmit its data after establishing a Radio Resource Control (RRC) connection successfully with an eNodeB in LTE-A networks, which requires a prior allocation of periodic resources dedicated to its Random Access (RA) preamble⁵.

Now, with a huge number of expected M2M devices, the eNodeB should serve loads of RRC connection requests simultaneously. As a result, and by reaching the cut-off point, a RACH procedure overload problem could lead to an unacceptable performance degradation in the LTE-A network. To address this problem, we should have a clear description of the RACH procedure overload problem, as explained in the next section.

3.3 RACH procedure overload problem

The procedure is similar to any UE access procedure. For this purpose, the term UE will be used to represent either M2M device/MTC or H2H device/HTC.

In the frequency domain, each RA slot⁶ consists of six RBs and has 1.08 MHz bandwidth (6 x 180 KHz). In the time domain, the basic duration is equal to 1 ms, as shown in Figure 3-2. Using one of the 64 RA preambles provided by the eNodeB an UE can submit his access request in one RA slot [40].

On one hand, M2M devices transmit only their data in small packet sizes in most cases, but on the other hand, a huge amount of M2M devices are expected to contend in a higher frequency than H2H devices in order to establish data connections, especially the signaling and traffic load spikes caused by a sudden surge of the number of M2M devices trying to access the same eNodeB simultaneously (*e.g.*, a huge number

⁵A Random Access (RA) preamble is an unique signature chosen by the UE from a list of 64 preamble signatures provided by the eNodeB in each LTE-A cell.

⁶RA slot is the allowable time slot for an UE to transmit its access request.

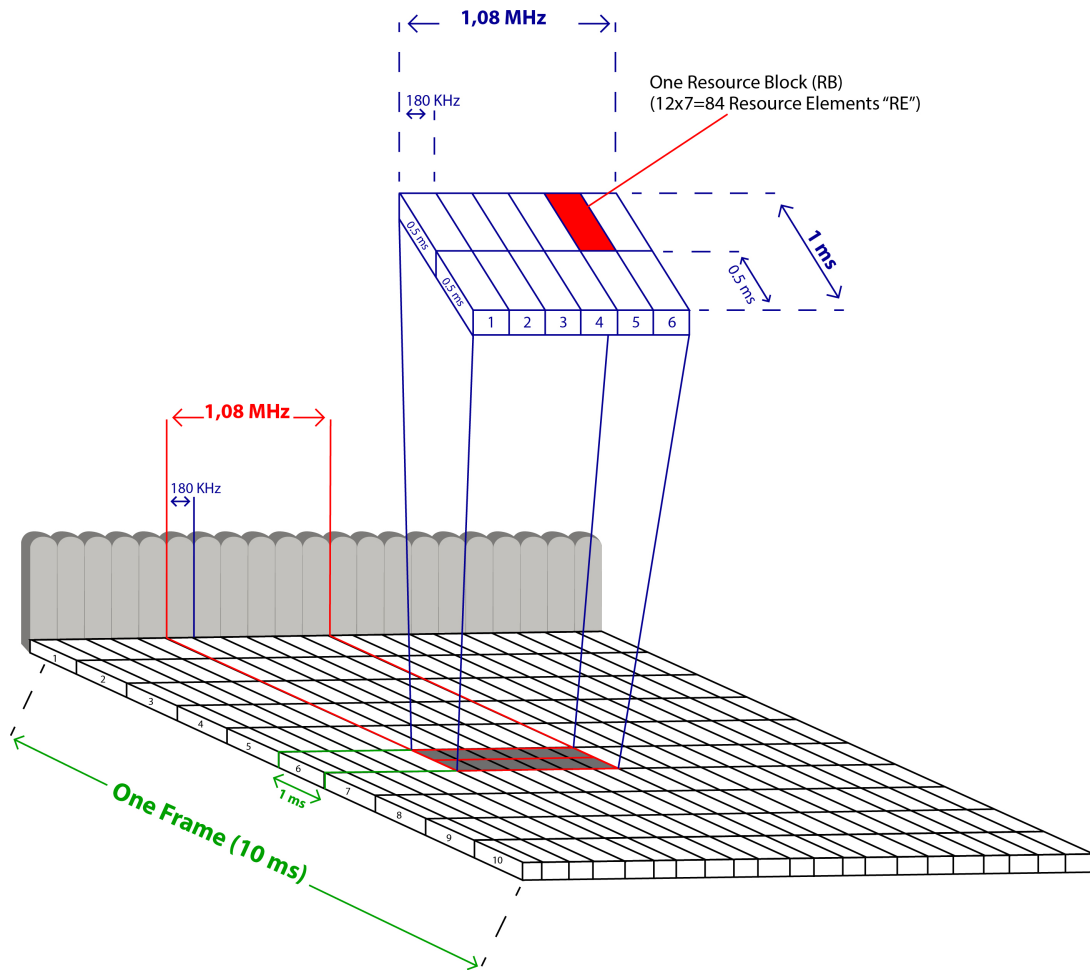


Figure 3-2: RA slot in a LTE-A frame.

of smart meters becoming active simultaneously after power outage), leading to a low random access success rate. In this sticky situation, a high network congestion in the RACH procedure occurs and many problems arise (*e.g.*, extra energy consumption, packet loss, *etc.*) causing in the end a service interruption.

To reduce the load on the RACH procedure, we can increase the number of access opportunities scheduled per frame, but this determines a reduction of the amount of resources available for data transmission.

Summing up, the standard LTE-A procedure for managing channel access requests will not properly scale in the presence of massive access attempts by a large number of UEs. As a result, a sharp degradation of the quality offered to the conventional services arises because of long access delay and high access failure rate [40].

Addressing this issue requires having a close look at the access procedures and how the *Contention-based RACH procedure* is affected by the M2M traffic, as explained in the coming section 3.3.2.

3.3.1 RACH procedures

To transmit packets, an UE performs a random access during an allowable time-frequency slot, called RA slot. The RACH procedure should be initiated in two cases:

- An UE is in "idle mode"; therefore it does not have an uplink radio resources.
- An UE is in "connected mode": either the UE is moving from a previous coverage area to another one during a Handover process [43], or after a radio link failure which requires recovery.

A contention-based or a contention-free RACH procedure starts relatively as soon as one of the two previous cases is detected. The contention-free RACH procedure, shown in Figure 3-3, is under the full control of the eNodeB in order to avoid delayed-constrained access requests with high success requirements, such as those related to Handover [40]. Obviously, no contention is required to be resolved in this procedure, hence, it is not affected by the M2M traffic; Therefore, we will not focus on this case (for more details about the contention-free RACH procedure refer to [35]).

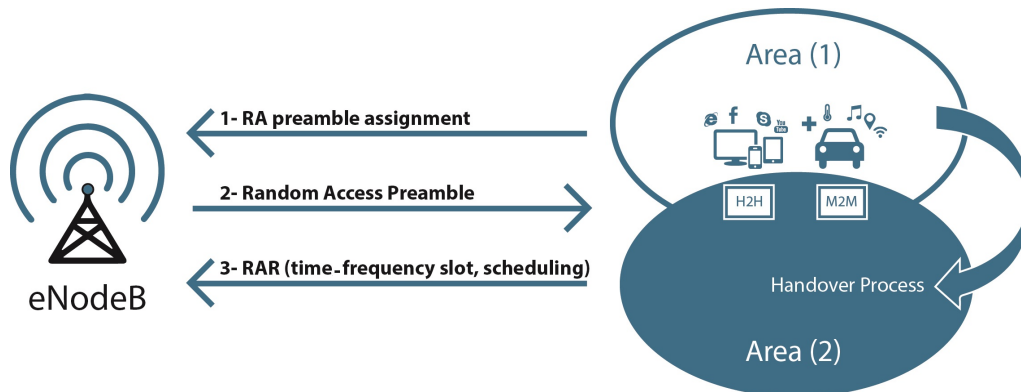


Figure 3-3: Contention-free RACH procedure.

On the flip side, the contention-based RACH procedure is much more susceptible to M2M traffic, as it is discussed in the next section.

3.3.2 Contention-based RACH procedure

In this procedure, an UE initiates a contention-based random access by choosing randomly any RA preambles (less than 64 preambles) initially provided by the eNodeB [40]. Unfortunately, because of the expected huge amount of UEs, it is more likely that more than one UE choose the same RA preamble, which requires a contention resolution procedure to solve this issue [35].

The contention-based RACH procedure, shown in Figure 3-4, consists of the following four steps:

1. *Random Access preamble assignment*: An UE chooses one RA preamble provided by the eNodeB and sends it during the RA slot. When two UEs select the same RA preamble a collision occurs, which requires a contention resolution.
2. *Random Access Response (RAR)*: When an eNodeB receives a RA preamble, it replies with a RAR message containing the appropriate configurations (*e.g.*, time-frequency slot, uplink scheduling, *etc.*) for further communication between the candidate UE and its eNodeB. As soon as an UE receives these configurations, it synchronizes its uplink timing and proceeds to the next step.
3. *Connection request*: After receiving the aforementioned RAR, the candidate UE transmits a request RRC message to the eNode in order to establish a connection.
4. *Contention resolution*: If the eNodeB can decode any request message from the previous step, it replies with an identifier. This identifier can be detected to a unique UE owner which acknowledges the message. Therefore, the connection is established and the UE gains access to the network and transmits its data successfully. The remaining colliding UEs try to access the network by triggering

a new RACH procedure in a second attempt after waiting for a random back-off period [35].

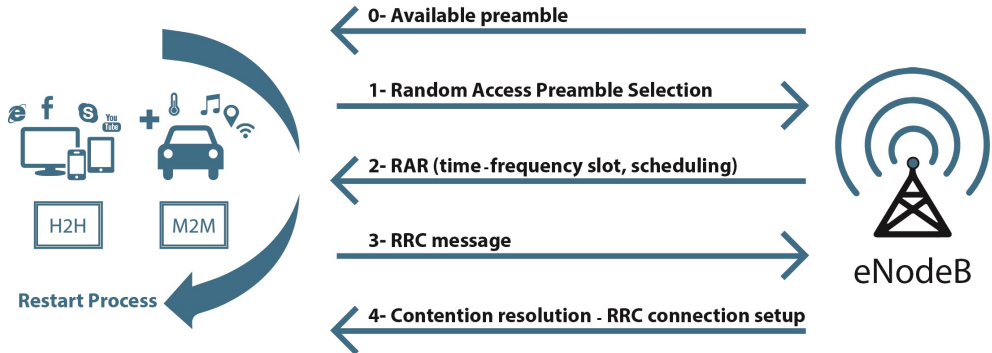


Figure 3-4: Contention-based RACH procedure.

As result of the collisions and their resolutions, many challenges are expected to arise. These challenges and their existing solutions in the literature are outlined in the next section.

3.3.3 RACH overload control mechanisms in the literature

In this section, the different mechanisms to control the RACH overload problem caused by M2M traffic in LTE-A networks are outlined as follows:

1. *Access Class Barring (ACB) scheme*: ACB can define 16 Access Classes (AC); "AC 0" to "AC 9" represents normal device, while "AC 10" represents an emergency call, and "AC 11" to "AC 15" represents specific high-priority services [24]. Each class is assigned an Access Probability Factor (APF) and a Barring Timer (BT). The devices belonging to a certain AC are allowed to transmit their RA preambles in a RA slot only by drawing a random number lower than the APF. Otherwise, the access is barred and the devices have to wait for a random back off time which is determined according to the BT of that class, before attempting a new access [40].
2. *RACH resource separation scheme*: Two approaches are suggested in this scheme in order to allocate RACH resources to M2M devices different than H2H de-

- VICES. In the first approach, available RA preambles are split into two groups; The first group is dedicated for M2M devices and the second one is dedicated for H2H devices. While, in the second approach, although the available RA preambles are split into two groups, but the first group is dedicated to M2M devices and the second one is shared by H2H and M2M devices [40].
3. *Slotted access scheme*: Initially M2M devices are in "sleep mode", but in specific radio frames and in specific RA-slots, M2M devices are allowed to send their RA preambles based on this mechanism. The radio frames and the RA-slots are calculated by M2M devices based on their identity and RA-cycle [46].
 4. *Dynamic resource allocation between M2M and H2H devices*: With an adaptive resource allocation, there is no dedicated resources neither for M2M devices nor for H2H devices. All resources are shared in a dynamic mechanism based on the network access requests. When excessive access attempts initiated by loads of M2M devices are detected by the network, it allocates additional RACH resources for M2M devices automatically [41].
 5. *Pull-based scheme*: Usually, M2M devices receive paging messages sent from the eNodeB, which triggers responses from M2M devices towards the eNodeB by initiating random access requests. Following a centralized mechanism helps the eNodeB in controlling the paged devices based on the network congestion level and the remaining available resources [47].
 6. *M2M-specific back off scheme*: This mechanism implements different delays on the random access attempts based on the device type M2M or H2H device. The back-off time assigned to M2M devices is 48 folds H2H devices (960 ms for M2M devices vs. 20 ms for H2H devices) [24].
 7. *Grouping or Clustering scheme*: UEs are grouped based on QoS applications or based on their geographical locations. A "*coordinator*" for each group is selected as a relay agent for the whole group members with their associated eNodeB [35].

In order to compare the previous different mechanisms, we conduct a comparative analysis among these mechanisms by highlighting on the strength and weakness points in each one of them.

3.3.4 Comparative analysis of RACH overload control mechanisms

In this section, we analyze the proposed solutions found in the literature to alleviate the overload problem in M2M communications and list the gaps needed to be filled in order to fully support the M2M paradigm:

1. *Access Class Barring scheme:* In [24], it is concluded that using the ACB mechanism the eNodeB can deal with the RACH overload by lowering the value of APF, but this could cause longer random access delays to some devices. While in [35], they shed the light on using EAB (Extended Access Barring) in which delay-tolerant devices are not allowed to perform a random access when a M2M device is labeled as "EAB device". Meanwhile in [40], they stress the fact that ACB can alleviate the M2M massive access issue by defining a dedicated class for M2M devices with higher AP and a lower BT. One main drawback appears when many M2M devices need to access the channel in a short time interval as result of a sudden event (*e.g.*, earthquakes, flooding, *etc.*). This issue needs more studies on how to combine the ACB mechanism with other techniques.
2. *RACH resource separation scheme:* In [35, 47], the authors used the same mechanism in which the total available number of RA preambles is split into two groups based on two different approaches:
 - a) Dedicated H2H and Dedicated M2M preambles.
 - b) Dedicated H2H and Shared H2H-M2M preambles.

The RACH congestion problem could be solved especially when an ACB mechanism previous to the selected approach is implemented first, then UEs can send

their RA preambles (or by adopting a "Game Theory Scheme" presented in [48]). In this scheme, the eNodeB selection method and the back-off procedure are neglected. Therefore, it is preferable to do extra efforts to delve more into finding a sub-solution to this issue and by mixing all together, we can tackle in the end an ideal solution (*e.g.*, "Q-learning" solution presented in [49]). In [41], two ways of resource separation are proposed:

- a) Radio resource separation in the same frequency-band; One for H2H devices and the other one for M2M devices.
- b) Out-of-band dedicated frequency-band (*e.g.*, below 1 *MHz*); This band is dedicated to the M2M devices only.

Both suggested ways need additional research and design modifications [41]. In [40], the authors proposed two different approaches to distinguish M2M resources from H2H resources by either splitting the RA preambles into two groups or by allocating different RA slots for each group. We can notice a drawback to this solution: When the number of reserved resources in each group of devices doesn't reflect the actual demand causing low performance. This scheme needs to be coupled with other mechanisms to switch dynamically between both groups in order to fulfill the requested needs (*e.g.*, using "SOOC" Self-Optimizing Overload Control presented in [46]).

3. *Slotted access scheme*: In [24, 35, 40, 47], the authors present the same previous mechanism without any additional explanation. Moreover, no proposed improvements are suggested, but one weakness point "long access latency" was mentioned in [35]: In dense areas, where a massive number of M2M devices are attempting to access the network simultaneously, the total number of unique access slots do not fulfill the excessive access needs causing a contention among M2M devices to seize shared access slots which leads inevitably to many collision incidents. We can resolve this issue by extending the RA cycle but this can cause a huge delay in RA requests and require searching for an appropri-

- ate solution especially with delay-constrained M2M applications (*e.g.*, alarms) [40].
4. *Dynamic resource allocation between M2M and H2H devices:* In [24, 35, 47], the authors outline the same mechanism, but they spot on two remaining challenges:
 - a) This technique is still limited by the available resources.
 - b) The adjustment decision is not clear enough: When? or how to make it?
 5. *Pull-based scheme:* In [24, 35, 40, 47], the same scheme was analyzed in each article, and the following issues arise:
 - a) The scheme cannot deal with unexpected surge of M2M access requests.
 - b) Managing a channel access with a regular pattern is not supported by this scheme (a "QoS-based clustering" proposed in [50] can be useful in this case).
 - c) The eNodeB selection problem was not addressed: In [35], a reinforcement learning-based eNodeB selection was proposed, in which an eNodeB, maximizing the QoS performance of M2M devices, can be chosen in an overlapping area where multiple eNodeBs could be found.
 6. *M2M back-off scheme:* In [24, 35, 40, 47], "M2M back-off" improves the performance in low channel overload but cannot solve the congestion problem in high overload situations, when more devices perform the RA mechanism simultaneously.

3.4 Simulations, results and discussions

We conducted many simulations using various scenarios. This section shows an example of the M2M traffic load in an emergency event (*e.g.*, earthquakes, fire, terrorist attacks, *etc.*). In such emergency events, besides of the regular H2H network traffic

(Voice over Internet Protocol (VoIP), Video Streaming and file transfer), an additional M2M surge traffic attempts to access the network caused by the consequences of the emergency event.

The core of the scenario uses the open source network modeler SimuLTE [15], see appendix C, number of M2M requests attempting to access the LTE-A network simultaneously in a 1 second interval.

The SimuLTE scenario settings are given in Table 3.2.

Table 3.2: Simulation parameters.

Parameter	Value
Simulation length	300 sec
Min./Max. (eNodeB-UE distance)	35 m / 300 m
Terminal velocity	120 Km/h
Mobility model	Linear mobility
Transmission bandwidth	5 MHz (for Down-Link (DL) and Down-Link (UL) each)
No. of PRBs	25 (for DL and UL each)

The different LTE-A traffics: VoIP, Video Streaming, file transfer and M2M are shown in Table 3.3.

VoIP Model	Parameter	Setting
	Application Packet	40 Bytes
	Interval	20 ms
	Talkspurts and Silences	Default settings
Video Streaming Model	Parameter	Setting
	Video Size	10 MB
	Packet Length	1000 Bytes
	Frame Interval	75 ms
M2M Model	Parameter	Setting
	Packet Size	128 Bytes
	Interval	1 sec
FTP Model	Parameter	Setting
	File Size	20 MB

Table 3.3: LTE-A traffic Models used in two different platforms.

The above scenario is simulated using the open-source network modeler SimuLTE 0.9.1 in an environment of OMNeT++ 4.6 [51], see appendix A, with the open-source model library INET Framework 2.3.0 [52], see appendix B, in two different platforms:

1. *PC platform:* Intel(R) Core(TM) i7-6700 HQ processor at 2.60 GHz, with 12 GB of RAM, and "Windows 10-64 bits" operating system.
2. *Cluster platform:* AMD Opteron(TM) processor (6274x58) at 2.2 GHz, with 24 GB of RAM, and Ubuntu 16.04-64 bits operating system.

In all scenarios, the number of VoIP-UL, VoIP-DL, video streaming, File Transfer Protocol (FTP)-UL and FTP-DL users is 10 each.

The number of M2M increases till the peak of M2M is reached as shown in Table 3.4.

Table 3.4: Maximum number of M2M devices.

M2M traffic	16 B/1 sec	128 B/1 sec	6 KB/1 sec
Cluster platform	above 1000	above 1000	above 1000
PC platform	800	600	800

By exceeding the maximum number of M2M on the Personal Computer (PC) platform, an error appears: *"Error in module (TCP) server M2M.tcp Model error: Address already in use: there is already a connection listening on IP address: Port Number"*. Meanwhile, this error didn't appear while simulating using the same parameters on the cluster platform. This result sheds the light on the importance of the robustness of the platform.

Furthermore, if we consider the results of [30, 31] which show that the performance of voice users remain unaffected by the additional users, while file upload and M2M traffic experience a significant delay around four times higher. Moreover, if we continue to the contents of [27, 28, 29] in which the authors spot on the maximum number of UEs that could overload an eNodeB especially when loads of M2M devices are contending to access the network in dense areas. The answer of the maximum number of UEs vary as follows: 250 UEs [27], 320 UEs [29] and 400 [28], which show a kind of contradiction in between the aforementioned results.

This contradiction needs additional analysis according to the different components and various parameters among several scenarios and platforms, as spotted in the coming chapters.

In the end, we want to mention that part of the aforementioned results have been published in [14].

3.5 Conclusion

Certainly, IoT will take place in every part of our lives with loads of innovative applications. M2M communications emerged with LTE-A networks are becoming the more candidate infrastructure to fulfill these needs. As result, a surge of M2M devices should be connected via LTE-A networks in order to fully automate our daily lives.

Our aim in this chapter is to shed the light on the coming overload congestion problem caused by the ubiquity of M2M communications which shall arise in the near future. Furthermore, a survey of the main solutions proposed in the literature to overcome this issue is presented. Additionally, an analysis has been conducted here as a result of RACH procedure limitations. Although, many proposed solutions appear to be optimized on (time, frequency) but for the moment the overload congestion problem is still a talking point with no clear solution.

Finally, different results mentioned in Table 3.4 are concluded according to two different platforms in an emergency event full of H2H and M2M devices, which require extra investigations using different parameters among several scenarios and platforms, which will be discussed, implemented and analyzed in the coming chapters.

Chapter 4

Adaptive Evolved Node B (A-eNB) in LTE-M

Glossary

3GPP: 3rd Generation Partnership Project

A-eNB: Adaptive evolved Node B

A-eNB-EMG: Adaptive evolved Node B EMerGency stage

BGP: Border Gateway Protocol

BL: Bandwidth Limit

Cat: Category

CMMPP: Coupled Markov Modulated Poisson Processes

DL: Down-Link

DRX: Discontinuous Reception

eMTC: Enhanced Machine Type Communication

FDD: Frequency Division Duplex

FTP: File Transfer Protocol

GERAN: GSM/EDGE Radio Access Network

H2H: Human-to-Human

IoT: Internet of Things

IP: Internet Protocol

LTE: Long Term Evolution
LTE-A: Long Term Evolution-Advanced
LTE-M: Long Term Evolution for Machines
M2M: Machine-to-Machine
MIMO: Multiple Input Multiple Output
OFDMA: Orthogonal Frequency-Division Multiple Access
OSPF: Open Shortest Path First
PPP: Point to Point Protocol
PRB: Physical Resource Block
QPSK: Quadrature Phase-Shift Keying
QoS: Quality of Service
RB: Resource Block
RE: Resource Element
SNR: Signal-to-Noise Ratio
TCP: Transmission Control Protocol
TCP/IP: Transmission Control Protocol/Internet Protocol
UDP: User Datagram Protocol
UE: User Equipment
UL: Up-Link
VoIP: Voice over Internet Protocol

4.1 Introduction

Machine-to-Machine (M2M) communication has increasingly become an attractive area for both researchers and industry. It is a novel communication technology whereby a large number of connected devices can exchange information and perform actions without any direct human intervention. The creativity of this new era is boundless with novel potentials. The future M2M devices should sense and communicate via Internet of Things (IoT) technology, command and control applications in a universal ecosystem network making the human's life much easier.

Although, M2M devices transmit small-sized packages in different time intervals, but due to their specificity and functionality they send their payloads in form of synchronized storms, unlike traditional Human-to-Human (H2H) communications. Therefore, one of the prominent challenges that will face the mobile operators is the fast saturation of the limited bandwidth of Long Term Evolution for Machines (LTE-M) with the rise of M2M devices especially in many emergency cases in which all these devices are requesting to send their data simultaneously (*e.g.*, terrorist attacks, tsunamis, power outages, *etc.*).

The saturation problem leads inevitably to a remarkable impact on both M2M and H2H traffics, services and applications. According to Heavy Reading [53], mobile operators are spending 20 billion dollars a year to overcome network outages and service degradation.

In this chapter, we propose a novel scheme called Adaptive evolved Node B (A-eNB) for LTE-M to address this challenge. M2M overload congestion is solved gradually by adapting the number of resources reserved for M2M till soaking-up all storms. Meanwhile, a threshold number of resources has been set in order to maintain H2H traffic not to be affected as result of M2M problem resolution, knowing that partial results of this chapter have been published in [16].

4.2 LTE-M networks

4.2.1 LTE-M overview

In the upcoming years, a ubiquitous coverage, a long battery life, a low device and deployment cost, and a massive number of devices support are mandatory to satisfy the IoT requirements, as explained in section 1.1.2.

To reach this goal, 3rd Generation Partnership Project (3GPP) adopts its legacy Long Term Evolution-Advanced (LTE-A) technology by proposing an enhanced ver-

sion called Enhanced Machine Type Communication (eMTC) also known as *LTE-M*, which uses a reduced bandwidth of 1.4 *MHz* using only 6 Physical Resource Blocks (PRB)s. This bandwidth is selected to allow the LTE-M User Equipment (UE) to follow the same cell search and random access procedures as legacy UEs, which use the channels and signals that occupy six PRBs [54].

Additionally, trying to seize the maximum portion of the market, LTE-M adopts all its specifications based on the IoT requirements as follows:

- Long battery life: Enhance the Discontinuous Reception (DRX) cycle¹ in LTE-M to allow for longer inactivity periods and thus optimize battery life by achieving up to 10 years of battery operation for LTE-M with a daily update of 200 bytes, while taking into account leakage current and battery self discharge [4].
- Low device cost: A significant cost reduction is achieved compared to category 1 (Cat-1)² when 3GPP added a new Category 0 (Cat-0)³ [56] in 3GPP Release 12.

The key reductions agreed in 3GPP Release 12 [57] are:

- a) Using half-duplex Frequency Division Duplex (FDD) operation which makes it possible to operate Long Term Evolution (LTE) FDD time multiplexed avoiding the duplex filter.
- b) Reducing the device receiving bandwidth to 1.4 *MHz* which allows for a substantial complexity reduction.
- c) Using a single receiver chain which allows to remove the dual receiver chain for Multiple Input Multiple Output (MIMO).

¹The main feature to reduce power consumption from the radio perspective is Discontinuous Reception (DRX). A UE configured with a DRX cycle can avoid monitoring the control channel continuously, enabling the UE to switch off parts of the circuitry to reduce power consumption.

²Cat-1: it was included in the LTE specifications in Release 8. A UE (Cat-1) can achieve 10 Mbps downlink and 5 Mbps uplink channel data rates. No MIMO is supported but the UE should still have 2 receiver antennas.

³Cat-0: it is one of the newest standardized categories from Release 12. According to [55], UEs (Cat-0) are intended for IoT use cases, and provide 1 Mbps data rates for both uplink and downlink. Cat-0 UEs have reduced complexity by up to 50% compared to Cat-1; requirements include only one receiver antenna and support of half-duplex operation, providing ways for the manufacturers to significantly reduce the modem cost compared to more advanced UE categories.

- d) Using lower data rates which helps in significant reductions in complexity and cost for both processing power and memory.
- Low deployment cost: LTE-M has many competitive advantages of using the same hardware and a share spectrum by making them compatible with the legacy LTE-A. This allows the LTE-M deployment with the existing infrastructure just by applying a software update [54].
 - Ubiquitous coverage: The presence of devices in extreme coverage conditions (*e.g.*, a meter in a basement) requires the UEs to operate with much lower Signal-to-Noise Ratio (SNR). LTE-M targets 15 dB coverage enhancement for the device category [54]. Additionally, control and data signals can be repeated to reach the required coverage enhancements.
 - Massive number of M2M devices support: LTE-M can support up to 52 k M2M devices.

In the near future, LTE-M technology is expected to attract a huge amount of future IoT market if it can offer a better IoT platform by allowing customers to scale and manage their business requirements more efficiently [4].

4.2.2 LTE-M Data-rate

The previously computed LTE-A maximum data-rate, in section 3.2.1, is usually dedicated for H2H users. However, in 3GPP Release 13, LTE-M technology dedicates a 1.4 *MHz* of the total bandwidth for M2M communications as shown in Figure 4-1.

Following equation (3.1), but with 6 PRBs, Quadrature Phase-Shift Keying (QPSK) modulation and half-duplex mode; the maximum data-rate is reduced to 1 Mbps in Up-Link (UL) and 1 Mbps in Down-Link (DL) for M2M traffic.

4.2.3 LTE-M bandwidth limitation (case study)

In real life, the emergency events such as natural disasters, multiple accidents and terror attacks are unfortunately not predictable at all. With vague scenarios and lack

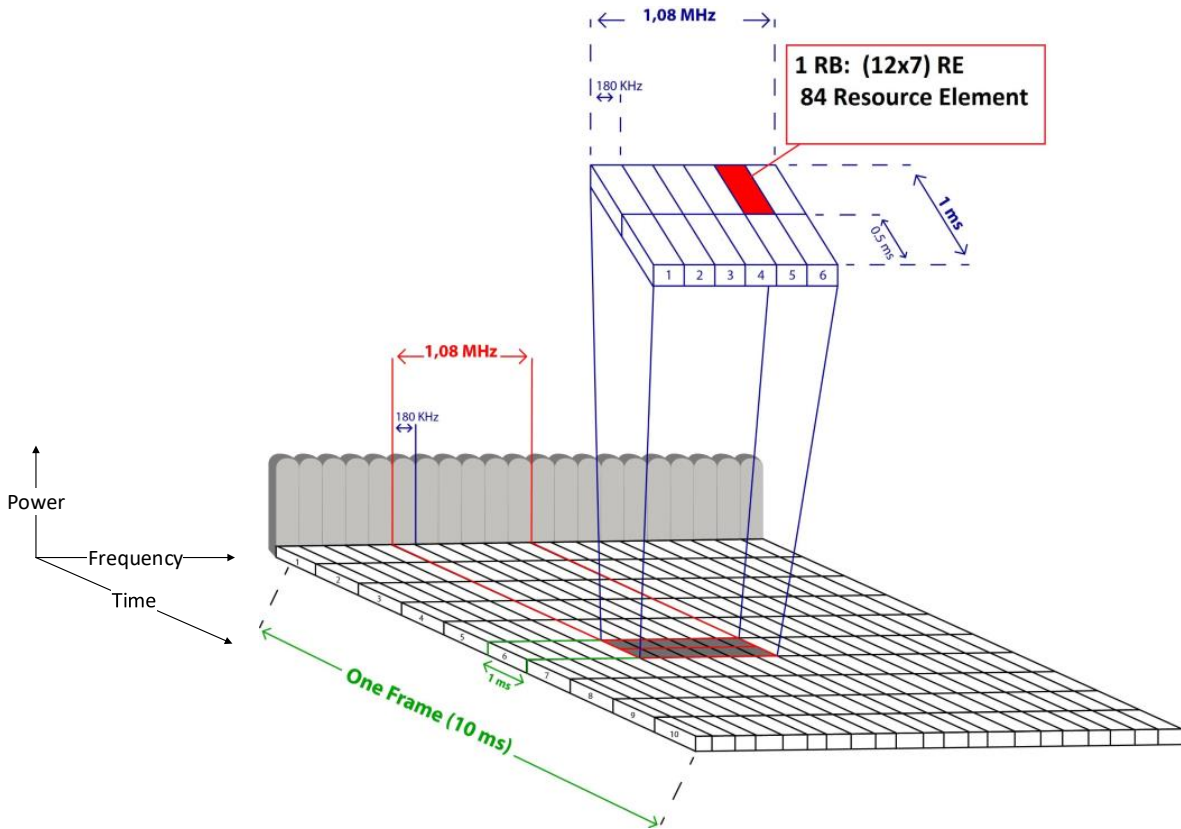


Figure 4-1: Bandwidth Limitation for LTE-M carrier within LTE-A carrier, where Resource Block (RB), Resource Element (RE).

of statistics and researches about the behavior of M2M devices throughout emergency scenarios, we build, in this section, a case study based on some use-cases and 3GPP technical reports.

In [58], a use-case, in which LTE-M technology is expected to fulfill M2M requests efficiently with a cut-off point of 80K devices per sector for an interval of 4 upload and 4 download transfers per day (with full security) in normal scenarios.

Supporting more than 52000 M2M devices per cell is one of the LTE-M targets in order to scale to the IoT requirements [5]. This is the reason why, we consider in our use-case three different groups selected according to the parameters set to different models proposed by 3GPP GSM/EDGE Radio Access Network (GERAN) Technical Report [59], as follows:

- (a) Group1 contains 20000 Environmental monitoring devices, sending 200 Bytes with a rate of 1 message per hour or 4800 Bytes per day.

- (b) Group2 consists of 20000 Assisted Medical devices, dispatching 100 Bytes with a rate of 8 messages per day or 800 Bytes per day.
- (c) Group3 contains 20000 Asset tracking devices, transmitting 50 Bytes with a rate of 100 messages per day or 5000 Bytes per day.

By analyzing the behavior of M2M devices in normal scenarios, and knowing that a day contains $24 \times 3600 = 86400$ sec, we realize that M2M devices send their payloads with an average data-rate equal to:

$$(4800 + 800 + 5000) \text{ Bytes} / 86400 \text{ sec} \times 20000 = 2453.7 \text{ Bytes per second} = 2453.7 \times 8 \text{ bits per second} = 19629.6 \text{ bits per second} = 0.0187 \text{ Mbps.}$$

Consequently, If we compare it with the maximum data-rate in LTE-M (1 Mbps), we conclude that in normal cases, LTE-M can work efficiently without any congestion problem.

In emergency scenarios, M2M devices may send their payloads in a synchronize manner. So, we may expect to have an enormous M2M traffic in a split second in form of three M2M group types (20000 M2M devices each). Assuming that each group is divided into 5 sub-groups, each sub-group consists of 20% of M2M devices (4000 M2M devices for each) sending their synchronized payload simultaneously. Consequently, we might expect 3 groups of storms in different intervals as follows:

- (a) Group1-Storm: The first group sends its payload (200 Bytes) with a rate of 1 message per hour. Consequently, the total rate is equal to: $(200 \text{ Bytes} * 4K \text{ devices}) \text{ per second} = 6.1 \text{ Mbps}$. Because M2M Group1 devices are sending their payloads with a rate of 1 message per hour, so this type of storms will be repeated $24 \times 5 = 120$ storm/day.
- (b) Group2-Storm: following the same equation with a payload = 100 Bytes, the storm can reach 3 Mbps total payload rate. As a result, we expect to receive $8 \times 5 = 40$ storm/day, if we take into consideration that M2M Group2 devices interval is 8 messages/day.

- (c) Group3-Storm: similar to the two previous storm calculations, but with 50 Bytes payload, this storm peaks at 1.5 Mbps as total payload rate. If we know that the repetition of group3 devices is 100 messages/day, we conclude that this storm repetition is about $100 \times 5 = 500$ storm/day.

If we compare the different storms in emergency cases, with the maximum up-link data-rate in LTE-M (1 Mbps) calculated in section 4.2.2, we conclude that the available bandwidth will suffer from a huge degradation in both cases.

As a conclusion, we emphasize the need of a coexistence study on how to allow both M2M and H2H traffics to access the network efficiently in emergency events with a minimal impact one towards the other.

4.3 Literature review

In [60], the authors concluded that the reduced bandwidth allows for a substantial complexity reduction 81% but the fast depletion of the bandwidth when facing an expected M2M storm in an emergency scenario wasn't discussed. Moreover in [61], a cross-layer solution was proposed to increase the number of devices that can be served by one eNodeB. The solution combines reduction of the Transmission Control Protocol/Internet Protocol (TCP/IP) overhead with buffering and clustering concepts in order to maximize efficiency of the transmission of small payloads by a high number of devices. Although, the proposal enables to serve up to 65K devices by one eNodeB in case of a 10 MHz bandwidth, but it didn't show the cost on both M2M and H2H traffics in a congested eNodeB. Another approach was proposed in [62], where a source modelling was proposed based on Coupled Markov Modulated Poisson Processes (CMMPP) to overcome the massive number of devices problem. Although, the proposed model demonstrates the parallel deployment of 30K M2M devices with reasonable efforts, but it couldn't scale to 52K M2M devices recommended in the IoT requirements. In [63], despite the mathematical model for LTE downlink bandwidth allocation that was proposed with the aim of providing a good QoS for each UE, the coexistence between LTE-M and LTE-A systems and the bandwidth adaptation

are not spotted. In [64], a proposed cognitive-based radio access strategy with a priority queuing scheme is applied in LTE-A networks with M2M/H2H coexistence distinguishing M2M devices based on their traffic QoS requirements. Although an analytical model is developed in normal scenarios, but the expected surge number of M2M devices which might have higher priority during disaster scenarios and how to deal with this sticky situation hadn't been discussed.

To sum up, in the literature, many approaches in [60, 61, 62, 63, 64] are trying to find an ideal solution for the massive access problem and its consequences. But, it can be noticed that the existing architectures provide only preliminary solutions, without any projection to the expected exponential growth of M2M storms especially in disaster scenarios. To the best of our knowledge, our adaptive solution is the first solution to address this problem from its roots in a smart, adaptive and robust methodology, which can result in a clean LTE traffic free from M2M congestions.

4.4 Adaptive eNodeB for LTE-M

Previous sections shed the light on the importance of an intelligent solution operating effectively to absorb any expected signaling storm. To this end, we propose a promising approach which extends the classical functionality of eNodeB and includes an adaptive control of the bandwidth based on the M2M load status throughout the network. The proposed Adaptive eNodeB manages automatically both H2H and M2M traffics. Based on the instantaneous requests of M2M, a part of the total bandwidth will be dedicated to fulfill these sudden needs. The Adaptive eNodeB for LTE-M networks depicted in Figure 4-2, is designed to operate in different conditions, depending on the number of M2M connected devices C_d throughout the network:

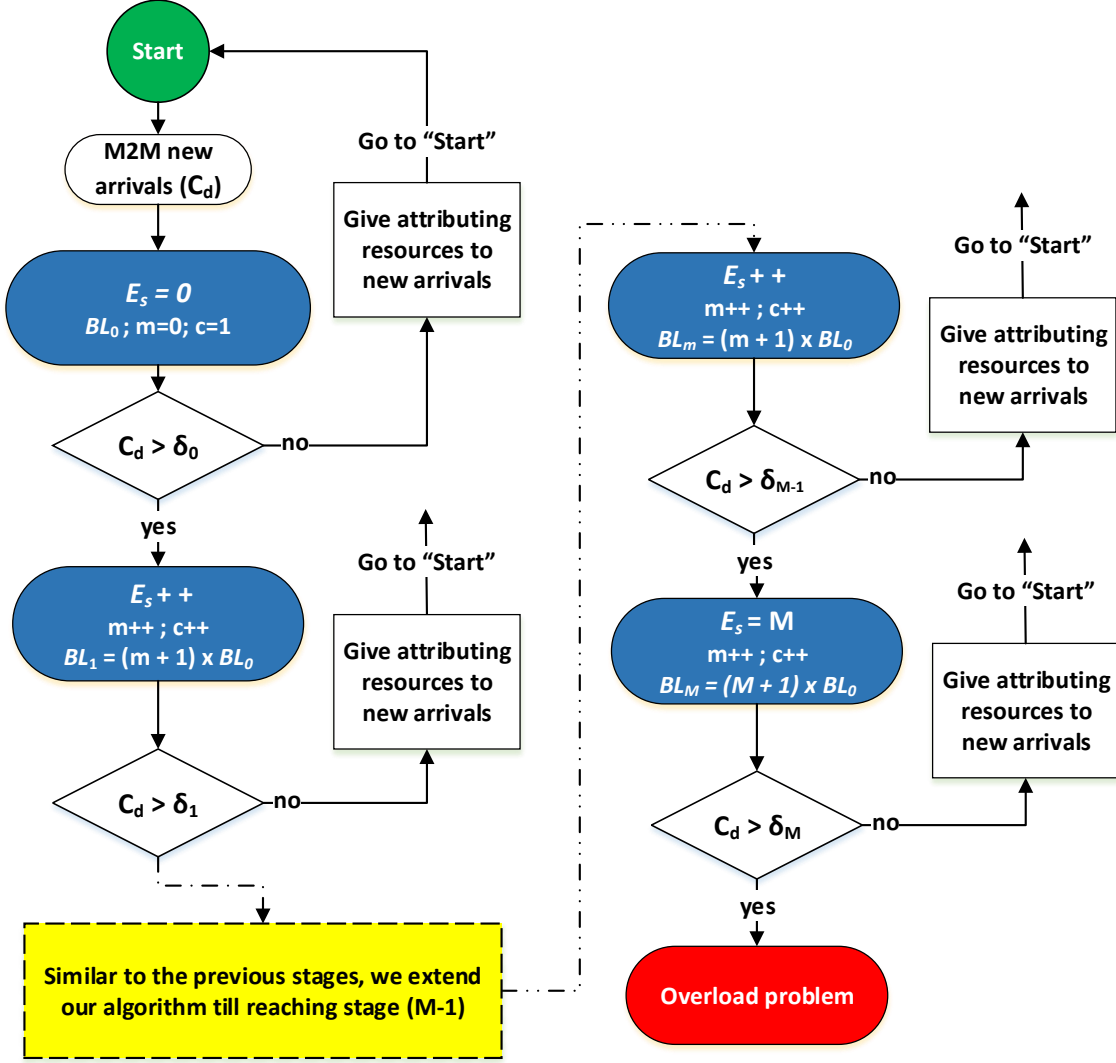


Figure 4-2: Flow chart for LTE-M Adaptive eNode-B over normal cycle, where E_s is the emergency stage number, $BL_m =$ Bandwidth Limited at the m^{th} stage, $BL_0 = 1.4$ MHz is the initial bandwidth, m stands for current stage number, M is the maximum stage number, C_d represents the number of M2M connected devices, $\delta_m =$ threshold at the m^{th} stage and c stands for the number of system resources. This flowchart is extracted from our published work in [17].

1) Initial stage "A-eNB-INI"

In normal situations, an A-eNB works similar to any legacy eNodeB in LTE-A/LTE-M networks. A part of the total bandwidth is dedicated to LTE-M traffic. For

instance, a total of 20 MHz (100 PRBs) will be divided into a basic Bandwidth Limited ($BL_0 = 1.4 \text{ MHz}$) reserved for M2M devices (6 PRBs), and the remaining bandwidth (94 PRBs) are dedicated for H2H devices, as represented in Figure 4-3.

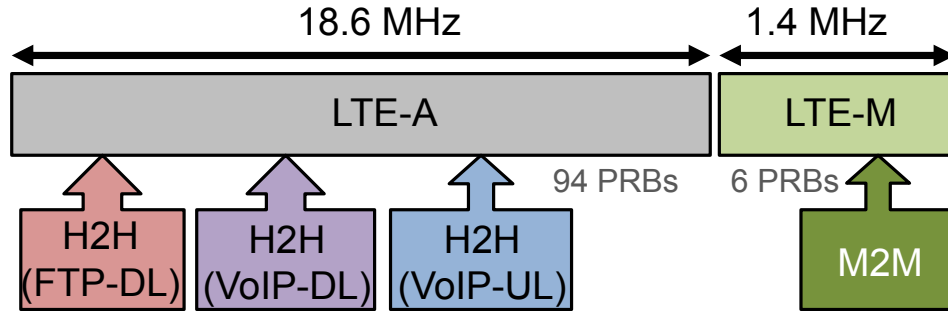


Figure 4-3: Initial Stage - $BL_0 = 1.4 \text{ MHz}$.

During this stage, M2M devices are allowed to access BL_0 reserved for LTE-M to send their payloads via different IoT applications, while H2H devices use the remaining bandwidth to exchange their information.

2) Emergency stages "A-eNB-EMG(1)" to "A-eNB-EMG(M)"

In the case of a disaster, a huge number of M2M devices saturate the initial LTE-M bandwidth (BL_0) by their signaling storm briefly. When C_d crosses the initial saturation threshold δ_0 (configurable), the bandwidth dedicated to M2M starts to increase from $BL_0 = 1.4 \text{ MHz}$ to $BL_1 = 2.8 \text{ MHz}$ in order to allow more M2M devices to access the network.

If C_d crosses a next saturation threshold δ_m , caused by an additional M2M signaling storm, the A-eNB adapts gradually its bandwidth from BL_1 till it reaches BL_M :

$$BL_M = [(M + 1) \times BL_0], \text{ as shown in Figure 4-4.}$$

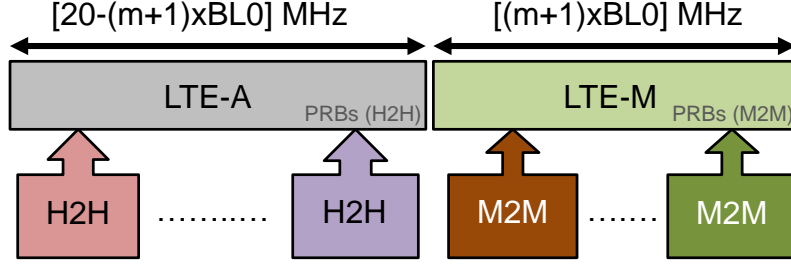


Figure 4-4: Adaptive bandwidth during emergency stages: "A-eNB-EMG(1)" to "A-eNB-EMG(M)", m is the current stage number (0 to M), PRB represents the Physical Resource Block.

For a given stage m , we can calculate:

- Total number of PRBs reserved for M2M devices:

$$p_{m2m} = (m + 1) \times 6 \text{ PRB}$$

- Maximum M2M data-rate: following equation (3.1) the maximum M2M data-rate could be estimated as follows:

$$R_{m2m} = n \times o \times l \times p_{m2m} \times e \quad (4.1)$$

where R_{m2m} is the maximum data rate for M2M traffic, n represents the number of sub-carriers, o is the number of OFDMA symbols, l is the number of slots, p represents the number of PRBs reserved for M2M traffic, e is the number of carried bits per RE.

- Total number of PRBs reserved for H2H devices:

$$p_{h2h} = 100 - p_{m2m}$$

- Maximum H2H data-rate: similar to equation (4.1):

$$R_{h2h} = n \times o \times l \times p_{h2h} \times e \quad (4.2)$$

where R_{h2h} is the maximum data rate for H2H traffic, n represents the number of sub-carriers, o is the number of OFDMA symbols, l is the number of slots, p represents the number of PRBs reserved for H2H traffic, e is the number of carried bits per RE.

3) Reset stage "A-eNB-RST"

Once the number of M2M connections starts decreasing, A-eNB reduces iteratively the LTE-M maximum bandwidth BL_M until achieving the initial state with BL_0 at the end of the disastrous events, and consequently, it starts operating similar to any traditional eNodeB.

Following this methodology, the A-eNB has the adaptability and robustness to lend progressively a temporary bandwidth: $(M+1)$ folds of 1.4 MHz to M2M devices tentative use during disaster scenario. Inevitably, this solution can soak up most of M2M storms with a minimal impact on H2H devices.

4.5 Simulation scenarios

A new eNodeB architecture was proposed in section 4.4 taking into account disastrous and normal conditions. The purpose of this section is to evaluate the A-eNB by measuring the cost on both M2M and H2H traffics.

To this end, we use the open-source network simulator *SimuLTE* Modeler 0.9.1 because of his flexibility which lets researchers simulate and benchmark their solutions on an easy-to-use *OMNeT++*. *OMNeT++* 4.6 framework is an extensible, modular, component-based C++ simulation library and framework, primarily for building network simulators. Moreover, it can be integrated with other modules from the *INET* 2.3.0 Framework which contains models for the Internet stack (Transmission Control Protocol (TCP), User Datagram Protocol (UDP), Internet Protocol (IP)-v4, IP-v6, Open Shortest Path First (OSPF) , Border Gateway Protocol (BGP), *etc.*), wired and wireless link layer protocols (Ethernet, Point to Point Protocol (PPP), IEEE 802.11, *etc.*), support for mobility, several application models, and many other protocols and components. The softwares: *OMNeT++*, *INET* and *SimuLTE* are introduced respectively in appendices A, B and C.

The considered simulation settings are:

- Total simulation time is 200 sec.

- Maximum distance between eNodeB and UE is 300 m.
- A linear mobility of UEs.
- Terminal velocity of 120 Km/h.
- Total bandwidth is 20 *MHz*.

4.6 Results and discussions

In all our next scenarios, we assume that:

- H2H Traffic: The scenario consists of 30 H2H users (10 File Transfer Protocol (FTP)-DL, 10 Voice over IP (VoIP)-UL, 10 VoIP-DL) during normal conditions. In an emergency event, additional 90 H2H users start to operate (30 FTP-DL, 30 VoIP-UL, 30 VoIP-DL) as consequence of the emergency scenario. The H2H traffic models are represented in Table 4.1.

Table 4.1: H2H traffic models.

<i>Model</i>	<i>VoIP</i> [65]	<i>FTP</i>
Packet Size [Bytes]	40	536
Interval [ms]	20	1000

- M2M Traffic: The M2M traffic models are mapped to three selected groups of IoT applications according to the parameters set by 3GPP GERAN Technical Report [59] as mentioned in section 4.2.3.

4.6.1 Regular eNodeB

At first, the impact of M2M on H2H traffic in a LTE-A network with a legacy eNodeB is experimented. To this end, the simulated architecture is composed of:

- A fixed number of H2H traffic (40 FTP-DL, 40 VoIP-UL, 40 VoIP-DL).

- A variable number of M2M devices (100, 200, 300), connected all together to a legacy eNodeB.

The measured network performance is depicted in Figure 4-5. The VoIP traffic does not reveal a considerable effect of the increasing M2M traffic load within the LTE-A network, as the priority of voice traffic to access the network is higher than the M2M communication. However, the file transfer traffic suffers from a significant degradation of the download performance; For only 300 M2M devices, the data downloaded by FTP-DL traffic decreases by 41% comparing to 100 M2M devices. This is also due to the network priority; FTP has a lower priority than VoIP and the same priority as M2M traffic which results in a significant congestion when accessing the network.

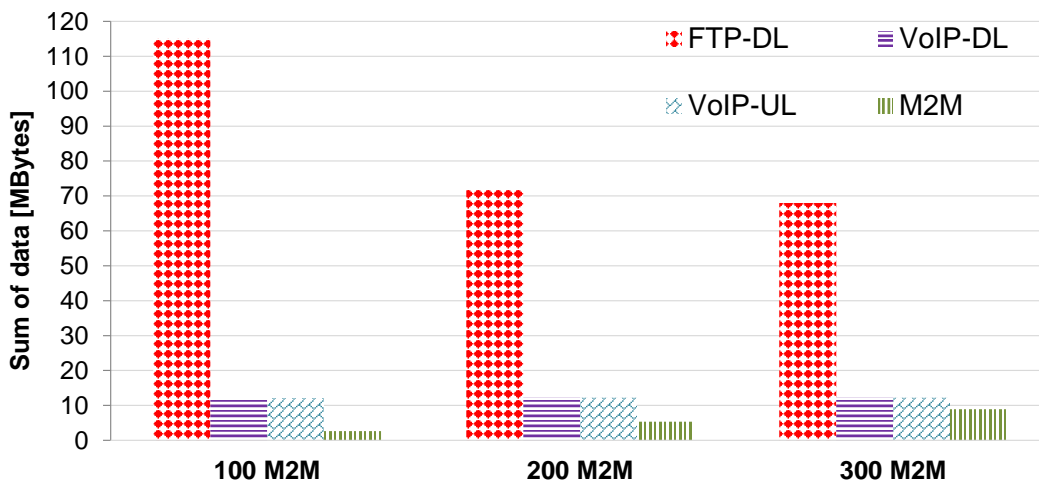


Figure 4-5: Impact of M2M traffic on H2H traffic in a LTE-A legacy scenario for different traffics: FTP-DL: File download traffic, VoIP: Voice over IP traffic, M2M: Machine-to-Machine traffic.

4.6.2 A-eNB Initial stage

At this stage, we model the LTE-M network with the aim to measure the improvement on both M2M and H2H traffics. To this end, we consider the same traffics as the previous simulation, but with a dedication of 1.4 MHz for M2M devices ($BL_0 = 1.4$ MHz). The results show that similar to the previous results in the regular eNodeB,

the VoIP traffic is not affected by M2M traffic, but a significant improvement in the FTP traffic can be noticed as shown in Figure 4-7; For example, with 300 M2M devices the FTP traffic has a gain of 82% comparing to the previous results in regular eNodeB as depicted in Figure 4-6.

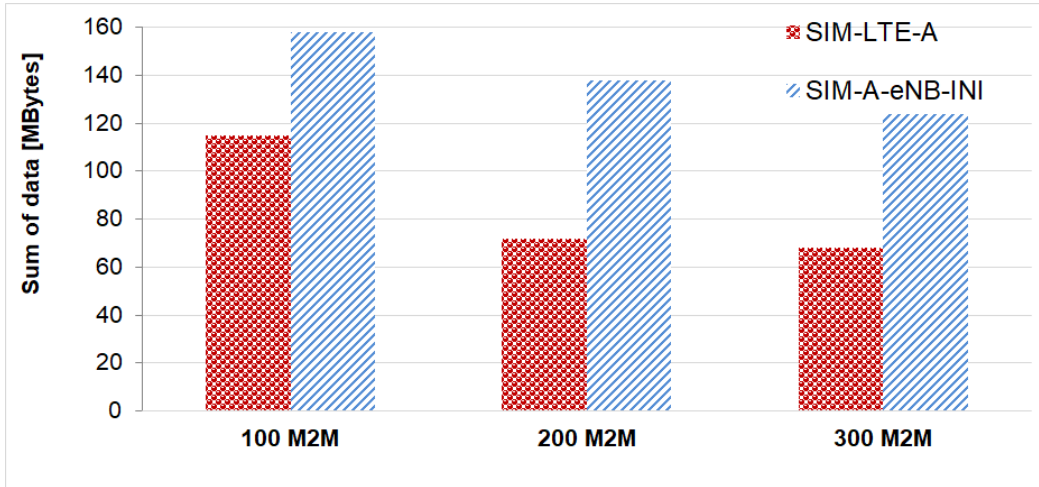


Figure 4-6: FTP-DL improvement in LTE-M comparing to LTE-A., where SIM-LTE-A represents the results in a regular eNodeB/LTE-A and SIM-A-ENB-INI is the results during an initial stage of the A-eNB/LTE-M.

Meanwhile, the M2M traffic reaches its peak (4.9 MB) starting from 300 M2M devices as depicted in Figure 4-7.

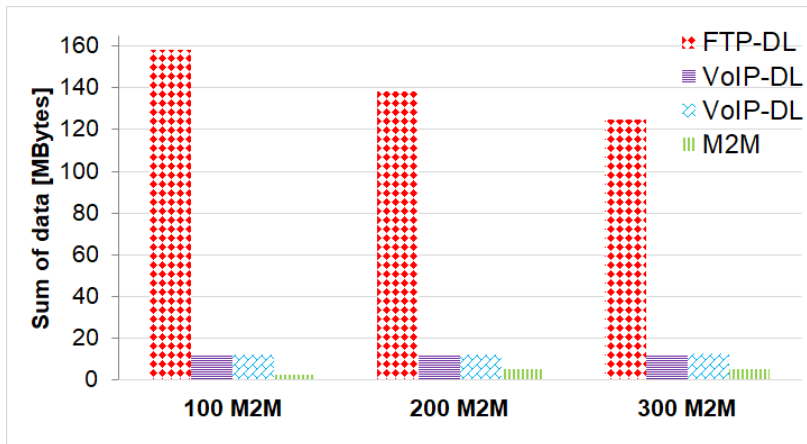


Figure 4-7: Improvement on FTP-DL in a LTE-M ($BL_0 = 1.4 \text{ MHz}$), where: FTP-DL: File download traffic, VoIP: Voice over IP traffic, M2M: Machine-to-Machine traffic.

Indeed, LTE-M improves the FTP traffic significantly, but a bandwidth saturation problem appears, limiting the number of M2M devices attempting to access the network.

4.6.3 A-eNB Emergency stages

In order to corroborate the performance of our proposed solution on both the FTP traffic and the number of M2M devices accessing the network, we simulate our proposed A-eNB in its first act toward a M2M storm by switching from its initial state, in which $BL_0 = 1.4 \text{ MHz}$, to the first emergency state, in which $BL_1 = 2.8 \text{ MHz}$. Then, we evaluate the different traffic performance during this emergency stage. To this end, a fixed number of H2H traffic is considered (40 FTP-DL, 40 VoIP-UL, 40 VoIP-DL) with an increasing number of M2M devices (300, 400, 500, 600, 700), connected all together to the A-eNB. The network performance are depicted in Figure 4-8.

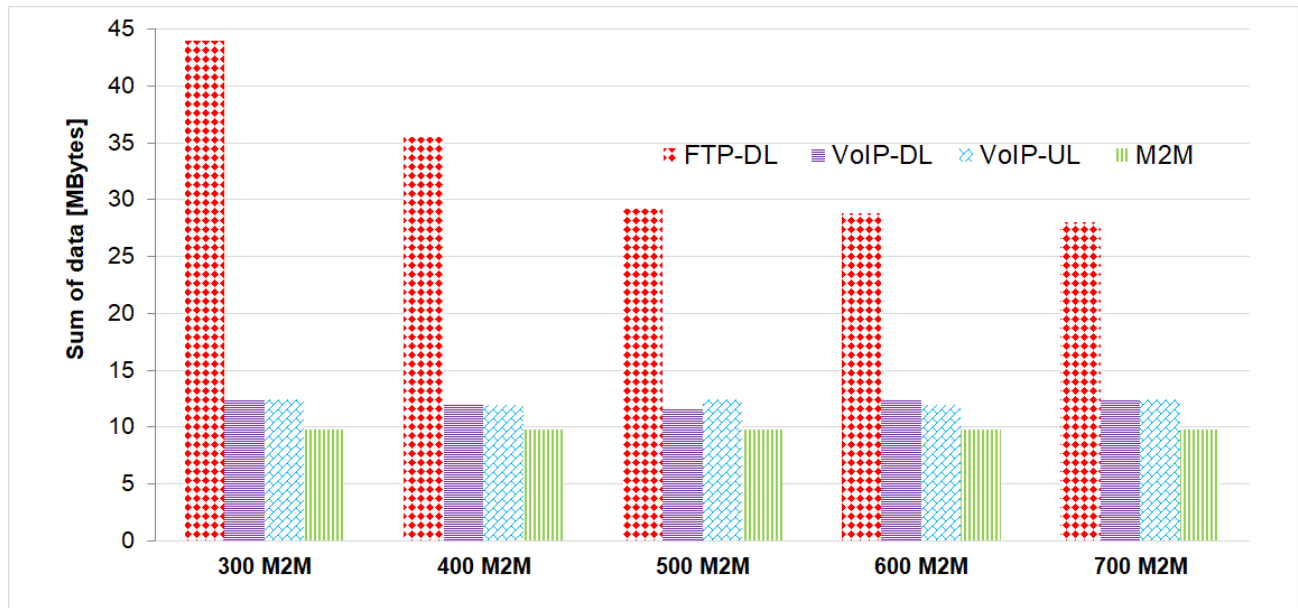


Figure 4-8: Improvement on M2M traffic in emergency stage1 ($BL_1 = 2.8 \text{ MHz}$), where: FTP-DL: File download traffic, VoIP: Voice over IP traffic, M2M: Machine-to-Machine traffic.

The VoIP traffic is not affected by the increasing M2M traffic, similarly to previous two scenarios in regular eNodeB and the initial stage of A-eNB. We recall that our target is to maximize the number of M2M devices accessing the network and to minimize the impact on FTP traffic. In this emergency stage, the M2M traffic is significantly improved, comparing to the results in the initial stage of A-eNB (*e.g.*, in 300 M2M devices the improvement is 100%). Meanwhile 36% of FTP traffic is affected, if we compare the FTP traffic in 300 M2M devices to the FTP traffic in 700 M2M devices. This degradation is due to the new bandwidth allocation, deduction of 1.4 MHz from H2H to M2M bandwidth ($BL_1 = 2.8 \text{ MHz}$). It is important to recall that this degradation is a temporary degradation during the emergency event only, which worth the FTP traffic sacrifice toward the M2M traffic gain as we can realize in the next paragraph. The M2M bandwidth reaches its cut-off point (9.8 MB) starting from 300 M2M devices because of the saturation of BL_1 , which requires additional actions from the A-eNB to absorb the increasing M2M storm.

4.6.4 A-eNB Predictive emergency scenarios

In this predictive scenario, we estimate the expected actions taken by the A-eNB toward a continuous escalated storm by switching from its first emergency state till it reaches the last stage, by increasing $BL_1 = 2.8 \text{ MHz}$ till $BL_M = [(M + 1) \times BL_0] \text{ MHz}$, we estimate a gain = M in the M2M traffic comparing to the initial stage results as depicted in Figure 4-9 (supposing M = 7).

Also, we expect a stability in the VoIP traffic with good QoS and a minimum FTP request fulfillment.

To sum up, it is clear that the proposed A-eNB in an emergency scenario absorbs gradually the M2M storm while keeping the QoS of H2H devices within the acceptable standards.

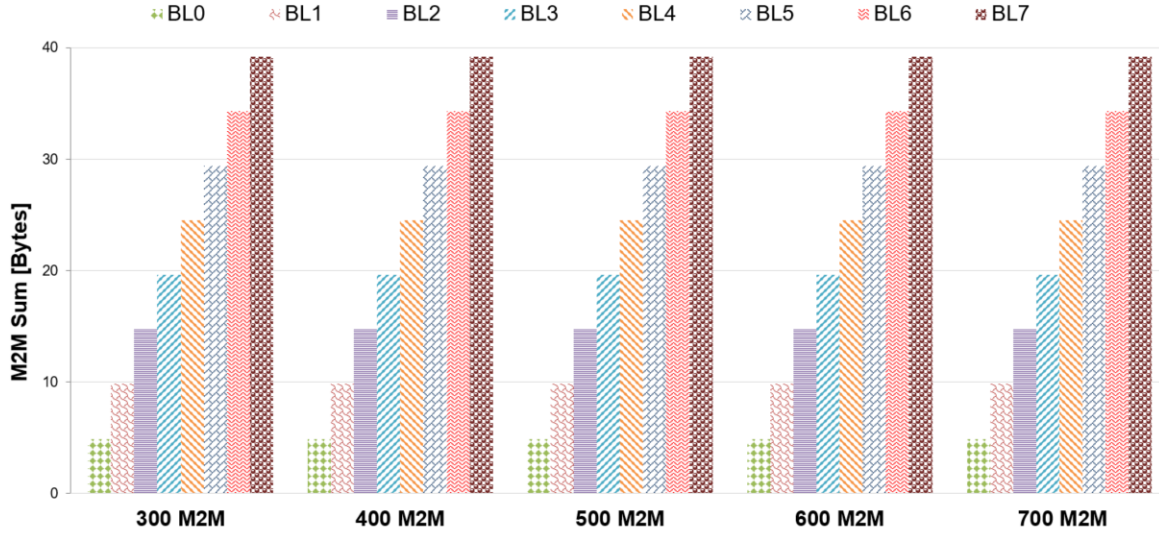


Figure 4-9: M2M saturation points comparing to different A-eNB stages, where BL_0 to BL_7 are the bandwidths during different stages from 0 to $M = 7$.

4.7 Conclusion

The support of M2M communications in an IoT environment requires a parallel establishment of many new features. In this chapter, we have proposed an effective solution so called A-eNB as an extension to the classic eNodeB in a LTE-M networks. The classic eNodeB in a LTE-M network leads to a fast depletion of its bandwidth when huge number of M2M access requests are sent due to its non-adaptability and limited bandwidth $BL_0 = 1.4 \text{ MHz}$. In our solution A-eNB we implement a flexible reallocation of bandwidth which can be adapted based on the increase or decrease of different traffics without affecting the QoS of various traffics. Moreover, during emergency events, this solution A-eNB, apply its flexible reallocation of bandwidth to absorb presumable M2M storms. Based on SimuLTE modeler, the results show that by leasing some folds of LTE-M classic bandwidth, we can achieve a significant gain in M2M traffic.

Chapter 5

CTMC modeling for H2H/M2M coexistence in LTE-A/LTE-M

Glossary

3GPP: 3rd Generation Partnership Project

CANAL: Coexistence Analyzer and Network Architecture for Long term evolution

CTMC: Continuous-Time Markov Chain

DTMC: Discrete Time Markov Chain

eNodeB: Evolved Node B

FIFO: First Input First Output

H2H: Human-to-Human

HP: High Priority

IoT: Internet of Things

LIFO: Last Input First Output

LP: Low Priority

LTE: Long Term Evolution

LTE-A: Long Term Evolution-Advanced

LTE-M: Long Term Evolution for Machines

M2M: Machine-to-Machine

M2M-NRT: Machine-to-Machine Non-Real-Time

M2M-RT: Machine-to-Machine Real-Time

MAE: Mean Absolute Error

QCU: Queuing Control Unit

QoS: Quality of Service

RAC: Resource Allocation Control

RACH: Random Access CHannel

RB: Resource Block

RMSE: Root Mean Squared Error

SCR: Service Completion Rate

5.1 Introduction

Traditionally, Long Term Evolution-Advanced (LTE-A) network was developed to better serve Human-to-Human (H2H) services such as voice calls, video-streaming and data traffics. But with the novel paradigm called Machine-to-Machine (M2M) communications (*e.g.* actuators, sensors, smart meters, *etc.*), an unprecedented innovation for the current LTE-A becomes a must [66]. Offering an attractive M2M services in an utopian autonomous IoT (Internet of Things) world seems tempting, but with massive connectivity, significant challenges arise on how to manage large number of devices, typically transmitting only small payloads, across wide range applications. In 2020, there will be around 50 Billion connections with unavoidable coexistence among H2H and M2M traffics in one LTE-A network [67]; Consequently, an efficient radio access strategy becomes one of the most challenges for mobile operators, researchers and 3rd Generation Partnership Project (3GPP) community [68]. This community sounds ultimately keen on conducting several studies and researches to identify the mutual impact among M2M and H2H communications. In this chapter, we mathematically characterize the key performance characteristics of M2M and H2H communications, then we propose an analytical methodology using Continuous-Time Markov Chain (CTMC) model with the following objectives:

- Studying the mutual impact of both M2M and H2H traffics.

- Measuring and analyzing the M2M congestion solutions.
- Modeling a new framework called Coexistence Analyzer and Network Architecture for Long term evolution (CANAL).

The CANAL framework optimizes prioritize radio resource allocation procedures in a LTE-A/Long Term Evolution for Machines (LTE-M) network and achieves understanding of the system performance to reach good balance between M2M and H2H communications. Then, we simulate different dense area and disaster scenarios, which can be studied, analyzed and measured. Finally, by comparing the calculated results with the simulated ones, we come to the conclusion that with a noticeable results correlation we can validate our assumptions, models and the proposed architecture.

5.2 Continuous-Time Markov Chain (CTMC)

Markov chains represent a system of elements moving from one state to another over time. The changes of state of the system are called transitions. The probabilities associated with various state changes are called transition probabilities, which are used to describe the Markov chains behavior. The process is characterized by a state space and a transition matrix describing the probabilities of particular transitions [69].

Markov chains have many applications as statistical models of real-world processes, such as studying cruise control systems in motor vehicles, queues or lines of customers arriving at an airport, exchange rates of currencies, storage systems, and population growths [70].

In this thesis, we used a Markov chain with the queuing theory to analyze telecommunication system performance, and to explore queuing solutions for the network optimization.

There are two types of Markov chains [71]:

- Discrete Time Markov Chain (DTMC): It is a sequence of random variables with Markov property for a process with a discrete set of times.

- Continuous Time Markov Chain (CTMC): Used when the system evolves in a continuous set of times.

To study the H2H/M2M coexistence in LTE-A/LTE-M, it is essential to model the M2M and H2H request arrivals to LTE-A/LTE-M networks. These requests have no specific arrival times, they might arrive at any time along with continuous time intervals. Additionally, they could be served depending on the available service rate provided by the network at a certain moment. Therefore, the probability of transitioning from one state to another depends on the service completion rate for each service. For this reason, we have chosen the CTMC in our project in order to study, measure and evaluate the Quality of Service (QoS) for both traffics H2H and M2M in case of coexistence.

5.3 Existing solutions found in the literature

Since the dawn of Long Term Evolution (LTE) in Release 8 [72] till Release 15 [73], the 3GPP has made major enhancements on LTE perspectives in term of reducing latency, power consumption, improving measurement accuracy, improving random access reliability and small cell support [74].

Supporting H2H communications (*e.g.*, web-browsing, phone calls, internet, *etc.*) is the main target of creating LTE-A networks. While M2M devices (*e.g.*, smart cities, mobile health, smart meters, *etc.*), with their periodic and tiny packet sizes, dispatch their messages into the same LTE-A channel but with different goals. Sharing the same bandwidth causes many drawbacks for both traffics M2M and H2H which should be avoided at all costs. To solve this issue, 3GPP introduces a licensed spectrum technology called LTE-M to serve M2M traffics, which occupies 1.4 *MHz* from the LTE-A bandwidth. But with different perspectives and a diversity of applications of M2M and H2H traffics, many challenges are expected as result of this coexistence.

Now, with a close look to many works found in the literature, many research flaws can be spotted: Random Access CHannel (RACH) congestion in a LTE-A/LTE-M network, high and low priority for both M2M and H2H strategies, queuing for high and

low priority M2M and H2H mechanisms, adaptive resource allocation and emergency scenario solutions.

As for to the congestion and overload problems, which may occur when a large number of M2M devices attempt to access the LTE-A network using RACH, a serious degradation of performance for both M2M and H2H devices is more likely to happen [75]. In section 3.3.3, we compare the most common mechanisms found in the literature that deal with the RACH procedure issues and challenges by analyzing the existing solutions and approaches to avoid RACH overload congestion in M2M communications.

In [40], the main challenges raised by the M2M vision are spotted and by focusing in particular on the problems related to the support of massive M2M access in current cellular communication systems. Then, the most common approaches proposed in the literature to enable the coexistence of H2H and M2M services in the current and next generation of cellular wireless systems are analyzed. In [76], the authors study the coexistence of H2H traffic and M2M traffic originating from wireless sensors. By using experimental measurements of real-world smart grid applications, the impact of variable H2H traffic loads on the sensor end-to-end delay performance are investigated. The obtained results show that a conventional Ethernet may cause a bottleneck and increase the delay for both H2H and M2M traffics. In [77], the author addresses the high overhead signaling load problem of M2M devices using a priority-based data aggregation scheme at the M2M gateway to maintain a good trade-off between the power consumption and delay requirement. According to this work, three types of priority are assigned to M2M devices based on their related applications. Therefore, the data from M2M devices with higher priorities are served faster than the ones with lower priorities. An analytical model considering the idle and busy states behavior of the aggregator using preemption queuing with priority disciplines is developed and analyzed. In [78], a shared channel resource allocation in a H2H/M2M coexistence scenario is considered in order to formulate the resource sharing problem among M2M and H2H communications. In [64], the author proposes a cognitive-based radio access model with a priority queuing scheme. The model is applied for LTE-A networks

with M2M/H2H coexistence distinguishing M2M devices based on their traffic QoS requirements, in which M2M communications have Real-Time (M2M-RT) and Non-Real-Time (M2M-NRT) traffic. Radio access gives the highest priority to H2H, while M2M-RT has higher priority than M2M-NRT. Although an analytical methodology is developed in normal scenarios, but the expected surge number of M2M devices which might have higher priority during disaster scenarios and how to deal with this sticky situation hasn't been discussed.

Additionally, Markov chains are also found in the literature to characterize M2M and/or H2H traffics in LTE-A/LTE-M networks, due to the specificity of LTE-A/LTE-M networks as time variant networks. Thus many works can be spotted based on this stochastic method. In [79], an efficient radio access strategy is proposed to manage an LTE network system where M2M devices and H2H users coexist. In this paper, a CTMC model is developed to evaluate the system performance in terms of service completion rate, blocking and forced termination probabilities and mean queuing delay of the M2M traffic. Although the proposed model can be used to improve the system performance of M2M communication with different priorities, but the model design doesn't allow the analysis of the system behavior when facing different H2H applications with different priorities.

Based on all previous work outcomes, we are motivated to propose a CTMC model to study the H2H and M2M coexistence through a mathematical framework. Additionally, a new architecture is proposed which can help in studying and analyzing the mutual impact between M2M and H2H traffic coexistence while considering high and low priority traffics for both M2M and H2H devices. Moreover, an adaptive resource allocation is proposed also to scale the network bandwidth especially during disaster scenarios.

5.4 Coexistence Analyzer and Network Architecture for Long term evolution (CANAL)

The coexistence of H2H and M2M traffic sharing the same network is a hot topic because of its specificity and the increasing number of M2M connections. As a result, the mutual impacts on M2M and H2H traffic become inevitable. They have already been approached in several works [80] with some success for wireless connections [81] or for traffic regulation [82] but also with some limits.

Our aim is to measure the mutual impact of M2M and H2H traffics in different scenarios while maintaining a sustainable level of services with minimal congestion during emergency events by allowing both M2M and H2H devices to access the network resources efficiently. In this context, we propose a new architecture called "Coexistence Analyzer and Network Architecture for Long term evolution (CANAL)", which extends the classical functionality of the evolved Node B (eNodeB) and includes an adaptive control of bandwidth based on the M2M load status throughput, as shown in Figure 5-1.

The proposed CANAL architecture manages automatically both H2H and M2M traffics. Based on the instantaneous requests of M2M, a part of the total bandwidth is dedicated to fulfill these sudden needs. To this end, the model operates at different conditions depending on the M2M load state throughout the network. The CANAL architecture consists of a single uplink cell in one eNodeB dedicated for a LTE-A/LTE-M network.

Let RB_{H2H} be the maximum Resource Blocks (RB) reserved for LTE-A network and RB_{M2M} be the one reserved for LTE-M network. There are two types of traffics: M2M and H2H traffics. Each traffic has Low Priority (LP) and High Priority (HP) arrival rates (λ_i) assumed to be following a Poisson's distribution [62]. Let λ_{HH} and λ_{HL} the average arrival rates for H2H HP and H2H LP respectively. Similarly, $\lambda_{MH}/\lambda_{ML}$ the average arrival rates for M2M HP and M2M LP respectively.

The CANAL architecture contains also a Queuing Control Unit (QCU) consisting of four different queues: Two for H2H HP and LP traffics (H2H LP-Q, H2H HP-Q)

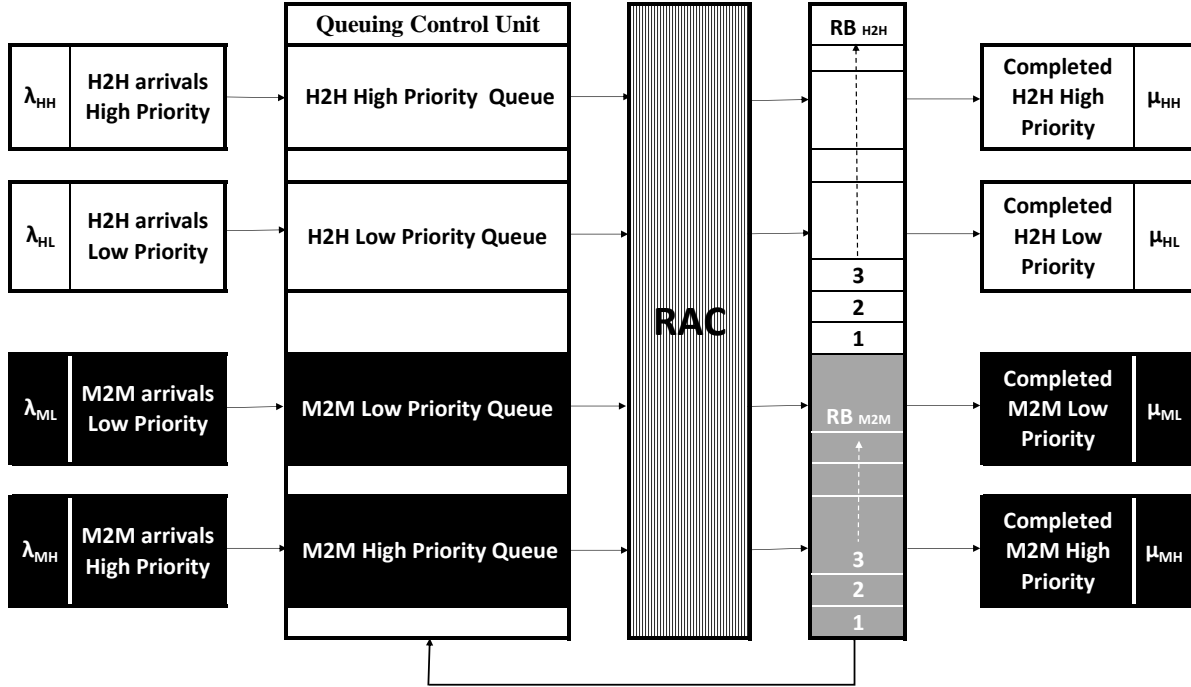


Figure 5-1: CANAL Architecture.

with queue sizes n and o respectively. Similarly, two queues for M2M Low and High Priority (M2M LP-Q, M2M HP-Q) with queue sizes p and q respectively.

Additionally, the CANAL architecture includes a Resource Allocation Control (RAC) unit used to control, manage and grant access the network requests based on the available resources. The RAC has a vital role when congestion occurs especially during emergency scenarios by dominating the resources adaptively in order to fulfill the excessive requests of M2M devices, while keeping the H2H requests within the acceptable norms.

Finally, service rates $(\mu_{HH}, \mu_{HL}, \mu_{MH}, \mu_{ML})$ will be accomplished for each traffic respectively (H2H HP, H2H LP, M2M HP and M2M LP). All notations used in our model architecture are summarized in Table 5.1.

5.4.1 Queuing Control Unit (QCU)

The CANAL architecture contains four feedback queues for the four traffic types: H2H LP, H2H HP, M2M LP and M2M HP. By applying a special priority strategy, we end

up with an efficient radio management which dominates the H2H/M2M coexistence.

The system is considered in the "full state" when the number of reserved resource blocks in a LTE-A network peaks to RB_{H2H} and/or in a LTE-M network peaks to RB_{M2M} . In Figure 5-1, the queuing process is initiated when the system reaches its cut-off point (the full state) and a new M2M/H2H device request contends to access the system. The new arrival request might be either H2H/M2M HP or H2H/M2M LP. The contention resolution is the core of the QCU by differentiating between two cases:

- *H2H/M2M LP new arrival*: If the system receives an arrival with a priority type H2H/M2M LP, the QCU checks for available resources.

If there is at least one resource available, the request is granted access to the network. If not, the request is enrolled in H2H/M2M LP-Q queue, unless if

Table 5.1: Description of notations used in CANAL architecture.

Notation	Description
<i>CANAL</i>	Coexistence Analyzer and Network Architecture for Lte
RB_{H2H}	Maximum resource blocks reserved for LTE-A
RB_{M2M}	Maximum resource blocks reserved for LTE-M
λ_{HH}	Average arrival rate for H2H High priority
λ_{HL}	Average arrival rate for H2H Low priority
λ_{MH}	Average arrival rate for M2M High priority
λ_{ML}	Average arrival rate for M2M Low priority
QCU	Queuing Control Unit
HP	High Priority
LP	Low Priority
HP-Q	High Priority Queue
LP-Q	Low Priority Queue
RAC	Resource Allocation Control unit
μ_{HH}	Completed H2H High priority service rate
μ_{HL}	Completed H2H Low priority service rate
μ_{MH}	Completed M2M High priority service rate
μ_{ML}	Completed M2M Low priority service rate
n	H2H Low priority queue size
o	H2H High priority queue size
p	M2M Low priority queue size
q	M2M High priority queue size

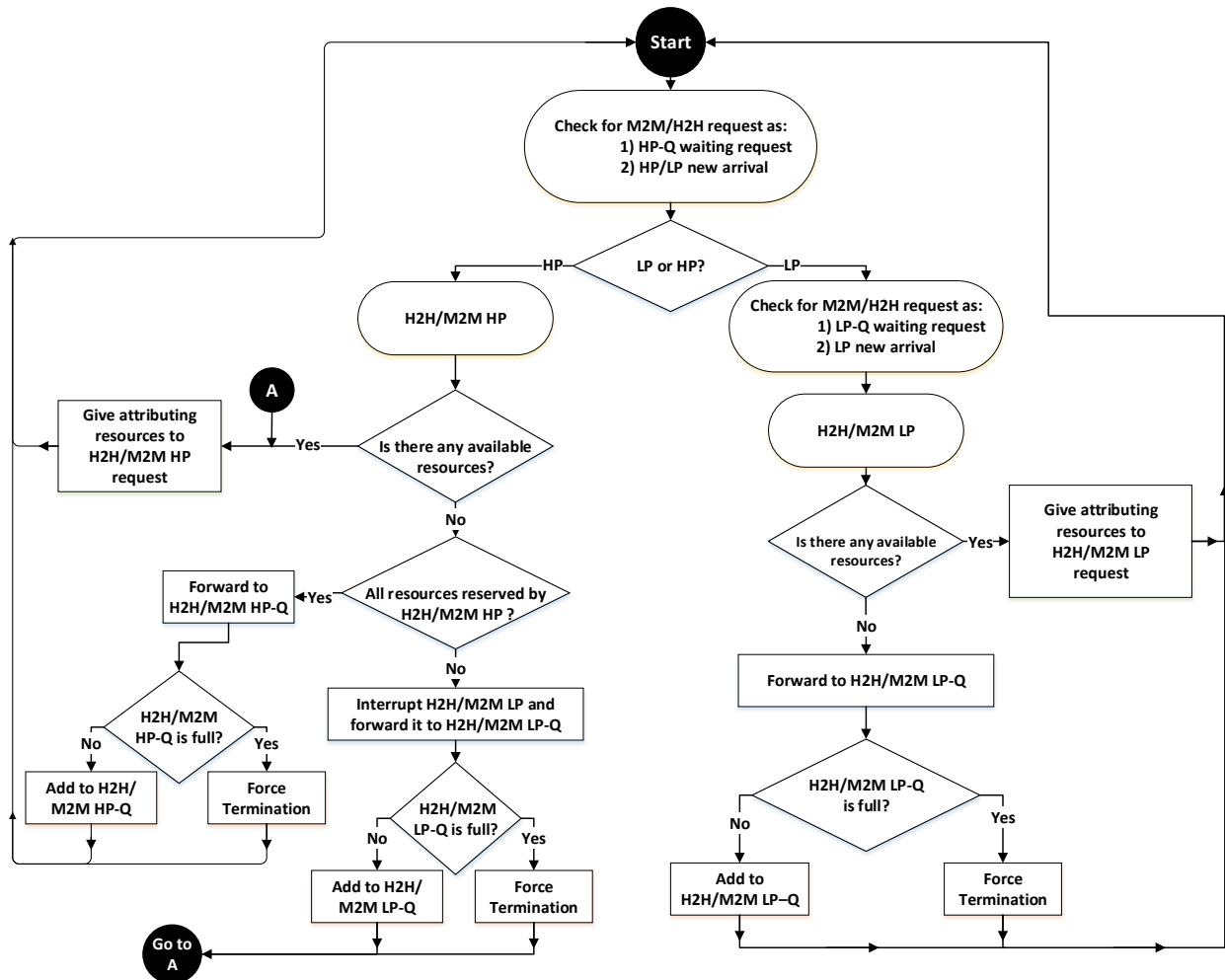


Figure 5-2: H2H/M2M HP and LP Queuing.

the queue is in the "full state" which requires a termination for the request, as shown by the flowchart presented in Figure 5-2.

- *H2H/M2M HP new arrival*: The QCU checks for available resources at each arrival with a priority type H2H/M2M HP. If there is at least one resource available, then the request is accepted and the UE is granted access to the network. If there is no available resources, the QCU verifies whether all resources are reserved by H2H/M2M HP or not;
 - If "yes", the request will be forwarded to the H2H/M2M HP-Q. However, if the queue is "full", then the request is terminated, otherwise the request is enrolled in the appropriate queue.

- If the answer is "No", an interruption command will be initiated asking for an immediate evacuation of a H2H/M2M LP resource reservation which will be pushed to the H2H/M2M LP-Q. Similarly, if the queue is "full" the request is terminated; Otherwise, the request is enrolled in the appropriate queue, as shown by the flowchart presented in Figure 5-2.

5.4.2 Resource Allocation Control (RAC) unit

The CANAL architecture contains an important part called "Resource Allocation Control (RAC)" unit, which plays a crucial role in granting access, interrupting services and managing resources for both LTE-A and LTE-M networks, especially, when it comes to an increasing storm of requests during normal and emergency scenarios caused by the synchronization behavior of M2M devices. A balancing act must be kept continuously; On one hand, the RAC might fulfill the excessive requests of M2M devices, but on the other hand any major QoS degradation on H2H traffic is not tolerated at all costs. Thus, we are in need of an efficient strategy to dominate network resources in an adaptive, stimulative and rescued way which might help to master a new eNodeB that can go ahead of the curve, as shown in Figure 5-3.

In normal situations, the CANAL architecture works similar to any legacy eNodeB in LTE-A/LTE-M networks as initial phase (INI). An initial Bandwidth-Limited (BL_0) dedicated for LTE-M traffic from the total bandwidth by reserving RB_{M2M} for M2M devices. Meanwhile, the remaining resources are reserved for H2H devices denoted by RB_{H2H} and connected to a LTE-A network.

In the case of an emergency EMG(1), a huge number of M2M Connected Devices (CD) saturates the initial LTE-M bandwidth (BL_0) by their storm briefly. When the CD data rate reaches a certain initial threshold $Th(0)$, the RAC increases the bandwidth from BL_0 to $BL_1 = 2 \times BL_0$ to allow more CD accessing the network. Similarly, if the CD data rate reaches the next threshold $Th(1)$ again caused by an additional M2M storm, the CANAL architecture adapts gradually BL_1 till reaching: $BL_M = (M + 1) \times BL_0$

Finally, if the CD data rate exceeds the last threshold $Th(M)$ an overload problem

occurs. Once the number of M2M connections starts receding, the RAC keeps receding the LTE-M bandwidth iteratively until reaching the initial phase at the end of the disastrous events, consequently, resumes operating similar to any legacy eNodeB. Following this methodology, the CANAL architecture has the adaptability to lend progressively a temporary bandwidth up to $BL_M = (M + 1) \times BL_0$ of the total H2H bandwidth to M2M devices tentative use. By implementing this proposed solution for the limited bandwidth saturation in LTE-M networks, an adaptive reallocation of the bandwidth leads for an acceptable resolution for any presumable M2M storm.

5.5 CTMC analytical methodology

Our CTMC analytical methodology consists of four steps:

1. We use CTMC as a stochastic method to describe the sequence of possible events M2M or H2H built on different states and various priorities.

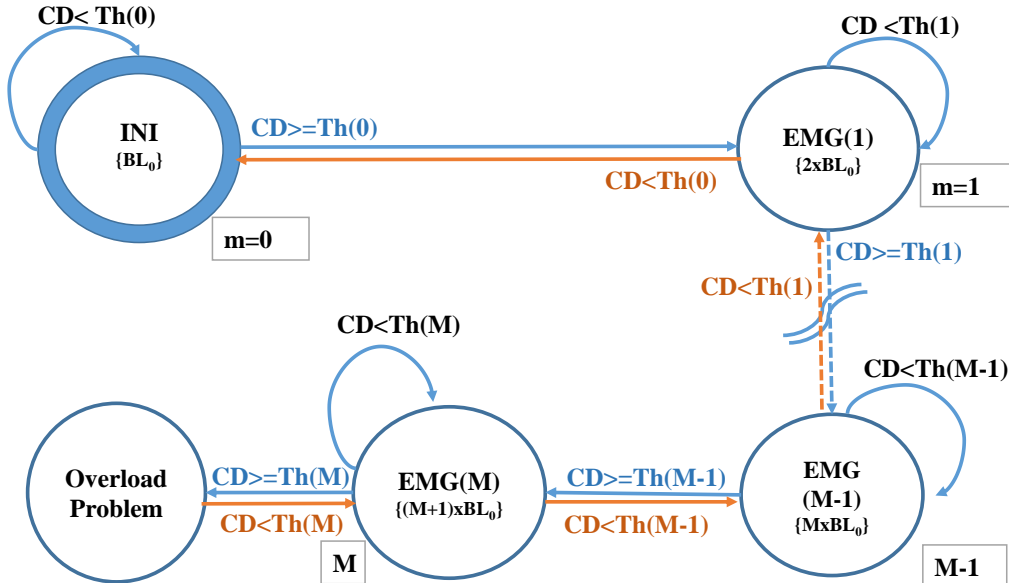


Figure 5-3: RAC phases for CANAL architecture: where CD : Connected devices, Th : Threshold, INI: Initial state, EMG: Emergency state, BL : Bandwidth Limited, m : current state number, M : Maximum state number.

2. We generate the equilibrium equations which rule the probability of transitioning from one state to another.

3. We turn these probabilities and states into a linear system and we can calculate each state probability under certain conditions.

4. Using the aforementioned probabilities, we propose some metrics to characterize the performance of the network, with the aim to evaluate M2M and H2H traffics.

5.5.1 Representing the system as a set of states

In our system states, two variables (i,j) are considered to denote the number of two ongoing services one is high priority traffic and the other is low priority traffic respectively. The system moves from one state to another if an event occurs (increase/decrease of i or j). The two traffics have two average arrival rates (λ_1, λ_2) assumed to follow a Poisson distribution and two service rates (μ_1, μ_2) assumed to follow an exponential distribution. The generic CTMC model is shown in Figure 5-4. All notations used in the analytical methodology are summarized in Table 5.2.

Table 5.2: Description of notations used in the analytical methodology.

Notation	Description
i	Number of ongoing services for HP traffic
j	Number of ongoing services for LP traffic
λ_1	Average arrival rate for HP traffic
λ_2	Average arrival rate for LP traffic
μ_1	Completed service rate for HP traffic
μ_2	Completed service rate for LP traffic
c	Number of resource blocks used in the network
s_t	Number of states
$\pi_{(i,j)}$	The probability to be in the state $S(i, j)$
\mathbf{A}	The coefficient matrix in a linear system
Π	The steady-state probability vector
SCR_{HP}	Service Completion Rate for HP traffic
SCR_{LP}	Service Completion Rate for LP traffic
R_u	Resource utilization
$E(i)$	The expected value of ongoing services for HP traffic
$E(j)$	The expected value of ongoing services for LP traffic

The number of states calculated based on the number of resources denoted (c) should meet the following condition:

$$i + j \leq c \quad (5.1)$$

The number of states (s_t) can be derived as follows [17]:

$$s_t = \sum_{i=0}^c (c - i + 1) = \frac{(c + 1)(c + 2)}{2} \quad (5.2)$$

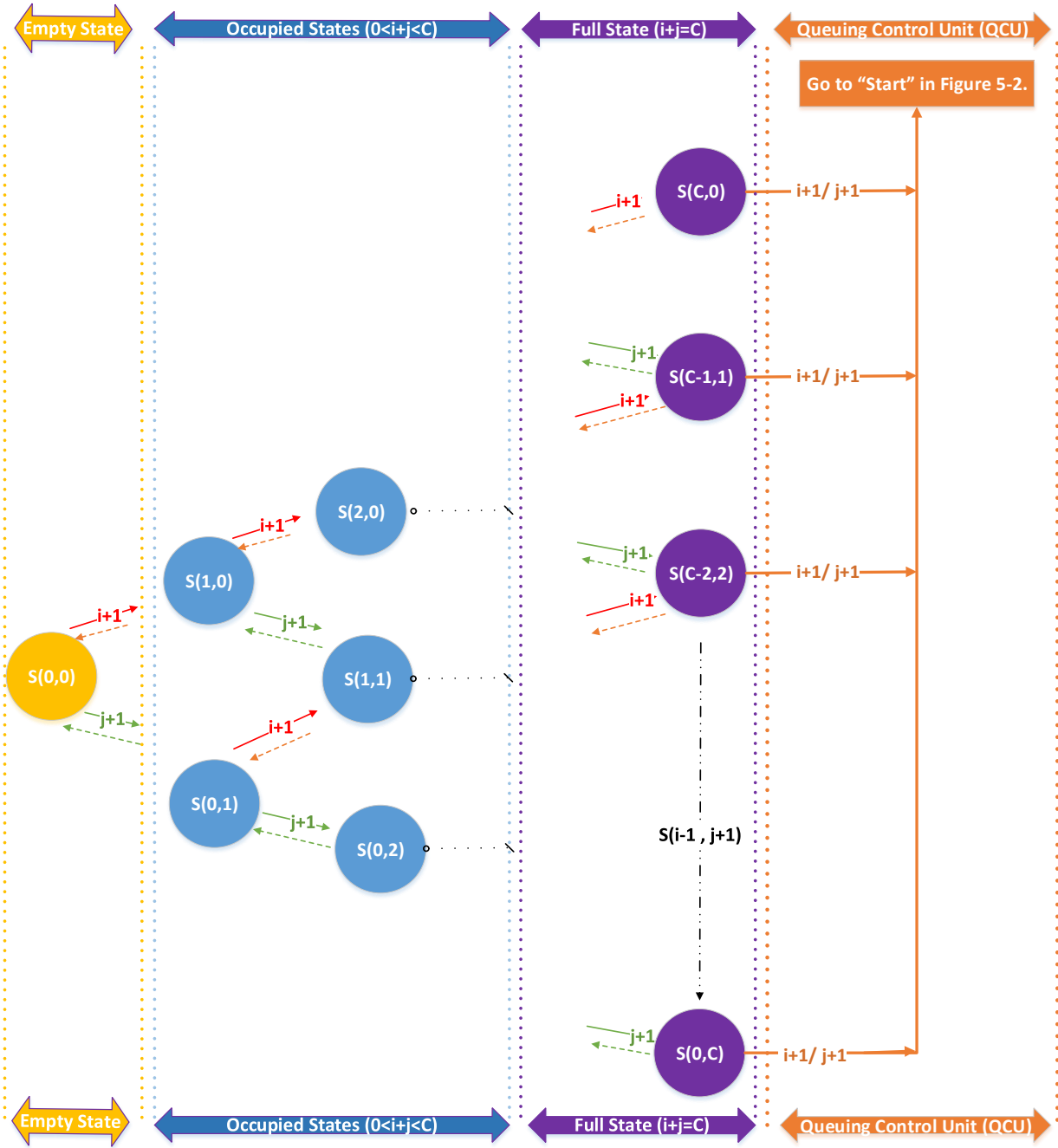


Figure 5-4: Generic CTMC Model, where i : number of ongoing services for HP requests, j : number of ongoing services for LP requests, c : maximum number of resource blocks, $S(i, j)$: a certain state denoted by i and j .

5.5.2 Generating the equilibrium equations

In the following, we assume that the observation time intervals are so small that at one time interval only one event $(i + 1, i - 1, j + 1, j - 1)$ may occur.

In this case, the system falls into one of the following three cases:

1. Case 1: "Empty State", where $i = j = 0$, includes one state $S(0, 0)$ and has the following equilibrium relationship:

$$(\lambda_1 + \lambda_2)\pi_{(0,0)} = \mu_1\pi_{(1,0)} + \mu_2\pi_{(0,1)} \quad (5.3)$$

Where $\pi_{(i,j)}$ is the probability to be in the state $S(i, j)$.

2. Case 2: "Occupied State", where $0 < i + j < c$, includes $\frac{(c-1)(c+2)}{2}$ states and has the following equilibrium equation:

$$\begin{aligned} (\lambda_1 + \lambda_2 + i\mu_1 + j\mu_2)\pi_{(i,j)} = & \lambda_1\pi_{(i-1,j)} + \lambda_2\pi_{(i,j-1)} \\ & + (i+1)\mu_1\pi_{(i+1,j)} + (j+1)\mu_2\pi_{(i,j+1)} \end{aligned} \quad (5.4)$$

3. Case 3: "Full State", where $i + j = c$, includes $(c+1)$ states and has the following equilibrium equation:

$$(i\mu_1 + j\mu_2)\pi_{(i,j)} = \lambda_1\pi_{(i-1,j)} + \lambda_2\pi_{(i,j-1)} \quad (5.5)$$

The above three equations can be summarized by the following generic equation:

$$\begin{aligned} (\alpha\lambda_1 + \alpha\lambda_2 + i\beta\mu_1 + j\beta\mu_2)\pi_{(i,j)} = & \beta\lambda_1\pi_{(i-1,j)} + \beta\lambda_2\pi_{(i,j-1)} \\ & + (i+1)\alpha\mu_1\pi_{(i+1,j)} + (j+1)\alpha\mu_2\pi_{(i,j+1)} \end{aligned} \quad (5.6)$$

Where $\alpha = 0$ indicates the "Full state" (otherwise $\alpha = 1$) and $\beta = 0$ represents the "Empty state" (otherwise $\beta = 1$).

5.5.3 Linear system solution

To recall, in our system states, (i, j) denotes the number of ongoing services, i.e. H2H and M2M. The system moves from one state to another, when a service is achieved or a new request is arriving (by increasing or decreasing i or j) with a steady-state probability $\pi_{(i,j)}$:

$$\sum_{i=0}^c \sum_{j=0}^{c-i} \pi_{(i,j)} = 1 \quad (5.7)$$

$$0 \leq \pi_{(i,j)} \leq 1 \quad (5.8)$$

The general equation (5.6) can be written in linear form:

$$\mathbf{A}\Pi = 0 \quad (5.9)$$

where the square matrix \mathbf{A} represents the coefficients of a linear system, and Π represents the steady-state probability vector:

$$\Pi = \left(\pi_{(0,0)} \quad \pi_{(0,1)} \quad \dots \quad \pi_{(c,0)} \right)^T \quad (5.10)$$

By replacing the first row of \mathbf{A} by the coefficients of equation (5.7), we obtain the following modified system:

$$\mathbf{B}\Pi = \left(1 \quad 0 \quad \dots \quad 0 \right)^T \quad (5.11)$$

Equation (5.6) can be solved while considering the number of states s_t . s_t could be

calculated using equation (5.2) in order to know the dimensions for the steady-state probability vector Π . \mathbf{B} is a $(s_t \times s_t)$ full rank matrix.

5.5.4 Performance and metrics

We use two performance metrics to validate our model:

1. *Service Completion Rate (SCR)*: It gives the number of completed requests per time interval and it is based on the service rate μ and the number of ongoing requests for a certain traffic (*e.g.* SCR_{HP} and SCR_{LP} which represent the Service Completion Rate for HP/LP traffics [64]):

$$SCR = \sum_{ij} i\mu\pi_{(i,j)} \quad (5.12)$$

Because we consider the service rate (μ) as constant in our model, (5.12) can be written:

$$SCR = \mu \sum_i i \sum_j \pi_{(i,j)} \quad (5.13)$$

As the sum of conjoint probability leads to the marginal probability, the previous equations can be re-written as follows:

$$\begin{aligned} SCR &= \mu \sum_i i\pi_{(i)} \\ &= \mu E(i) \end{aligned} \quad (5.14)$$

Where $E(i)$ represents the expected value of an ongoing service denoted by (i) .

2. *Resource utilization (R_u)*: This metric gives the probability of the system to be busy serving the arrivals in terms of the number of utilized RBs in each state:

$$\begin{aligned} R_u &= \sum_{ij} \frac{i+j}{c} \pi_{(i,j)} \\ &= \frac{\sum_{ij} i\pi_{(i,j)} + \sum_{ij} j\pi_{(i,j)}}{c} \\ &= \frac{E(i) + E(j)}{c} \end{aligned} \quad (5.15)$$

5.6 Simulations, results and discussions

In this section, we present our developed simulation model which can generate both H2H and M2M traffics with full flexibility to add queuing or priority for any traffic in order to study the mutual impact for H2H and M2M traffics.

5.6.1 M/M/1 Queuing Model

A M/M/1 Queuing Model¹ consists of a single server, in which arrivals are determined by a Poisson process and job service times which have an exponential distribution.

The proposed architecture is based on the approach presented in [83] and explained in appendix E. Many enhancements are made in order to match our CANAL architecture presented previously in section 5.4:

- The architecture consists of a single-queue and single-server system with a double traffic sources. In order to omit the role of the queuing for the moment, we consider an infinite storage capacity (in this case the queue sizes $n = o = p = q = 0$ for the different traffics, see Table 5.1).
- The architecture follows a Continuous Time Markov Chain with transition rate matrix similar to the one in section 5.5.
- Arrivals occur at rate λ according to a Poisson process and move the process from state (i) to $(i + 1)$ and/or (j) to $(j + 1)$.
- Service times have an exponential distribution with rate parameter μ in the M/M/1 queue, where $\frac{1}{\mu}$ is the mean service time.
- A single server serves H2H and M2M traffics one at a time by allowing one event only (either i or j) in the smallest time interval.

¹In a M/M/1 Queuing Model: The first part represents the input process, the second the service distribution, and the third the number of servers.

- The queue could be configured to work according to a First Input First Output (FIFO), Last Input First Output (LIFO) or by priority discipline, with a flexible buffer size.
- When the service is completed, the served request leaves the system and the number of ongoing services in the system reduces by one (i) to ($i - 1$) and/or (j) to ($j - 1$).
- The architecture makes it easy to compare empirical results with the corresponding theoretical results.

Our methodology to reach our final simulation model is to build the simplest model with one traffic first, then we develop a basic H2H and M2M traffic model in which we can fix a starting experimental point with some highlighted flaws. In the end, by working on the drawbacks of the basic model, many enhancements could be suggested as result of adding priority and queuing strategies and compare it with the basic model.

5.6.2 Single traffic simulations and results

As a first step, we start by experimenting the behavior of a single traffic system with the aim to proof the accuracy of the proposed CTMC model. Additionally, a comparison for both analytical and simulation results are conducted in order to validate the accuracy of the system by calculating:

- The Mean Absolute Error (MAE):

$$MAE = \frac{1}{n} \sum_{t=1}^n |x_{c(t)} - x_{s(t)}|$$

- The Root Mean Squared Error ($RMSE$):

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (x_{c(t)} - x_{s(t)})^2}$$

The analytical and simulation results for the probability to be in a given state $\pi_{(i)}$ and with a number of Resource Blocks $c = 6$ RBs are shown in Figure 5-5,

which shows that the accuracy of the proposed system falls into the error limit with a $MAE_{max} = 0.0592$.

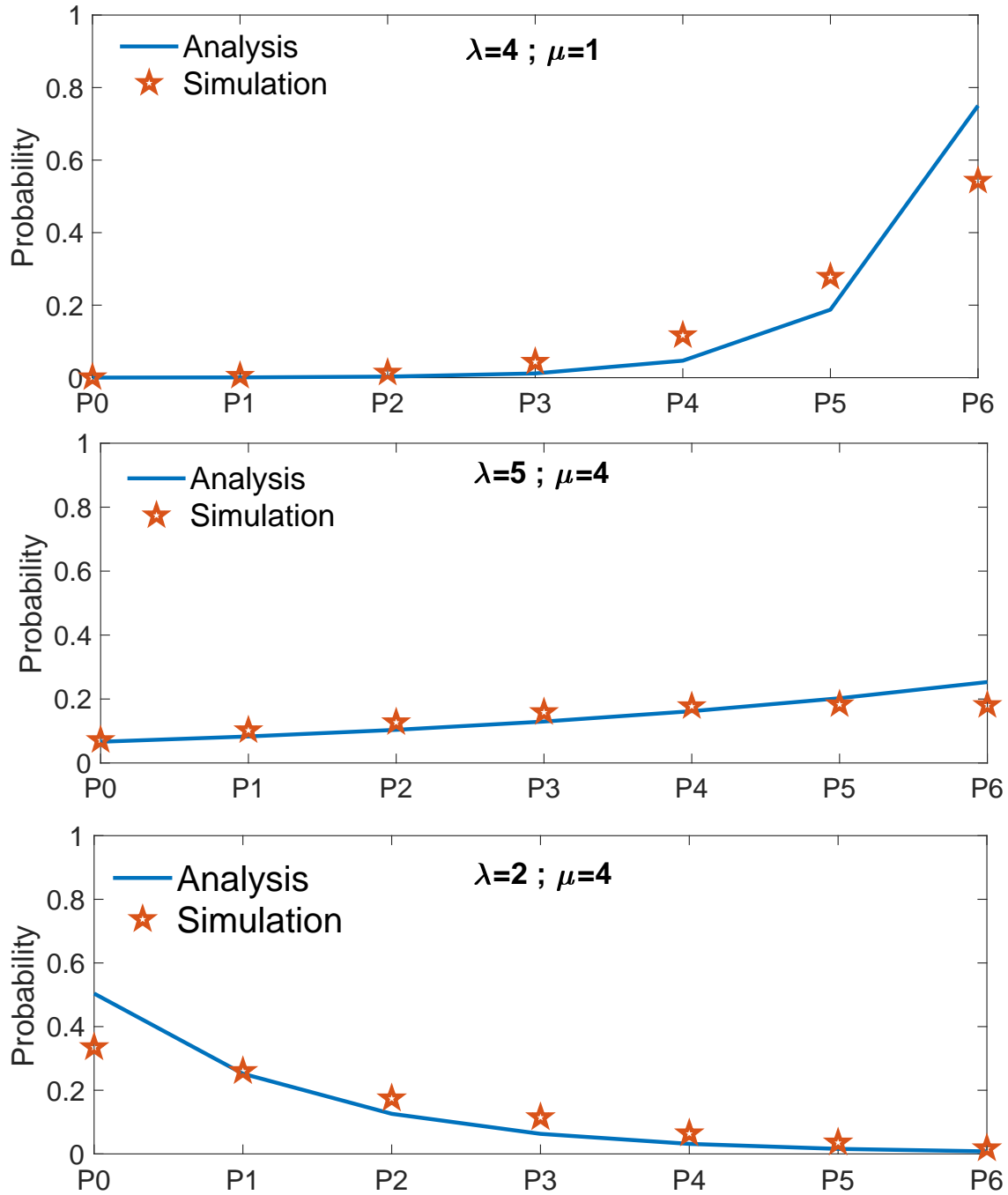


Figure 5-5: $\pi(i)$ result comparison for $c = 6$: $\pi(i)$: the probability to be in a given state $S(i, j)$, c : the maximum number of resource blocks, λ_1 : the average arrival rate, μ_1 : completed service rate.

Moreover, a slight mismatch between the analysis and simulation results could be spotted especially in the full state P6² (explained previously in section 5.4.1). This mismatch is unavoidable because in our simulator, we added two random generators for two arrivals (i and j) following a Poisson distribution. In some cases, $(i+j)$ exceeds the maximum number of available resources c . These cases should be removed in order to respect the constraint: $(i + j \leq c)$ which affects our random variables supposed following the Poisson distribution. So, we solve this issue by introducing a saturation function in our simulations. Consequently, an acceptable approximation is considering the cases (where $i + j \geq c$) as $(i + j = c)$, knowing that in our example ($c = 6$). In this case, the probability to be in the full state is more likely to happen with a noticed difference comparing to the analysis results. This non-linear behavior can explain the small mismatch.

The results obtained with ($c = 6$) drive us to do more exploration by increasing c to 25 RBs. The simulated results are shown in Figure 5-6, in which, we end up with a $MAE_{max} = 0.0177$.

To sum up, all previous exploration experiments are able to achieve a good accuracy with the aforementioned MAE and RMS, which might be considered as a valid starting point to take up our simulation model by upgrading the single traffic simulation to a basic simulation for H2H and M2M traffics.

²A full state in this case is considered when the arrival average peaks to the maximum system capacity 6.

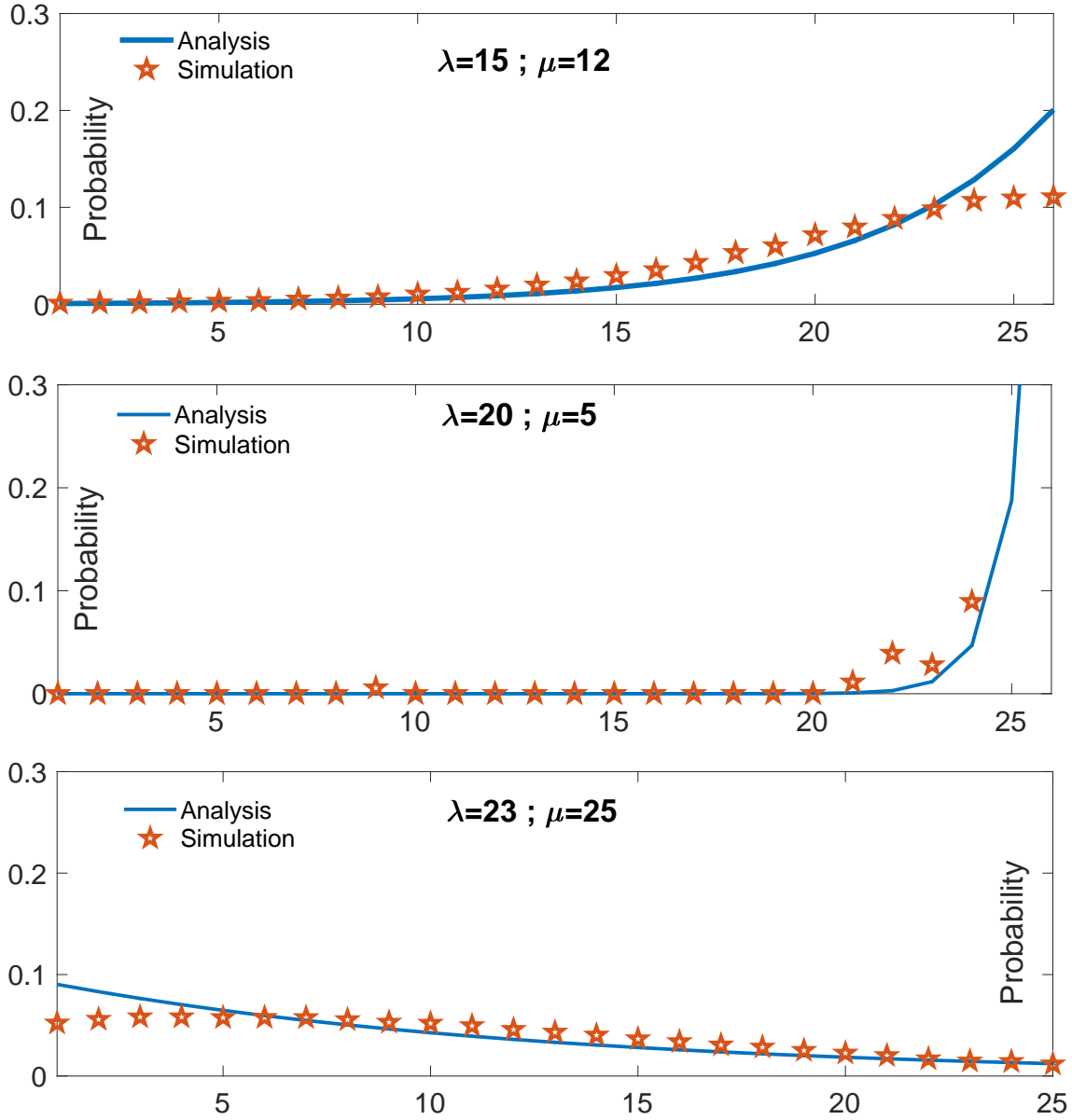


Figure 5-6: $\pi(i)$ result comparison for $c = 25$ RB: $\pi(i)$: the probability to be in a given state $S(i, j)$, c : the maximum number of resource blocks, λ_1 : the average arrival rate, μ_1 : completed service rate.

5.6.3 Basic simulations and results

In order to study the mutual impact of H2H and M2M traffics in dense area or emergency scenarios, we conduct several simulations based on the proposed architecture in [83], with the following parameters:

- We assume to have one LTE-A network using a bandwidth of 5 MHz ($c = 25$ RB) in order to stress H2H and M2M traffics to the maximum.
- Each traffic has an average arrival rate (λ_1, λ_2) with a service rate (μ_1, μ_2) .
- H2H and M2M traffics have the same priority.
- A FIFO queue type is used with an infinite capacity size.
- Simulation duration = 1000 Seconds.

1. **Dense area scenario:** In urban and dense areas, a huge number of M2M and H2H devices may exchange their payloads exponentially. Consequently, we consider a fixed arrival rate of H2H requests ($\lambda_1 = 5$) and incremental arrival rate of M2M requests

$2 < \lambda_2 \leq 20$, while $\mu_1 = 0.5$ and $\mu_2 = 1$. From the results shown in Figure 5-7 and by calculating the percentage of served requests comparing to the total arrivals, we can conclude the following:

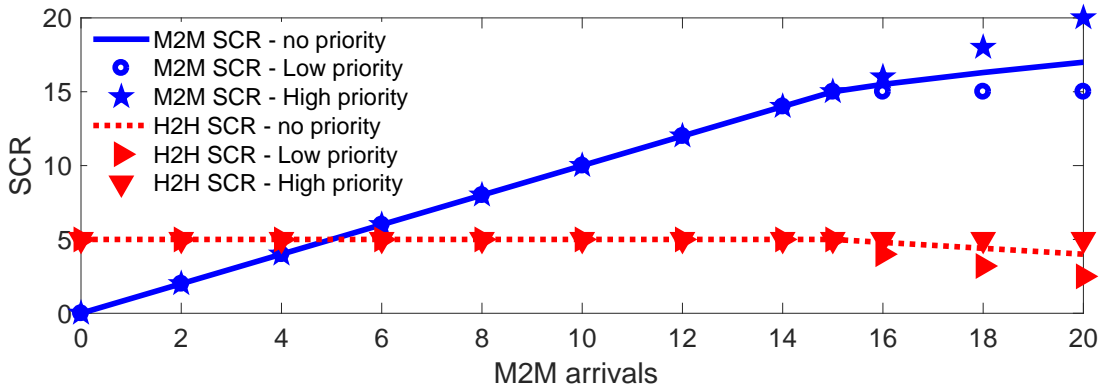


Figure 5-7: SCR_{H2H} vs SCR_{M2M} in dense area cases.

- When $\lambda_2 \leq 15$: The system is able to serve all M2M (dotted line) and H2H (straight line) requests, because H2H arrivals, with $\lambda_1 = 5$ and a service rate $\mu_1 = 0.5$, occupy an average of 5 resources from the 25 total resources for two consecutive time intervals, which means on the second time interval it occupies 10 resources (5 previous arrivals and 5 new arrivals). Consequently, it remains around 15 resources per time interval for M2M arrivals with a maximum $\lambda_2 = 15$ and a service rate $\mu_2 = 1$.
- When $15 < \lambda_2 \leq 20$: A degradation on both M2M and H2H service completion rate can be realized, because in our assumption, they have similar priority and the total number of requests are much more than available resources $c = 25$. At the peak ($\lambda_2 = 20$), only 4K of 5K total H2H requests are served (80%) and 17K of 20K total M2M requests are served (85%). Because an average of 4 from the 25 total resources will be occupied by H2H traffic for two consecutive time intervals (actually 4 previous arrivals and 4 new arrivals), and the remaining 17 resources will be used by M2M traffic.
- The high utilization of the system ($R_u = 100\%$) requires a queuing strategy to fulfill the excessive requests and to minimize the number of forced terminated services, as suggested in section 5.4.1.

2. **Emergency scenario:** In emergency cases, a M2M signaling storm is definitely expected as a result of connected devices, which leads inevitably to a M2M congestion. This congestion caused by the simultaneous synchronization affects both H2H and M2M devices. In this scenario, we consider the following system parameters: $\lambda_1 = 50$, $50 \leq \lambda_2 \leq 200$, $\mu_1 = 0.5$ and $\mu_2 = 1$. From the results shown in Figure 5-8, when $\lambda_2 = 50$ a degradation on both M2M (rectangles) and H2H (dotted line) service completion rate can be realized, because they have similar priority and the total number of requests are much more than the available resources $c = 25$. Only 8K of 50K total H2H requests are served (16%) and 8K of 50K total M2M requests are served (16%), because an average of 8

from the 25 total resources will be occupied by H2H traffic for two consecutive time intervals, and the remaining 8 resources will be used by M2M traffic each time interval.

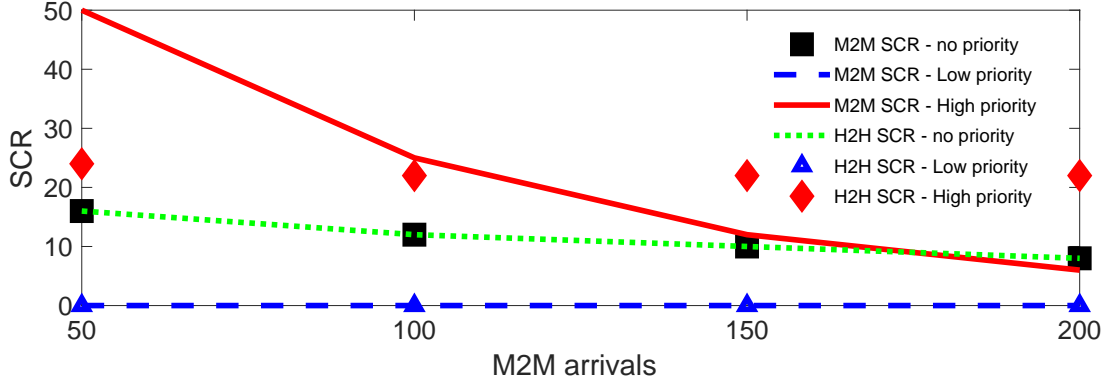


Figure 5-8: SCR_{H2H} vs SCR_{M2M} in emergency cases.

By increasing $\lambda_2 = 100, 150, 200$, we end up with the results shown in Table 5.3.

Table 5.3: Emergency scenario results.

λ_2	$SCR_{H2H/M2M}$	$Completed_{H2H}$	$Completed_{M2M}$
100	12%	6K	12K
150	10%	5K	15K
200	8%	4K	16K

With a close look to the results in Table 5.3, we realize the huge degradation on the SCR for both H2H and M2M; Only [8-12]% of the services could be accomplished during emergency scenarios.

5.6.4 M/M/1 results

In the previous section 5.6.3, two different scenarios (dense area and emergency) are conducted to measure the H2H and M2M mutual impact but without neither considering a prioritize system nor applying queuing strategy to ensure the QoS of

different traffics. Now, it is time to consider two scenarios with similar parameters using a prioritize system as follows:

1. **Prioritize dense area scenario:** In this simulation, we emphasis similar case to the one considered in section 1; While assigning a higher priority for one traffic, we study the trade off between cost or gain on H2H and M2M traffics. From results shown in Figure 5-7, we realize that the system is able to serve all M2M and H2H requests as long as $\lambda_2 \leq 15$. After this point, the system maintains the higher priority traffic towards a degradation on the lower priority traffic.
2. **Prioritize emergency scenario:** Similar to the case in section 2 but with different M2M and H2H priorities, the results shown in Figure 5-8 spot that the system maintains the higher priority traffic with a total elimination of the lower priority traffic.

To sum up, the simulation results show that using a prioritize LTE-A system for both M2M and H2H traffics could handle more requests in dense area cases by maintaining the higher priority traffic without any degradation while reducing the completion rate of the lower priority traffic. Meanwhile, in emergency cases, the higher priority traffic takes precedence over the lower priority traffic with an improvement on its completion rate while totally eliminating the low priority traffic; In such cases, the non priority traffic strategy has a competitive advantage because it keeps H2H and M2M traffics working at the same time.

Finally, we want to recall that a part of the results of this chapter has been published in [18].

5.7 Conclusion

In this chapter, we address the saturation problem caused by the expected huge number of M2M devices which leads to remarkable impacts on both M2M and H2H

traffics. We have proposed an enhanced architecture designed for LTE-A/LTE-M networks in order to fulfill H2H/M2M traffic coexistence supported with various priority strategies to satisfy the QoS for each traffic. A queuing strategy is enforced by a queuing control unit, which rules four queues allocated for different traffic types. A CTMC model is proposed as a stochastic process tool to characterize the H2H/M2M coexistence based on analytical equations. Our simulation results matched with the CTMC analytical model which corroborates the proposed model. Moreover, they show that using a prioritize LTE-A system for both M2M and H2H traffics is more convenient in dense area scenarios; While in emergency cases, it is more appropriate to use a non prioritize traffic strategy to keep both H2H and M2M traffics working properly at the same time.

In this chapter, a homogeneous number of resources has been considered for all applications; In the next chapter, we will study the heterogeneity in different applications which reflects the actual behavior of various IoT applications.

Chapter 6

CTMC modeling for M2M/H2H coexistence in an NB-IoT Adaptive eNodeB

Glossary

3GPP: 3rd Generation Partnership Project

A-eNB: Adaptive evolved Node B

CP: Control Plan

CTMC: Continuous-Time Markov Chain

GSM: Global System for Mobile

H2H: Human-to-Human

HARQ: Hybrid Automatic Repeat Request

IoT: Internet of Things

LTE: Long Term Evolution

LTE-A: Long Term Evolution-Advanced

M2M: Machine-to-Machine

MAE: Mean Absolute Error

MAR: Mobile Autonomous Reporting

NB-IoT: Narrow Band-IoT

OFDMA: Orthogonal Frequency-Division Multiple Access

PRB: Physical Resource Block

QPSK: Quadrature Phase-Shift Keying

QoS: Quality of Service

RE: Resource Element

RMSE: Root Mean Squared Error

RRC: Radio Resource Control

SDT: Small Data Transmission scheme

UE: User Equipment

6.1 Introduction

Throughout the last decade, Internet of Things (IoT) has paved the way to a prominent game changer known as Machine-to-Machine (M2M) communications. This new trend influences the human lives by delegating control in many fields to autonomous and semi-autonomous machines. Promoting dummy machines to smart adaptive cognitive-machines, allowing them to monitor, track and control various parts in civil lives over distance, thus reducing the human resources, mistakes and mood-swing drawbacks in future businesses. Following this aspect, the IoT becomes a promising technology with around 50 Billion connections in 2020 which enforces an unavoidable coexistence among Human-to-Human (H2H) and M2M traffics in one Long Term Evolution-Advanced (LTE-A) network [67]; Consequently, an efficient radio access strategy becomes one of the most challenges for mobile operators, researchers and the 3rd Generation Partnership Project (3GPP) community [68]. They sound ultimately keen on conducting several studies and researches to identify the mutual impact among M2M and H2H communications.

LTE-A was coined initially to support H2H communications (*e.g.*, web-browsing, phone calls, internet televisions, *etc.*). M2M devices (*e.g.*, smart cities, mobile health, smart meters, *etc.*), with their sporadic small packet sizes, pour their payloads into the same LTE-A channel but with different aspects and specifications.

On one hand, H2H traffic enforces telecommunication operators to broaden the LTE-A bandwidth, thus providing higher data rates which implies a significant complexity on the future model design. On the other hand, M2M traffic requires a narrower bandwidth due to low data rates which paves the way to low complexity models. Due to these differences, it would be an imperfect match for these two traffics to share the same LTE-A bandwidth. 3GPP introduces a licensed spectrum technology called Narrowband-Internet of Things (NB-IoT). NB-IoT technology occupies 180 *KHz* within the LTE-A bandwidth, which helps in a significant reduction on model complexity consequently a minimum cost model. But with a diversity of applications, a flood of devices trying to access the network and send their payloads becomes unavoidable.

Inevitably, the synchronized manner of certain devices serving the same type of applications will cause a performance degradation on NB-IoT services.

In this chapter, we propose the concept of an Adaptive eNodeB (A-eNB) for NB-IoT technology while addressing all aforementioned challenges. Our proposed A-eNB can solve the M2M overload congestion gradually, while maintaining the Quality of Service (QoS) of H2H traffic within the acceptable standards. The network adaptation is provided through a dynamic NB-IoT bandwidth re-allocation with the objective of maximizing the number of M2M connections while maintaining H2H traffic.

The main contributions of this chapter are the following:

- A new Adaptive eNodeB architecture for NB-IoT technology.
- An evaluation of the A-eNB on H2H and M2M traffics.
- A Continuous Time Markov Chain (CTMC) model used as stochastic process tool to characterize H2H/M2M coexistence based on analytic equations.
- Many simulations constructed to validate our model.

6.2 NB-IoT networks

6.2.1 NB-IoT overview

To recall, in section 1.1.2, the IoT requirements are: deployment flexibility, low device complexity, long battery lifetime, support of massive numbers of devices in a cell and significant coverage extension.

In Release 13 [84], 3GPP introduces NB-IoT in order to satisfy the above IoT requirements [5] as follows:

- Deep coverage: By increasing the number of repetitions, an enhancement on NB-IoT coverage could be achieved with a maximum coupling loss 20 dB higher than LTE Release 12.
- A long battery life: For a device with 164 dB coupling loss, a 10-year battery life can be reached if the User Equipment (UE) transmits 200 bytes of data per day on average [59].
- A low complexity and low device cost: NB-IoT enables low complexity UE implementation by the designs highlighted below:
 - a) Significantly reduced transport block sizes.
 - b) Only single antenna is required at the UE.
 - c) Support only single Hybrid Automatic Repeat Request (HARQ) process.
 - d) No Connected mode mobility measurement is required, because a UE only needs to perform mobility measurement during an idle mode.
 - e) Allow only a half-duplex frequency-division duplexing operation.
- A massive number of devices support: NB-IoT with one Physical Resource Block (PRB) supports more than 52,500 UEs per cell [59].

NB-IoT was designed to achieve excellent coexistence performance with the legacy Global System for Mobile (GSM) and Long Term Evolution (LTE) technologies with a requirement of 180 kHz as a system bandwidth. This choice enables a number of deployment options. A GSM operator can replace one GSM carrier (200 kHz) with

NB-IoT as a standalone carrier. A LTE operator can deploy NB-IoT inside a LTE carrier by allocating one of the PRBs of 180 kHz to NB-IoT either within the LTE spectrum or in the guard band. These different deployment scenarios are illustrated in Figure 6-1.

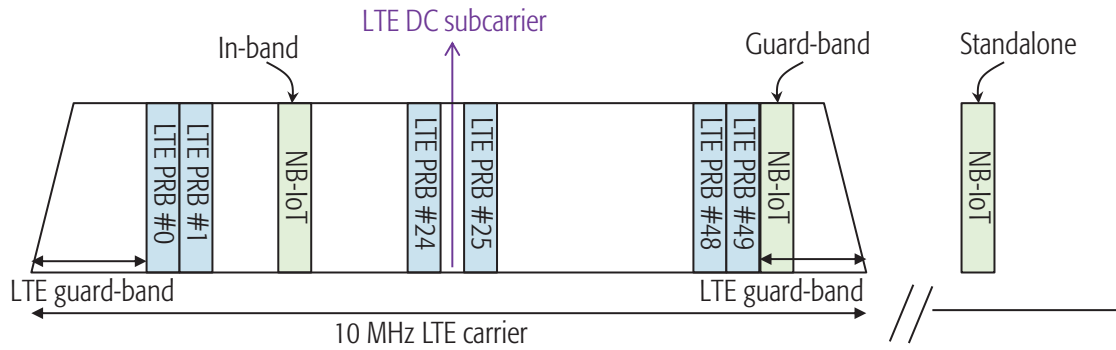


Figure 6-1: Examples of NB-IoT stand-alone deployment and LTE in-band and guard-band deployments [5].

6.2.2 NB-IoT Data-rate

The previously computed LTE-A maximum data-rate, in section 3.2.1, is usually dedicated to H2H users. However, NB-IoT technology in 3GPP Release 13, dedicates a 180 kHz of the total bandwidth for M2M communications, as shown in Figure 6-2.

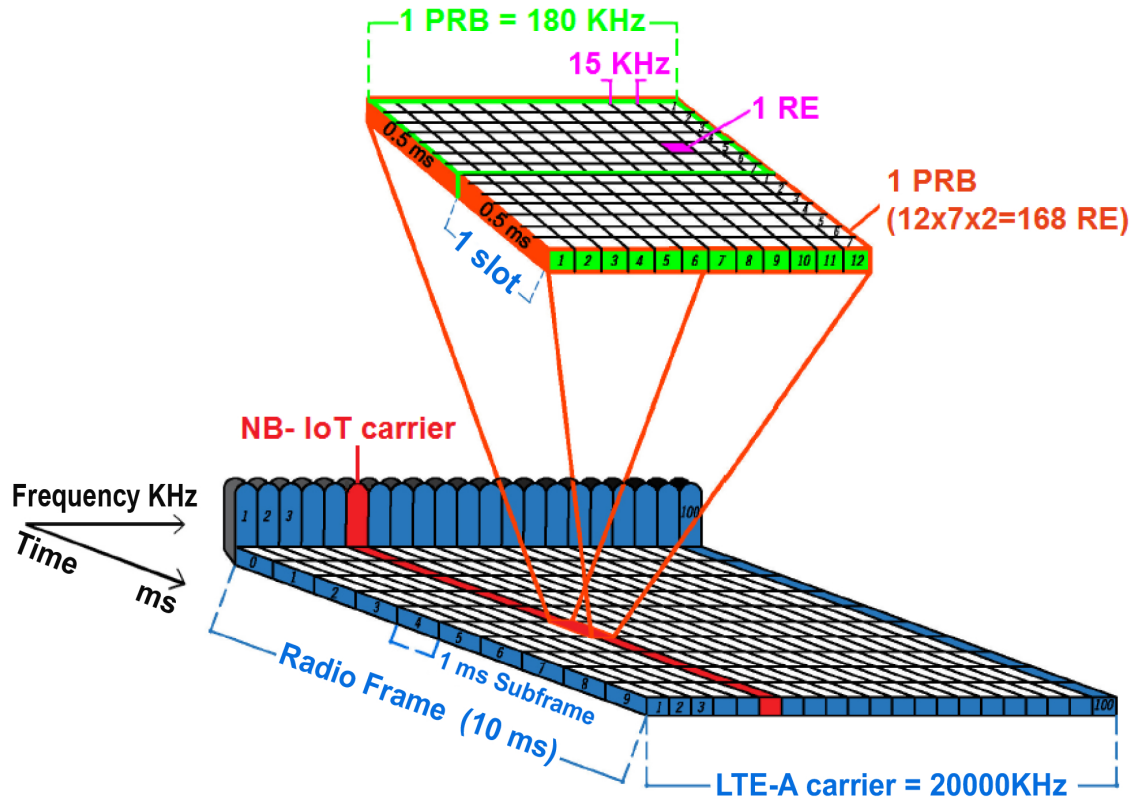


Figure 6-2: Bandwidth Limitation (BL) for NB-IoT within LTE-A: PRB: Physical Resource Block, RE: Resource Element.

Following similar calculation as by equation (3.1), but with 1 PRB, Quadrature Phase-Shift Keying (QPSK) modulation and in half-duplex mode; the maximum data-rate in NB-IoT technology is reduced to 150 Kbps for the M2M traffic.

6.2.3 NB-IoT bandwidth limitations

If we compare the storm rates in the case-study presented in section 4.2.3 and summarized in Table 6.1 to the maximum data-rate in NB-IoT technology (150 Kbps), we conclude that the available bandwidth can suffer from a huge degradation which requires an efficient solution as discussed in section 6.4.

Table 6.1: Case-study M2M storms.

Group #	M2M device Type	Message size (Bytes)	Rate (msg/day)	Number of devices (K)	Storm rate (Kbps)	Number of storms (Storm/day)
1	Asset tracking	50	100	20	1600	500
2	Assisted medical	100	8	20	3200	40
3	Env. monitoring	200	24	20	6400	120

6.3 Existing solutions found in the literature

As we mentioned before, in the near future, it is expected to have billions of global connections around the world [85]. When the number of devices peaked exponentially, a transformation effect on the bandwidth becomes a must. This effect is accumulative either as a result of connected devices load with higher bandwidths for huge time horizons, or as consequence of a burst of autonomous devices with their sporadic payloads.

In [86], a proposed Small Data Transmission scheme (SDT) has been introduced to enable devices in an idle state to transmit a small data packet without following a Radio Resource Control (RRC) connection setup process. Although, an improvement on the maximum supported devices which have insufficient radio resources in NB-IoT can be achieved, but on the flip side, the larger the number of devices the greater the bandwidth needed, that wasn't discussed.

In [58], a real drive test measurement was conducted with an application based on a UE performing one Up-Link (UL) transmission with a payload of 128 or 256 bytes similar to the Mobile Autonomous Reporting (MAR) traffic model. The authors of [58] concluded that NB-IoT technology can provide coverage for more than 95% of the devices due to its maximum coupling loss (164 dB).

Based on MAR traffic models, the results in [12] show that more than 72K devices can be supported by NB-IoT cells. However, the 72K result triggers a response of many questions:

- What if all these devices are trying to send their payloads simultaneously during a disaster?
- What is the saturation point of the limited NB-IoT bandwidth?

- How can we deal with the expected saturation during a disaster situation?

In [84], a Control Plan (CP) solution has been specified in NB-IoT technology, where a device in an idle state transmits a data packet through the RRC connection setup procedure to reduce the signaling overhead. But if we take into consideration that:

- a) The maximum number of repetition for each RRC message is 128 repetitions during one transmission [87].
- b) The signaling overhead on the up-link radio resources can be significantly increased by a large number of devices in NB-IoT networks, we conclude clearly that the radio resources to support loads of devices placed in deep coverage or cell edge can be insufficient and requires additional research to address this problem efficiently.

In [88], the authors propose an energy-efficient resource allocation algorithm to optimize the energy efficiency while guaranteeing the QoS provisioning through a joint channel selection and a power allocation design. First, the available resources are scheduled by each UE in an energy-efficient way. Since UEs are only interested in improving their own performance, a game-theory approach is proposed with the aim to model the distributed resource allocation problem as a non-cooperative game. In [89], an adaptive dropout deep computation model with crowd-sourcing is presented for big data feature learning in industrial Internet of Things. The dropout is proposed to prevent over-fitting for deep learning models. It is especially effective to learn the large-scale deep neural networks with a small number of training samples.

One of the targets of 3GPP to introduce NB-IoT technology, is to serve better deep coverage IoT devices (*e.g.* basement). Using a repetition transmission scheme in NB-IoT technology can be considered as a serious burden to its limited bandwidth. All previous studies focus on the capacity and the way to connect a huge number of devices in a NB-IoT system effectively. This is an important step towards IoT requirements, but there is a prominent aspect that merits to be improved; It is bandwidth and its limitation. Therefore, it can be essential to study the efficient use of

the bandwidth in NB-IoT technology or to find an adaptive way to extend it with a minimal cost on LTE-A bandwidth. In this chapter, we are motivated to study the limitation of a NB-IoT bandwidth while proposing an effective solution for it.

6.4 Adaptive eNodeB for NB-IoT

All the above calculations shed the light on the importance of an intelligent solution which operates effectively to absorb any expected signaling storm. To this end, we propose a promising approach, which extends the classical functionality of eNodeB and includes an adaptive control of the bandwidth based on the M2M load status throughout the network. The proposed A-eNB manages automatically both H2H and M2M traffics. Based on the instantaneous requests of M2M, a part of the total bandwidth will be dedicated to fulfill these sudden needs.

The Adaptive eNodeB for NB-IoT networks depicted in Figure 6-3, is designed to operate in different conditions, depending on the number of M2M connected devices C_d throughout the network.

A) Initial stage "INI"

In normal situations, an A-eNB works in a similar way to any legacy eNodeB in LTE-A/NB-IoT networks. A part of the total bandwidth is dedicated to NB-IoT traffic. For instance, a total of 20 MHz (100 PRBs) will be divided into a basic Bandwidth-Limited ($BL_0 = 200KHz$) reserved for M2M devices and the remaining bandwidth *i.e.* 19.8 MHz (99 PRBs) are dedicated for H2H devices.

During this stage, M2M devices are allowed to access BL_0 reserved for NB-IoT to send their payloads via different IoT applications, while H2H devices use the remaining bandwidth to exchange their information.

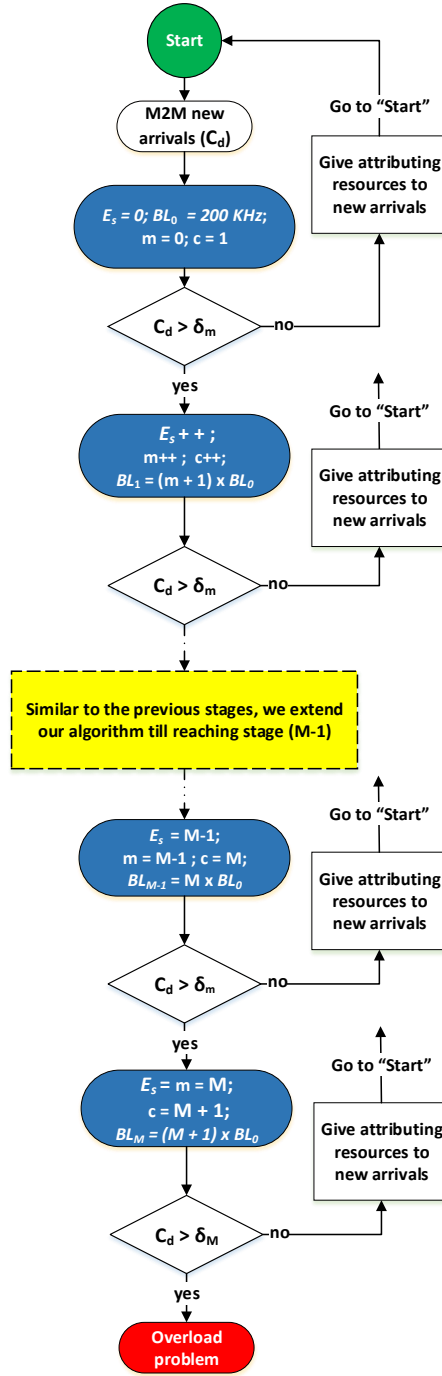


Figure 6-3: Flow chart for NB-IoT Adaptive eNode-B over a normal cycle: E_s : Emergency stage number, BL_m : Bandwidth Limited at the m^{th} stage, $BL_0 = 200KHz$: Initial Bandwidth Limited, m : current stage number, M : Maximum stage number, C_d : number of M2M connected devices, δ_m : threshold at the m^{th} stage, c : number of system resources.

B) Emergency stages "EMG(m)"

In the case of a disaster or an emergency, a huge number of M2M devices saturates the initial NB-IoT bandwidth (BL_0) by their signaling storm briefly. When C_d crosses the initial saturation threshold δ_0 (configurable), the bandwidth dedicated to M2M starts to increase from $BL_0 = 200KHz$ to $BL_1 = 400KHz$, to allow more M2M devices to access the network.

If C_d crosses a next saturation threshold δ_1 , caused by an additional M2M signaling storm, the A-eNB adapts gradually its bandwidth from BL_1 till it reaches:

$BL_{(M)} = [(M + 1) \times BL_0]$, as shown in Figure 6-4.

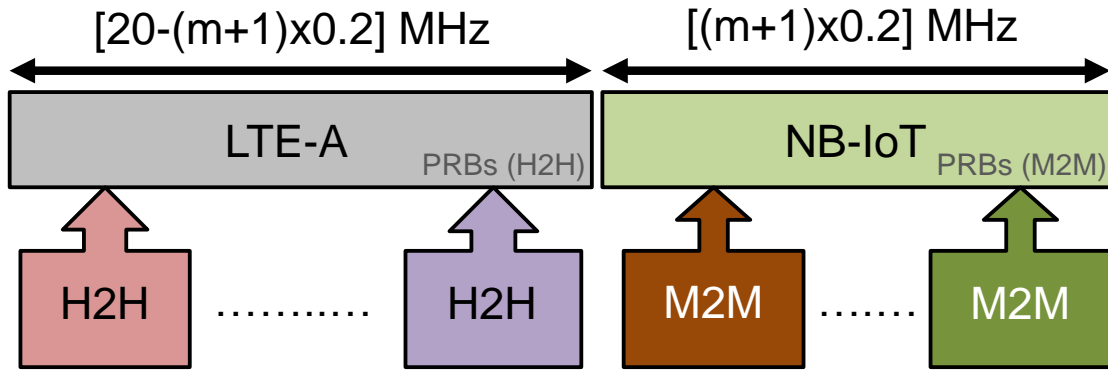


Figure 6-4: Adaptive bandwidth during an emergency scenario: m : current stage number (0 to M), PRB: Physical Resource Block.

For a given stage m , we can calculate:

- Total number of PRBs reserved for M2M devices:

$$p_{m2m} = m + 1$$

- Maximum M2M data-rate: following equation (3.1) the maximum M2M data-rate could be estimated as follows:

$$R_{m2m} = n \times o \times l \times p_{m2m} \times e \quad (6.1)$$

Where R_{m2m} is the maximum data rate for M2M traffic, n represents the number of sub-carriers, o is the number of Orthogonal Frequency-Division Multiple Access (OFDMA) symbols, l is the number of slots, p_{m2m} represents the number of PRBs reserved for M2M traffic, e is the number of carried bits per Resource Element (RE).

- Total number of PRBs reserved for H2H devices:

$$p_{h2h} = 100 - p_{m2m}$$

- Maximum H2H data-rate: similar to equation (6.1):

$$R_{h2h} = n \times o \times l \times p_{h2h} \times e \quad (6.2)$$

where R_{h2h} is the maximum data rate for H2H traffic, p_{h2h} represents the number of PRBs reserved for H2H traffic.

C) Reset stage "RST"

Once the number of M2M connections starts decreasing, A-eNB reduces iteratively the NB-IoT bandwidth BL_m until reaching the initial state, BL_0 , at the end of the disastrous events and consequently starts similarly operating to any traditional e-NodeB.

Following this methodology, the A-eNB has the adaptability and robustness to lend progressively a temporary bandwidth: $(M+1)$ folds of 200 KHz to M2M devices tentative use during disaster scenario. Inevitably, this solution can soak up most of M2M storms with a minimal impact on H2H devices.

6.5 CTMC analytical model

As we previously discussed in section 5.5, a Markov chain is a stochastic model describing a sequence of possible events for a system that could be represented as a set

of states. Each state has a steady-state probability, $\pi_{(i,j,k,etc.)}$, which represents the probability to be in a given state $s(i, j, k, etc.)$.

In our CTMC model, any state $s(i, j, k, etc.)$ represents the system state after receiving an ongoing request or a termination of services for an application.

Our methodology to build this model consists of four steps:

1. We use CTMC as a stochastic method to describe the sequence of possible events built on different states and probabilities for various applications (*e.g.*, H2H and M2M requests).
2. We generate the equilibrium equations which rule the probability of transitioning from one state to another.
3. By using these probabilities putting in a linear system, we calculate all steady-state probabilities.
4. Using the aforementioned probabilities, we propose some metrics to characterize the performance of the network, with the aim to evaluate M2M and H2H traffics.

6.5.1 CTMC model for one traffic

As a preliminary step, we use a CTMC model for a single traffic which mimics a simulation model proposed in [83]. Then, in section 6.6.2, an analysis for both the analytical and simulation results are conducted in order to validate the accuracy of the simulation model that will be used in our scenarios. In our CTMC model, we consider the following assumptions:

- Arrivals suppose to follow a Poisson's distribution with an average arrival rate (λ) and move the process from state (i) to $(i + 1)$.
- Service times assumed to follow an exponential distribution with a rate parameter (μ).
- A single server ($c = 1$) can only serve one event in a smallest time interval¹.

¹The time interval is considered as the minimum period required by the fastest unit to accomplish a request.

- When the service is completed, the served request leaves the system and the number of ongoing services in the system reduced by one from (i) to $(i - 1)$.

Representing the model as a set of states

The system moves from one state to another if an event occurs (increasing/decreasing of i), as shown in Figure 6-5.

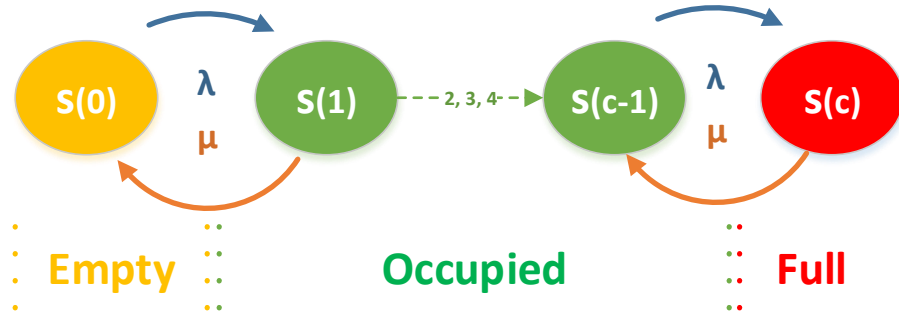


Figure 6-5: Simple CTMC states: s : state, c : system resources, λ : average arrival rate, μ : service rate.

Generating the equilibrium equations

By considering new arrival events with an average rate λ and a service rate μ , we can create the equilibrium equations based on the model proposed in [90]. The system can be in one of the following three cases:

- Empty state: $i = 0$, includes one state only and its equilibrium equation is given:

$$\lambda\pi_{(i)} = \mu\pi_{(i+1)} \tag{6.3}$$

$\pi_{(i)}$: is the steady-state probability in a given i state.

- Occupied state: $0 < i < c$, when the equilibrium equation depends on two states (entering the state and leaving the state):

$$(\lambda + \mu)\pi_{(i)} = \lambda\pi_{(i-1)} + \mu\pi_{(i+1)} \quad (6.4)$$

- Full state: $i = c$, includes one state and its equilibrium equation is given as follows:

$$\mu\pi_{(i)} = \lambda\pi_{(i-1)} \quad (6.5)$$

The three previous equations can be represented in a general equation as follows:

$$(\alpha\lambda + \beta\mu)\pi_{(i)} = \beta\lambda\pi_{(i-1)} + \alpha\mu\pi_{(i+1)} \quad (6.6)$$

Where $\alpha = 0$ indicates the Full state (otherwise $\alpha = 1$) and $\beta = 0$ represents the Empty state (otherwise $\beta = 1$).

Linear system solution

The previous general equation (6.6) can be written similar to (5.9).

where $\Pi = [\pi_{(0)}, \pi_{(1)}, \dots, \pi_{(c)}]^T$: represents the steady-state probability vector, and \mathbf{A} is a $(c + 1) \times (c + 1)$ matrix.

The linear system in (5.9) should be solved while respecting the following constraints:

$$\sum_{i=0}^c \pi_{(i)} = 1 \quad (6.7)$$

$$0 \leq \pi_{(i)} \leq 1 \tag{6.8}$$

According to the nature of equilibrium equation system, shown in equations (6.3), (6.4), (6.5) and (6.6), we can consider that: $Rank(\mathbf{A}) = c$.

By replacing the first row of \mathbf{A} by the coefficient of equation (6.7), we can easily obtain the same equation as (5.11). Where \mathbf{B} is a full rank $(c + 1) \times (c + 1)$ matrix.

Indeed, following a huge number of simulations constructed with $c = 3$ and $c = 6$ while changing $\lambda_1, \lambda_2 \in [1, 2, \dots, 10]$ and $\mu_1, \mu_2 \in [1, 2, \dots, 10]$ in each simulation, we observed a full rank \mathbf{B} matrix and that the solution obtained from (5.11) satisfies the constraint (6.8).

6.5.2 CTMC model for M2M/H2H traffics

Now, we extend our model to deal with two variables (i, j) denoting the number of ongoing services for two traffics H2H and M2M respectively.

Representing the system as a set of states

The system moves from one state to another if an event occurs (increase or decrease for i or j). The two traffics asumed to follow a Poisson distribution with two different average arrival rates (λ_1, λ_2) while the two service rates (μ_1, μ_2) assumed to follow an exponential distribution. The general CTMC model is shown in Figure 6-6.

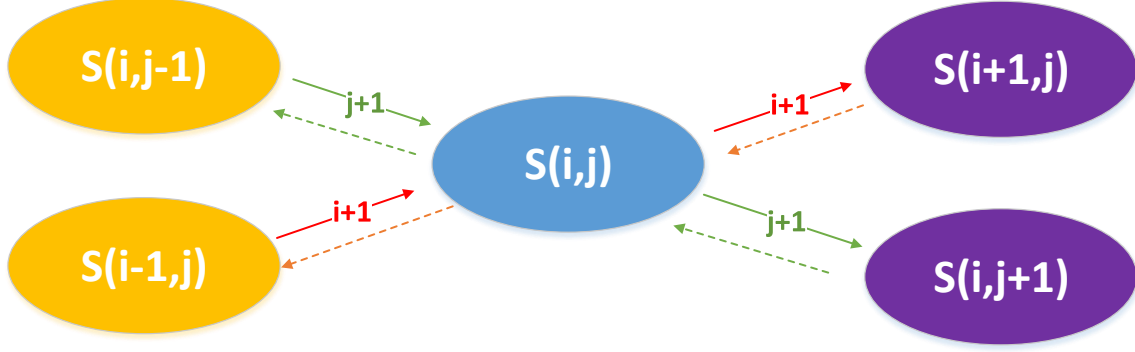


Figure 6-6: Generic CTMC states: s : state, i : number of ongoing service for H2H requests, j : number of ongoing service for M2M requests.

Generating the generic equilibrium equation

Similar to section 6.5.1 but while considering two ongoing services instead of one, we end up with the following general equation:

$$(\alpha\lambda_1 + \alpha\lambda_2 + i\beta\mu_1 + j\beta\mu_2)\pi_{(i,j)} = \beta\lambda_1\pi_{(i-1,j)} + \beta\lambda_2\pi_{(i,j-1)} + (i+1)\alpha\mu_1\pi_{(i+1,j)} + (j+1)\alpha\mu_2\pi_{(i,j+1)} \quad (6.9)$$

Where $\alpha = 0$ indicates the Full state (otherwise $\alpha = 1$) and $\beta = 0$ represents the Empty state (otherwise $\beta = 1$).

Linear system solution

The previous general equation (6.9) can be solved similarly to section 6.5.1, while considering s_t as the dimension for the steady-state probability vector Π and $(s_t) \times (s_t)$ as dimensions for the square matrix \mathbf{D} . s_t can be calculated as follows: $s_t = \frac{(c+1) \times (c+2)}{2}$

Because $0 \leq i + j \leq c$, we can only consider the upper triangle part of the transition probability matrix $i \times j$:

$$\mathbf{P} = \begin{bmatrix} \pi_{(0,0)} & \pi_{(0,1)} & \pi_{(0,2)} & \pi_{(0,3)} \\ \pi_{(1,0)} & \pi_{(1,1)} & \pi_{(1,2)} & \pi_{(1,3)} \\ \pi_{(2,0)} & \pi_{(2,1)} & \pi_{(2,2)} & \pi_{(2,3)} \\ \pi_{(3,0)} & \pi_{(3,1)} & \pi_{(3,2)} & \pi_{(3,3)} \end{bmatrix}$$

Then, we build the steady-state probability vector Π ($s_t = 10$) as its dimension with the following order:

$$\Pi = \left[\pi_{(0,0)} \quad \pi_{(0,1)} \quad \pi_{(0,2)} \quad \pi_{(0,3)} \quad \pi_{(1,0)} \quad \pi_{(1,1)} \quad \pi_{(1,2)} \quad \pi_{(2,0)} \quad \pi_{(2,1)} \quad \pi_{(3,0)} \right]$$

\mathbf{D} should be arranged to follow the same order. Finally, using equation (5.11), we can solve the linear system.

Performance metrics

We propose hereinafter two performance metrics in order to identify how far the degradation on the traffic could occur during emergency scenarios and the impact on the network. Using our proposed solution A-eNB, we show how congestion problems could be alleviated by implementing an adaptive bandwidth whenever a bottleneck is reached.

- Service Completion Rate (S_c): It gives the number of completed requests per time interval and it is based on the service rate μ and the number of ongoing requests for a certain application (i):

$$S_c = \sum_{ij} i\mu\pi_{(i,j)} \quad (6.10)$$

- Resource Utilization (R_u): This metric gives the probability of the system to be busy serving the arrivals in terms of the number of utilized PRBs in each state:

$$R_u = \sum_{ij} \frac{i+j}{c} \pi_{(i,j)} \quad (6.11)$$

In the next section, we translate the previous methodology into a case-study, then we simulate different scenarios with the aim to evaluate the A-eNB and its response throughout critical events.

6.6 Simulation scenarios and Result discussions

In our simulated scenarios, we built our use-cases, test the accuracy of our model, study the mutual impact among H2H and M2M traffics and test the robustness of A-eNB in solving an expected congestion.

6.6.1 Case-study:

Real use-cases show that we cannot predict the occurrence, consequences and influence of disaster events. Consequently, we built a case-study in a previous section 4.2.3 based on some use-cases such the one found in [58] and 3GPP technical reports [59] as summarized previously in Table 6.1.

By comparing again the different previous storms with the maximum up-link data-rate in NB-IoT (150 Kbps), we conclude that the available bandwidth can suffer from a huge degradation which requires an effective solution for this problem.

6.6.2 Model accuracy

We conducted many simulations using the model proposed in [83] which can generate similar traffic under the same conditions and parameters with some modifications to make it suitable to our scenarios. Then, an analysis for both analytical and simulation results are conducted in order to validate the accuracy of the model by calculating:

- The Mean Absolute Error (*MAE*):

$$MAE = \frac{1}{n} \sum_{t=1}^n |x_{c(t)} - x_{s(t)}|$$

where: $x_{s(t)}$ represents the analytical result and $x_{c(t)}$ denotes the simulation result.

- The Root Mean Squared Error (*RMSE*):

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (x_{c(t)} - x_{s(t)})^2}$$

The comparison, shown in Table 6.2, reveals that the proposed model results fall into the acceptable error limits comparing to the analytical results with a $MAE_{max} = 0.0592$ when $c = 6$ and a $MAE_{max} = 0.0177$ when $c = 25$.

Table 6.2: MAE and RMSE for analytical and simulation results

Simulation	c	λ	μ	MAE	RMSE
1	6	4	1	0.0592	0.0903
2	6	5	4	0.0262	0.0330
3	6	2	4	0.0483	0.0709
4	25	15	12	0.0128	0.0227
5	25	20	5	0.0177	0.0522
6	25	23	25	0.0082	0.0117

6.6.3 Simulation parameters:

The purpose of this sub-section is to evaluate the A-eNB by measuring the service completion rate (S_c) for different arrival rates in different cases. To this end, we use the simulation model² presented in [83] to create different scenarios that validate our work.

Based on our case-study in section 6.6.1, in which we consider three different groups selected according to the parameters set to different models, we turn the group storms into different average arrival rates $\lambda_1 = 1.6$ for the first group "Group 1" storm, $\lambda_2 = 3.2$ for "Group 2" storm, $\lambda_3 = 6.4$ for "Group 3" storm and $\lambda_w = 11.2$ for worst-case storm (the three storms simultaneously).

At first, we assume that all storms have the same service rate: $\mu_1 = \mu_2 = \mu_3 = \mu_w = 1$ to enhance the impact of the other parameters. Additionally, we assume that the observation time intervals are small enough to have one arrival or departure at one time interval with a total simulation time equals to 1000 time intervals.

²The model consists of a single-queue and single-server system with a single traffic source and an infinite storage capacity.

6.6.4 Comparison and result discussions:

Using an eNode-B

To mimic a disaster scenario in our use-case, we consider only a regular eNodeB with a fixed number of resources ($c = 1$) reserved for M2M traffic, while considering different storms with $\lambda_1 = 1.6$, $\lambda_2 = 3.2$, $\lambda_3 = 6.4$ and $\lambda_w = 11.2$. In the results shown in Figure 6-7, we realize a huge degradation in the service completion rate when moving from Group1 storm ($\lambda_1 = 1.6$) with a 56% completion rate till reaching the worst-case scenario ($\lambda_w = 11.2$) with only 9% completion rate.

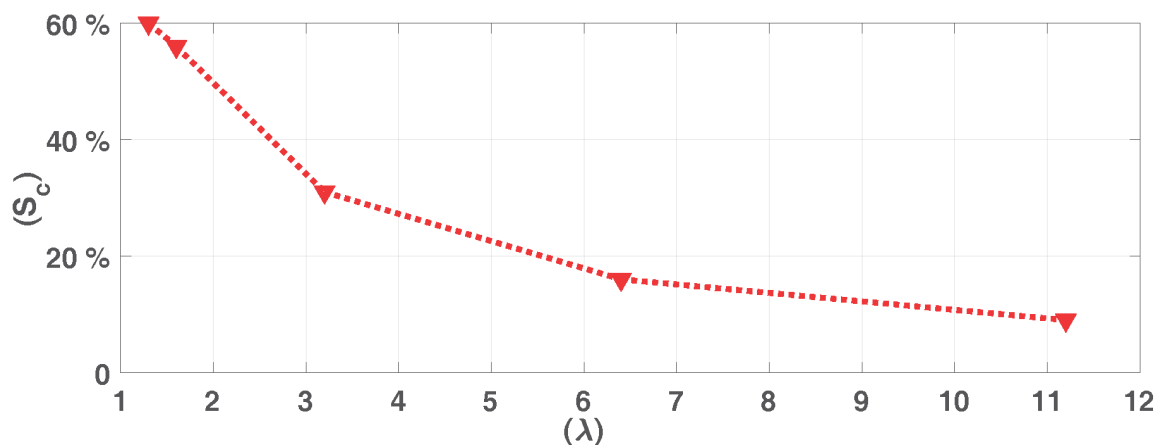


Figure 6-7: Overload problem in a regular eNodeB during emergency scenario: while ($c = 1$), c : represents the number of system resource, S_c : Service completion rate, λ : average arrival rate.

Using an A-eNB

To test the adaptability of our A-eNB during disaster scenarios, we consider the same different storms with $\lambda_1 = 1.6$, $\lambda_2 = 3.2$ and $\lambda_3 = 6.4$ while enabling the flexibility of the bandwidth till it absorbs all previous storms.

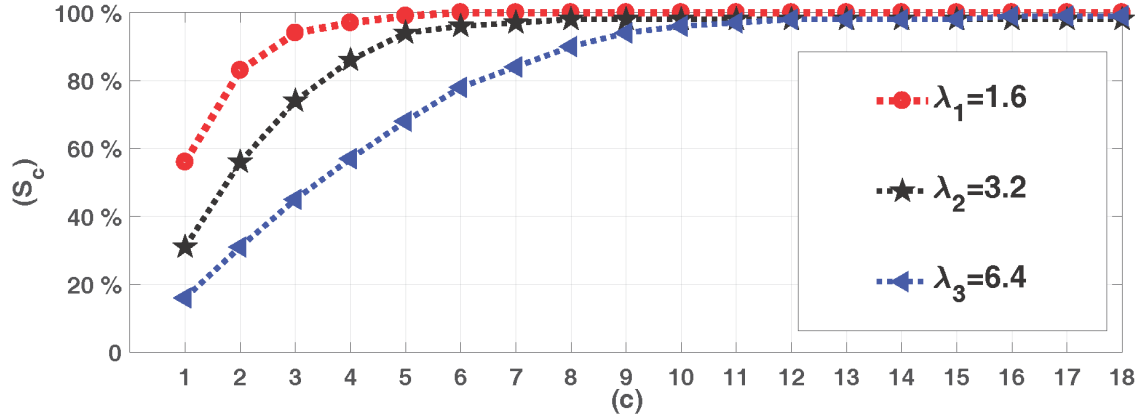


Figure 6-8: Resolving M2M storms by an A-eNB during an emergency scenario: A-eNB: Adaptive eNodeB, λ_1 , λ_2 and λ_3 represent the arrival rates for "Group 1", "Group 2" and "Group 3" storms.

From the results shown in Figure 6-8, we can realize:

- In the beginning of the simulation, the system uses one resource block ($c = 1$) similar to any regular eNodeB. With one resource block, the eNodeB can fulfill the requests of 58% from "Group 1" ($\lambda = 1.6$) because the eNodeB receives 1.6 average arrivals and can terminate one service in each time interval. While in "Group 2" ($\lambda = 3.2$), the eNodeB receives 3.2 average arrivals with one terminated service only, which leads to a 35% service completion rate. Meanwhile, in "Group 3" ($\lambda = 6.4$), the eNodeB ends up with an 18% service completion rate as result of receiving 6.4 average arrivals accompanied with one terminated service in one time interval.
- With this bottleneck, the A-eNB takes precedence over the problem and solve it by an incremental number of resources till it reaches a total resolution with a service completion rate 98% using only nine resources ($c = 9$). This result sounds ultimately fair, because there is no doubt that 9 resources are more than enough to accomplish a task with an average arrival rate 6.4 and one service completion per time interval.

To sum up, the above results reveal that by adapting the NB-IoT bandwidth from ($c = 1$) used usually in a regular eNodeB, to ($c = 9$) promoted in an A-eNB, we can eliminate all M2M storms with a completion rate of 98%.

6.6.5 Worst-case scenario:

Due to the synchronized behavior of M2M devices, we expect to have worst-case scenarios in which the three M2M storms influence the network simultaneously with a $\lambda_w = 11.2$. By simulating this case, the results spot that the service completion rate reach its lowest value with 9%. Meanwhile, it peaks to 98% when implementing the A-eNB while ($c = 18$) only.

6.6.6 M2M and H2H coexistence scenarios:

After the validation of the proposed model for a single traffic, we extend our exploration to include M2M and H2H traffics, with the aim to study the mutual impact of both traffics in normal, dense area and emergency scenarios.

The proposed simulation is based on the simulation model proposed in [83], but with the following modifications:

- We assume to have one LTE-A network which serves both H2H and M2M traffics. Each traffic has an average arrival rate (λ_1, λ_2) used as parameters for the Poisson distribution. Each H2H and M2M traffic has a service rate (μ_1, μ_2) respectively.
- Both H2H and M2M traffics have the same priority.
- A bandwidth of 5 *MHz* ($c = 25$) is used in order to stress the traffics to the maximum.

Normal Scenario

In rural areas, we assume to have low requests for both M2M and H2H devices by considering the following parameters: $0 < \lambda_1 \leq 2$, $\lambda_2 = 1$, $\mu_1 = 0.5$ and $\mu_2 = 1$.

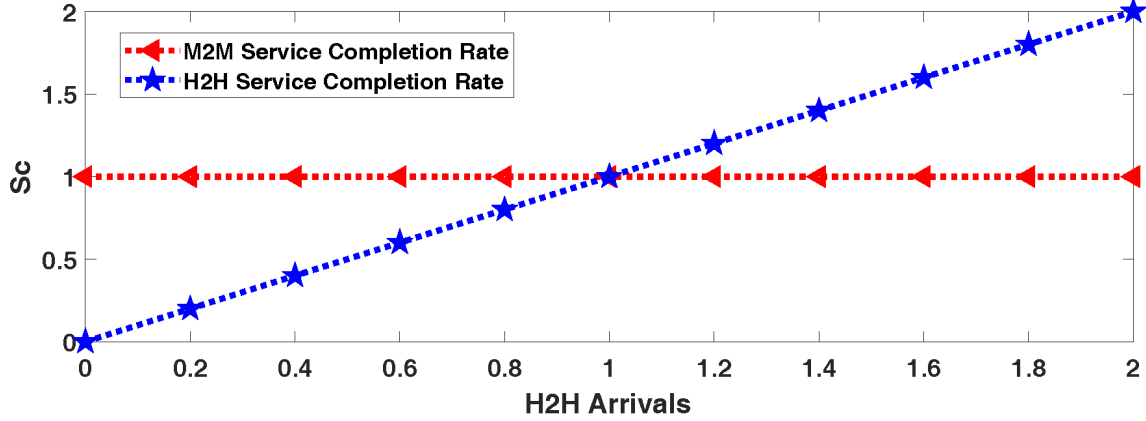


Figure 6-9: $M2M-S_c$ vs $H2H-S_c$ in Normal cases.

From the results shown in Figure 6-9, we conclude the following:

- M2M Traffic: All requests can be fulfilled throughout the whole simulation because the eNodeB has extra resources ($c = 25$) while having increasing requests which ranges from 1 request at the beginning till reaching 3 requests at the end of the simulation.
- H2H Traffic: Although the average arrival rate increases from 0 to 2 ($\mu = 0.5$) along with a constant M2M average arrival 1 ($\mu = 1$), but because of the excessive extra resources, the eNodeB could serve all requests during the whole simulation time.

To sum up, having 25 RBs per time interval is more than enough for such arrival rates. Moreover, the resource utilization reaches 20% only, which shows the ability of the system to receive much more requests with no noticeable shortage.

Dense Area Scenario

In urban and dense areas, a huge number of M2M and H2H devices are dispatching their payloads. Consequently, to study the impact of M2M traffic on H2H traffic we consider a fixed arrival rate of H2H requests $\lambda_1 = 5$ and incremental arrival rate of M2M requests $0 < \lambda_2 \leq 20$, while $\mu_1 = 0.5$ and $\mu_2 = 1$.

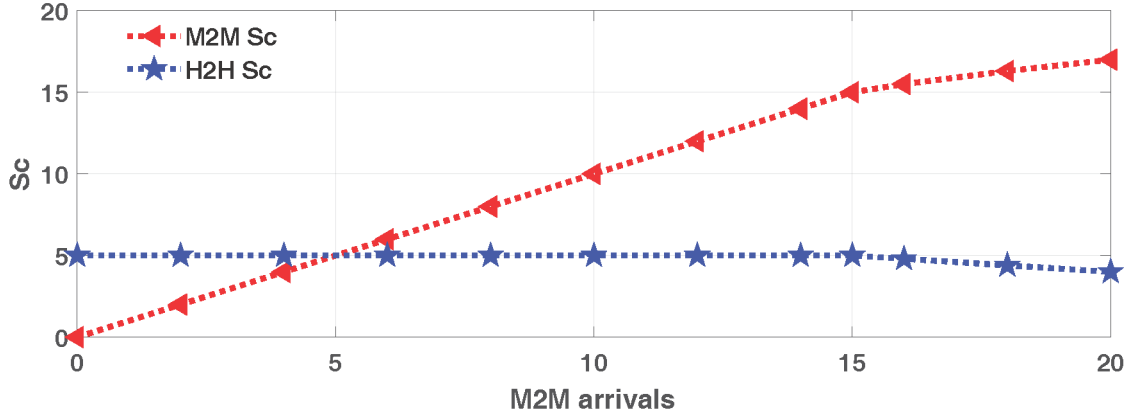


Figure 6-10: $M2M-S_c$ vs $H2H-S_c$ in Dense area cases.

From the results shown in Figure 6-10, we can conclude the following:

- *M2M Service Completion Rate ($M2M-S_c$):*

The system is not able to serve all of M2M requests when $\lambda_2 \geq 15$. Meanwhile, at the peak, only 80% from the M2M requests are served.

- *H2H Service Completion Rate ($H2H-S_c$):*

The H2H do not suffer from any degradation till $\lambda_2 \geq 15$. At the peak, 20% from the H2H requests aren't served.

- A high utilization of the system ($R_u = 90\%$) which requires a queuing strategy to fulfill the excessive requests and to minimize the number of forced terminated services.

If we want to study the impact of H2H traffic on M2M traffic by considering the same parameters, it is obvious to have the same results because H2H and M2M devices have the same priority.

Emergency Scenario

In emergency cases, a M2M signaling storm is definitely expected as a result of connected devices. This congestion caused by the simultaneous synchronization affects

both H2H and M2M devices. In this scenario, we consider the following system parameters: $\lambda_1 = 50$, $50 \leq \lambda_2 \leq 200$, $\mu_1 = 5$ and $\mu_2 = 1$.

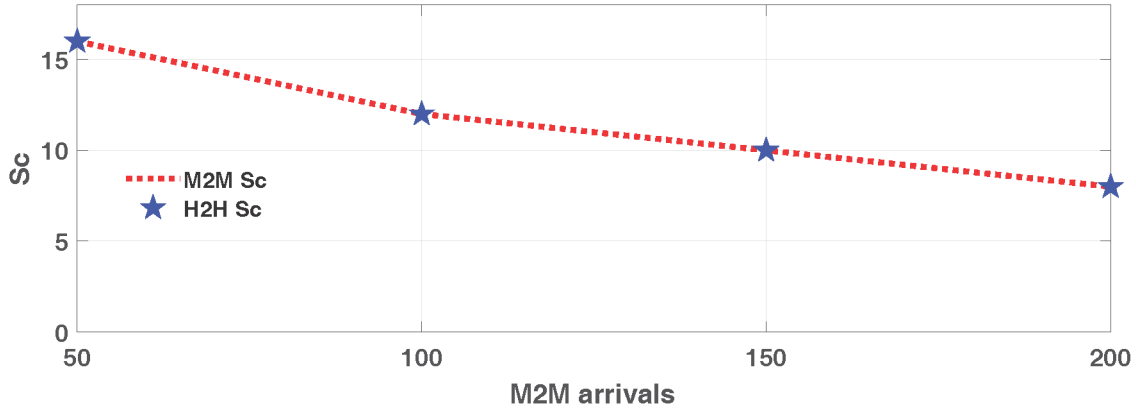


Figure 6-11: $M2M-S_c$ vs $H2H-S_c$ Emergency cases.

From the results shown in Figure 6-11, we conclude the following:

- $M2M-S_c$ or $H2H-S_c$: The results shows that at the beginning only 22% of the requests are served. Meanwhile, at the peak only 12% from the requests are served.
- The system utilization peaks to its cut-off point with a $R_u = 100\%$.

Consequently, using a fixed number of resources (*i.e.*, $c = 25$) results in a huge degradation on both M2M and H2H services during emergency events which stresses the fact of the need of an adaptive resource allocation as explained in section 6.6.4.

6.7 Conclusion

H2H and M2M traffic coexistence sharing the same LTE-A networks becomes a hot topic due to their different specificity and functionality. Consequently, mutual impacts on both M2M and H2H traffics become unavoidable.

In Release 13 [84], 3GPP addresses this problem with its new NB-IoT technology dedicating a limited bandwidth for M2M traffic. Unlike H2H communications, M2M

devices transmit small-sized packages in different time intervals in form of synchronized storms. Therefore, one of the prominent challenges facing mobile operators is the fast saturation of the limited bandwidth with the rising of M2M devices especially in emergency cases where all these devices are attempting to send their payloads simultaneously and causing inevitably overload congestion problems.

In this chapter, we have proposed a new concept called A-eNB as an extension to the classical eNodeB in NB-IoT networks. Our A-eNB keeps broadening the NB-IoT bandwidth gradually while shrinking LTE-A bandwidth till soaking-up all M2M storms. Moreover, this solution provides an essential resolution of presumable M2M storms especially during disaster and emergency events.

In order to study M2M and H2H traffic coexistence, a CTMC is proposed to model the system behavior and study the impact on the traffics and the network.

Our simulation results show that the network will be facing a huge degradation in the service completion rate when using the classical eNodeB. This degradation reaches 9% in the worst-case scenario ($\lambda_w = 11.2$). But using our A-eNB solution and by leasing a maximum 18 PRBs for M2M traffic of the total bandwidth in LTE-A/NB-IoT networks, can result a completion rate of 98% on all simulated M2M storms throughout emergency scenarios.

Chapter 7

Heterogeneous traffic modeling for M2M traffic with H2H coexistence

Glossary

3GPP: 3rd Generation Partnership Project

A-eNB: Adaptive evolved Node B

CANAL: Coexistence Analyzer and Network Architecture for Long term evolution

CMMPP: Coupled Markov Modulated Poisson Processes

DL: Down-Link

EDP: Event-Driven Pattern

eNodeB: evolved Node B

FIFO: First Input First Output

FTP: File Transfer Protocol

H2H: Human-to-Human

IoT: Internet of Things

IPP: Interrupted Poisson Process

LoRa: Long Range

LPWAN: Low Power Wide Area Network

LTE-A: Long Term Evolution-Advanced

LTE-M: Long Term Evolution for Machines

M2M: Machine-to-Machine
MMPP: Markov Modulated Poisson Processes
OPNET: Optimized Network Engineering Tool
OMNeT: Objective Modular NeTwork
PC: Personal Computer
PUP: Periodic Update Pattern
QCU: Queuing Control Unit
QoS: Quality of Service
RAC: Resource Allocation Control
UL: Up-Link
VoIP: Voice over Internet Protocol

7.1 Introduction

As we have explained in chapter 2, Machine-to-Machine (M2M) communication is expected to take a major part in any futuristic wireless network along with its counterpart Human-to-Human (H2H) communication. Although M2M and H2H communications share complementary objectives in term of techno-civilization in different fields (*e.g.*, civil transportation, electric power grid, medical treatment, *etc.*), but M2M communication is expected to play a vital role as a proxy that might replace/limit many human interventions via intelligent systems. If we take into account that M2M characteristics should live up to any rejuvenating technology requirements, the differences among H2H and M2M traffic characteristics might distract the unprecedented evolution deployment. By analyzing H2H and M2M traffics, two conspicuous differences can be spotted [62]:

- H2H traffic is heterogeneous whereas M2M traffic is highly homogeneous (all machines running similar applications behave similarly).
- H2H devices are uncoordinated on different timescales, while M2M devices can be mainly coordinated (many devices react on an external global event in a synchronized fashion).

Thus, most of traffic models designed for H2H communications should be adapted to be able to meet tight M2M characteristic requirements. With a close look on M2M traffic, two patterns can be distinguished [62]:

- Periodic Update Pattern (PUP): This non-real time pattern is usually generated when M2M devices transmit regularly application status reports (*e.g.*, gas meter, electricity meter, water meter, *etc.*) to a certain server.
- Event-Driven Pattern (EDP): This real-time pattern is triggered when a sudden event occurs (*e.g.*, tsunami alerts, health emergency notifications, terrorist attack alarms, *etc.*) causing certain parameter thresholds to be crossed.

In reality, M2M applications are often a patchwork of both aforementioned patterns. Hence, building new models based on PUP and EDP will enhance the accuracy of M2M communication models.

In addition, H2H and M2M traffic coexistence with their mutual influences implies variety of challenges that might be encountered on a shared network which reduces its effectiveness. One of the major reasons is the incompatibility patterns among H2H and M2M traffics. Unlike H2H traffic, M2M traffic is highly homogeneous in most of the cases. But with its synchronization behavior, an accumulative traffic is expected to be received, which turns it to a heterogeneous traffic.

In this chapter, we address the heterogeneity of M2M traffic along with H2H traffic and its consequences. Additionally, our proposed concept called Adaptive eNodeB (A-eNB) [17] is tested as an extension to the classical evolved Node B (eNodeB) in LTE-M networks. Our A-eNB keeps broadening the Long Term Evolution for Machines (LTE-M) bandwidth gradually while shrinking Long Term Evolution-Advanced (LTE-A) bandwidth till soaking-up all M2M storms. Moreover, this solution provides an essential resolution of presumable M2M storms especially during disaster and emergency events. At the end, a Markov Modulated Poisson Processes (MMPP) model is used as a framework to model the system behavior and to study the impact on both H2H and M2M traffics along with the network efficiency.

7.2 Traffic modeling

"Traffic modeling" can be represented by a stochastic processes that match the behavior of physical quantities of measured data traffic [91]. Traffic models are classified as *Source traffic models* (*e.g.*, video, data and voice) and *Aggregated traffic models* (*e.g.*, backbone networks, internet and high-speed links).

A source traffic simulator (*e.g.*, SimuLTE simulator [15], OPtimized Network Engineering Tool (OPNET) [92], Objective Modular NeTwork (OMNeT) [51], *etc.*) generates packets with certain sizes and intervals to reflect real traffic behaviors. In [93], an analysis of several typical source traffic models, including the ON/OFF, Interrupted Poisson Process (IPP) and two-state Markov Modulated Poisson Processes (MMPP) models are presented. The authors of [93] used OPNET to validate their model. In our previous work in [14], our scenarios focus on M2M traffic load in emergency events (*e.g.*, earthquakes, fire, terrorist attacks, *etc.*). In such emergency events, besides of H2H traffics (Voice over Internet Protocol (VoIP), video streaming and File Transfer Protocol (FTP)), an additional M2M surge traffic attempts to access the network caused by the consequences of the emergency event. All scenarios considered use a source traffic simulator such as SimuLTE to focus on the ability of an eNodeB to deal with a fixed number of H2H traffics (FTP-UpLink (UL), FTP-DownLink (DL), VoIP-UL, VoIP-DL, Video Streaming) with an increasing number of M2M requests attempting to access a LTE-A network simultaneously. In the same paper [14], we conclude that by exceeding 800 M2M connected devices on a Personal Computer (PC) platform, an error appears causing the interruption of the simulation. By trying various platforms, we end up with different errors due to the heaviness of generating different traffics by a huge number of devices. This modeling flaw needs additional exploration and analysis which highlights on a hot research topic that compares various scenarios built according to different components and parameters among several simulators. Eventually, if we recall that according to [5], it is expected to have more than 52K devices per cell trying to send their payloads simultaneously during a disaster event, we end up that *Source traffic models* become extremely heavy

to be executed in such cases which requires switching to *Aggregated traffic modeling* development.

Aggregated traffic models (*i.e.*, SimuLink simulator [83]) consist of finding a good approximation of the arrival process of multiple devices respecting a good balancing act towards accuracy and simulation efficiency [94]. As an example, in our previous work in [18], we study the mutual impact of H2H and M2M traffics in dense area or emergency scenarios. Additionally, we conduct several simulations based on the proposed architecture in [83], while assuming to have one LTE-A network and average arrival rates (λ_1, λ_2) with service rates (μ_1, μ_2) for H2H and M2M traffics. The simulation results show that using a prioritize LTE-A system for both M2M and H2H traffics could handle more requests in dense area cases by maintaining the higher priority traffic without any degradation while reducing the completion rate of the lower priority traffic. Meanwhile, in emergency cases, the higher priority traffic takes precedence over the lower priority traffic with an improvement on its completion rate while totally eliminating the low priority traffic; In such cases, the non priority traffic strategy has a competitive advantage because it keeps H2H and M2M traffics working at the same time.

According to [95], the 3rd Generation Partnership Project (3GPP) presents two models: Model 1 representing non-synchronized M2M traffic and Model 2 for synchronized M2M traffic, as shown in Fig. 7-1.

The two different traffic models are assumed in order to evaluate the network performance under different access intensities:

- *3GPP model 1* can be considered as a normal scenario in which M2M devices access the network uniformly throughout a certain period of time (*i.e.*, a non-synchronized way).
- *3GPP model 2* can be considered as a disaster scenario in which a large number of M2M devices access the network in an extremely synchronized manner (*i.e.*, after a power outage).

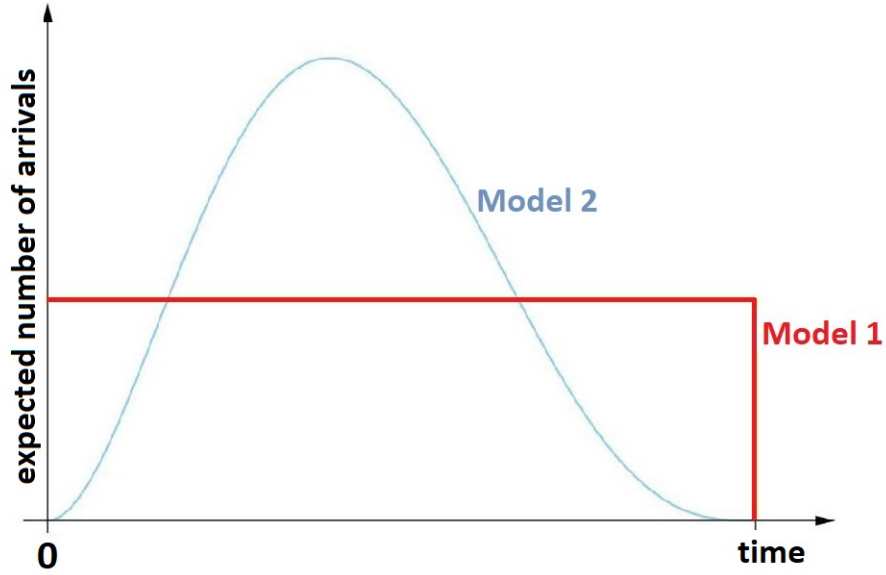


Figure 7-1: The expected arrival rate over time for synchronized and non-synchronized M2M traffics using 3GPP Model 1 and Model 2 [96].

In [96], the authors propose an approach combining the advantages of both modeling paradigms (Source traffic modeling and Aggregated traffic modeling), namely, Coupled Markov Modulated Poisson Processes (CMMPP) framework. It demonstrates the feasibility of source traffic modeling for M2M, being enabled by only linearly growing complexity. Compared to aggregated M2M traffic models, such as proposed by the aforementioned 3GPP model in [95], CMMPP enhances the accuracy and the flexibility at the cost of moderate computational complexity.

In [77], Data aggregation is used to decrease power consumption as M2M devices have disparate traffic types with different delay requirements. A priority-based data aggregation scheme at the M2M gateway is proposed which effectively maintains a good trade-off between the power consumption and delay requirement. An analytical model considering the idle and busy states behavior of the aggregator using preemption queuing with priority disciplines is developed and analyzed. The performance measures in terms of system delay and power consumption are derived and used to quantify the trade-off between M2M delay sensitive traffic and low power consumption provisions. Based on the numerical and simulation results, the proposed scheme provides a good trade-off between delay and power consumption.

In [94], it was spotted that a performance evaluation of VoIP applications requires reliable traffic models. Although good traffic models for VoIP applications exist, but the simulation of thousands of simultaneous connections turns the simulation studies very heavy. Aggregate traffic models are badly required to achieve faster simulations without losing accuracy. In this work, the characterization of the coexistence of homogeneous and heterogeneous VoIP applications under heavy and light traffic intensities are envisioned. Moreover, simple aggregate traffic models for VoIP applications are suggested while showing the performance limits of these models in network simulation environment.

In [97], the authors model the traffic generated by Internet of Things (IoT) devices connected through Low Power Wide Area Networks (LPWAN) technologies. However, the traffic can be classified as either periodic, event-triggered, or a combination of both, but due to diverse applications of IoT, it is not trivial to have a single traffic model to represent all of them. To this end, an evaluation of the performance of Long Range (LoRa) network, one of LPWAN technologies, is conducted in the presence of a hybrid of both traffic types, where the event propagates spatially over time. Thereby, when an event occurs, spatial and temporal correlation in the traffic rate can be spotted due to the natural phenomena. Finally, the CMMPP model presented in [97] is used to represent such characteristic traffic from independent IoT devices triggered by an event.

In [98], a comparative analysis for the three different models (3GPP model 1, 3GPP model 2 and CMMPP model), using M2M traffic, has been performed with variable arrival rates under the assumption that the LTE network has limited resources. The results show the characteristics of the M2M traffic in a more realistic manner pinpointing the differences from the standard traffic in cellular network.

To sum up, from all previous works, an essential question merits seeking for a comprehensive answer is: whether it is feasible to model the traffic of a large amount of autonomous machines simultaneously using the source traffic modeling, which is in general more accurate than the aggregated traffic modeling (*i.e.*, treating the accumulated data from all M2M devices as single stream).

Actually, for multiple access and enormous capacity evaluations, we do not need knowledge about the behavior of a single device, (*i.e.*, simulate every single connection between a device and its eNodeB). As a result, the general behavior of M2M traffic may fit into the *Aggregated traffic models* with a simple Poisson process representation.

However, due to the synchronization behavior of M2M traffic, the average arrival rate (λ) is expected to vary over time $\lambda(t)$ [99, 100]. Nowadays, *Aggregated traffic models*, such as heterogeneous with time-varying arrival rate $\lambda(t)$, becomes a hot topic [96] that merits to be studied and compared along with the homogeneous traffic with a constant average arrival rate (λ), as explained in the next section.

7.3 Markov Modulated Poisson Processes Model

In this section, a MMPP model is used as a framework to accurately model M2M traffic sources by analyzing an event when loads of M2M devices behave in a synchronized fashion, as suggested in [101].

Our MMPP model consists of a constant average arrival rate $\lambda_{(h)}$ for H2H traffic and a variable Poisson process modulated rate $\lambda_{(m)}$ for M2M traffic which is modulated in each time space Δt by a β distribution, and is determined by the state of a Markov chain $s_{(m)}$ (m denotes the index of Markov states which varies from 0 to M , where M is the total number of states), as shown in Fig. 7-2.

Further, the overall average rate $\lambda_{(g)}$ of the MMPP is given by:

$$\lambda_{(g)} = \sum_{m=0}^M \lambda_{(m)} \pi_{(m)} \quad (7.1)$$

where $\pi_{(m)}$ is the probability for the system to be in a certain state $s_{(m)}$ at a given time space Δt , $\sum_{m=0}^M \pi_{(m)} = 1$.

A basic example for a M2M device modeled by a MMPP would be a $(M+1)$ states with the first state s_0 representing a "Normal state", the remaining states represent the variable expected arrival rate over time $\lambda_{(m)}$ during different "Emergency states"

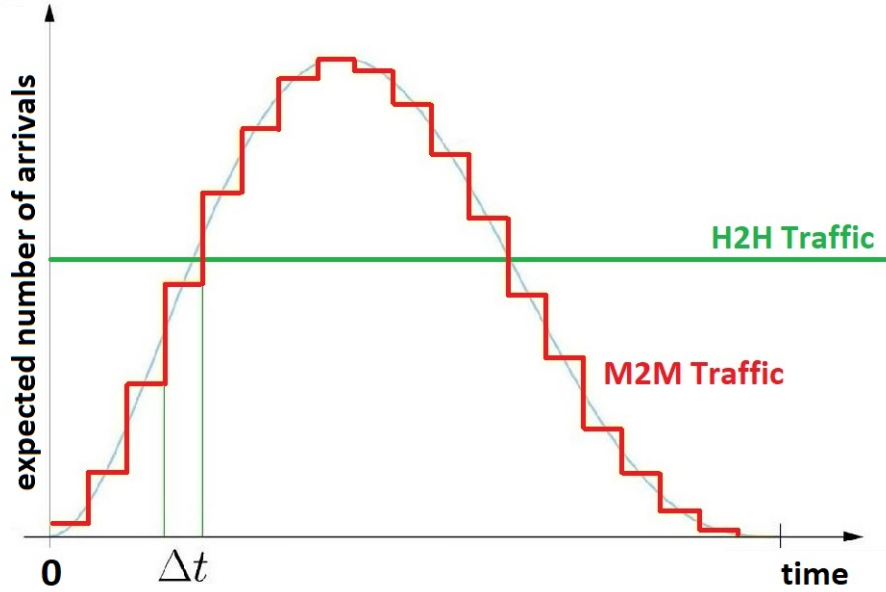


Figure 7-2: The interpretation as MMPP for a M2M traffic, where $\lambda_{(m)}$ the mean arrival rate of a Poisson process [62].

till reaching the "Worst-scenario state" ($m = M$). This principle is depicted in Fig. 7-3, where $p_{(m,m+1)}$ are the transition probabilities between the states of the chain from $s_{(m)}$ to $s_{(m+1)}$.

The system falls into one of the following three cases:

1. Case 1: Normal state, " $s_{(0)}$ ", includes the initial state and has the following equilibrium relationship:

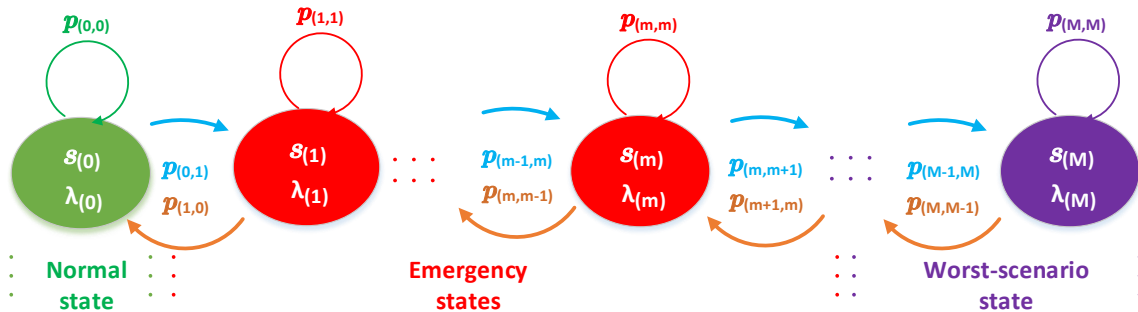


Figure 7-3: Generic MMPP model: M is the total number of states, $p_{(m,m+1)}$ are the transition probabilities between the states of the chain from $s_{(m)}$ to $s_{(m+1)}$, each Markov chain states $s_{(m)}$ represents a mean arrival rate $\lambda_{(m)}$.

$$p_{(0,1)}\pi_{(0)} - p_{(1,0)}\pi_{(1)} = 0 \quad (7.2)$$

2. Case 2: Emergency states, " $s_{(m)}$ ", includes $(M - 1)$ states:

$$p_{(m-1,m)}\pi_{(m-1)} - [p_{(m,m-1)} + p_{(m,m+1)}]\pi_{(m)} + p_{(m+1,m)}\pi_{(m+1)} = 0 \quad (7.3)$$

where $\pi_{(m)}$ is the probability to be in the state $s_{(m)}$, and $m \in [1, 2, 3, \dots, M - 1]$.

3. Case 3: Worst-scenario state, " $s_{(M)}$ ", includes the final state and has the following equilibrium relationship:

$$p_{(M-1,M)}\pi_{(M-1)} - p_{(M,M-1)}\pi_{(M)} = 0 \quad (7.4)$$

The above three equations can be summarized by the following generic equation:

$$\alpha p_{(m-1,m)}\pi_{(m-1)} - [\alpha p_{(m,m-1)} + \xi p_{(m,m+1)}]\pi_{(m)} + \xi p_{(m+1,m)}\pi_{(m+1)} = 0 \quad (7.5)$$

where:

- $\alpha = 0$ in the "Normal state", otherwise $\alpha = 1$.
- $\xi = 0$ in the "Worst-scenario state", otherwise $\xi = 1$.

The state probabilities $\pi_{(m)}$ resides in the state probability vector $\mathbf{\Pi}$ according to:

$$\mathbf{\Pi} = \left(\pi_{(0)} \quad \pi_{(1)} \quad \cdot \quad \cdot \quad \cdot \quad \pi_{(M)} \right)^T \quad (7.6)$$

Additionally, the transition probabilities can be condensed into a state transition matrix \mathbf{P} . Consequently, the equations for the \mathbf{M} state can be represented as a linear system in the following form:

$$\mathbf{P}\boldsymbol{\Pi} = \mathbf{0} \quad (7.7)$$

The linear system in (7.7) should be solved while respecting the following two constraints:

$$\sum_{m=0}^M \pi_{(m)} = \mathbf{1} \quad (7.8)$$

$$\mathbf{0} \leq \pi_{(m)} \leq \mathbf{1} \quad (7.9)$$

$$\forall m \in \{0, 1, 2, 3, \dots, M\}.$$

Finally, by solving the above linear system in (7.7), we can calculate $\boldsymbol{\pi}_{(0)}$ to $\boldsymbol{\pi}_{(M)}$ and $\boldsymbol{\lambda}_{(g)}$ using (7.1).

7.4 Case-study

In our case-study in section (4.2.3), we assume to have a variable Poisson process modulated rate $\boldsymbol{\lambda}_{(m)}$ for M2M traffic with three steps only which represent three different groups; assuming that the observation time intervals are small enough Δt to have one step forward or one step backward at one time interval. Then, we turn the three correlated storms into different average arrival rates $\boldsymbol{\lambda}_{(1)}$ for Group₍₁₎ storm, $\boldsymbol{\lambda}_{(2)}$ for Group₍₂₎ storm, $\boldsymbol{\lambda}_{(3)}$ for Group₍₃₎ storm, as summarized in Table 6.1.

So, in our case-study we have four cases ($M = 3$):

- "Normal case": it represents the regular traffic when all M2M devices are sending there payloads uniformly:

$$\boldsymbol{\lambda}_{(0)} = \frac{\boldsymbol{\lambda}_{(1)} + \boldsymbol{\lambda}_{(2)} + \boldsymbol{\lambda}_{(3)}}{M}$$

- First "Emergency" storm: when Group₍₁₎ submits its data as result of a sudden event: $\lambda_{(E1)} = \lambda_{(1)}$
- Second "Emergency" storm: when Group₍₁₎ and Group₍₂₎ send their data all together: $\lambda_{(E2)} = \lambda_{(1)} + \lambda_{(2)}$
- "Worst-case" storm: it occurs when the three storms dispatch their data simultaneously: $\lambda_{(W)} = \lambda_{(1)} + \lambda_{(2)} + \lambda_{(3)}$

We can extract four balance equations using equations (7.2), (7.3), (7.4) and (7.6).

Then, we brief the equations in a state transition matrix \mathbf{P} :

$$\mathbf{P} = \begin{bmatrix} p_{(0,1)} & -p_{(1,0)} & 0 & 0 \\ p_{(0,1)} & -[p_{(1,0)} + p_{(1,2)}] & p_{(2,1)} & 0 \\ 0 & p_{(1,2)} & -[p_{(2,1)} + p_{(2,3)}] & p_{(3,2)} \\ 0 & 0 & p_{(2,3)} & -p_{(3,2)} \end{bmatrix}$$

By solving the linear system in (7.7), we can calculate the state probabilities $\pi_{(0)}$, $\pi_{(1)}$, $\pi_{(2)}$, $\pi_{(3)}$, then using (7.1), we can calculate $\lambda_{(g)}$.

7.5 Model and metrics

7.5.1 CANAL Model

To recall, our aim is to measure the mutual impact of M2M heterogeneous traffic on H2H traffic in different scenarios while maintaining a sustainable level of services with minimal congestion during emergency events by allowing both M2M and H2H devices to access the network resources efficiently. In this context, we use our proposed architecture in section (5.4) so-called Coexistence Analyzer and Network Architecture for Long term evolution (CANAL), which extends the classical functionality of the eNodeB and includes an adaptive control of bandwidth based on the M2M connected devices.

To recall, the proposed CANAL architecture manages automatically both H2H and M2M traffics. Based on the instantaneous requests of M2M, a part of the total

bandwidth is dedicated to fulfill these sudden needs. To this end, the model operates at different conditions depending on the M2M load state throughput. The CANAL architecture consists of a single uplink cell in one eNodeB with a total number of resource blocks used in the network (\mathbf{c}).

There are two types of traffics: M2M heterogeneous traffic and H2H traffic with arrival rates $\lambda_{(m)}$, $\lambda_{(h)}$ respectively, assumed to be following Poisson's distributions [62]. Let $\mu_{(m)}$, $\mu_{(h)}$ the service rates for M2M and H2H traffics respectively assumed to follow an exponential distribution.

The CANAL architecture includes a Resource Allocation Control (RAC) unit used to control, manage and grant access the network requests based on the available resources. The RAC has a vital role when congestion occurs especially during emergency scenarios by dominating the resources adaptively in order to fulfill the excessive requests of M2M devices, while keeping the H2H requests within the acceptable norms. Let $\mathbf{rb}_{(h)}$ is the maximum resource blocks reserved for LTE-A network and $\mathbf{rb}_{(m)}$ is the one reserved for LTE-M network, where $\mathbf{rb}_{(h)} + \mathbf{rb}_{(m)} = \mathbf{c}$. In normal cases, both $\mathbf{rb}_{(h)}$ and $\mathbf{rb}_{(m)}$ reserve the initial number of resources (*i.e.*, in a LTE-A network with a bandwidth = 10 **MHz**: $\mathbf{rb}_{(h)} = \mathbf{50}$ and in a LTE-M network: $\mathbf{rb}_{(m)} = \mathbf{6}$). When the utilization of $\mathbf{rb}_{(m)}$ peaks to the maximum as result of a sudden event, the RAC lends an additional 6 resources to fulfill the temporary requests from H2H resources. So, the number of resources reserved for M2M will increase by 6 to be: $\mathbf{rb}_{(m)} = \mathbf{12}$ and the number of resources reserved for H2H will decrease by 6 to be: $\mathbf{rb}_{(h)} = \mathbf{44}$ (in the case of LTE-A bandwidth = 10 **MHz**). Lending more and more resources will be repeated until absorbing the whole storm or when the sudden events recedes.

Finally, the CANAL architecture contains also a Queuing Control Unit (QCU) consisting of two different queues: one for H2H and one for M2M (H2H-Q, M2M-Q) with queue sizes n and o respectively. The system is considered in the "Full State" when the number of reserved resource blocks for a LTE-A network peaks to $\mathbf{rb}_{(h)}$ and/or in a LTE-M network peaks to $\mathbf{rb}_{(m)}$. In this case, the queuing process is initiated when the system reaches its cut-off point and a new M2M/H2H device request contends to access the system following the QCU strategy explained in [18].

7.5.2 Performance and metrics

We use two performance metrics to measure the mutual impact between H2H and M2M traffics:

- a) Service completion rate (\mathbf{scr}): It gives the number of completed requests per time interval and it is based on the service rate $\boldsymbol{\mu}$ and the average arrival requests for a certain traffic (e.g. $\mathbf{scr}_{(h)}$ and $\mathbf{scr}_{(m)}$ which represent the Service Completion Rate for H2H or M2M traffics [64]).
- b) H2H/M2M Resource utilization ($\mathbf{ru}_{(h)}/\mathbf{ru}_{(m)}$): This metric gives the probability of the system to be busy serving H2H/M2M arrivals in terms of the number of utilized $\mathbf{rb}_{(h)}/\mathbf{rb}_{(m)}$ in each state, comparing to the total number of resource blocks used in the network (\mathbf{c}).

7.6 Simulations

7.6.1 Simulator

In this section, we present our developed simulation model which can generate both H2H and M2M traffics with full flexibility to add queuing or priority for any traffic H2H or M2M.

Suppose that we have a LTE-A network with a total number of resources ($\mathbf{c} = 100$). Six resources of the total number of resources are dedicated to LTE-M network to serve M2M requests ($\mathbf{rb}_{(m)} = 6$). As result, the remaining 94 resources are reserved to deal with H2H traffic ($\mathbf{rb}_{(h)} = 94$).

In order to study the mutual impact of H2H and M2M traffics in normal or emergency scenarios, we conduct several simulations based on the proposed architecture in [83] which uses the queuing theory to provide exact theoretical results for some performance measures that facilitates comparing empirical results with the corresponding theoretical results.

Many enhancements are made on the model in order to match our CANAL architecture presented in section (7.5) such as:

- The architecture consists of two servers with two traffic sources (H2H and M2M), where arrivals are determined by Poisson processes with the two parameters $\lambda_{(h)}$ and $\lambda_{(m)}$ respectively.
- Service times have an exponential distribution with rate parameter μ , where $\frac{1}{\mu}$ is the mean service time.
- Assuming that H2H traffic has a fixed average arrival rate $\lambda_{(h)}$, with a service rate $\mu_{(h)} = 1$.
- We assume also that M2M traffic has five different variable average arrival rates: $\lambda_{(m)} \in \{5, 10, 15, 20, 25\}$, with a service rate $\mu_{(m)} = 1$.
- H2H and M2M traffics have the same priority.
- A First Input First Output (FIFO) queue type is used with an infinite capacity size (in order to omit the role of the queuing for the moment, we consider queue sizes $n = o = 0$ for the two traffics H2H and M2M).
- Simulation duration = 1000 Seconds.

7.6.2 Regular eNodeB scenarios, results and discussions

Normal scenario

Initially, we consider only a regular eNodeB with the five aforementioned groups and different variable average arrival rates. The uniform arrival rate during normal cases is considered as $\lambda_{(m_0)} = \frac{5+10+15+20+25}{5} = 15$. Meanwhile, a constant average arrival rate for H2H traffic is considered $\lambda_{(h)} = 25$.

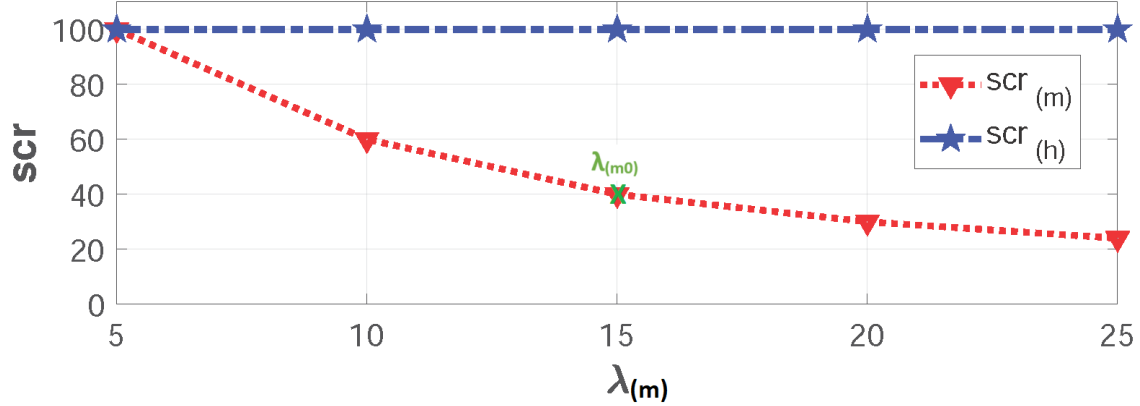


Figure 7-4: Regular eNodeB in a normal scenario, where " $\lambda_{(m)}$ " is the average arrival rate for M2M traffic, "scr" is the service completion rate.

In the results shown in Fig. 7-4, we realize:

- a) In normal operation, a uniform average arrival rate is expected $\lambda_{(0)} = 15$ with a 40% completion rate ($scr_{(m)} = 40\%$) and $ru_{(m)} = 100\%$ as result of having only 6 resources to serve 15 instantaneous requests.
- b) When receiving a single storm from a synchronized group (Group₍₁₎ to Group₍₅₎), a huge degradation in the service completion rate is spotted when moving from $\lambda_{(1)} = 5$ with a 100% completion rate till reaching $\lambda_{(5)} = 25$ with a 24% completion rate only. These results are obvious as the network has only a fixed number of resources $rb_{(m)} = 6$ reserved for M2M traffic while having an increasing demand on M2M services: $scr_{(m)} = \{100\%, 60\%, 40\%, 30\%, 24\%\}$. Additionally, the system utilization peaks to its cut-off point with a $ru_{(m)} = 100\%$ due to the high load of arrivals comparing to the available resources in the system except for Group₍₁₎ $ru_{(m)} = 83\%$.
- c) H2H traffic doesn't suffer from any limitation as the network reserves the major amount of resources to H2H traffic $rb_{(h)} = 94$ while receiving only an average of 25 request per time-interval ($\lambda_{(h)} = 25$) with $scr_{(h)} = 100\%$ and $ru_{(h)} = 26.5\%$

Disaster scenario

During a disaster, we assume to receive a double H2H traffic with a constant average arrival rate $\lambda_{(h)} = 50$ as result of a disastrous event. Meanwhile, the five groups are gradually synchronized:

- First "Emergency" storm: when Group₍₁₎ submits its data as result of a sudden event: $\lambda_{(E1)} = \lambda_{(1)} = 5$
- Second "Emergency" storm: when Group₍₁₎ and Group₍₂₎ dispatch their payloads simultaneously: $\lambda_{(E2)} = \lambda_{(1)} + \lambda_{(2)} = 15$
- Third "Emergency" storm: when Group₍₁₎, Group₍₂₎ and Group₍₃₎ send their data at the same time:
 $\lambda_{(E3)} = \lambda_{(1)} + \lambda_{(2)} + \lambda_{(3)} = 30$
- Forth "Emergency" storm: when Group₍₁₎, Group₍₂₎, Group₍₃₎ and Group₍₄₎ send their payloads all together:
 $\lambda_{(E4)} = \lambda_{(1)} + \lambda_{(2)} + \lambda_{(3)} + \lambda_{(4)} = 50$
- "Worst-case" storm: it occurs when the five storms dispatch their data simultaneously:
 $\lambda_{(W)} = \lambda_{(1)} + \lambda_{(2)} + \lambda_{(3)} + \lambda_{(4)} + \lambda_{(5)} = 75$

In the results shown in Fig. 7-5, we realize that:

- a) A huge degradation in the service completion rate can be spotted when receiving the five synchronized groups gradually while moving from Emergency₍₁₎ storm ($\lambda_{(E1)} = 5$) with a 100% completion rate till reaching Emergency₍₅₎ storm ($\lambda_{(W)} = 75$) with a 8% completion rate only. These results are obvious as the network has only a fixed number of resources $rb_{(m)} = 6$ reserved for M2M traffic while having an increasing demand on M2M services ($scr_{(m)} = \{100\%, 40\%, 20\%, 12\%, 8\%\}$). Additionally, the system utilization peaks to its cut-off point with a $ru_{(m)} = 100\%$ due to the high load of

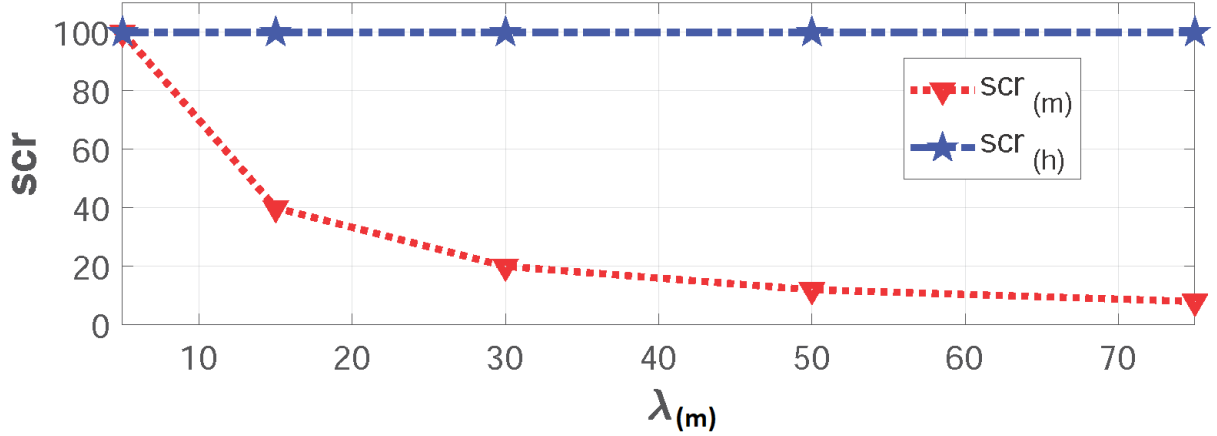


Figure 7-5: Regular eNodeB in a emergency scenario, where " $\lambda_{(m)}$ " is the average arrival rate for M2M traffic, "scr" is the service completion rate.

arrivals comparing to the available resources in the system except for Group₍₁₎ as mentioned previously.

- b) H2H traffic doesn't suffer from any limitation as the network reserves the major amount of resources to H2H traffic $rb_{(h)} = 94$ while receiving only an average of 50 request per time-interval

($\lambda_{(h)} = 50$) with $scr_{(h)} = 100\%$ and $ru_{(h)} = 53\%$

7.6.3 Adaptive eNodeB scenarios, results and discussions

As a solution for the service completion rate degradation shown in Fig. 7-4 and Fig. 7-5, we will test our proposed solution in [17], the A-eNB, which operates in different modes depending on M2M traffic intensity to tackle the limitation problem of the network.

Normal scenario

To test the adaptability of our A-eNB during normal scenarios, we consider the same different storms $\lambda_{(m)} \in \{5, 10, 15, 20, 25\}$ and with a constant H2H average arrival rate $\lambda_{(h)} = 25$ while enabling the flexibility of the bandwidth till it absorbs all previous storms.

The results shown in Table 7.1 reveal that by adapting the number of resources from ($\mathbf{rb}_{(m)} = \mathbf{6}$) used usually in a regular LTE-M network, to ($\mathbf{rb}_{(m)} = \mathbf{30}$) promoted in an A-eNB, we can eliminate all M2M storms with a completion rate of 100% ($\mathbf{scr}_{(m)} = \mathbf{100\%}$). Meanwhile, H2H traffic doesn't reveal any degradation with a constant ($\mathbf{scr}_{(h)} = \mathbf{100\%}$) due to having enough resources reserved for H2H traffic.

Table 7.1: The adaptivity of an A-eNB in normal cases

<i>Group #</i>	$\lambda_{(m)}$	$\mathbf{rb}_{(m)}$	$\mathbf{rb}_{(h)}$	$\mathbf{scr}_{(h)}$	$\mathbf{scr}_{(m)}$
1	5	6	94	100	100
2	10	12	88	100	100
3	15	18	82	100	100
4	20	24	76	100	100
5	25	30	70	100	100

Disaster scenario

During emergency cases, we are expecting to receive similar previous storms but gradually $\lambda_{(m)} \in \{\mathbf{5}, \mathbf{15}, \mathbf{30}, \mathbf{50}, \mathbf{75}\}$ and we assume receiving a double constant H2H average arrival rate $\lambda_{(h)} = \mathbf{50}$ while enabling the flexibility of the bandwidth till it absorbs all previous storms. The results shown in Table 7.2 reveal that by adapting the number of resources from ($\mathbf{rb}_{(m)} = \mathbf{6}$) used usually in a regular LTE-M network, to ($\mathbf{rb}_{(m)} = \mathbf{72}$) promoted in an A-eNB, we can eliminate all M2M storms with a completion rate of 96%-100%. Meanwhile, H2H traffic doesn't reveal any degradation from E1-E4 with a constant ($\mathbf{scr}_{(h)} = \mathbf{100\%}$) due to having enough resources reserved for H2H traffic. Only one drawback can be spotted for H2H traffic in the "worst-case" scenario, when its completion rate reaches 56% ($\mathbf{scr}_{(h)} = \mathbf{56\%}$) as the system fulfills 96% of M2M requests. But this flaw is considered acceptable since the adaptation of the resources will resume to its initial state once the emergency case recedes.

Table 7.2: The adaptivity of an A-eNB in emergency cases

#	$\lambda_{(m)}$	$rb_{(m)}$	$rb_{(h)}$	$scr_{(h)}$	$scr_{(m)}$
E1	5	6	94	100	100
E2	15	18	82	100	100
E3	30	30	70	100	100
E4	50	48	52	100	96
Worst-case	75	72	28	56	96

7.7 Conclusion

As result of H2H and M2M traffic coexistence with their mutual influences, a variety of challenges may be encountered on a shared network which reduces its effectiveness. One of the major reasons is the incompatibility patterns among H2H and M2M traffics. Unlike H2H traffic, M2M traffic is highly homogeneous in most of the cases. But with its synchronization behavior, an accumulative traffic is expected to be received, which turns it to a heterogeneous traffic. Therefore, one of the prominent challenges facing mobile operators is the fast saturation of the limited bandwidth for LTE-M networks, especially with the exponential booming of M2M devices in emergency cases where all these devices may attempt to send their payloads simultaneously and causing inevitably overload congestion problems. In this chapter, we have simulated the heterogeneous M2M traffic along with H2H traffic and its consequences on both traffics. Additionally, we have tested our proposed concept called A-eNB as an extension to the classical eNodeB in LTE-M networks. Our A-eNB keeps broadening the LTE-M bandwidth gradually while shrinking LTE-A bandwidth till soaking-up all M2M storms. Moreover, this solution provides an essential resolution of presumable M2M storms especially during disaster and emergency events. In order to study the heterogeneous M2M traffic, we have used the Markov chain concept to model the system behavior and study the impact on the traffics and the network.

Our simulation results show that using the classical eNodeB, the network will be facing a huge degradation in the service completion rate for M2M and H2H by reaching 8% in the worst-case scenario ($\lambda_w = 75$). But using our A-eNB solution and by leasing a maximum 72 resources reserved for M2M traffic from the total network resources, can result a completion rate of 96% throughout the worst-case scenario.

Chapter 8

General conclusion and perspectives

Cellular networks like 4th Generation (4G) and future 5th Generation (5G) are extensively designed for Mobile Broad-Band (MBB) services. But, with the rapid growth of cellular IoT devices many challenges arise such as: supporting a massive number of devices, bandwidth limitation, dealing with emergency cases, Human-to-Human (H2H) and Machine-to-Machine (M2M) coexistence, overload problems, M2M congestion. In this aspect, we have focused in this thesis on the expected problems that might be faced as result of the aforementioned challenges while envisioning the proposed mechanisms and methods to alleviate these problems.

Our thesis journey starts with an exploration on the random access procedures with their related channel called *Random Access CHannel* (RACH). A first bottleneck is expected on the RACH as result of the massive number of M2M requests attempting to access the network simultaneously which causes an overload problem inevitably. To this end, a survey with a comparative analysis of the main solutions proposed in the literature to overcome this issue is presented. Although, many proposed solutions appear to be optimized on (time, frequency) but for the moment the overload congestion problem is still a talking point with no clear solution. For example, in the literature many publications are trying to explore the expected number of User Equipment (UE)s that could be handled by an evolved Node B (eNodeB) without overloading it especially when loads of M2M devices are contending to access the network in dense areas like in [27, 28, 29]. The answers vary from 250 to

400 UEs, which show a kind of contradiction in between the aforementioned results. This contradiction needs additional analysis according to the different components and various parameters among several scenarios and platforms. In our future work, we plan to extend our preliminary study on RACH by improving the proposed initial access mechanism in Long Term Evolution (LTE) networks.

After granting any UE access to the network, a part of the total bandwidth should be reserved to submit its payload. Based on different literature reviews such that in [24, 35, 47], we find out that although the market is heading towards splitting bandwidth direction (a part of the total bandwidth is dedicated to M2M traffic use and the remaining bandwidth is dedicated to H2H traffic use); But, we believe that it is not a quite enough resolution because of the limited bandwidth reserved for M2M devices that will be depleted due to the synchronization behavior of M2M devices or as a result of any event that triggers M2M simultaneous responses. To solve the bandwidth limitation problem, we have proposed an effective solution as an extension to the classic eNodeB in Long Term Evolution for Machines (LTE-M) networks called Adaptive eNodeB (A-eNB). By implementing this novel solution, an adaptive reallocation of the bandwidth leads for an essential resolution for many presumable M2M storms. Based on SimuLTE modeler, the results proof that by leasing some folds of LTE-M classic bandwidth, we can achieve a significant gain in M2M traffic.

Our third station in the thesis journey continues by studying the mutual impact of M2M and H2H traffic coexistence sharing the same bandwidth or by testing two separate bandwidth one for M2M traffic and the other one dedicated for H2H traffic. Although many proposed models can be found in the literature to improve the system performance of M2M traffic with the coexistence of H2H traffic such in [62, 64, 77, 79, 96], but these models do not allow the analysis of the system behavior when facing different M2M/H2H applications with different priorities. To this end, we propose an enhanced network architecture called Coexistence Analyzer and Network Architecture for Long term evolution (CANAL) designed for Long Term Evolution-Advanced (LTE-A)/LTE-M networks that helps studying H2H/M2M traf-

fic coexistence while supporting the ability to apply various priority strategies to satisfy the Quality of Service (QoS) for each traffic. Additionally, a queuing strategy is enforced by a queuing control unit, which rule all queues allocated for different traffic types. To validate our model, a Continuous-Time Markov Chain (CTMC) model is proposed as a stochastic process tool to characterize the H2H/M2M coexistence based on analytical equations. Our simulation results matched with the CTMC analytical model which corroborates the proposed model. Moreover, they show that using a prioritize LTE-A system for both M2M and H2H traffics is more convenient in dense area scenarios; while in emergency cases, it is more appropriate to use a non prioritize traffic strategy to keep both H2H and M2M traffics working properly at the same time.

Next, we improved our model by extending our study to include H2H/M2M mutual impact in LTE-A/Narrow Band-IoT (NB-IoT). Additionally, a significant research gap can be spotted in the literature that models NB-IoT networks especially during disaster scenarios. To this end, many simulations were performed along with a case-study that describes an expected emergency event build from different reports. Our results show that the network will be facing a huge degradation in the service completion rate when using the classical eNodeB. This degradation reaches 9% in the worst-case scenario where the average arrival rate $\lambda_w = 11.2$. But using our A-eNB solution in NB-IoT network and by leasing a maximum 18 PRBs for M2M traffic of the total bandwidth in LTE-A/NB-IoT networks, it can result a completion rate of 98% on all simulated M2M storms throughout emergency scenarios.

At the end of the journey, we realize that one of the major reasons of H2H/M2M mutual impact is the incompatibility patterns among H2H and M2M traffics. Unlike H2H traffic, M2M traffic is highly homogeneous in most of the cases because M2M traffic uses small chunks of data in a low transmission rate, mostly with predictable communication times and durations [13]. But with M2M synchronization behavior along with diverse applications from different domains that have different payloads, timings and data-rates an accumulative traffic is expected to be received from different sources, which forms an overall heterogeneous traffic. So, at the end of

our project, we address the heterogeneity of M2M traffic along with H2H traffic and its consequences using a Markov Modulated Poisson Processes (MMPP) framework that models heterogeneous M2M traffic with H2H traffic using Markov chains as a stochastic process tool to characterize the H2H/M2M coexistence based on different analytical equations. Our simulation results show that using the classical eNodeB, the network will be facing a huge degradation in the service completion rate for M2M and H2H by reaching 8% in the worst-case scenario where the average arrival rate $\lambda_w = 75$. But using our A-eNB solution and by leasing a maximum 72 resources reserved for M2M traffic from the total network resources, it can result a completion rate of 96% throughout the worst-case scenario.

In this thesis, we have focused on 3rd Generation Partnership Project (3GPP) technologies (LTE-A, LTE-M and NB-IoT) as presumable Internet of Things (IoT) enablers. In our future work, we tend to study non-3GPP technologies (*e.g.*, Long Range (LoRa), SigFox, *etc.*) and the behavior of M2M and H2H traffics which use different Low Power Wide Area Network (LPWAN) during a disaster while maintaining a satisfactory QoS.

Part I

References

List of Publications

International Journals

- **Ahmad Hani El Fawal**, Mohamad Najem, Ali Mansour, Frédéric Le Roy and Denis Le Jeune, "*CTMC modeling for H2H/M2M coexistence in LTE-A/LTE-M networks*", Journal of Engineering (JoE), Accepted on 15 October 2018.
- **Ahmad Hani El Fawal**, Ali Mansour, Mohamad Najem, Frédéric Le Roy and Denis Le Jeune, "*M2M Heterogeneous traffic modeling with H2H coexistence during disaster events*", Institution of Engineering and Technology (IET) Communications Journal, Submitted on 30 September 2018.

International Conferences

- **Ahmad Hani El Fawal**, Ali Mansour, Mohamad Najem, Frédéric Le Roy and Denis Le Jeune, "*CTMC modeling for M2M/H2H coexistence in a NB-IoT Adaptive eNodeB*", The 2018 IEEE International Conference on Internet of Things (iThings 2018), Halifax, Canada, July 2018.
- **Ahmad Hani El Fawal**, Ali Mansour, Mohamad Najem, Frédéric Le Roy and Denis Le Jeune, "*LTE-M Adaptive eNodeB for Emergency Scenarios*", The 8th International Conference on Information and Communication Technology Convergence (ICTC), Jeju Island, Korea, October 2017.
- **Ahmad Hani El Fawal**, Ali Mansour, Frédéric Le Roy, Denis Le Jeune and Ali Hamié, "*RACH Overload Congestion Mechanism for M2M Communication in*

LTE-A: Issues and Approaches", International Symposium on Networks Computers and Communication (ISNCC), Marrakech, Morocco, May 2017.

Bibliography

- [1] S. Haller, S. Karnouskos, and C. Schroth, “The internet of things in an enterprise context,” in *the first Future Internet Symposium (FIS)*, PP. 14-28, Vienna, Austria, September 2008.
- [2] Machina Research, Annual guidance on the growth of the IoT market worldwide, Accessed: 18-September-2016, <https://machinaresearch.com/news/press-release-global-internet-of-things-market-to-grow-to-27-billion-devices-generating-usd3-trillion-revenue-in-2025/>.
- [3] GSMA, Realising the benefits of mobile enabled IoT solutions, Tech. Rep., March 2015, Accessed: 02-October-2018, <https://www.gsma.com/iot/news/realising-the-benefits-of-mobile-iot-solutions/>.
- [4] NOKIA, LTE-M - Optimizing LTE for the Internet of Things, White Paper, Accessed: 24-Dec-2016, <https://novotech.com/docs/default-source/default-document-library/lte-m-optimizing-lte-for-the-internet-of-things.pdf?sfvrsn=0>.
- [5] Y. P. E. Wang, X. Lin, A. Adhikary, A. Grovlen, Y. Sui, Y. Blankenship, J. Bergman, and H. S. Razaghi, “A Primer on 3GPP Narrowband Internet of Things,” *IEEE Communications Magazine*, vol. 55, no. 3, pp. 117–123, March 2017.
- [6] 3GPP - Release 13 analytical view version Sept. 9th (2015). Accessed: 20-May-2017. [Online]. Available: <http://www.3gpp.org/release-13>
- [7] NOKIA, LTE evolution for IoT connectivity, Accessed: 14-Feb-2017, <http://resources.alcatel-lucent.com/asset/200178>.
- [8] M. Centenaro, L. Vangelista, A. Zanella, and M. Zorzi, “Long-range communications in unlicensed bands: the rising stars in the IoT and smart city scenarios,” *IEEE Wireless Communications*, vol. 23, no. 5, pp. 60–67, October 2016.
- [9] L.-f. Semtech. Semtech Lora FAQ. (Accessed: 20-Feb-2017). [Online]. Available: <http://www.semtech.com/wireless-rf/lora/LoRa-FAQs.pdf>

- [10] A. Díaz-Zayas, C. A. García-Pérez, Á. M. Recio-Pérez, and P. Merino, “3GPP Standards to Deliver LTE Connectivity for IoT,” in *2016 IEEE First International Conference on Internet-of-Things Design and Implementation (IoTDI)*, pp. 283-288, Berlin, Germany, April 2016.
- [11] B. Reynders, W. Meert, and S. Pollin, “Range and coexistence analysis of long range unlicensed communication,” in *IEEE 23rd International Conference on Telecommunications (ICT)*, pp. 1-6, Thessaloniki, Greece, May 2016.
- [12] R. Ratasuk, B. Vejlgaard, N. Mangalvedhe, and A. Ghosh, “NB-IoT system for M2M communication,” in *IEEE Wireless Communications and Networking Conference*, pp. 1-5, Doha, Qatar, April 2016.
- [13] A. Elmangoush, A. A. Corici, R. Steinke, M. Corici, and T. Magedanz, “A Framework for Handling Heterogeneous M2M Traffic,” *Procedia Computer Science*, vol. 63, pp. 112–119, 2015.
- [14] A. H. El Fawal, A. Mansour, F. Le Roy and D. Le Jeune, “RACH Overload Congestion Mechanism for M2M Communication in LTE-A: Issues and Approaches,” *IEEE International Symposium on Networks, Computers and Communications (ISNCC)*, Marrakech, Morocco, May 2017.
- [15] SimuLTE Modeler Version (0.9.1). Accessed: 21-July-2018. [Online]. Available: <http://www.simulte.com>
- [16] A. H. El Fawal, A. Mansour, M. Najem, F. Le Roy and D. Le Jeune, “Adaptive LTE-M eNodeB for Emergency Scenarios,” *The 8th International Conference on Information and Communication Technologies Convergence ICTC*, Jeju Island, Korea, October 2017.
- [17] A. H. El Fawal, A. Mansour, M. Najem, F. Le Roy, D. Le Jeune, “CTMC modeling for M2M/H2H coexistence in a NB-IoT Adaptive eNodeB,” in *The 11th IEEE International Conference on Internet of Things (iThings 2018)*, Halifax, Canada, July 2018.
- [18] A. H. El Fawal, M. Najem, A. Mansour, F. Le Roy, D. Le Jeune, “CTMC modeling for M2M/H2H coexistence in LTE-A/LTE-M networks,” in *Journal of Engineering (JoE)*, submitted on 13 June 2018, status: accepted under minor revision.
- [19] A. H. El Fawal, A. Mansour, M. Najem, F. Le Roy and D. Le Jeune, “M2M Heterogeneous traffic modeling with H2H coexistence during disaster events,” *Institution of Engineering and Technology (IET) Communications*, Submitted on 30 September 2018.
- [20] M. Chen, J. Wan, and F. Li, “Machine-to-machine communications: Architectures, standards and applications.” *KSII transactions on internet & information systems*, vol. 6, no. 2, 2012.

- [21] P. K. Verma, R. Verma, A. Prakash, A. Agrawal, K. Naik, R. Tripathi, M. Alsabaan, T. Khalifa, T. Abdelkader, and A. Abogharaf, "Machine-to-machine (m2m) communications: A survey," *Journal of Network and Computer Applications*, vol. 66, issue C, pp. 83-105, March 2016.
- [22] Y. Zhang, R. Yu, M. Nekovee, Y. Liu, S. Xie, and S. Gjessing, "Cognitive machine-to-machine communications: visions and potentials for the smart grid," *IEEE network*, vol. 26, no. 3, 2012.
- [23] N. Pandey, M2M communication concept, White Paper, Accessed: 28-July-2018, https://www.researchgate.net/publication/291337307_M2M_communication_concept.
- [24] F. Ghavimi and H.-H. Chen, "M2M communications in 3GPP LTE/LTE-A networks: architectures, service requirements, challenges, and applications," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 2, pp. 525–549, 2015.
- [25] E. Kartsakli, A. S. Lalos, A. Antonopoulos, S. Tennina, M. D. Renzo, L. Alonso, and C. Verikoukis, "A survey on m2m systems for mhealth: A wireless communications perspective," *Sensors*, vol. 14, no. 10, pp. 18 009–18 052, 2014. [Online]. Available: <http://www.mdpi.com/1424-8220/14/10/18009>
- [26] Cisco Visual Networking Index, Forecast and Methodology, 2016-2021, Accessed: 2-August-2018, <https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/complete-white-paper-c11-481360.html>.
- [27] P. Masek, J. Hosek, and M. Dubrava, "Influence of M2M communication on LTE networks," in *the 10th International IEEE Conference*, Zvule, Czech Republic, August 2014.
- [28] D. Kovac, P. Masek, and J. Hosek, "Simulation-Based Study on Capacity Performance of 4G Mobile Network for M2M Services," in *the International Conference "Technical Universities: Integration with European and World Systems of Education"*, Izhevsk, Russia, April 2014.
- [29] Y.-C. Chang, "Study of Overload Control Problem for Intelligent LTE M2M Communication System," *Advances in Smart Systems Research*, vol. 3, no. 3, pp. 44–48, 2013.
- [30] T. Pötsch, S. N. K. K. Marwat, Y. Zaki, and C. Gorg, "Influence of future M2M communication on the LTE system," in *Wireless and Mobile Networking Conference (WMNC)*, Pages 1-4, Dubai, United Arab Emirates, April 2013.
- [31] T. Pötsch, S. N. K. Marwat, Y. Zaki, and C. Görg, "Performance Evaluation of Machine-to-Machine Communication on Future Mobile Networks in Disaster Scenarios." Springer, Berlin, Germany, 2013, pp. 270–273.

- [32] Y. Mehmood, S. Marwat, T. Pötsch, F. Ahmad, C. Görg, and I. Rashid, "Impact of M2M traffic on LTE data traffic performance," in *Logistics Dynamics International Conference (LDIC)*, Bremen, February 2014.
- [33] Y. Mehmood, C. Görg, and A. Timm-Giel, "A radio resource sharing scheme for iot/m2m communication in lte-a downlink," in *IEEE International Conference on Communications Workshops (ICC)*, pp. 296-301, Kuala Lumpur, Malaysia, May 2016.
- [34] J. Song, A. Kunz, M. Schmidt, and P. Szczytowski, "Connecting and managing M2M devices in the future Internet," *Mobile Networks and Applications*, vol. 19, no. 1, pp. 4–17, 2014.
- [35] M. Hasan, E. Hossain, and D. Niyato, "Random access for machine-to-machine communication in LTE-advanced networks: Issues and approaches," *IEEE Communications Magazine*, vol. 51, no. 6, pp. 86–93, 2013.
- [36] A. Farhan, Z. Safdar, M. Yasir, and G. Carmelita, "Tailoring LTE-advanced for M2M communication using wireless inband relay node," in *Proceedings of World Telecommunications Congress (WTC 2014)*, pp. 1-3, Berlin, Germany, June 2014.
- [37] A. Farhan, M. Safdar, Z. Yasir, M. Yasir, and G. Carmelita, "Machine-to-machine sensor data multiplexing using LTE-advanced relay node for logistics," in *Dynamics in Logistics International Conference (LDIC)*, pp. 247-257, Bremen, Germany, December 2016.
- [38] Y. Mehmood, S. N. K. Marwat, C. Görg, Y. Zaki, and A. Timm-Giel, "Evaluation of M2M data traffic aggregation in LTE-A uplink," in *20 ITG Mobile Communication Conference*, pp. 24-29, Osnabrück, Germany, August 2015.
- [39] J. Hosek, P. Masek, D. Kovac, and F. Kröpfel, "M2M gateway: the centerpiece of future home," in *6th International Congress Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, pp. 190-197, Petersburg, Russia, October 2014.
- [40] A. Biral, M. Centenaro, A. Zanella, L. Vangelista, and M. Zorzi, "The challenges of M2M massive access in wireless cellular networks," *Digital Communications and Networks*, vol. 1, no. 1, pp. 1–19, 2015.
- [41] K. Zheng, S. Ou, J. Alonso-Zarate, M. Dohler, F. Liu, and H. Zhu, "Challenges of massive access in highly dense LTE-advanced networks with machine-to-machine communications," *IEEE Wireless Communications*, vol. 21, no. 3, pp. 12–18, 2014.
- [42] LTE-A Resource Grid, https://www.tutorialspoint.com/lte/lte_ofdm_technology.htm, Accessed: 07-July-2018.

- [43] S. Sesia, I. Toufik, and M. Baker, *LTE-the UMTS Long Term Evolution, 2011*. Wiley Online Library: <https://www.wiley.com/en-us/LTE+The+UMTS+Long+Term+Evolution>
- [44] C.-Y. Tu, C.-Y. Ho, and C.-Y. Huang, “Energy-efficient algorithms and evaluations for massive access management in cellular based machine to machine communications,” in *IEEE Vehicular Technology Conference (VTC Fall)*, pp. 1-5, San Francisco, USA, September 2011.
- [45] M. A. Uddin, A. Mansour, D. L. Jeune, M. Ayaz, and e.-H. M. Aggoune, “Uav-assisted dynamic clustering of wireless sensor networks for crop health monitoring,” *Sensors*, vol. 18, no. 2, 2018.
- [46] A. Lo, Y. W. Law, M. Jacobsson, and M. Kucharzak, “Enhanced LTE-advanced random-access mechanism for massive machine-to-machine (M2M) communications,” in *27th World Wireless Research Forum (WWRF) Meeting*. pp. 1-5, Dusseldorf, Germany, October 2011.
- [47] Y. Mehmood, C. Görg, M. Muehleisen, and A. Timm-Giel, “Mobile M2M communication architectures, upcoming challenges, applications, and future directions,” *EURASIP Journal on Wireless Communications and Networking*, vol. 2015, no. 1, pp. 1–37, 2015.
- [48] Y.-C. Pang, S.-L. Chao, G.-Y. Lin, and H.-Y. Wei, “Network access for M2M/H2H hybrid systems: a game theoretic approach,” *IEEE Communications Letters*, vol. 18, no. 5, pp. 845–848, 2014.
- [49] P. M. L. Bello and D. Grace, “Application of Q-learning for RACH access to support M2M traffic over a cellular network,” in *20th European Wireless Conference*, pp. 1-6, Barcelona, Spain, May 2014.
- [50] S.-Y. Lien, K.-C. Chen, and Y. Lin, “Toward ubiquitous massive accesses in 3GPP machine-to-machine communications,” *IEEE Communications Magazine*, vol. 49, no. 4, pp. 66–74, 2011.
- [51] OMNET++ Modeler, Accessed: 21-July-2018, <https://www.omnetpp.org>.
- [52] INET Framework, Accessed: 21-July-2018, <https://inet.omnetpp.org>.
- [53] Heavy Reading (2016), Mobile Network Outages Service Degradations, Survey Analysis. Accessed: 24-June-2017. [Online]. Available: <http://www.spirentfederal.com>
- [54] A. Rico-Alvarino, M. Vajapeyam, H. Xu, X. Wang, Y. Blankenship, J. Bergman, T. Tirronen, and E. Yavuz, “An overview of 3GPP enhancements on machine to machine communications,” *IEEE Communications Magazine*, vol. 54, no. 6, pp. 14–21, 2016.

- [55] ERICSSON. Ericsson Research Blog, Cellular IoT alphabet soup. Accessed: 20-Aug-2018. [Online]. Available: <https://www.ericsson.com/research-blog/cellular-iot-alphabet-soup/>
- [56] User Equipment (UE) radio access capabilities, Evolved Universal Terrestrial Radio Access (E-UTRA), 3GPP Technical Specification TS 36.306, v15.0.0, April 2018.
- [57] 3GPP. Technical Specifications and Technical Reports for a UTRAN-based 3GPP system, Release 12. (Accessed: 29-June-2018). [Online]. Available: <http://www.3gpp.org/specifications/releases/68-release-12>
- [58] M. Lauridsen, I. Z. Kovacs, P. Mogensen, M. Sorensen, and S. Holst, "Coverage and Capacity Analysis of LTE-M and NB-IoT in a Rural Area," in *2016 IEEE 84th Vehicular Technology Conference (VTC-Fall)*, Montreal, Canada, September 2016.
- [59] G. T. 45.820. (2015, November) Cellular System Support for Ultra Low Complexity and Low Throughput Internet of Things, v. 13.1.0, http://www.3gpp.org/ftp/Specs/archive/45_series/45.820/45820-d10.zip, accessed Oct. 6, 2016.
- [60] V. Saxena, J. Bergman, Y. Blankenship, A. Wallen, and H. S. Razaghi, "Reducing the Modem Complexity and Achieving Deep Coverage in LTE for Machine-Type Communications," in *2016 IEEE Global Communications Conference (GLOBECOM)*, pp. 1-7, Washington DC, USA, December 2016.
- [61] J. Plachy, Z. Becvar, and E. C. Strinati, "Cross-layer approach enabling communication of high number of devices in 5G mobile networks," in *IEEE 11th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, pp. 809–816, Abu Dhabi, United Arab Emirates, October 2015.
- [62] M. Laner, P. Svoboda, N. Nikaein, and M. Rupp, "Traffic models for machine type communications," in *Proceedings of the Tenth International Symposium on Wireless Communication Systems (ISWCS)*, Ilmenau, Germany, August 2013.
- [63] G. Sergiy, A. D. A. M. Khodayer, and A. J. H. Dheyaa, "Results of development of model for bandwidth management in lte downlink with resource allocation type 1," in *The Experience of Designing and Application of CAD Systems in Microelectronics*, pp. 409–413, Lviv–Polyana, Ukraine, February 2015.
- [64] S. Alqahtani, "Performance analysis of cognitive-based radio resource allocation in multi-channel LTE-A networks with M2M/H2H coexistence," *IET Communications*, vol. 11, no. 5, pp. 655–663, 2017.
- [65] D.-H. Nguyen, H. Nguyen, and E. Renault, "WE-MQS-VoIP priority: An enhanced LTE downlink scheduler for voice services with the integration of VoIP

- priority mode,” *International Journal of Advanced Computer Science and Applications IJACSA*, vol. 7, no. 7, pp. 560–567, 2016.
- [66] J. Shi, C. Liu, X. Hong, and C.-X. Wang, “Coexistence of delay-sensitive MTC/HTC traffic in large scale networks,” *Science China Information Sciences*, vol. 60, no. 10, 2017.
- [67] A. Mansour, R. Mesleh, and M. Abaza, “New challenges in wireless and free space optical communications,” *Optics and Lasers in Engineering*, vol. 89, pp. 95–108, 2017.
- [68] I. Gudkova, K. Samouylov, I. Buturlin, V. Borodakiy, M. Gerasimenko, O. Galinina, and S. Andreev, “Analyzing impacts of coexistence between M2M and H2H communication on 3GPP LTE system,” in *International Conference on Wired/Wireless Internet Communications*, pp. 162–174, Paris, France, May 2014.
- [69] A. Shamshad, M. Bawadi, W. W. Hussin, T. Majid, and S. Sanusi, “First and second order markov chain models for synthetic generation of wind speed time series,” *Energy*, vol. 30, no. 5, pp. 693–708, 2005.
- [70] S. P. Meyn and R. L. Tweedie, *Markov chains and stochastic stability*. Springer Science & Business Media, 2012.
- [71] M. T. Howard and S. Karlin, “An introduction to stochastic modeling, Third edition,” *San Diego Academic Press*, 1998.
- [72] 3GPP. Technical Specifications and Technical Reports for a UTRAN-based 3GPP system, Release 8. (Accessed: 20-Dec-2017). [Online]. Available: <http://www.3gpp.org/specifications/releases/72-release-8>
- [73] 3GPP. Technical Specifications and Technical Reports for a UTRAN-based 3GPP system, Release 15. (Accessed: 20-Dec-2017). [Online]. Available: <http://www.3gpp.org/release-15>
- [74] R. Ratasuk, N. Mangalvedhe, Z. Xiong, M. Robert, and D. Bhatoolaul, “Enhancements of narrowband IoT in 3GPP Rel-14 and Rel-15,” in *IEEE Conference on Standards for Communications and Networking (CSCN)*, pp. 60–65, Helsinki, Finland, September, 2017.
- [75] Mistic Jelena, Mistic Vojislav and Khan Nargis, “Sharing It My Way: Efficient M2M Access in LTE/LTE-A Networks,” *IEEE Transactions on Vehicular Technology*, vol. 66, no. 1, pp. 696–709, 2016.
- [76] M. Lévesque, F. Aurzada, M. Maier, and G. Joós, “Coexistence analysis of H2H and M2M traffic in FiWi smart grid communications infrastructures based on multi-tier business models,” *IEEE Transactions on Communications*, vol. 62, no. 11, pp. 3931–3942, 2014.

- [77] S. Alqahtani, "Analysis and modelling of power consumption-aware priority-based scheduling for M2M data aggregation over long-term-evolution networks," *IET Communications*, vol. 11, no. 2, pp. 177–184, 2017.
- [78] S. Hamdoun, A. Rachedi, and Y. Ghamri-Doudane, "A flexible M2M radio resource sharing scheme in LTE networks within an H2H/M2M coexistence scenario," in *IEEE International Conference on Communications (ICC)*, pp. 1-7, Kuala Lumpur, Malaysia, May 2016.
- [79] S. Alqahtani, "Cognitive-based resource allocation scheme in LTE-A networks with M2M/H2H coexistence," in *Wireless Telecommunications Symposium (WTS)*, pp. 1-7, New York City, USA, 2017.
- [80] F. Behmann and K. Wu, *Collaborative internet of things (C-IoT): For future smart connected life and business*. John Wiley & Sons, 2015.
- [81] H. Florencio and A. D. Neto, "Probabilistic Inference of the Packet Delivery Ratio in Industrial Wireless Sensor Networks," in *IEEE International Joint Conference on Neural Networks (IJCNN)*, pp. 2015-2023, Rio, Brazil, July 2018.
- [82] F. Pacheco, E. Exposito, J. Aguilar, M. Gineste, and C. Baudoin, "A novel statistical based feature extraction approach for the inner-class feature estimation using linear regression," in *IEEE International Joint Conference on Neural Networks (IJCNN)*, pp. 1005-1013, Rio, Brazil, July 2018.
- [83] Mathworks website, M/M/1 Queuing System. Accessed: 19-December-2017. [Online]. Available: <https://www.mathworks.com/help/simevents/examples/m-m-1-queuing-system.html>
- [84] 3GPP. Overall description, Stage 2 (Release 13), 3GPP TS 36.300, v13.4.0, June 2016. (Accessed: 02-October-2018). [Online]. Available: https://www.etsi.org/deliver/etsi_ts/136300_136399/136300/13.04.00_60/ts_136300v130400p.pdf
- [85] ERICSSON. (2015) Ericsson mobility report, on the pulse of the networked society. Accessed: 20-May-2017. [Online]. Available: <http://www.ericsson.com/res/docs/2015/mobility-report/ericsson-mobility-report-nov-2015.pdf>
- [86] S. M. Oh and J. Shin, "An Efficient Small Data Transmission Scheme in the 3GPP NB-IoT System," *IEEE Communications Letters*, vol. 21, no. 3, pp. 660–663, March 2017.
- [87] 3GPP. Radio resource control (RRC), Protocol specification (Release 13), 3GPP TS 36.331, v13.2.0, June 2016. (Accessed: 02-October-2018). [Online]. Available: https://www.etsi.org/deliver/etsi_ts/136300_136399/136331/13.00.00_60/ts_136331v130000p.pdf

- [88] Z. Zhou, M. Dong, K. Ota, G. Wang, and L. T. Yang, "Energy-Efficient Resource Allocation for D2D Communications Underlying Cloud-RAN-Based LTE-A Networks," *IEEE Internet of Things Journal*, vol. 3, no. 3, pp. 428–438, June 2016.
- [89] Q. Zhang, L. T. Yang, Z. Chen, P. Li, and F. Bu, "An Adaptive Dropout Deep Computation Model for Industrial IoT Big Data Learning with Crowdsourcing to Cloud Computing," *IEEE Transactions on Industrial Informatics*, January 2018.
- [90] G. Grimmett, D. Stirzaker, *Probability and random processes*. Oxford university press, Oxford, U.S.A., 2001.
- [91] A. Adas, "Traffic models in broadband networks," *IEEE Communications Magazine*, vol. 35, no. 7, pp. 82–89, 1997.
- [92] OPNET Modeler. Accessed: 3-Sep.-2018. [Online]. Available: <https://www.riverbed.com/sg/products/steelcentral/opnet.html>
- [93] C. Xinjie, T. Tan, and K. Subramanian, "Source traffic modeling in OPNET," in *Proceedings of the OPNETWORK*, Washington DC., USA, September 1999.
- [94] H. Hassan, J. Garcia, and C. Bockstal, "Aggregate traffic models for VoIP applications," in *IEEE International Conference on Digital Telecommunications (ICDT'06)*, PP. 70, Côte d'Azur, France, August 2006.
- [95] 3GPP. Study on RAN Improvements for Machine-type communications, Technical report, TR 37.868, 2012. (Accessed: 08-August-2018). [Online]. Available: <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx>
- [96] M. Laner, N. Nikaein, D. Drajić, P. Svoboda, M. Popovic, and S. Krco, *Traffic models for machine-to-machine (M2M) communications: types and applications*. Book chapter of "Machine-to-machine communications, architecture, performance and applications?", Edited by M. Dolher and C. Anton, Woodhead Publishing, 2014. [Online]. Available: <http://www.eurecom.fr/publication/4265>, Accessed: 02 September 2018.
- [97] V. Gupta, S. K. Devar, N. H. Kumar, and K. P. Bagadi, "Modelling of IoT Traffic and Its Impact on LoRaWAN," in *IEEE Global Communications Conference (GLOBECOM 2017)*, Singapore, December 2017.
- [98] K. Smiljkovic, V. Atanasovski, and L. Gavrilovska, "Machine-to-Machine traffic characterization: Models and case study on integration in LTE," in *4th International Conference on Wireless Communications, Vehicular Technology, Information Theory and Aerospace & Electronic Systems (VITAE)*, Aalborg, Denmark, May 2014.

- [99] H. Heffes and D. Lucantoni, "A Markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance," *IEEE Journal on selected areas in communications*, vol. 4, no. 6, pp. 856–868, 1986.
- [100] R. C. Paiva, R. D. Vieira, and M. Saily, "Random access capacity evaluation with synchronized MTC users over wireless networks," in *IEEE 73rd Vehicular Technology Conference (VTC2011-Spring)*, pp. 1-5, Budapest, Hungary, May 2011.
- [101] T. Le-Ngoc and S. Shah-Heydari, "MMPP Modeling of Aggregated ATM Traffic," in *IEEE Canadian Conference on Electrical and Computer Engineering*, pp. 129-132, Waterloo, Canada, May 1998.

Part II

Appendix

Appendix A

OMNeT++

A.1 Introduction

Objective Modular Network Testbed in C++ (OMNeT++) [51] is a modular, component-based C++ simulation library and framework, primarily for building network simulators.

The OMNeT++ simulation adds functionality for creating and configuring models (NEtwork Description (NED) and Initialize (INI) files), performing batch executions and analyzing the simulation results. It has a generic architecture, so it can be used in various problem domains:

- Modeling of wired and wireless communication networks.
- Protocol modeling.
- Modeling of queueing networks.
- Modeling of multiprocessors and other distributed hardware systems.
- Validating of hardware architectures.
- Evaluating performance aspects of complex software systems.

- In general, modeling and simulation of any system where the discrete event approach is suitable, and can be conveniently mapped into entities communicating by exchanging messages.

OMNeT++ itself is not a simulator of anything concrete, but rather provides infrastructure and tools for writing simulations. One of the fundamental ingredients of this infrastructure is a component architecture for simulation models. Models are assembled from reusable components termed modules. Well-written modules are truly reusable, and can be combined in various ways like LEGO blocks.

Modules can be connected with each other via gates (other systems would call them ports), and combined to form compound modules. The depth of module nesting is unlimited. Modules communicate through message passing, where messages may carry arbitrary data structures.

Modules can pass messages along predefined paths via gates and connections, or directly to their destination; the latter is useful for wireless simulations, for example. Modules may have parameters that can be used to customize module behavior and/or to parameterize the model's topology. Modules at the lowest level of the module hierarchy are called simple modules, and they encapsulate model behavior. Simple modules are programmed in C++, and make use of the simulation library.

Finally, OMNEST is the commercially supported version of OMNeT++. OMNeT++ is free only for academic and non-profit use; for commercial purposes, one needs to obtain OMNEST licenses from Simulcraft Inc.

A.2 Installation

To install OMNET++, apply the following steps:

- 1) Download the OMNeT++ source code from <http://omnetpp.org>. Make sure you select the Windows-specific archive, named `omnetpp-4.6-src-windows.zip`.
- 2) The package is nearly self-contained: in addition to OMNeT++ files it includes a C++ compiler, a command-line build environment, and all libraries and programs

required by OMNeT++.

- 3) Copy the OMNeT++ archive to the directory where you want to install it.
- 4) Choose a directory whose full path does not contain any space; for example, do not put OMNeT++ under Program Files. Extract the zip file. To do so, right-click the zip file in Windows Explorer, and select "Extract All from the menu". You can also use external programs like Winzip or 7zip.
- 5) Rename the resulting directory to omnetpp-4.6. When you look into the new omnetpp-4.6 directory, should see directories named doc, images, include, tools, etc., and files named *mingwenv.cmd*, *configure*, *Makefile*, and others.
- 6) To configure and build OMNeT++, start "mingwenv.cmd" in the omnetpp-4.6 directory by double-clicking it in Windows Explorer. It will bring up a console with the *MSYS bash shell*, where the path is already set to include the omnetpp-4.6/bin directory. Enter the following commands:

```
./configure  
make
```

The build process will create both debug and release binaries.
- 7) To start OMNeT++, use the command "omnetpp".

A.3 OMNeT++ components

Any OMNeT++ project consists of the following components:

- NED files: The user describes the structure of a simulation model in the NED language. NED stands for Network Description. NED lets the user declare simple modules, and connect and assemble them into compound modules. NED files have the ".ned" extension.
- INI files: In OMNeT++, simulation models are parameterized and configured for execution using configuration files with the ".ini" extension, called INI files.

INI files are text files, which can be edited using any text editor. However, OMNeT++ 4.x introduces a tool expressly designed for editing INI files.

- cc files: A simple module (".cc" extension) is nothing more than a C++ class which has to be sub-classed with one or more virtual member functions redefined to define its behavior.

A.4 Simulation Launcher

The easiest way to launch a simulation is by selecting a project, folder, ini or NED file in Project Explorer, and clicking the *Run* button on the toolbar.

- If a folder is selected and it contains a single INI file, OMNeT++ will use this file to start the simulation.
- If an INI file is selected, it will be used during the launch as the main INI file for the simulation.
- If a NED file is selected which contains a network definition, the OMNeT++ will scan for INI files in the active projects and will try to find a configuration that allows this network to start.

A.5 User interface

As shown in Figure A-1, the top of the window contains the following elements below the menu bar:

- Toolbar: The toolbar lets you access the most frequently used functions, such as stepping, running and stopping the simulation.
- Status bar: Two rows of various fields and gauges, displaying the current event number, simulation time, information about the next simulation event, and other details. When the simulation is running, it displays performance data like

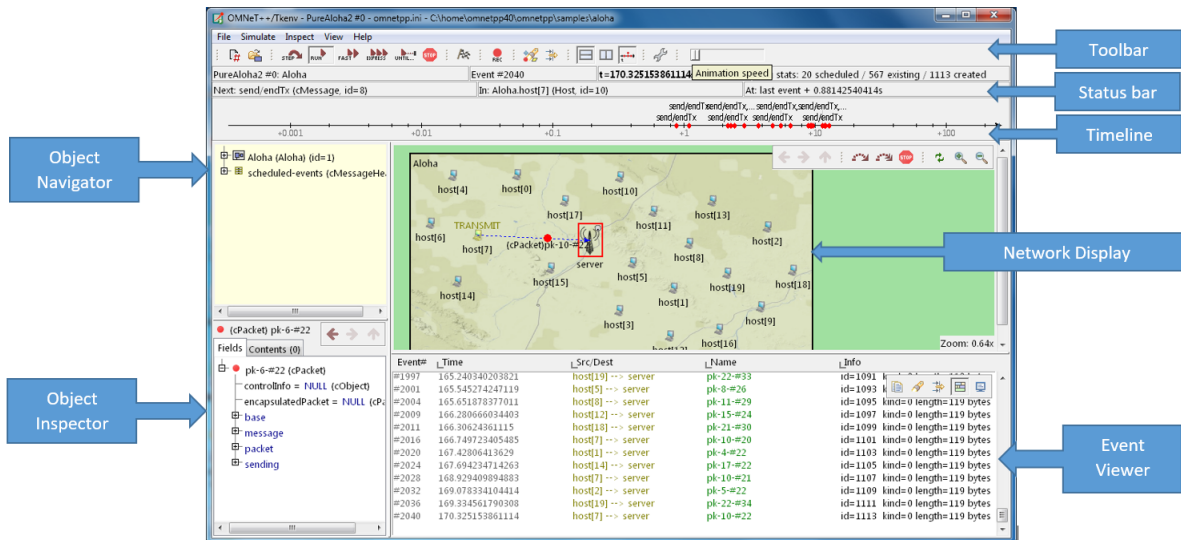


Figure A-1: The main window of OMNeT++ user interface.

the number of events processed per second. The second row can be turned off to free up vertical space.

- Timeline: Displays the contents of the future events set.

The main window is divided into the following areas:

- Object Navigator: Displays the hierarchy of objects in the current simulation.
- Object Inspector: Displays the contents and properties of the selected object.
- Network Display: Displays the network or any module graphically. This is also where animation takes place.
- Log Viewer: Displays the log of packets or messages sent between modules, or log messages output by modules during simulation.

A.6 Tic-Toc example

The Tic-Toc example is available in the OMNeT++ installation under the directory `samples/tictoc`. Tic-Toc is the most basic example in this appendix and it provides

a quick overview on how to use and understand the "sequence chart". Start the simulation and choose the simplest configuration, "Tictoc1", which specifies only two nodes called "Tic" and "Toc". During initialization, one of the nodes will send a message to the other. From then on, every time a node receives the message, it will simply send it back. This process continues until you stop the simulation. In Figure A-2, "Tic-Toc with two nodes" you can see how this is represented on a Sequence Chart. The two horizontal lines correspond to the two nodes and are labeled "Tic" and "Toc". The arrows represent message sends and the circles at the end of each arrow represent events. It is easy to see that all message sends take 100 milliseconds and that the first sender is the node "Tic".

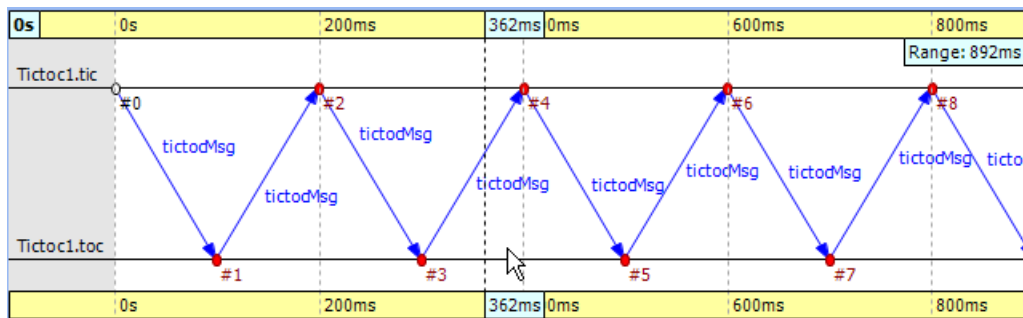


Figure A-2: Tic-Toc with two nodes.

In Figure A-3, there are six nodes tossing a message around until it reaches its destination. You can see how the message goes from one node to another, starting from node '0' and passing through it twice more, until it finally reaches its destination, node '3'.

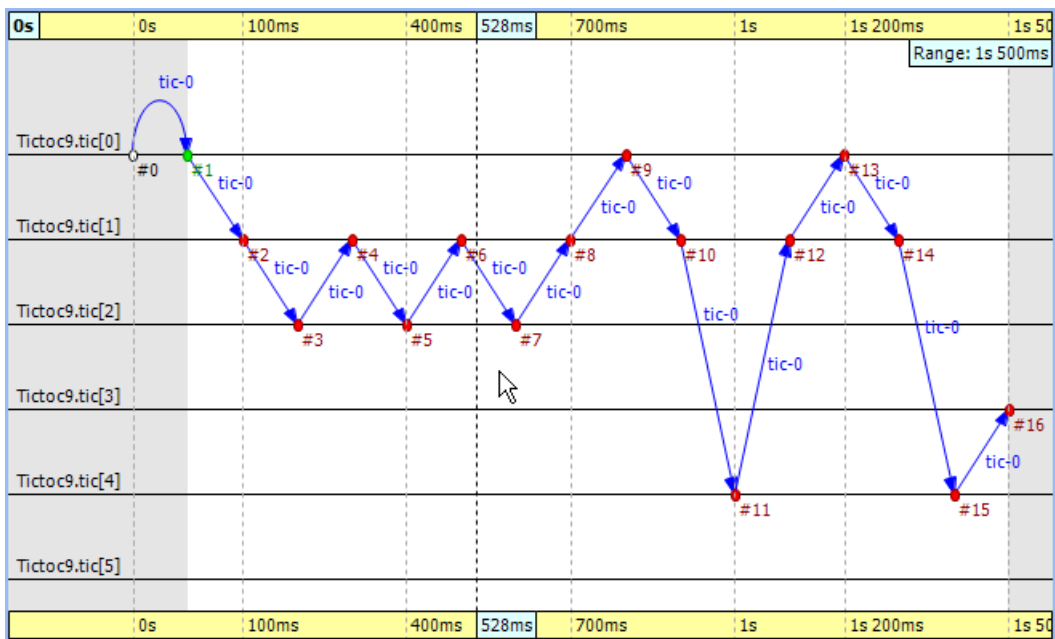


Figure A-3: Tic-Toc with six nodes.

Appendix B

INET

B.1 Introduction

INET Framework [52] is an open-source model library for the OMNeT++ simulation environment. It provides protocols, agents and other models for researchers and students working with communication networks. INET is especially useful when designing and validating new protocols, or exploring new scenarios.

INET contains models for the Internet stack (TCP, UDP, IPv4, IPv6, OSPF, BGP, etc.), wired and wireless link layer protocols (Ethernet, PPP, IEEE 802.11, etc), support for mobility, Mobile Ad Hoc Network (MANET) protocols, several application models and many other protocols and components.

Several other simulation frameworks take INET as a base, and extend it into specific directions, such as vehicular networks, overlay/peer-to-peer networks, or Long Term Evolution (LTE).

INET is built around the concept of modules that communicate by message passing. Agents and network protocols are represented by components, which can be freely combined to form hosts, routers, switches, and other networking devices. New components can be programmed by the user, and existing components have been written so that they are easy to understand and modify.

INET benefits from the infrastructure provided by OMNeT++. Beyond making use of the services provided by the OMNeT++ simulation kernel and library (com-

ponent model, parameterization, result recording, etc.), this also means that models may be developed, assembled, parameterized, run, and their results evaluated from the comfort of the OMNeT++ Simulation, or from the command line.

Some INET features:

- Open System Interconnection (OSI) layers implemented (physical, link-layer, network, transport, application).
- Pluggable protocol implementations for various layers.
- Internet Protocol (IP)-v4/IP-v6 network stack (or build your own network layer).
- Transport layer protocols: Transmission Control Protocol (TCP), User Datagram Protocol (UDP), Stream Control Transmission Protocol (SCTP) Routing protocols (ad-hoc and wired).
- Wired/wireless interfaces (Ethernet, Point-to-Point Protocol (PPP), Institute of Electrical and Electronics Engineers (IEEE) 802.11, etc.).
- Wide range of application models.
- Network emulation support.
- Mobility support.
- Supports the modeling of the physical environment.
- Visualization support.

B.2 Installation

There are several ways to install INET Framework:

- Let OMNeT++ download and install it for you. This is the easiest way. Just accept the offer to install INET in the dialog that comes up when you first start OMNeT++.

- From INET Framework website, <http://inet.omnetpp.org>. It always installs the last stable version compatible with your version of OMNeT++. If you need some other version, they are available for download from the website.
- From GitHub. If you have experience with git, clone the INET Framework project (`inet-framework/inet`), check out the revision of your choice, and follow the `INSTALL` file in the project root.

The INET Framework builds upon OMNeT++ and uses the same concept: modules that communicate by message passing. Hosts, routers, switches and other network devices are represented by OMNeT++ compound modules. These compound modules are assembled from simple modules that represent protocols, applications, and other functional units. A network is again an OMNeT++ compound module that contains host, router and other modules.

B.3 Practical example in INET

In this example we show the configuration of two hosts communicating through a wireless network using INET Framework.

In the first step, we want to create a network that contains two hosts, with one host sending a UDP data stream wirelessly to the other. Our goal is to keep the physical layer and lower layer protocol models as simple as possible. The scenario is depicted in Figure B-1.

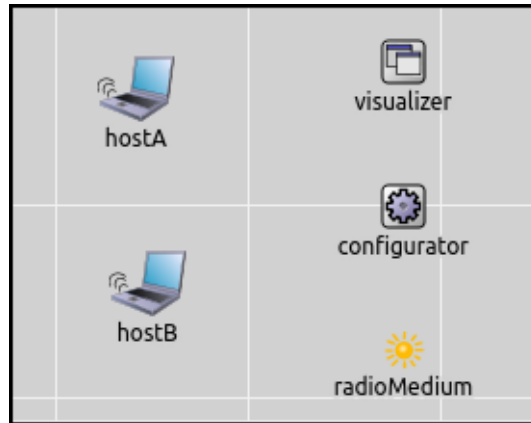


Figure B-1: Wireless network in INET Framework.

The NED source of the network is shown in Figure B-2.

NED source explanation:

- The playground: The model contains a playground of the size 500x650 meters, with two hosts spaced 400 meters apart. These numbers are set via "display" strings.
- The hosts: In INET, hosts are usually represented with the *StandardHost* NED type, which is a generic template for TCP/IP hosts. It contains protocol components like TCP, UDP and IP, slots for plugging in application models, and various Network Interface Cards (NIC)s. *StandardHost* has some variations in INET, for example *WirelessHost*, which is basically a *StandardHost* preconfigured for wireless scenarios.

As you can see, the hosts' type is parametric in this NED file (defined via a *hostType* parameter and the *INetworkNode* module interface). This is done so that in later steps we can replace hosts with a different NED type. The actual NED type here is *WirelessHost*, and later steps will override this setting using *omnetpp.ini*.

- Address assignment: IP addresses are assigned to hosts by an *Ipv4NetworkConfigurator* module, which appears as the *configurator* sub-module in the network. The hosts also need to know each others' Media Access Control (MAC)

addresses to communicate, which in this model is taken care of by using per-host *GlobalArp* modules instead of real Address Resolution Protocol (ARP).

- Traffic model: In the model, host A generates UDP packets which are received by host B. To this end, host A is configured to contain a *UdpBasicApp* module, which generates 1000-byte UDP messages at random intervals with exponential distribution, the mean of which is 12ms. Therefore the app is going to generate 100 kbyte/s (800 kbps) UDP traffic, not counting protocol overhead. Host B contains a *UdpSink* application that just discards received packets. The model also displays the number of packets received by host B.

The text is added by the *@figure[rcvdPkText]* line, and the subsequent line arranges the figure to be updated during the simulation.

- Physical layer modeling: Let us concentrate on the module called *radioMedium*. All wireless simulations in INET need a radio medium module. This module represents the shared physical medium where communication takes place. It is responsible for taking signal propagation, attenuation, interference, and other physical phenomena into account.

In this model, we configure the chosen physical layer model (*UnitDiskRadioMedium* and *UnitDiskRadio*) as follows. The communication range is set to 500 m. The radio data rates are set to 1 Mbps.


```

network Wirelessly
{
  parameters:
    @display("bgb=650,500;bgg=100,1,grey95");
    @figure[title](type=label; pos=0,-1; anchor=sw; color=darkblue);

    @figure[rcvdpkText](type=indicatorText; pos=380,20; anchor=w; font=,18; textFormat="packets received: %g"; initialValue=0);
    @statistic[packetReceived](source=hostB.app[0].packetReceived; record=figure(count); targetFigure=rcvdpkText);

  submodules:
    visualizer: <default("IntegratedCanvasVisualizer")> like IntegratedVisualizer if hasVisualizer() {
      @display("p=580,125");
    }
    configurator: IPv4NetworkConfigurator {
      @display("p=580,200");
    }
    radioMedium: <default("UnitDiskRadioMedium")> like IRadioMedium {
      @display("p=580,275");
    }
    hostA: <default("WirelessHost")> like INetworkNode {
      @display("p=50,325");
    }
    hostB: <default("WirelessHost")> like INetworkNode {
      @display("p=450,325");
    }
  }
}

```

Figure B-2: The NED source of a wireless network in INET Framework.

Results: When we run the simulation, Host A's *UdpBasicApp* generates UDP packets at random intervals. These packets are sent down via UDP and IPv4 to the network interface for transmission. The network interface queues packets, and transmits them as soon as it can. As long as there are packets in the network interface's transmission queue, packets are transmitted back-to-back, with no gaps between subsequent packets.

These events can be followed on OMNeT++'s run-time GUI. We can see a UDP packet being sent down from the *udpApp* sub-module, traversing the intermediate protocol layers, and being transmitted by the wlan interface.

Appendix C

SimuLTE

C.1 Introduction

SimuLTE [15] is an innovative simulation tool enabling complex system level performance-evaluation of Long Term Evolution (LTE) and LTE Advanced networks for the Objective Modular Network Testbed in C++ (OMNeT++) framework.

SimuLTE is written in C++ and is fully customizable with a simple pluggable interface. One can also develop new modules implementing new algorithms and protocols.

SimuLTE is an open source project building on top of OMNeT++ and INET Framework. The idea behind SimuLTE is to let researchers simulate and benchmark their solutions on an easy-to-use framework. It borrows the concept of modularity from OMNeT++ thus it is easy to extend.

C.2 Installation

Prerequisites:

SimuLTE requires the installation of OMNeT++ and INET Framework prior to its installation:

- Make sure your OMNeT++ installation works fine (e.g. try running the samples) and it is in the path (to test, try the command "which nedtool").

On Windows, open a console with the "mingwenv.cmd" command. The PATH and other variables will be automatically adjusted for you. Use this console to compile and run INET and SimuLTE.

- Install and test INET and be sure to check if the INET examples are running fine before continuing.

Building SimuLTE:

- a) Extract the downloaded SimuLTE tarball from its website (<http://simulte.com>) next to the INET directory.
- b) Change to the "simulte" directory.
- c) Type "make makefiles". This should generate the makefiles.
- d) Type "make" to build the SimuLTE executable.
- e) You can run examples by changing into a directory under 'simulations', and executing "./run"

C.3 Single Cell example

We are ready for setting up our first simulation scenario. In the omnetpp.ini file, we create a new configuration called SingleCell:

```
"/[Config SingleCell]  
network = lte.simulations.networks.SingleCell"
```

The network model (defined in the "omnetpp/simulations/networks" directory) contains one eNodeB serving a variable number of UEs.

UEs need to communicate with a server located on the internet. Thus, we configure the server so as it runs as many application as the number of UEs. UEs must be associated with the eNodeB, by setting "macCellId" and "masterId" parameters (eNodeB IDs are assigned progressively, starting from 1).

```

*.numUe = ${numUEs=5}

##### Amount of applications #####
*.ue[*].numUdpApps = 1
*.server.numUdpApps = ${numUEs}

# connect each UE to the eNB
**.ue[*].macCellId = 1
**.ue[*].masterId = 1

##### Positioning and mobility #####
*.eNodeB.mobility.initFromDisplayString = false
*.eNodeB.mobility.initialX = 300m
*.eNodeB.mobility.initialY = 300m
*.ue[*].mobility.constraintAreaMaxX = 600m
*.ue[*].mobility.constraintAreaMaxY = 600m
*.ue[*].mobility.constraintAreaMinX = 0m
*.ue[*].mobility.constraintAreaMinY = 0m
*.ue[*].mobility.initFromDisplayString = false
*.ue[*].mobility.initialX = uniform(0m,600m)
*.ue[*].mobility.initialY = uniform(0m,600m)
*.ue[*].mobility.speed = 1mps
*.ue[*].mobilityType = "LinearMobility"
#-----#

```

Figure C-1: Single cell configuration.

Mobility parameters (i.e., initial position and mobility type) are also specified both for the eNodeB and the UEs, as shown in Figure C-1.

Actually, till now we have not yet defined the type of the traffic. To do so, we create two different configurations by extending the previous one in Figure C-1: one for the downlink direction and one for the uplink direction, as shown in Figure C-2.

In this example, we use VoIP traffic model. The main difference between downlink and uplink is the modules where the "VoIPSender" and "VoIPReceiver" applications are located. Destination addresses and port numbers must be set such that one sender application can unequivocally identify the receiver application. In Figure C-3 a screen-shot of the graphical run-time environment.

```

[Config SingleCell-DL]
extends=SingleCell

*.ue[*].udpApp[0].typename = "VoIPReceiver"
*.server.udpApp[*].destAddress = "ue["+string(ancestorIndex(0))+"]"
*.server.udpApp[*].localPort = 3088+ancestorIndex(0)
*.server.udpApp[*].typename = "VoIPSender"
*.server.udpApp[*].startTime = uniform(0s,0.02s)
#-----#

[Config SingleCell-UL]
extends=SingleCell

*.server.udpApp[*].typename = "VoIPReceiver"
*.server.udpApp[*].localPort = 3000+ancestorIndex(0)
*.ue[*].udpApp[*].destAddress = "server"
*.ue[*].udpApp[*].destPort = 3000+ancestorIndex(1)
*.ue[*].udpApp[*].localPort = 3088
*.ue[*].udpApp[*].typename = "VoIPSender"
*.ue[*].udpApp[*].startTime = uniform(0s,0.02s)
#-----#

```

Figure C-2: Single cell traffic.

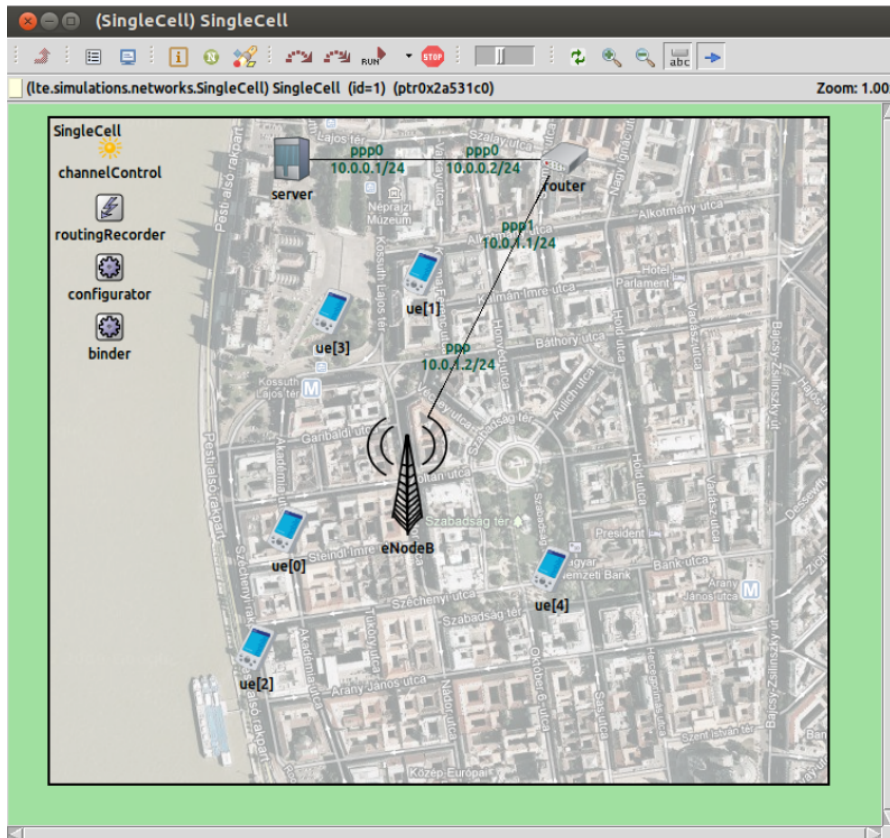


Figure C-3: Single cell graphical run-time environment.

C.4 ENSTAUL model

In our model, we use the open-source network simulator SimuLTE Modeler 0.9.1, in an environment of OMNeT++ 4.6 and INET 2.3.0, as shown in in Figure C-4.

We built our model "ENSTAUL" based on the previous example named "single cell". All files related to this example could be found in the following path:

"omnetpp/samples/simulte/simulations"

Our model consists of the following components:

- One evolved Node B (eNodeB): which provides connectivity to a User Equipment (UE); in another word it works as an access device in the LTE-A network.
- User Equipment (UE): In the Universal Mobile Telecommunications System (UMTS) and the 3rd Generation Partnership Project Long Term Evolution (LTE), a UE is any device used directly by an end-user to communicate. It can be a mobile telephone, a laptop computer equipped with a mobile broadband adapter, or any other device.

In our model we consider two types of traffic H2H and M2M traffics.

H2H traffic sub-categories are:

- a) ue_ftpd: A UE performing a File Transfer (FTP) download from a server "server_ftpd".
- b) ue_ftpu: A UE performing a FTP upload to a server "server_ftpu".
- c) ue_voipd: A UE performing a Voice over IP (VoIP) download from a server "server_voipd".
- d) ue_voipu: A UE performing a VoIP upload to a server "server_voipu".
- e) ue_video: A UE performing a video streaming download from a server "server_video".

M2M traffic is represented via M2M devices "u2_m2m" which perform uploads for their data to a M2M server "server_M2M".

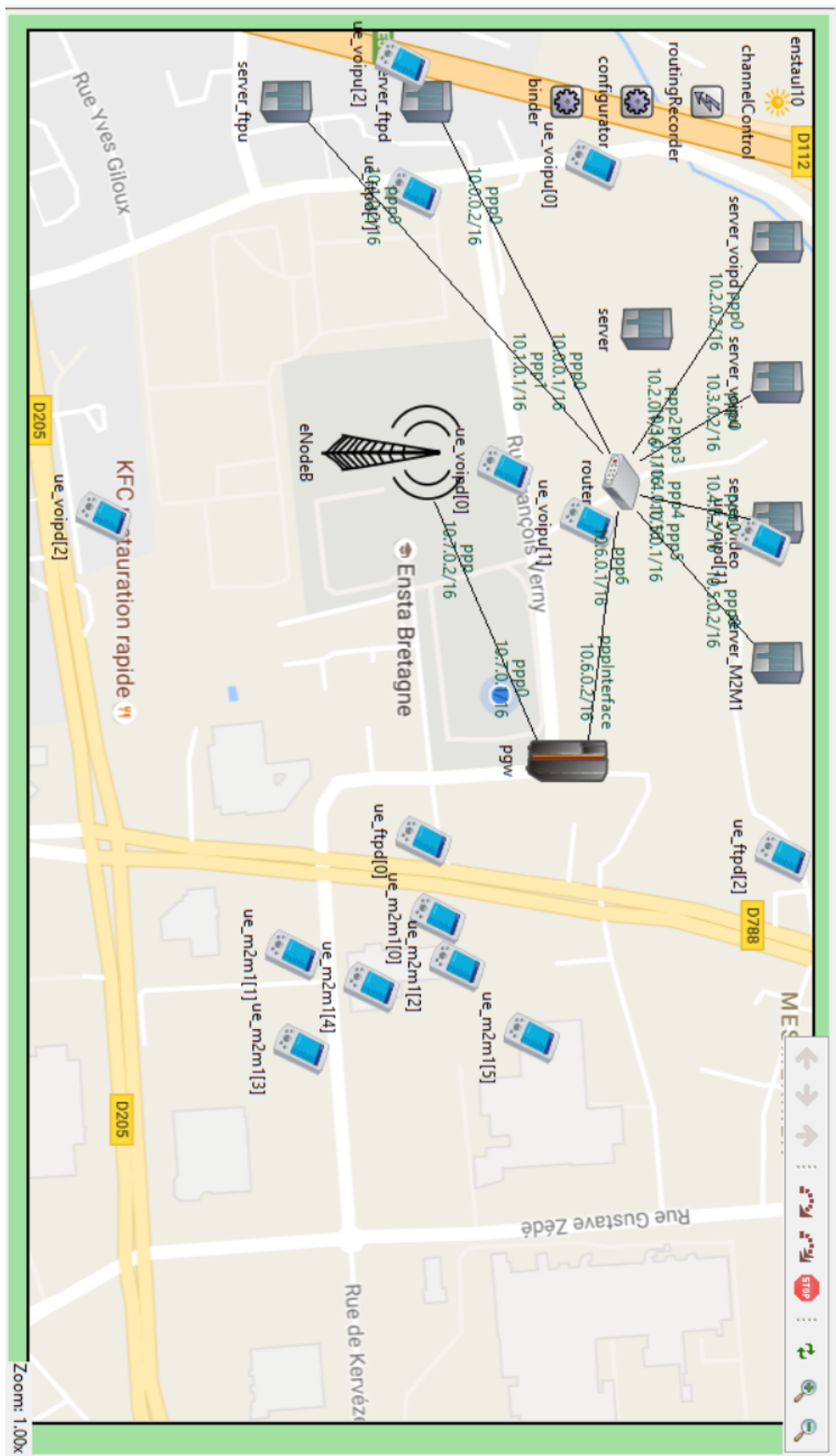


Figure C-4: ENSTAUL model

Our scenarios show an example of the M2M traffic load in an emergency event (e.g., earthquakes, fire, terrorist attacks, etc.). In such emergency events, besides of the regular H2H network traffic (VoIP, Video Streaming and file transfer), an additional M2M surge traffic attempts to access the network caused by the consequences of the emergency event.

The core of the scenario uses the SimuLTE Modeler to focus on the ability of an eNodeB to deal with a fixed number of H2H traffics (FTP-UL, FTP-DL, VoIP-UL, VoIP-DL, Video Streaming "10 each") with an increasing number of M2M requests attempting to access the LTE-A network simultaneously in 1 sec interval.

The SimuLTE scenario settings are given in Table C.1.

Table C.1: Simulation Parameters

Parameter	Value
Simulation Length	300 sec
Min./Max. (eNodeB-UE distance)	0 m / 300 m
Terminal velocity	120 Km/h
Mobility model	Linear Mobility
Transmission bandwidth	5 MHz (for DL and UL each)
No. of PRBs	25 (for DL and UL each)

The different traffic: VoIP, Video Streaming, file transfer and M2M are shown in Table C.2.

Table C.2: Traffic Models

VoIP Model	Parameter	Setting
	Application Packet	40 Bytes
	Interval	20 ms
	Talkspurts and Silences	Default settings
Video Streaming Model	Parameter	Setting
	Video Size	10 MB
	Packet Length	1000 Bytes
	Frame Interval	75 ms
M2M Model	Parameter	Setting
	Packet Size	128 Bytes
	Interval	1 sec
FTP Model	Parameter	Setting
	File Size	20 MB

We adjust all configurations in "omnetpp.ini" to simulate our scenario as follows:

```

1 [General]
2 tkenv image path=../../images
3 tkenv plugin path = ../../../../inet/etc/plugins
4 tkenv default config =
5 output scalar file append = true
6 **.vector recording = true
7
8 #===== Simulation parameters =====
9 sim time limit=300s
10 warmup period=2s
11 repeat = 1
12
13 #===== Channel parameters =====

```

```

14 **.channelControl.pMax = 10W
15 **.channelControl.alpha = 1.0
16 **.channelControl.carrierFrequency = 2100e+6Hz
17
18 #===== PhyLayer & Feedback parameters
    =====
19 **.nic.phy.channelModel=xmldoc("config_channel.xml")
20 **.feedbackComputation = xmldoc("config_channel.xml")
21 **.configurator.config = xmldoc("demo.xml")
22 **.fbDelay = 1
23
24 #===== Mobility parameters =====
25 **.mobility.constraintAreaMinZ = 0m
26 **.mobility.constraintAreaMaxZ = 0m
27
28 #===== AMC module parameters =====
29 **.rbAllocationType = "localized"
30 **.deployer.numRbDl = 25
31 **.deployer.numRbUl = 25
32 **.numBands = 6
33
34 #===== Statistics =====
35 output scalar file = ${resultdir}/${configname}/${repetition
    }.sca
36 output vector file = ${resultdir}/${configname}/${repetition
    }.vec
37 seed set = ${repetition}
38
39 #####
40 # SingleCell Configuration #

```

```

41 #####
42 [ Config mySingle ]
43 network = lte.simulations.networks.enstaui10
44
45 #Number of UEs in the network
46
47 #Change the number of UEs for every application in this part
48
49 #Number of UEs downloading voip from the server in group
    Group_VoIP_DL
50 *.numVoip_download = ${voipd=10}
51
52 #Number of UEs uploading voip to server in group
    Group_VoIP_UL
53 *.numVoip_upload = ${voipu=10}
54
55 #number of UEs watching video in group Group_Videostreaming
56 *.numvideo = ${videostreaming=10}
57
58 #number of UEs downloading FTP in group Group_FTP_DL
59 *.numFtp_download = ${ftpdownload=10}
60
61 #Number of UEs uploading FTP in group Group_FTP_UL
62 *.numFtp_upload = ${ftpupload=10}
63
64 #Number of UEs sending ue_m2m signals    First Group of M2M
65 *.num_m2m1 = ${m2m1= 300}
66
67 #Number of External Cells in the Network
68 *.numExtCells = ${extcells = 0}

```

```

69
70 #===== Amount of applications =====
71
72 #numUDP apps on UE
73 *.ue_voipd[*].numUdpApps = 1
74 *.ue_voipu[*].numUdpApps = 1
75 *.ue_vid[*].numUdpApps = 1
76
77 # numTcp apps on UE
78 *.ue_ftpd[*].numTcpApps = 1
79 *.ue_ftpu[*].numTcpApps = 1
80 *.ue_m2m1[*].numTcpApps = 1
81
82 #===== Amount of applications =====
83
84 # connect each UE to the eNB
85 **.ue*[*].macCellId = 1
86 **.ue*[*].masterId = 1
87
88 #server Apps
89
90 #FTP:TCP apps on Server
91 *.server_ftpd.numTcpApps = ${ftpdownload}
92 *.server_ftpu.numTcpApps = ${ftpupload}
93 *.server_M2M1.numTcpApps = ${m2m1}
94
95 #UDP APPS on Server
96 *.server_voipd.numUdpApps = ${voipd}
97 *.server_voipu.numUdpApps = ${voipu}
98 *.server_video.numUdpApps = ${videostreaming}

```

99

100

101 #===== Positioning and mobility =====

102 *.eNodeB.mobility.initFromDisplayString = false

103 *.eNodeB.mobility.initialX = 300m

104 *.eNodeB.mobility.initialY = 300m

105 #voipd location

106 *.ue_voipd*.mobility.constraintAreaMaxX = 600m

107 *.ue_voipd*.mobility.constraintAreaMaxY = 600m

108 *.ue_voipd*.mobility.constraintAreaMinX = 0m

109 *.ue_voipd*.mobility.constraintAreaMinY = 0m

110 *.ue_voipd*.mobility.initFromDisplayString = false

111 *.ue_voipd*.mobility.initialX = uniform(0m,600m)

112 *.ue_voipd*.mobility.initialY = uniform(0m,600m)

113 #voipu location

114 *.ue_voipu*.mobility.constraintAreaMaxX = 600m

115 *.ue_voipu*.mobility.constraintAreaMaxY = 600m

116 *.ue_voipu*.mobility.constraintAreaMinX = 0m

117 *.ue_voipu*.mobility.constraintAreaMinY = 0m

118 *.ue_voipu*.mobility.initFromDisplayString = false

119 *.ue_voipu*.mobility.initialX = uniform(0m,600m)

120 *.ue_voipu*.mobility.initialY = uniform(0m,600m)

121 #video location

122 *.ue_vid*.mobility.constraintAreaMaxX = 600m

123 *.ue_vid*.mobility.constraintAreaMaxY = 600m

124 *.ue_vid*.mobility.constraintAreaMinX = 0m

125 *.ue_vid*.mobility.constraintAreaMinY = 0m

126 *.ue_vid*.mobility.initFromDisplayString = false

127 *.ue_vid*.mobility.initialX = uniform(0m,600m)

128 *.ue_vid*.mobility.initialY = uniform(0m,600m)

```

129 #ftpd location
130 *.ue_ftpd*.mobility.constraintAreaMaxX = 600m
131 *.ue_ftpd*.mobility.constraintAreaMaxY = 600m
132 *.ue_ftpd*.mobility.constraintAreaMinX = 0m
133 *.ue_ftpd*.mobility.constraintAreaMinY = 0m
134 *.ue_ftpd*.mobility.initFromDisplayString = false
135 *.ue_ftpd*.mobility.initialX = uniform(0m,600m)
136 *.ue_ftpd*.mobility.initialY = uniform(0m,600m)
137 #ftpu location
138 *.ue_ftpu*.mobility.constraintAreaMaxX = 600m
139 *.ue_ftpu*.mobility.constraintAreaMaxY = 600m
140 *.ue_ftpu*.mobility.constraintAreaMinX = 0m
141 *.ue_ftpu*.mobility.constraintAreaMinY = 0m
142 *.ue_ftpu*.mobility.initFromDisplayString = false
143 *.ue_ftpu*.mobility.initialX = uniform(0m,600m)
144 *.ue_ftpu*.mobility.initialY = uniform(0m,600m)
145 #m2m1 location
146 *.ue_m2m1*.mobility.constraintAreaMaxX = 800m
147 *.ue_m2m1*.mobility.constraintAreaMaxY = 600m
148 *.ue_m2m1*.mobility.constraintAreaMinX = 0m
149 *.ue_m2m1*.mobility.constraintAreaMinY = 0m
150 *.ue_m2m1*.mobility.initFromDisplayString = false
151 *.ue_m2m1*.mobility.initialX = uniform(600m,800m)
152 *.ue_m2m1*.mobility.initialY = uniform(200m,400m)
153 #*.extCell*.position_x= uniform(0m,600m)
154 #*.extCell*.position_y = uniform(0m,600m)
155
156
157 *.ue*.mobility.speed = 33.333 mps
158 *.ue*.mobilityType = "LinearMobility"

```



```

159 #                                     #
160
161 #####
162 # SingleCell Configuration
163 #                                     #
164 #####
165 [ Config new_mix ]
166 extends = mySingle
167
168 #Group_VoIP_DL downloading voip from server_voipd
169 *.ue_voipd[*].udpApp[0].typename = "VoIPReceiver"
170 *.server_voipd.udpApp[*].destAddress = "ue_voipd["+string(
171     ancestorIndex(0))+"]"
172 *.server_voipd.udpApp[*].localPort = 9999+ancestorIndex(0)
173 *.server_voipd.udpApp[*].typename = "VoIPSender"
174 *.server_voipd.udpApp[*].startTime = uniform(0s,0.02s)
175
176 #Group_VoIP_UL uploading voip on server_voipu
177 *.server_voipu.udpApp[*].typename = "VoIPReceiver"
178 *.server_voipu.udpApp[*].localPort = 9500+ancestorIndex(0)
179 *.ue_voipu[*].udpApp[0].destAddress = "server_voipu"
180 *.ue_voipu[*].udpApp[0].destPort = 9500+ancestorIndex(1)
181 *.ue_voipu[*].udpApp[0].localPort = 3088
182 *.ue_voipu[*].udpApp[0].typename = "VoIPSender"
183 *.ue_voipu[*].udpApp[0].startTime = uniform(0s,0.02s)
184
185 #Group_Videostreaming watching a video
186 *.ue_vid[*].udpApp[0].typename = "UDPVideoStreamCli"

```

```

187 *.ue_vid[*].udpApp[0].serverAddress = "server_video"
188 *.ue_vid[*].udpApp[0].localPort = 9999
189 *.ue_vid[*].udpApp[0].serverPort = 3088+ancestorIndex(1)
190
191
192 *.server_video.udpApp[*].typename = "UDPVideoStreamSvr"
193 *.server_video.udpApp[*].destAddress = "ue_vid["+string(
    ancestorIndex(0))+"]"
194 *.server_video.udpApp[*].localPort = 3088+ancestorIndex(0)
195 *.server_video.udpApp[*].sendInterval = 0.2ms
196
197 *.server_video.udpApp[*].videoSize = 10MiB
198 *.server_video.udpApp[*].startTime = uniform(0s,0.02s)
199 *.server_video.udpApp[*].packetLen = ${packetLen = 1000B}
200 *.mtu = 10000B
201
202 #Group_FTP_DL downloading from server_ftpd
203 *.ue_ftpd[*].tcpApp[0].typename = "TCPSinkApp"
204 *.server_ftpd.tcpApp[*].connectAddress = "ue_ftpd["+string(
    ancestorIndex(0))+"]"
205 *.server_ftpd.tcpApp[*].localPort = 3088+ancestorIndex(0)
206 *.server_ftpd.tcpApp[*].typename = "TCPSessionApp"
207 *.server_ftpd.tcpApp[*].sendBytes = 20MiB
208 *.server_ftpd.tcpApp[*].startTime = 0s
209
210
211 #Group_FTP_UL uploading ftp to server_ftpu
212 *.server_ftpu.tcpApp[*].typename = "TCPSinkApp"
213 *.server_ftpu.tcpApp[*].localPort = 5000+ancestorIndex(0)
214 *.ue_ftpu[*].tcpApp[0].typename = "TCPSessionApp"

```

```

215 *.ue_ftpu[*].tcpApp[0].connectAddress="server_ftpu"
216 *.ue_ftpu[*].tcpApp[0].connectPort=5000+ancestorIndex(1)
217
218 *.ue_ftpu[*].tcpApp[0].sendBytes = 20MiB
219
220 #Group_ue_m2m1 uploading ftp to server_ue_m2m First Group
221 *.server_M2M1.tcpApp[*].typename = "TCPSinkApp"
222 *.server_M2M1.tcpApp[*].localPort = 12000+ancestorIndex(0)
223 *.ue_m2m1[*].tcpApp[0].typename = "TCPSessionApp"
224 *.ue_m2m1[*].tcpApp[0].connectAddress="server_M2M1"
225 *.ue_m2m1[*].tcpApp[0].connectPort=12000+ancestorIndex(1)
226 *.ue_m2m1[*].tcpApp[0].topen = 0s
227 *.ue_m2m1[*].tcpApp[0].sendBytes = 0B
228 #*.ue_m2m1[*].tcpApp[0].sendScript = "0.800626 128;1.69727
    128;2.78977 128;3.76491 128;4.97632 128;5.86337
    128;7.01861 128;7.9706 128;8.8163 128;9.68996 128;10.9842
    128;11.8437 128;12.6482 128;13.714 128;14.8149 128;15.698
    128;16.7234 128;17.5519 128;18.7436 128;19.8035
    128;21.0415 128;22.3194 128;23.3891 128;24.4201
    128;25.6512 128;26.841 128;28.1394 128;29.2452 128;30.1783
    128;31.3983 128;32.3862 128;33.5248 128;34.3292
    128;35.2672 128;36.3611 128;37.5799 128;38.6224
    128;39.7942 128;40.8232 128;41.9954 128;43.0949
    128;44.2624 128;45.3486 128;46.2244 128;47.237 128;48.2955
    128;49.4713 128;50.3557 128;51.4017 128;52.5516 128;"
229 include script.ini

```

We also created a M2M traffic generator to mimic the M2M behavior in the file "script.ini" which includes:

```

1 *.ue_m2m1[0..9].tcpApp[0].sendScript = "1.73767 128;2.59651

```

```

128;3.51298 128;4.43521 128;5.33747 128;6.25099 128;
..... 266.676 128;267.607 128;268.479 128;269.339
128;"
2 *.ue_m2m1[10..19].tcpApp[0].sendScript = "1.05365 128;1.96341
128;2.85797 128;3.78886 128;4.70285 128;5.5612 128;
..... 267.312 128;268.216 128;269.152 128;270.082
128;"
3 *.ue_m2m1[20..29].tcpApp[0].sendScript = "1.05282 128;1.93837
128;2.8085 128;3.70694 128;4.59328 128;5.45686 128;
..... 267.312 128;268.216 128;269.152 128;270.082
128;"
4 *.ue_m2m1[30..39].tcpApp[0].sendScript = "1.2727 128;2.16478
128;3.08617 128;3.94979 128;4.84642 128;5.74896 128;
..... 267.629 128;268.549 128;269.479 128;270.374
128;"
5 *.ue_m2m1[40..49].tcpApp[0].sendScript = "1.28242 128;2.18509
128;3.08105 128;4.0074 128;4.92072 128;5.8537 128;
..... 267.537 128;268.433 128;269.377 128;270.264
128;"
6 *.ue_m2m1[50..59].tcpApp[0].sendScript = "1.30966 128;2.22044
128;3.10559 128;3.96458 128;4.88599 128;5.81863
128;..... 267.775 128;268.698 128;269.597 128;270.535
128;"
7 *.ue_m2m1[60..69].tcpApp[0].sendScript = "1.28566 128;2.23374
128;3.12732 128;4.00701 128;4.92503 128;5.79823 128;
..... 268.275 128;269.209 128;270.145 128;271.03 128; "
8 *.ue_m2m1[70..79].tcpApp[0].sendScript = "1.86477 128;2.81391
128;3.6749 128;4.57069 128;5.50624 128;6.37552 128;
..... 266.685 128;267.609 128;268.534 128;269.461
128;"

```

```
9 *.ue_m2m1[80..89].tcpApp[0].sendScript = "1.09959 128;2.04002
    128;2.98646 128;3.84231 128;4.74318 128;5.61345 128;
    ..... 267.34 128;268.239 128;269.15 128;270.083 128; "
```

10

11 Because the file is too long, we cannot include **all** its
contents. So, we omit the lines **for** the group of M2M
devices: ue_m2m1[90..269].

12

```
13 *.ue_m2m1[270..279].tcpApp[0].sendScript = "1.05733
    128;1.92404 128;2.85593 128;3.80175 128;4.74835
    128;5.67609 128; ..... 268.364 128;269.228 128;270.167
    128;271.046 128; "
```

```
14 *.ue_m2m1[280..289].tcpApp[0].sendScript = "1.25304
    128;2.20236 128;3.11034 128;4.05617 128;4.9115 128;5.85033
    128; ..... 267.553 128;268.427 128;269.322 128;270.253
    128; "
```

```
15 *.ue_m2m1[290..299].tcpApp[0].sendScript = "1.72234
    128;2.64798 128;3.5326 128;4.48047 128;5.34227 128;6.20482
    128; ..... 267.695 128;268.575 128;269.463 128;270.331
    128; "
```

Appendix D

CTMC analysis code in MATLAB

D.1 CTMC analytical methodology

Our CTMC analytical methodology consists of four steps:

- First, we use CTMC as a stochastic method to describe the sequence of possible events M2M or H2H built on different states and various priorities.

- Second, we generate the equilibrium equations which rule the probability of transitioning from one state to another.

- Third, we turn these probabilities and states into a linear system and by solving it we can calculate each state probability under certain conditions.

- Fourth, using the aforementioned probabilities, we propose some metrics to characterize the performance of the network, with the aim to evaluate M2M and H2H traffics.

In our system states, two variables (i, j) are considered to denote the number of two ongoing services one is high priority traffic and the other is low priority traffic respectively. The system moves from one state to another if an event occurs (increase/decrease of i or j). The two traffics have two average arrival rates (λ_1, λ_2) assumed to follow Poisson distribution and two service rates (μ_1, μ_2) assumed to follow an exponential distribution.

D.2 Matlab code

```
1 %           Continuous Time Markov Chain
2 % % Note: The equations were checked and the test for C=3
   was validated
3 %           and compared with manual equation generation.
4 %           This program automatically generates equations for any
   C value.
5 %           It also computes the following performance metrics
6 %           * CR_M2M: Service completion rate for M2M
7 %           * CR_H2H: Service completion rate for H2H
8 %
9 % Assumptions
10 % Li == Lambda(i)
11 % Mi == Mue(i)
12 % Picj == P(i,j)
13 % NB: 'c' represent the comma to separate i and j
14 clear all;
15
16 % System inputs:
17 C = 25; % C>0
18 % syms L1 L2 M1 M2;
19 L1=50;
20 L2=200;
21 M1=5;
22 M2=1;
23
24 disp(['
   ']);
25 disp(['           Continuous Time Markov Chain model']);
```

```

26 disp(['      C = ', num2str(C)]);
27 disp(['
      ']);
28 %% System Specification: Balance equations from graph
29 %      > Case 1: This is not considered in the equations.
30 % Equation 1 : (L1+L2)*P0c0 == M1*P1c0 + M2*P0c1
31 %      > Case 2
32 % Equation 2: (L1+L2+i*M1+j*M2)*Picj == L1*Pi-1cj + L2*Picj-1
      + (i+1)*M1*Pi+1cj
33 % + (j+1)*M2*Picj+1
34 %      > Case 3
35 % Equation 3: (i*M1+j*M2)*Picj == L1*Pi-1cj + L2*Picj-1
36 %      > Norm Equation
37 % Equation 4: Sum(Pij) = 1
38
39
40 %% Equations generation
41 DEBUG = [];
42 v_P = [];
43 str_Picj = [];
44 v_Equations = [];
45 eqn_temp = '0';
46
47 disp([' * Equation generation ... ']);
48 % Generation of Pij + Equation 4
49 for i=1:C+1
50     for j=1:C+1
51         if(i+j-2>C)
52             continue;
53         end

```



```

54
55     % i 1: ongoing H2H & j 1: ongoing M2M
56     DEBUG = [DEBUG sprintf( '(%d,%d) ', i 1 , j 1 ) ];    % For
                    debugging.
57     % constructing the Pij string
58     str = strcat( 'P', num2str(i 1) , 'c', num2str(j 1) );
59     % Declaration as symbols
60     eval( sprintf( 'syms %s;', str ) );
61     % Update the v_P vector
62     eval( sprintf( 'v_P = [v_P %s]; ', str ) );
63     if (i==1) && (j==1)
64         str_Picj = 'P0c0';
65     else
66         str_Picj = strcat( str_Picj , ',' , str );
67     end
68     % Equation 4
69     eqn_temp = strcat( eqn_temp , ' + ' , str );
70
71     end
72 end
73
74 eqn_temp = strcat( eqn_temp , ' == 1 ' ); % Completing equation
75     4
76     v_Equations = [ v_Equations , eval( eqn_temp ) ]; % Adding the
77     equation to the vector of equations
78
79 % Generation of equations 2 and 3
80 for i=1:C+1
81     for j=1:C+1
82         % i 1: ongoing H2H & j 1: ongoing M2M

```

```

81     clear eqn_temp;
82
83     if (i+j 2 > C)
84         continue;
85     elseif ((i+j 2 < C) && (i+j 2 > 0)) % Equation 2
86         % (L1+L2+i*M1+j*M2)*Pij == L1*Pi 1 , j + L2*Pi , j 1 +
            (i+1)*M1*Pi+1 , j + (j+1)*M2*Pi , j+1
87         if (i 1 == 0)
88             eqn_temp = strcat(' (L1+L2+', num2str(i 1), '*M1+',
                num2str(j 1), '*M2)*P', num2str(i 1), 'c',
                num2str(j 1), ' == L2*P', num2str(i 1), 'c',
                num2str(j 2), '+', num2str(i), '*M1*P',
                num2str(i), 'c', num2str(j 1), '+', num2str(j
                ), '*M2*P', num2str(i 1), 'c', num2str(j)) ;
89         elseif (j 1 == 0)
90             eqn_temp = strcat(' (L1+L2+', num2str(i 1), '*M1+',
                num2str(j 1), '*M2)*P', num2str(i 1), 'c',
                num2str(j 1), ' == L1*P', num2str(i 2), 'c',
                num2str(j 1), '+', num2str(i), '*M1*P',
                num2str(i), 'c', num2str(j 1), '+', num2str(j
                ), '*M2*P', num2str(i 1), 'c', num2str(j)) ;
91         else
92             eqn_temp = strcat(' (L1+L2+', num2str(i 1), '*M1+',
                num2str(j 1), '*M2)*P', num2str(i 1), 'c',
                num2str(j 1), ' == L1*P', num2str(i 2), 'c',
                num2str(j 1), '+ L2*P', num2str(i 1), 'c',
                num2str(j 2), '+', num2str(i), '*M1*P',
                num2str(i), 'c', num2str(j 1), '+', num2str(j
                ), '*M2*P', num2str(i 1), 'c', num2str(j)) ;
93     end

```

```

94         v_Equations = [v_Equations eval(eqn_temp)];
95     elseif(i+j-2 == C) % Equation 3
96         % (i*M1+j*M2)*Pij == L1*Pi-1,j + L2*Pi,j-1
97         if (i-1 == 0)
98             eqn_temp = strcat('(' , num2str(i-1) , '*M1+' ,
99                             num2str(j-1) , '*M2)*P' , num2str(i-1) , 'c' ,
100                             num2str(j-1) , ' == L2*P' , num2str(i-1) , 'c' ,
101                             num2str(j-2));
102         elseif (j-1 == 0)
103             eqn_temp = strcat('(' , num2str(i-1) , '*M1+' ,
104                             num2str(j-1) , '*M2)*P' , num2str(i-1) , 'c' ,
105                             num2str(j-1) , ' == L1*P' , num2str(i-2) , 'c' ,
106                             num2str(j-1));
107         else
108             eqn_temp = strcat('(' , num2str(i-1) , '*M1+' ,
109                             num2str(j-1) , '*M2)*P' , num2str(i-1) , 'c' ,
110                             num2str(j-1) , ' == L1*P' , num2str(i-2) , 'c' ,
111                             num2str(j-1) , '+ L2*P' , num2str(i-1) , 'c' ,
112                             num2str(j-2));
113         end
114     v_Equations = [v_Equations eval(eqn_temp)];
115     else % Equation 1 is not considered in the solution
116         continue;
117     end
118 end
119 end
120 %% Equation Solve

```

```

114 % v_Equations and v_P are previously generated
115 % str_Picj: is a string that contains all Picj names
116
117 disp([' * Equation solve ... ']);
118 str= strcat('[' , str_Picj , ']= solve(v_Equations , v_P);');
119 str
120 eval(str);
121 %[P0c0, P0c1, P0c2, P0c3, P1c0, P1c1, P1c2, P2c0, P2c1, P3c0]
    = solve(v_Equations , v_P);
122 %[P0c0, P0c1, P0c2, P0c3, P1c0, P1c1, P1c2, P2c0, P2c1, P3c0
    param conditions]= solve(v_Equations , v_P, '
    ReturnConditions',true );
123
124 %% Performance Metrics > The system performances should be
    generalized also
125 % S: State Space
126 % CR_M2M = Service completion rate M2M: Sum(j*M2*Picj)
127 % CR_M2M = 0*M2*Solution.P00 + 1*M2*Solution.P01 + 2*M2*
    Solution.P02 + 3*M2*Solution.P03 + 0*M2*Solution.P10 + 1*
    M2*Solution.P11 + 2*M2*Solution.P12 + 0*M2*Solution.P20 +
    1*M2*Solution.P21 + 0*M2*Solution.P30;
128
129 disp([' * System computation ... ']);
130
131 CR_M2M = 0;
132 % Generation of equations 2 and 3
133 for i=1:C+1
134     for j=1:C+1
135         % i 1: ongoing H2H & j 1: ongoing M2M
136         if (i+j 2>C)

```

```

137         continue;
138     end
139     strM2M = strcat('CR_M2M = CR_M2M + ', num2str(j-1), '*
140         M2*P', num2str(i-1), 'c', num2str(j-1), ';');
141     eval(strM2M);
142 end
143 eval(CR_M2M)
144 % CR_H2H = Service completion rate M2M: Sum(i*M1*Pij)
145 % CR_H2H = 0*M1*Solution.P00 + 0*M1*Solution.P01 + 0*M1*
146     Solution.P02 + 0*M1*Solution.P03 + 1*M1*Solution.P10 + 1*
147     M1*Solution.P11 + 1*M1*Solution.P12 + 2*M1*Solution.P20 +
148     2*M1*Solution.P21 + 3*M1*Solution.P30;
149
150 CR_H2H = 0;
151 % Generation of equations 2 and 3
152 for i=1:C+1
153     for j=1:C+1
154         % i-1: ongoing H2H & j-1: ongoing M2M
155         if(i+j-2>C)
156             continue;
157         end
158         strH2H = strcat('CR_H2H = CR_H2H + ', num2str(i-1), '*
159             M1*P', num2str(i-1), 'c', num2str(j-1), ';');
160         eval(strH2H);
161     end
162 end
163 eval(CR_H2H)
164 % Perf2: Resource Utilization M2M and H2H: Sum (((i+j)/C)*Pij
165     );

```

```

161 % Perf2 = ((0+0)/C)*Solution.P00 + ((0+1)/C)*Solution.P01 +
      ((0+2)/C)*Solution.P02 + ((0+3)/C)*Solution.P03 + ((1+0)/C
      ) * Solution.P10 + ((1+1)/C)*Solution.P11 + ((1+2)/C)*
      Solution.P12 + ((2+0)/C)*Solution.P20 + ((2+1)/C)*Solution
      .P21 + ((3+0)/C)*Solution.P30;

162
163
164 %% Evaluation 1
165 disp([' * Evaluation 1... ']);
166
167 L1 = [0:0.1:2];
168 L2 = 1;
169 M1 = 0.5;
170 M2 = 0.5;
171
172 %disp(['Service completion rate of M2M: ', num2str(eval(Perf1
      ))]);
173 % disp(['Total resource utilization: ', num2str(eval(Perf2))
      ]);
174
175 figure , plot(L1, eval(CR_M2M), 'k' , L1, eval(CR_H2H), 'r');
176 % figure , plot(L1, eval(Perf2));
177
178 %% Evaluation 2
179 % L1 = 1;
180 % L2 = [0:0.1:2];
181 % M1 = 0.5;
182 % M2 = 0.5;
183
184 %disp(['Service completion rate of M2M: ', num2str(eval(Perf1

```

```

    )]);
185 % disp(['Total resource utilization: ', num2str(eval(Perf2))
    ]);
186
187 % figure , plot(L2, eval(CR_M2M), 'k' , L2, eval(CR_H2H), 'r');
188 % figure , plot(L1, eval(Perf2));
189
190
191 %% Saving workspace
192 save(sprintf('CTMM_workspace_C%s', num2str(C)));

```

Appendix E

M/M/1 Queuing Model in SimuLink

E.1 Overview

This example shows how to model a single-queue, single-server system with a double traffic source and an infinite storage capacity. In the notation, the M stands for Markovian; M/M/1 means that the system has a Poisson arrival process, an exponential service time distribution, and one server. Queuing theory provides exact theoretical results for some performance measures of an M/M/1 queuing system and this model makes it easy to compare empirical results with the corresponding theoretical results.

E.2 Structure

The proposed architecture in Figure E-1 is based on Matlab libraries presented in [83]. Many enhancements are made in order to match our CANAL architecture presented in chapter (5):

- The architecture consists of a single-queue and single-server system with a double traffic sources. In order to omit the role of the queuing for the moment, we consider an infinite storage capacity.
- Entity Generator block: Models a Poisson arrival process by generating entities. Arrivals occur at an average rate λ according to a Poisson process and move the process from state (i) to $(i + 1)$ and/or (j) to $(j + 1)$.

- Simulink Function "*exponentialArrivalTime()*": The "entity server" block models H2H and M2M traffics one at a time by allowing one event only (either \mathbf{i} or \mathbf{j}) in the smallest time interval. The service time has an exponential distribution with a rate parameter $\boldsymbol{\mu}$, where $\frac{1}{\boldsymbol{\mu}}$ is the mean service time.
- Entity Queue block: Stores entities that have yet to be served in First Input First Output (FIFO) order or by priority discipline, with a flexible buffer size.
- When the service is completed, the served request leaves the system and the number of ongoing services in the system reduces by one (\mathbf{i}) to $(\mathbf{i} - \mathbf{1})$ and/or (\mathbf{j}) to $(\mathbf{j} - \mathbf{1})$.
- The architecture makes it easy to compare empirical results with the corresponding theoretical results.

E.3 Results and Displays

The model includes these visual ways to understand its performance:

- Scopes labeled "Waiting Time" showing the empirical values of the waiting time in the queue, on a single set of axes.
- A scope labeled "Server Utilization" (\mathbf{su}) showing the utilization of the single server over the course of the simulation, where $\mathbf{su} = \frac{\boldsymbol{\lambda}}{\boldsymbol{\mu}}$, $\boldsymbol{\lambda}$ is the arrival rate and $\boldsymbol{\mu}$ is the service rate.

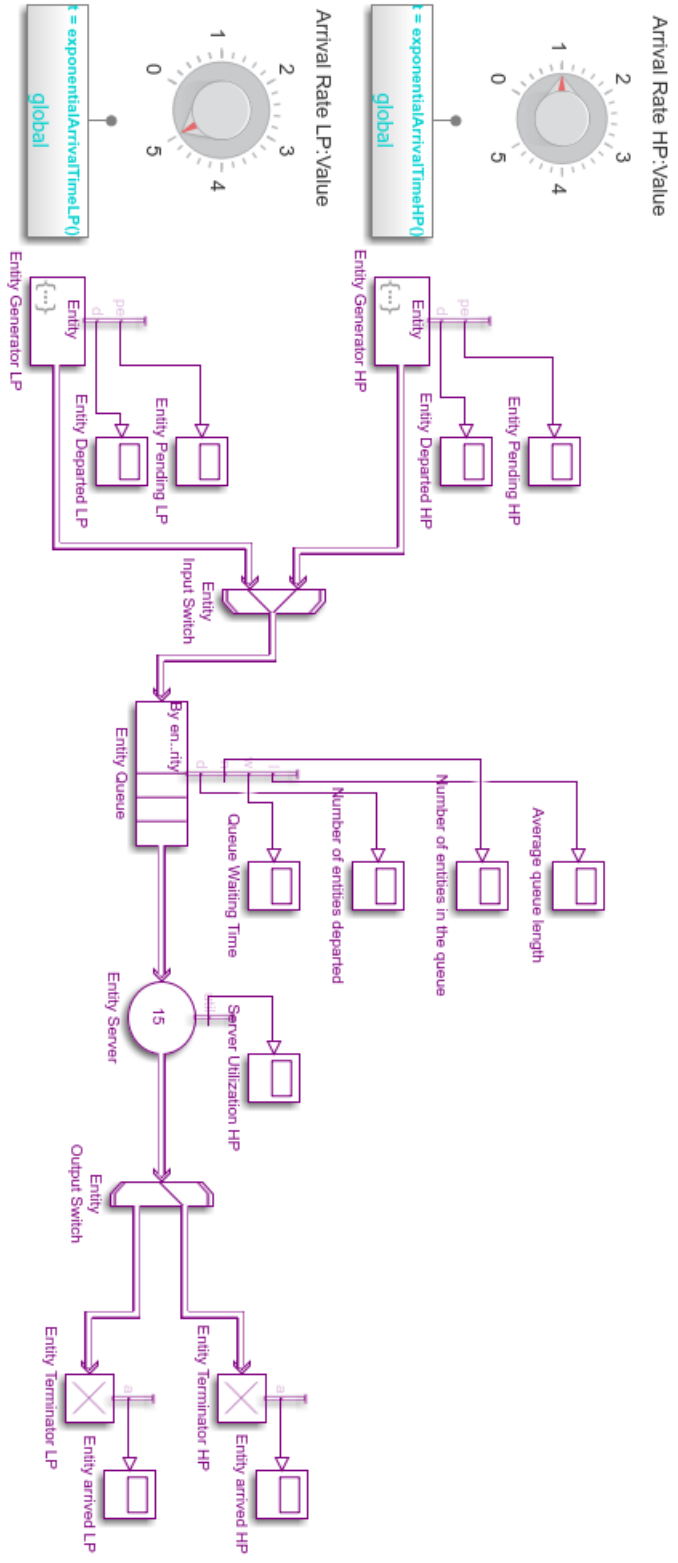


Figure E-1: M/M/1 Queuing System