

10-22-2018

Spectrum Sharing, Latency, and Security in 5G Networks with Application to IoT and Smart Grid

Imtiaz Parvez

Florida International University, iparv001@fiu.edu

Follow this and additional works at: <https://digitalcommons.fiu.edu/etd>

 Part of the [Electrical and Electronics Commons](#), [Power and Energy Commons](#), and the [Systems and Communications Commons](#)

Recommended Citation

Parvez, Imtiaz, "Spectrum Sharing, Latency, and Security in 5G Networks with Application to IoT and Smart Grid" (2018). *FIU Electronic Theses and Dissertations*. 3879.

<https://digitalcommons.fiu.edu/etd/3879>

This work is brought to you for free and open access by the University Graduate School at FIU Digital Commons. It has been accepted for inclusion in FIU Electronic Theses and Dissertations by an authorized administrator of FIU Digital Commons. For more information, please contact dcc@fiu.edu.

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

SPECTRUM SHARING, LATENCY, AND SECURITY IN 5G NETWORKS WITH
APPLICATION TO IOT AND SMART GRID

A dissertation submitted in partial fulfillment of the

requirements for the degree of

DOCTOR OF PHILOSOPHY

in

ELECTRICAL ENGINEERING

by

Imtiaz Parvez

2018

To: Dean John L. Volakis
College of Engineering and Computing

This dissertation, written by Imtiaz Parvez, and entitled Spectrum Sharing, Latency, and Security in 5G Networks with Application to IoT and Smart Grid, having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this dissertation and recommend that it be approved.

Kemal Akkaya

Leonardo Bobadilla

Mehdi Bennis

Ismail Guvenc, Co-Major Professor

Arif I. Sarwat, Co-Major Professor

Date of Defense: October 22, 2018

The dissertation of Imtiaz Parvez is approved.

Dean John L. Volakis
College of Engineering and Computing

Andres G. Gil
Vice President for Research and Economic Development
and Dean of the University Graduate School

Florida International University, 2018

© Copyright 2018 by Imtiaz Parvez

All rights reserved.

DEDICATION

I would like to dedicate this dissertation to my beloved parents, Md. Yousuf Hoassain and Papia Sultana, to my lovely wife Tanvira Nargish, and to my sister, Rumi, and my brother Raju. The completion of this work would not have been possible without their encouragement, support, understanding, patient, and most of all love.

ACKNOWLEDGMENTS

At first, I would like to express my sincere appreciation and gratitude to my major advisor, Arif I. Sarwat, for his support, outstanding guidance, and endless encouragement throughout my doctoral study. I would also like to thank my co-major advisor Dr. Ismail Guvenc, for his sincere and insightful direction, support and encouragement in my research.

I would also like to show my heartfelt gratitude to the members of my committee for their support, insightful comments, and constructive suggestions, in particular, Dr. Kemal Akkaya, Dr. Leonardo Bobadilla and Dr. Mehdi Bennis.

I would like to acknowledge the Presidential Fellowship (Fall 2013- Summer 2016) and Dissertation Year Fellowship (Spring 2017- Summer 2018) from FIU graduate school to support my PhD study. I also acknowledge the partial research support provided from the National Science Foundation (NSF) throughout the years of my doctoral research. Next, I would like to thank my colleagues in Energy, Power and Sustainability (EPS) Group, FIU, and Mobile, Pervasive, and Autonomous Communication Technologies (MPACT) Lab, NCSU for creating a wonderfully collaborative work environment.

I would also like to thank Florida Power and Light (FPL) for providing access to their lab facilities for validating some of my approaches.

Finally, I am thankful to the chair, faculty, and staffs of Electrical and Computer Engineering department at FIU, for their great commitment to student learning and services.

ABSTRACT OF THE DISSERTATION
SPECTRUM SHARING, LATENCY, AND SECURITY IN 5G NETWORKS WITH
APPLICATION TO IOT AND SMART GRID

by

Imtiaz Parvez

Florida International University, 2018

Miami, Florida

Professor Arif I. Sarwat, Co-Major Professor

Professor Ismail Guvenc, Co-Major Professor

The surge of mobile devices, such as smart phones, and tables, demands additional capacity. On the other hand, Internet-of-Things (IoT) and smart grid, which connects numerous sensors, devices, and machines require ubiquitous connectivity and data security. Additionally, some use cases, such as automated manufacturing process, automated transportation, and smart grid, require latency as low as 1 ms, and reliability as high as 99.99%. To enhance throughput and support massive connectivity, sharing of the unlicensed spectrum (3.5 GHz, 5GHz, and mmWave) is a potential solution. On the other hand, to address the latency, drastic changes in the network architecture is required. The fifth generation (5G) cellular networks will embrace the spectrum sharing and network architecture modifications to address the throughput enhancement, massive connectivity, and low latency.

To utilize the unlicensed spectrum, we propose a fixed duty cycle based coexistence of LTE and WiFi, in which the duty cycle of LTE transmission can be adjusted based on the amount of data. In the second approach, a multi-arm bandit learning based coexistence of LTE and WiFi has been developed. The duty cycle of transmission and down link power are adapted through the exploration and exploitation. This approach improves the aggregated capacity by 33%, along with cell edge and energy efficiency enhancement.

We also investigate the performance of LTE and ZigBee coexistence using smart grid as a scenario.

In case of low latency, we summarize the existing works into three domains in the context of 5G networks: core, radio and caching networks. Along with this, fundamental constraints for achieving low latency are identified followed by a general overview of exemplary 5G networks. Besides that, a loop free, low latency and local-decision based routing protocol is derived in the context of smart grid. This approach ensures low latency and reliable data communication for stationary devices.

To address data security in wireless communication, we introduce a geo-location based data encryption, along with node authentication by k-nearest neighbor algorithm. In the second approach, node authentication by the support vector machine, along with public-private key management, is proposed. Both approaches ensure data security without increasing the packet overhead compared to the existing approaches.

TABLE OF CONTENTS

CHAPTER	PAGE
1. Introduction	1
1.1 Motivation and Purpose of Research	4
1.1.1 Spectrum Sharing	5
1.1.2 Latency	7
1.1.3 Security	8
1.2 Dissertation Contributions	9
1.3 Organization of Dissertation	13
2. Literature Review	14
2.1 Spectrum Sharing	14
2.1.1 Coexistence between LTE and WiFi	14
2.1.2 Coexistence between LTE and ZigBee	18
2.2 Latency	19
2.3 Data Security in Smart Grid	22
3. LTE and ZigBee/WiFi Coexistence	25
3.1 Overview of LTE Systems	25
3.2 Overview of ZigBee Systems	27
3.3 Overview of WiFi Systems	29
3.4 Overview of Metering Infrastructure of Smartgrid	30
3.5 LTE-ZigBee Coexistence	32
3.5.1 System Model for LTE-ZigBee Coexistence	32
3.5.2 Simulation Results	35
3.6 LTE and WiFi Coexistence	40
3.6.1 System Model for LTE-WiFi Coexistence	40
3.6.2 Simulation Results	43
3.7 Conclusion	47
4. Coexistence of LTE-U and WiFi: A Multi-Armed Bandit Approach	49
4.1 System Model and Problem Formulation	51
4.1.1 Interference on DL and UL Transmissions	51
4.1.2 Duty Cycle of LTE-U	52
4.1.3 Capacity Calculation and Power Control	52
4.2 MAB Techniques for LTE-U WiFi Coexistence	56
4.3 Simulation Results	59
4.3.1 Aggregate Capacity with MAB	62
4.3.2 Cell Edge Performance	68
4.3.3 Energy Efficiency Performance	68
4.4 Conclusion	71

5. Low Latency Towards 5G: RAN, Core Network and Caching Solutions	72
5.1 Low Latency Services in 5G	74
5.2 Sources of Latency in a Cellular Network	79
5.3 Constraints and Approaches for Achieving Low Latency	83
5.4 RAN Solutions for Low Latency	85
5.4.1 Frame/packet structure	90
5.4.2 Advanced Multiple Access Techniques/Waveform	94
5.4.3 Modulation and Channel Coding	98
5.4.4 Transmitter Adaptation	100
5.4.5 Control Signaling	102
5.4.6 Symbol Detection	106
5.4.7 mmWave Communications	108
5.4.8 Location-Aware Communications for 5G Networks	110
5.4.9 QoS/QoE Differentiation	112
5.4.10 CRAN and Other Aspects	114
5.5 Core Network Solutions for low latency	119
5.5.1 Core Network Entities	119
5.5.2 Backhaul Solutions	125
5.6 Caching Solutions for Low Latency	129
5.6.1 Caching for cellular network	129
5.6.2 Fundamental Latency-storage trade-off in Caching	131
5.6.3 Existing Caching Solutions for 5G	136
5.7 Field Tests, Trials and Experiments	139
5.8 Open Issues, Challenges and Future Research Directions	143
5.8.1 RAN Issues	143
5.8.2 Core Network Issues	145
5.8.3 Caching Issues	147
5.9 Conclusion	148
6. RSS based Loop-free and Low Latency Compass Routing Protocol	150
6.1 Localization and Routing Techniques	150
6.1.1 Localization of Smart Meter	150
6.1.2 Compass Routing	153
6.2 Simulation Result	155
6.3 Conclusion	158
7. Key Management and Authentication based Data Security for Smart Grid	159
7.1 Architecture of the Proposed Secured AMI	160
7.1.1 Basic Notifications and Definitions	162
7.2 RSS Algorithm, OCSVM and Entropy of a Data Packet	163
7.2.1 Received Signal Strength (RSS) based localization	165
7.2.2 OCSVM algorithm	166
7.2.3 Entropy of a data packet	168

7.3	Privacy Scheme implementation and Data Traffic Flow	169
7.3.1	Attack Model	169
7.3.2	Data Traffic Flow	170
7.4	Simulation Results and Performance Analysis	174
7.4.1	Performance of Proposed Algorithm	174
7.4.2	Security Strength Analysis	181
7.5	Conclusion	182
8.	A Machine Learning and Geo-location based Data Security Scheme for Smart Grid	183
8.1	Challenges Faced by AMI Meters	184
8.2	Architecture of AMI	186
8.3	Localization Algorithm and Node Authentication Technique	188
8.3.1	Localization of Smart Meter by RSS method	188
8.3.2	Meter Authentication by kNN Algorithm	190
8.4	Encryption Process	192
8.5	Simulation Results	195
8.6	Security Strength Analysis	201
8.7	Conclusion	202
9.	Conclusion and Future Works	203
9.1	Concluding Remarks	203
9.2	Future Works	205
	BIBLIOGRAPHY	207
	VITA	243

LIST OF TABLES

TABLE	PAGE
3.1 IEEE 802.15.4 FREQUENCY NOMENCLATURE.	28
3.2 LTE MAC/PHY PARAMETERS.	35
3.3 ZIGBEE MAC/PHY PARAMETERS.	36
3.4 LTE MAC/PHY PARAMETERS.	43
3.5 WiFi MAC/PHY PARAMETERS.	44
4.1 LTE MAC/PHY PARAMETERS.	61
4.2 WiFi MAC/PHY PARAMETERS.	61
5.1 TYPICAL LATENCY AND DATA RATE REQUIREMENTS FOR DIFFERENT MISSION CRITICAL SERVICES.	76
5.1 TYPICAL LATENCY AND DATA RATE REQUIREMENTS FOR DIFFERENT MISSION CRITICAL SERVICES.	77
5.2 OVERVIEW OF TECHNIQUES IN RAN FOR LOW LATENCY.	86
5.2 OVERVIEW OF TECHNIQUES IN RAN FOR LOW LATENCY (CONTINUED).	87
5.2 OVERVIEW OF TECHNIQUES IN RAN FOR LOW LATENCY (CONTINUED).	88
5.3 PHY AND MAC LAYER BASED RADIO INTERFACE SOLUTIONS FOR LOW LATENCY.	93
5.4 PROPOSED MULTIPLE ACCESS TECHNIQUES FOR 5G.	95
5.5 WAVEFORM CONTENDER FOR 5G.	97
5.6 COMPARISON AMONG CHANNEL CODING SCHEMES FOR LOW LATENCY [1].	98
5.7 OVERVIEW OF SOLUTIONS IN TRANSMITTER ADAPTATION FOR LOW LATENCY.	101
5.8 OVERVIEW OF SOLUTIONS IN CONTROL SIGNALING FOR LOW LATENCY.	103
5.8 OVERVIEW OF SOLUTIONS IN CONTROL SIGNALING FOR LOW LATENCY (CONTINUED).	104
5.9 OVERVIEW OF SOLUTIONS IN SYMBOL DETECTION FOR LOW LATENCY.	107
5.10 OVERVIEW OF SOLUTIONS IN MMWAVE COMMUNICATIONS FOR LOW LATENCY.	109

5.11	LOCATION-AWARE COMMUNICATIONS FOR LOW LATENCY.	110
5.12	LITERATURE OVERVIEW RELATED TO QoS/QoE DIFFERENTIATION. . .	112
5.13	OVERVIEW OF TECHNIQUES IN CORE NETWORK FOR LOW LATENCY. . .	120
5.14	PROPOSED CORE NETWORK ENTITIES FOR 5G VISIONS.	124
5.15	OVERVIEW OF LITERATURE IN BACKHAUL SOLUTIONS TO ACHIEVE LOW LATENCY.	126
5.16	OVERVIEW OF LITERATURE IN CACHING.	130
5.17	DIFFERENT TYPES OF CACHING SCHEMES FOR THE CELLULAR NET- WORK.	132
5.18	DEFINITION OF METRICS USED FOR LATENCY EVALUATION IN CACHING SCHEMES.	132
5.19	SUMMARY OF THE WORKS ON LATENCY-STORAGE TRADE-OFF IN THE CACHING.	133
5.20	PLATFORM AND PERFORMANCE ON FIELD TESTS, TRIALS AND EXPER- IMENTS.	140
7.1	SYMBOL NOTATION.	162
7.2	MEAN SQUARE ERROR (MSE) IN LOCALIZATION FOR DIFFERENT NOISE VARIANCES	175
7.3	SMART METER REAL WORLD DATA SAMPLE	178

LIST OF FIGURES

FIGURE	PAGE
1.1 An exemplary 5G network supporting enhanced mobile broadband, massive machine type communication, and ultra reliable low latency communication.	2
1.2 IoT supporting numerous applications.	3
1.3 Smart grid connecting a variety of nodes through information communication technology.	4
1.4 Latency specification at the different generations of cellular network.	7
3.1 10 ms radio frame of LTE [2].	26
3.2 A typical example of superframe of ZigBee.	28
3.3 AMI communication architecture encompassing Home Area Network (HAN), Neighbor Area Network (NAN), data concentrator, and control center. . .	31
3.4 (a) Mean square error for different number of nodes and noise variances. (b) Mean square error for different number of nodes and path loss exponent.	33
3.5 LTE TDD subframe structure for two different configurations.	34
3.6 LTE DL capacity without ZigBee, and with ZigBee transmission at different traffic rates.	37
3.7 ZigBee capacity without LTE, and with LTE DL transmission at different traffic rates.	37
3.8 ZigBee capacity with all LTE DL or all LTE UL traffic in a 10 ms radio frame at different α , and at $\lambda_L = 1.5$	38
3.9 LTE capacity with all LTE DL or all LTE UL traffic in a 10 ms radio frame at different α , and at $\lambda_L = 1.5$	38
3.10 SINR distribution of LTE without ZigBee transmission, and with DL ZigBee transmission.	39
3.11 SINR distribution of ZigBee without LTE transmission, and with LTE DL transmission.	39
3.12 Cell layout for AMI of smart grid.	40

3.13	Protocol mapping between different entities of WiFi and LTE network. (AP: access point; EPC= Evolved packet core; APP= application; UDP= user datagram protocol; TCP= transmission control protocol; LLC= logic link control; IP= Internet protocol; PHY= physical; MAC= medium access control; GTP= GPRS tunneling protocol; RLC= radio link control; PDPC= packet Data Convergence Protocol)	41
3.14	Throughput performance of coexisted LTE/WiFi system at 60% duty cycle.	45
3.15	Throughput performance of coexisted LTE/WiFi system at 80% duty cycle.	45
3.16	SINR distribution of coexisted LTE/WiFi system at 60% duty cycle.	46
3.17	SINR distribution of coexisted LTE/WiFi system at 80% duty cycle.	47
4.1	DL and UL interference scenarios for LTE-U/WiFi transmissions.	53
4.2	Structure of the duty cycle for LTE-U transmissions.	54
4.3	Cellular coverage layout used in LTE-U and WiFi coexistence simulations.	60
4.4	LTE-U and WiFi UL/DL capacities.	62
4.5	Comparison of different UE and STA configurations.	63
4.6	Aggregate capacity of coexisting WiFi and LTE-U (80% duty cycle), MAB based coexisting LTE-U and WiFi, and stand-alone WiFi system for different ISDs.	64
4.7	WiFi capacity of coexisting WiFi and LTE-U (80% duty cycle), MAB based coexisting LTE-U and WiFi, and stand-alone WiFi system for different ISDs.	65
4.8	Aggregate capacity of coexisting WiFi and LTE-U (80% duty cycle), MAB based coexisting LTE-U and WiFi, and stand-alone WiFi system for different LTE traffic arrival rates.	65
4.9	Aggregate capacity of coexisting system of WiFi and LTE-U (80% duty cycle), MAB based coexisting LTE-U and WiFi, and stand-alone WiFi system for various energy detection threshold.	66
4.10	Scenario with two cases.	66
4.11	Capacity of 10 STAs or/and 10 UEs under stand-alone WiFi, stand-alone LTE, coexisting stand-alone WiFi and LTE-U (scenario 1) and MAB based coexisting LTE-U and WiFi (scenario 2) for different bandwidths and ISDs.	67
4.12	Capacity of MAB based coexistence for different UEs and STAs ratios and ISDs.	68

4.13	5th percentile throughput of MAB based coexisting LTE-U and WiFi for different UEs and STAs ratios.	69
4.14	Capacity of 10 UEs and 10 STAs under different PC configurations.	69
4.15	Energy efficiency under different PC configurations for various number of UEs (with 10 STAs).	70
5.1	Latency contribution in E2E delay of a packet transmission.	80
5.2	Categories of different solutions for achieving low latency in 5G.	84
5.3	Physical air interface (a) Conventional LTE radio frame, (b) Exemplary 5G radio frame with flexible time and frequency division for low latency [3] (GB: guard band; LLC: low latency communication).	89
5.4	Slot placment in GFDM, OFDM and SCFDM.	94
5.5	Transmission and symbol detection in cellular network.	106
5.6	Cloud-RAN architecture in 5G networks.	115
5.7	Simplified example architectures for core network (a) Architecture of SDN [4], and (b) Architecture of NFV [5] (APP: application; ODL: open-daylight platform; ONOS: open network operating system; DFE: dyna forming engineering; OSS: operations support systems; BSS: base station subsystem; EMS: element management system; NFVO: NFV orchestrator; VNFM: virtual network function manager; VIM: virtualized infrastructure manager).	117
5.8	Different types of caching in 5G.	131
6.1	Locating the new meter by known position meter (Anchor).	151
6.2	Routing diagram.	154
6.3	Loop in routing path.	154
6.4	The error from exact position for different number of nodes.	155
6.5	The relation between error and variance.	156
6.6	Degree of connectivity versus hop count.	157
6.7	Reliability of nodes (meters) at different packet size.	157
7.1	AMI architecture consisting of a cluster of mesh/hybrid connected meters, data concentrator, and control center.	160
7.2	Training and application of OCSVM model in smart meters	164

7.3	Data flow among various components of AMI.	171
7.4	Meter positions in Manhattan grid.	175
7.5	(a) Mean square error for different number of nodes and noise variances. (b) Mean square error for different number of nodes and path loss exponent.	176
7.6	Anomaly Detection with a) One Class SVM and b) Isolation Forest Algorithms	179
7.7	Confusion Matrix for a) One Class SVM, and b) Isolation Forest	180
7.8	ROC curve for OCSVM with different parameter settings. ($\text{Nu}=\nu$; $\text{Gamma}=\gamma$)	181
8.1	Proposed AMI architecture consists of HAN, NAN, Trusted Third Party (TTP), Data concentrator and Control center.	186
8.2	Localization of new meter (unknown positioned meter).	189
8.3	Encryption and data transmission process among various components of AMI.	191
8.4	Mapping of encryption key based on coordinate point.	194
8.5	Different shapes of AOI.	196
8.6	Number of nodes vs Mean Square Error.	196
8.7	Path loss vs Mean Square Error.	197
8.8	Variance vs Mean Square Error.	197
8.9	Performance of kNN algorithm for data set of size 100. a) Comparison of data value and their prediction by kNN algorithm. (b) Histogram of error distribution.	199
8.10	Performance of kNN algorithm for data set of size 200. a) Comparison of data value and their prediction by kNN algorithm. (b) Histogram of error distribution.	200
8.11	The relation of error between number of neighbors and data size.	201

LIST OF ACRONYMS

AS	Access stratum
AR	Augmented reality
BC	Broadcast channel
BLER	Block error rate
CCP	Communication control port
CCSE	Control channel sparse encoding
CSIT	Channel state information at the transmitter
DTB	Delivery time per bit
D2D	Device to device
DRB	Data radio bearer
eMBB	Enhanced mobile broadband
EPC	Evolved packet core
FFT	Fast Fourier transform
FDT	Fractional delivery time
FBMC	Filter bank multicarrier
GFDM	Generalized frequency division multiplexing
GP	Guard period
GTP	GPRS tunnel protocol
GGSN	Gateway GPRS (General Packet Radio Service) service node
HARQ	Hybrid Automatic-Repeat-Request
ICI	Inter carrier interference
IoT	Internet of Things
ITS	Intelligent transportation system
IFFT	Inverse first Fourier transform

IUT	International telecommunications union
ISI	Inter symbol interference
MEC	Mobile Edge Computing
MTP	Machine type communication
NGMN	Next generation mobile networks
MCC	Mission critical communication
MRC-ZF	Maximum ration combining zero forcing
MAC	Medium access control
mMTC	Massive machine type communication
MME	Mobility management entity
MISO	Multiple input single output
MDP	Markov decision process
NFV	Network virtualized function
NAS	Non access stratum
NDT	Normalized delivery time
OFDM	Orthogonal frequency division multiplexing
OOB	Out of bound
OW	Optical window
OLLA	Outer loop link adaption
PUCCH	Physical uplink control channel
P2P	Peer to peer
RAN	Radio access network
SC-FDMA	Single carrier frequency division multiple access
SRB	Signaling radio bearer
SCMA	Sparse code multiple access
SGW	Serving GPRS gateway

TDD	Time division duplex
RAT	Radio access technology
uRLLC	Ultra reliable low latency communication
UFMC	Universal filtered multi-carrier
UDN	Ultra dense network
VR	Virtual reality
VLC	Visible light communication
ZF	Zero forcing
5G	Fifth generation mobile network
5GETLA	5G flexible TDD based local area

CHAPTER 1

Introduction

The fifth generation (5G) cellular network is the communication solution to meet the wireless communication demands of beyond 2020. The goals of 5G networks will be to achieve 1000 times throughput, 100 billion connections and close to zero delays compared to current 4G LTE networks [6]. In particular, 5G will support latency/reliability-sensitive services and massive machine type devices, along with regular personal mobile communication supporting large throughput. Therefore, the requirements of 5G networks can be categorized into three broad factions: (1) Enhanced Mobile Broadband (eMBB), (2) massive Machine Type Communication (mMTC), and (3) Ultra Reliable Low Latency Communication (uRLLC). In case of eMBB, the peak data rate is 10-20 Gbps with spatial uniform data rates of 100 Mbps at the end users. In mMTC communication, a large number of smart devices, including utilities, manufacturing, health care, transportation, and consumer goods, need a low data rate uplink-oriented connection for a sporadic small packet transmission. A single cell may need to support up to 30,000 devices [7, 8]. In the case of uRLLC, not only the latency but also the reliability aspect needs to be addressed. More specifically, end-to-end latency of 1 ms, along with reliability on the order of 99.99%, is to be fulfilled. Moreover, the 5G will be an evolution of LTE complementing these features. An exemplary 5G network is illustrated in Fig. 1.1.

The 5G will play a critical role in the modern economy, affecting consumer, transportation, health, power system, education, logistics, and other major industries. The focus of 5G communication is to provide not only high-speed mobile Internet, but also the ubiquitous connectivity for numerous sensors, devices, and machines forming the Internet-of-Things (IoT). The IoT (as illustrated in Fig. 1.2), which connects anything to any other things anytime and anywhere, introduces many new use cases, such as smart transportation, virtual reality, tactile Internet and real-time control. These services put a

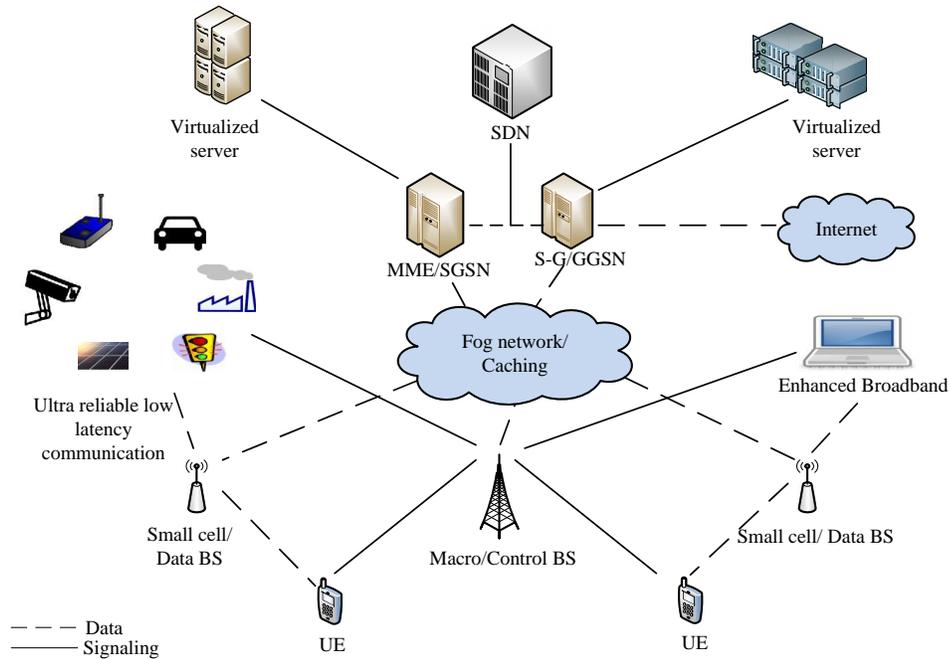


Figure 1.1: An exemplary 5G network supporting enhanced mobile broadband, massive machine type communication, and ultra reliable low latency communication.

new prerequisite of throughput, reliability, latency, and robustness on the network. Additionally, services are foreseen to have intermittent, as well as always-on hyper connections, in machine-type communications (MTC) supporting various applications such as connected vehicles, smart homes, smart meters, moving robots, and sensors working in an efficient and scalable approach. Besides, various emerging technologies, such as augmented reality, wearable devices, and full immersive experience (3D) are affecting the life styles of human end users and placing new requirements in the next generation networks. These use cases of 5G drive the 5G specifications on different dimensions, such as throughput, latency, network and device energy efficiency, reliability, connection density, mobility and traffic volume density. Current 4G telecommunication is not capable of fulfilling the technical requirements for these services.

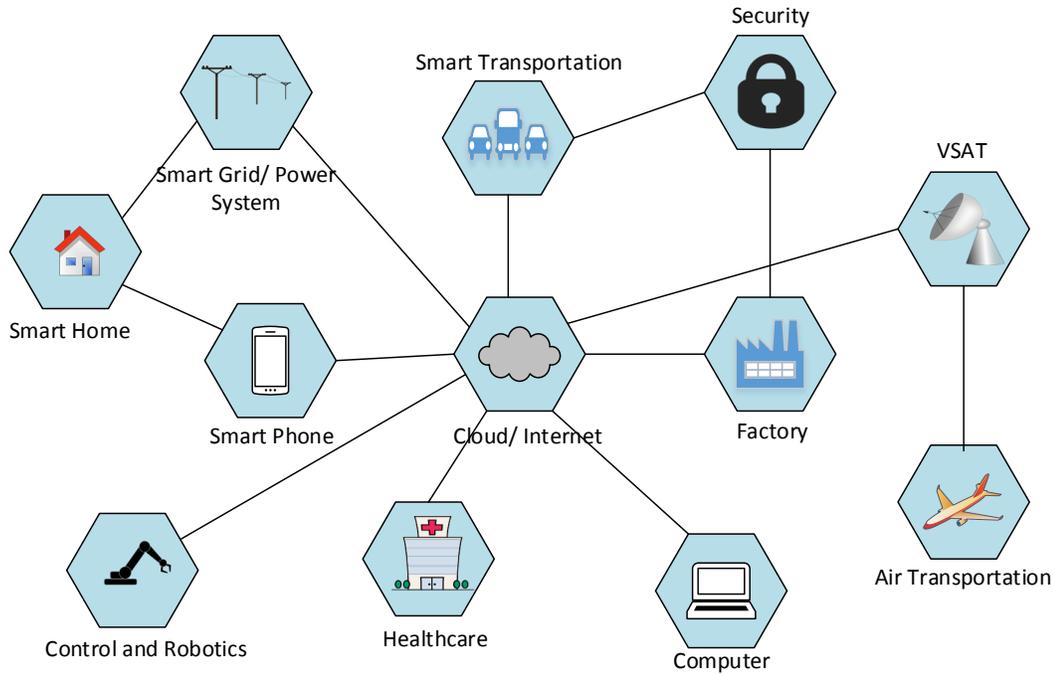


Figure 1.2: IoT supporting numerous applications.

Within the scope of IoT, smart grid (as illustrated in Fig. 1.3) is a prominent example of mMTC and uRLLC. Smart grid is the modern power system that utilizes bidirectional communication among its numerous nodes of generation, transmission, and distribution systems. In particular, metering infrastructure of smart grid that consists of millions of smart meters uses wireless communication to send energy consumption data periodically to its control center. Moreover, dynamic activation and deactivation of power suppliers, such as photo-voltaic and windmills allow round-trip latency on the order of 100 ms. Besides that, synchronous co-phasing of power suppliers (e.g., generators) has strict end-to-end latency requirement on the order of 1 ms. Latency more than 1ms may have serious consequence on the grid and devices in this case.

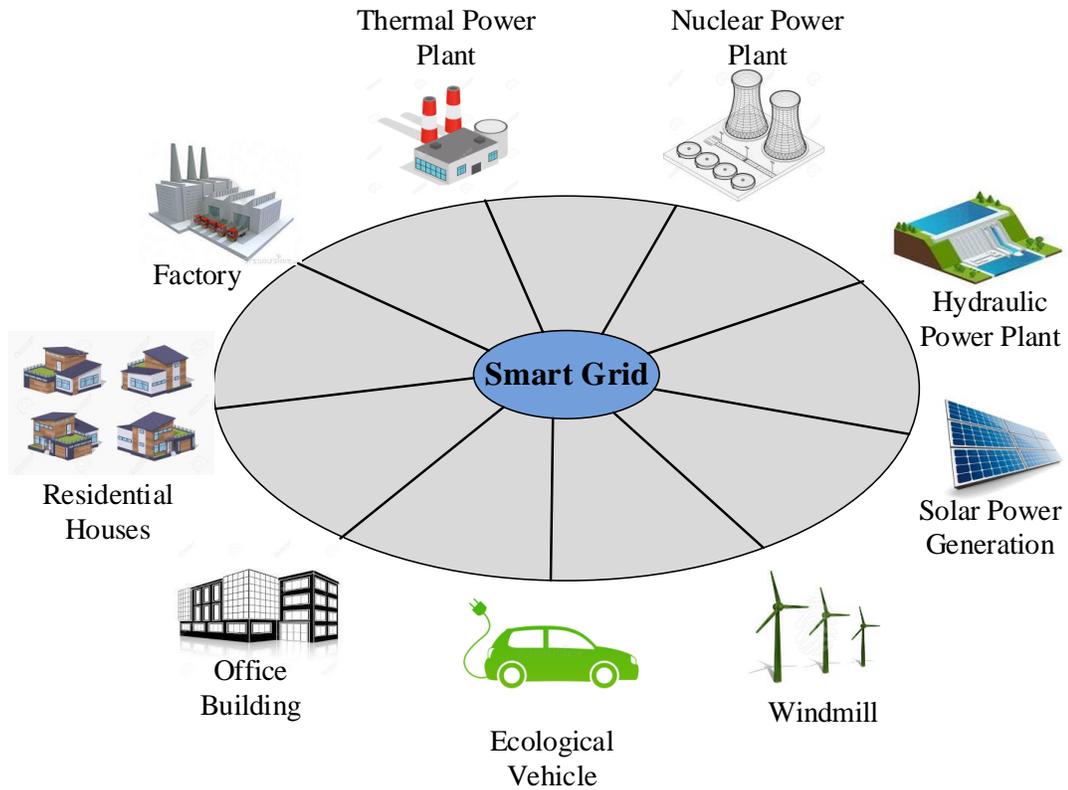


Figure 1.3: Smart grid connecting a variety of nodes through information communication technology.

1.1 Motivation and Purpose of Research

To enhance throughput and support massive connectivity in the 5G and its legacy LTE networks, spectrum sharing is a powerful tool. For spectrum sharing, fair and effective existence techniques are required. On the other hand, to support latency-critical services, latency reduction approaches need to be developed. For this, drastic changes in the network architecture, along with novel techniques, is to be conceived. Additionally, data security in the wireless communication is a prime concern. In the following subsections, we discuss the motivation of research on spectrum sharing and latency in detail.

1.1.1 Spectrum Sharing

The surge of mobile devices, such as smart-phones and tablets, along with diverse IoT applications, requires additional capacity. In some surveys, it is found that the exponentially increasing data is doubled approximately every year [9]. Moreover, the next generation networks need to support 30,000 devices per cell. The 5G and its legacy LTE networks is a powerful tool to fulfill this high data rate and massive connectivity complementing with other technologies such as WiFi and ZigBee. However, to support the massive connectivity and high data rates, larger bandwidth is required [10, 11]. But the licensed spectrum is limited, and expensive. Recently released 3.5 GHz, along with 5 GHz and mmWave, opens a new window of expanding cellular network operations in these unlicensed bands. These bands have a large amount of clean and free spectrum. However, cellular networks need to share these spectrum with others cellular operators, as well as variant technologies, such as WiFi and ZigBee. In this regard, we need effective coexistence techniques for fair and good neighborhood spectrum sharing.

In this dissertation, the following areas of spectrum sharing will be discussed and analyzed in detail.

1. Coexistence of LTE and ZigBee
2. Coexistence of LTE and WiFi

Coexistence of LTE and ZigBee

To expand the operation of LTE and foresee 5G in the unlicensed bands, including 868 MHz, 915 MHz, 2.4 GHz, 3.5 GHz and 5 GHz, it needs to coexist harmoniously with other technologies, such as ZigBee, WiFi, and Bluetooth. ZigBee, WiFi, and Bluetooth are already operating in most of these unlicensed bands. Therefore, the effect of one's operation on others needs to be investigated before developing a fair and effective coexistence technique. Simultaneous operation of WiFi, IEEE 802.15.4, and Bluetooth has been

investigated in [12, 13]. In [14], coexisting operation of UWB, WiFi and ZigBee is studied. The coexistence of LTE and WiFi is studied in [15]. However, coexisting operation of LTE and ZigBee in the same band has not been addressed.

The metering infrastructure of the smart grid, commonly named advance metering infrastructure (AMI), is a remarkable example of IoT. In AMI, millions of meters use the ZigBee network for meter-to-meter communication due to the self-organizing nature, capable of supporting numerous devices, and unlicensed band operation. On the other hand, the data collector or access point of AMI uses LTE networks to send collected data to a long distant control center [16]. Therefore, the study of the coexistence of LTE and ZigBee in a smart grid scenario carries a critical importance.

The purpose of this part of the research is to get insight into the interference effect on physical (PHY) and medium access control (MAC) layer operating regimes. Also, it will lead to developing effective and harmonious coexistence techniques.

Coexistence between LTE and WiFi

WiFi is the short range broad band communication operating in the unlicensed bands. The touted feature of WiFi is low cost, high data rate, easy set up, and short end-to-end delay. To have LTE coexist with WiFi in the same band, the main obstacle is the dissimilarity of two technologies. WiFi uses collision sensed multiple access collision avoidance (CSMA/CA) for channel access with a four-way handshaking [15]. On the other hand, LTE utilizes centralized scheduling for its users. However, it does not use the channel sensing mechanism before transmitting its packets. Therefore, the operation of LTE will put continuous interference on WiFi, which may block WiFi transmission in the typical case.

The purpose of this part of the research is to develop holistic adaptive techniques for fair and harmonious coexistence between 5G/LTE¹ and WiFi. In particular, the adaptive approach will utilize underused spectrum without hampering the regular operation of WiFi. Therefore, the approach will not only maintain the good neighborhood spectrum sharing, but also improve the aggregate capacity, cell edge performance, and energy efficiency. We also identify various use cases where LTE/WiFi coexistence can be applicable.

1.1.2 Latency

Latency is highly critical in several use cases, such as manufacturing industry, automated transportation, power system, health care, augmented/virtual reality, robotics and telepresence, entertainment, culture and education. In some applications, latency on the order of 1ms along with reliability of 99.99% is mandatory. In the current LTE systems, typical latency is 100ms. Therefore, we need drastic changes in the cellular network to achieve this low latency. The evolution of latency specification at the various generation of cellular networks is illustrated in Fig. 1.4.

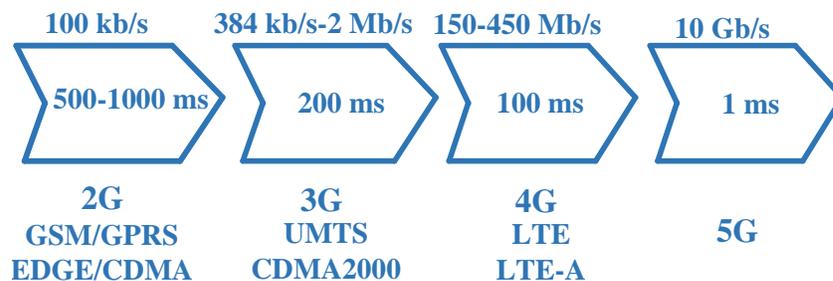


Figure 1.4: Latency specification at the different generations of cellular network.

To achieve low latency and high reliability, radical changes in the physical (PHY) and medium access control (MAC) layers is required, along with the upper layers. In particular, enhancements in the radio and core networks, including shorter transmission time

¹Since 5G is not fully standardized yet and it will be an evolution of LTE, coexistence technique of LTE and WiFi will be applicable to 5G and WiFi coexistence.

interval, small cell deployment, larger bandwidth utilization, new coding scheme, massive MIMO and introduction of new technologies, such as SDN, NFV and caching are the several prominent proposed modifications in the 5G networks. SDN and NFV enables dynamic management, and scaling of services and functions, where non access stratum and access stratum are coupled by the intelligent protocol. For separate provisioning of coverage and capacity, data plane and control could be split in the carrier frequency along with network architecture. Special network functionality as a service (XaaS) will cater on demand services, such as instantaneous resource pooling. Additionally, the introduction of information-centric networks, such as caching and fog networks/mobile edge computing, can reduce end-to-end latency and network operation costs significantly, along with throughput enhancement. The achievement of larger data rates and lower latency also leads to higher energy efficiency and longer device battery life.

The purpose of this research is to summarize the existing works to achieve low latency in the context of 5G networks. Along with this, fundamental constraints for achieving low latency are to be identified followed by a general overview of exemplary 5G network. Besides that, a loop free, low latency and local decision based routing protocol will be derived for fast and reliable data communication.

1.1.3 Security

Security is a prime concern in the wireless communication. In a typical scenario of metering infrastructure of the smart grid, ZigBee/WiFi is used for meter-to-meter communication and LTE is used for sending collected data to the control center. In LTE/LTE-A, there are five levels of security layers- (1) network access security, (2) network domain security, (3) user domain security, (4) application domain security, and (5) non 3GPP domain security [17]. Therefore, WiFi and ZigBee are not as much secured as LTE. Observing

the energy usage patterns of meters, an adversary/thief can predict the presence of the targeted consumers at home, which can be a threat to civil lives and privacy. Furthermore, from the fine-grained energy consumption data, the home appliance companies receive information about the life style patterns of consumers and the energy utilization of their home appliances. Thus competing companies can use this valuable information in their businesses. The consumers might want to alter the consumption data to reduce their electricity bill. The most crucial thing is that the opponent/hacker might jam or take over the AMI network by sending a false signal to meters on an unsecured system, which may cause a power outage in a wide area, as well as an imbalance in the demand generation model.

The purpose of this research is to develop robust security schemes for metering infrastructure of the smart grid. It will include not only packet encryption, but also node-to-node authentication. In this way, end-to-end data security will be ensured. Though, our approaches will use metering infrastructure as a use case, it can be implementable to any IoT applications.

1.2 Dissertation Contributions

The focus of this research is two important requirements for the next generation 5G networks: spectrum sharing and latency. Along with this, we address the security aspect of wireless communication with application to smart grid and IoT. It is to be noted that some of our approaches were validated using real world data. The research to be presented in this dissertation resulted in several publications in IEEE and other reputed conferences and journals [16, 18–32]. The key contributions of this dissertation can be summarized as follows.

1. Spectrum sharing:

- The coexistence of LTE and ZigBee is studied in the 2.4 GHz band under a smart grid scenario [19]. A time division duplexing (TDD)-LTE accompanied by ZigBee transmission is simulated for FTP traffic model, along with different combination of LTE down link (DL)/ uplink (UL) subframes. In the 10 ms subframe of LTE transmission, ZigBee transmits only during the guard band (blank subframe) time. The effect of LTE transmission on the throughput and signal-to-noise ratio (SINR) of WiFi and vice-versa is investigated. It helps to understand the interference effect on PHY/MAC operating regimes along with justification of using this coexistence system for metering infrastructure of smart grid.
- The performance of duty cycle based coexistence of LTE and WiFi is studied in the 3.5 band under a smart grid scenario [16]. A time division duplex (TDD)-LTE with various duty cycle, such as 60% and 80% duty cycle, is considered for the simultaneous operation with WiFi. In a 10ms transmission time interval, LTE transmits a fixed percentage of duty cycle leaving rest of the time for WiFi transmission. The throughput and signal to noise ratio (SINR) is studied for these variant duty cycles. The simulation results show that coexisting operation of LTE and WiFi can maintain harmonious and good neighborhood coexistence and therefore, can be a potential communication solution for smart grid.
- A multi-arm bandit (MAB) based adaptive duty cycle selection of LTE transmission is derived for coexisting with WiFi [21, 22]. MAB is a machine learning approach being used for maximizing reward over a time horizon. It is similar to gambling process where a gambler has a set of machines. Each time, the gambler pulls an arm and gets some rewards. The goal is to pull arms sequentially in such a way that the reward is maximized over the gambling time. However, there is a trade-off between

exploration and exploration i.e., to pull an arm for an immediate highest reward and long term accumulated maximum rewards. The adaptive down-link power control is also incorporated with the adaptive duty cycle selection. Finally aggregated capacity, cell edge performance, and energy efficiency are studied. The simulation results demonstrate that the aggregate capacity is improved by 33%, along with the enhancement of cell edge throughput for various user and base station densities.

2. Latency:

- The emerging technologies for achieving low latency is summarized covering three domains of 5G networks: (1) radio access network, (2) core network, and (3) caching [23]. Following this, a general overview of 5G cellular network composed of software defined network (SDN), network function virtualization (NFV), caching, and mobile edge computing (MEC) is presented, which will show that D2D, small cell access points, network cloud, and the Internet of Things can be a part of 5G cellular network architecture. The architecture can support latency critical services along with massive connectivity, and larger throughput. We also present promising results from the field tests, trials and experiments followed by open issues, challenges and future research directions.
- A compass algorithm based routing protocol is derived for a loop free, local decision, and low latency routing protocol [24]. For localization of the nodes, derivation of received signal strength by maximum likelihood and particle swarm optimization is proposed. The protocol is simulated under a smart grid scenario to investigate the latency, localization error, noise impact, scalability, and reliability. The simulation results show that it is loop free. Moreover, it ensures faster data routing along with less flooding, compared to conventional compass routing protocol.

3. Security:

- A geo-location based key management scheme is developed for data encryption with an application to smart grid [28, 30]. For localization of meters, received signal strength derivation using maximum likelihood and particle swarm optimization is proposed. Packets are encrypted with the keys associated with the geo-location and a random key index. For node-to-node authentication, k nearest neighbors (kNN) algorithm is utilized. Finally, security strength of the approach is analyzed using real world data of a local utility company. The results demonstrate that our approach ensures data security by implementing two layer security without increasing considerable packet overhead and complexity.
- Considering semi-trusted third party servers and untrustworthy/unreliable communication links, a two layer security schema is introduced with application to metering infrastructure of smart grid [29, 32]. The first layer implements an efficient novel encryption method for securing data exchange between meters and control center with the help of two partially trusted simple servers. One server is responsible for data encryption between the meter and control center/central database, and the other server manages the random sequence of data transmission. In the second layer, node-to-node authentication using one class support vector machine algorithm is proposed. The security strength of the approach is analyzed using the real world data, along with comparison to other schemes. This schema secures data communication, and imposes a comprehensive privacy throughout the system without considerably increasing the complexity of conventional key management scheme.

1.3 Organization of Dissertation

The remainder of this dissertation is organized as follows. Chapter II provides a literature review of the topics relevant to the issues discussed in the Introduction. Study of coexistence of LTE and ZigBee/WiFi has been presented in Chapter III. In Chapter IV, multi arm bandit for spectrum sharing of LTE and WiFi has been analyzed. Solutions for latency reduction have been presented in Chapter V. Compass algorithm based routing protocol has been presented in Chapter VI. In Chapters VII & VIII, machine learning based securing schemes for data communications in smart grid have been presented. Finally, Chapter IX lists the conclusion of this research work and also identifies some important directions for future research.

CHAPTER 2

Literature Review

In this chapter, we discuss about the state of art works on spectrum sharing, latency and smart grid data security.

2.1 Spectrum Sharing

The surge in smart phones, tablets, mobile APs, wearables and Internet-of-Things (IoT) has exponentially increased mobile data usage and wireless communication resulting in a huge explosion of traffic demand on the licensed frequency bands [18]. Similarly, the smart grid network has added more traffic to the existing channels which necessitate the utilization of more frequency bands to increase the capacity of the operators to fulfill the drastic demand of the traffic. Having large amounts of radio resources, the unlicensed spectrum has recently been treated as excellent supplementary frequency bands that augment the throughput of wireless communications [33]. LTE is generally divided into LTE-U and LTE-LAA when used in the unlicensed spectrum [34]. LTE-U was the early deployment, which has a simple mechanism and does not require alterations to existing LTE air interface protocols. It employs the LTE Release 10-12 aggregation protocol which does not require the Listen Before Talk (LBT) algorithm [21]. It is only applicable in the US, South Korea, India and China unlike the Europe and Japan where LBT is mandatory. LAA is ratified by the 3rd Generation Partnership (3GPP) as Release 13 which only aims on single global framework [35,36].

2.1.1 Coexistence between LTE and WiFi

Huge difference between LTE and Wi-Fi pose great challenges in the design of an effective coexistence mechanism [34]. The above factors should be carefully studied to design

a fair and efficient coexistence mechanism for LTE and Wi-Fi networks in unlicensed band.

Coexistence without LBT

In countries where there is no regulatory requirement for LBT, careful designed coexistence algorithm will guarantee a fair coexistence. Using the Release 10/11 LTE PHY/MAC standards, three mechanisms can be implemented to safeguard that LTEs coexistence in unlicensed band with Wi-Fi. Channel Selection permits small cells to choose the clearest channel based on Wi-Fi and LTE analysis [23]. If clear channel is found, LTE-U will occupy with full duty cycle for Secondary DL (SDL) transmission. If no clean channel is available, Carrier-Sensing-Adaptive Transmission (CSAT) is used to share the channel with Wi-Fi [17]. Depending on traffic demand, SDL carrier can be opportunistically retrieved. If there is low load, SDL carrier should be turned off and for higher load, channel selection should be executed again.

- **Channel Selection:**

In this mechanism, LTE-U small cells will scan the unlicensed band to search for the cleanest unused channels for the SDL carrier transmission. Given that there is an unused channel, the interference is avoided between the cells and its nearby Wi-Fi devices and other LTE-U small cells. Operating channel is monitored on an on-going basis by Channel Selection Algorithm. Measurements are usually completed at both the beginning power-up stage and later periodically at SDL operation stage. This period is usually at 10s of seconds. When interference is detected in the operating channel, LTE-U will attempt to switch to another clear channel with less interference based on LTE Release 10/11 procedures [33, 35, 37]. The interference level in a channel is measured by energy detection where initially the quantity

of interference sources and types are unknown. LTE and Wi-Fi measurements are engaged to augment interference detection.

In [38], Qualcomm presents an effective channel selection policy based on interference level. If the interference of the occupied channel exceeds a certain level, LTE-U/LAA changes the channel with provided that the interference is measured before and during the operation, and both at the equipment and network side. On the other hand, in [39] adaptive bandwidth channel allocation offered by LTE and Least Congested Channel Search(LCCS) has been suggested for channel selection.

- **Carrier-Sensing-Adaptive Transmission (CSAT):**

When no clean channel is available, LTE-U will be able to share the channel by implementing adaptive duty cycle or CSAT algorithm. The aim of the CSAT algorithm is to afford coexistence across different technologies in a Time Division Multiplexing (TDM) mode [40]. In general, the coexistence methods in unlicensed band are by using LBT or CSMA for Wi-Fi, which uses contention based access. For CSMA or LBT, the medium should be sensed and accessed if it is clear in order to implement TDM for coexistence. LTE-U radio continue measuring occupancy on a channel and decide how many frames to transmit or how many to stay quiet which is known as duty cycle. Duty cycle facilitate the interaction when LTE-U is ON and Wi-Fi is in OFF state. The LTE-U, which is on secondary cell is occasionally activated and de-activated using LTE MAC control elements.

In [41], blank subframe allocation by LTE has been proposed where LTE restrained from transmitting and WiFi keeps on transmission. A similar technique has been proposed in [42] where n of 5 sub-frames of LTE-U/LAA has been kept reserved for WiFi transmission. Qualcomm has proposed Carrier Sensing Adaptive Transmission (CSAT) for LTE-U/LAA MAC scheduling in which a fraction of TDD duty is used

for LTE-U transmission and the rest is used for other technologies. The cyclic on/off ration can be adaptively adjusted by based on activity of WiFi during the off period. In [43], Q-learning based duty cycle adjustment is presented to facilitate the sharing of the channel as well as to increase the overall throughput.

- **Opportunistic Supplementary Downlink (SDL):**

This mechanism is dependent on traffic and load demand. If the DL traffic of the small cell exceeds certain threshold and there exist active users within the unlicensed band spectrum, the SDL carrier can be turned on for offloading. On the other hand, when the primary carrier can easily manage the traffic demand and there is no user within the unlicensed band coverage the SDL is turned off. Opportunistic SDL decreases the interference from continuous RS transmission from LTE-U in unlicensed channel subsequently leading in noise reduction in and around a shared channel [34, 44].

Coexistence Based on LBT Mechanism

For Europe, Japan and India markets that requires a regulation in the unlicensed spectrum require a more robust equipment to periodically check for presence of other occupants in the channel (listen) before transmitting (talk) in millisecond scale. Two LBT mechanisms are employed in LTE-LAA mandated by European Telecommunications Standards Institute (ETSI). One is Frame based Equipment (FBE) and Load based Equipment (LBE) [45, 46].

- **FBE-Based LBT Mechanism:** In this [47–49], the equipment has a fixed frame period, where CCA is executed. When the current operating channel is dimmed to be clear, the equipment immediately can transmit for duration equivalent to the channel occupancy time. Similarly, if the operating channel is busy, the equipment

cannot transmit on the channel for the next fixed frame period. The channel occupancy time requirement is minimum 1ms and maximum 10ms and idle period accounting 5 of channel occupancy time. FBE-based LBT is simple for the design of reservation signal and requires less standardization.

- **LBE-Based LBT Mechanism:** In LBE [50,51], the equipment is required to define whether the channel is clear or not. Unlike FBE, LBE is demand-driven and not dependent on fixed time frame. In the case where the equipment discovers a clear operating channel, it will instantly transmit. If not, an Extended CCA (ECCA) is implemented, where the channel is observed for a period of random factor N multiplied by the CCA time slot [37, 50, 51]. N is the quantity of clear slots so that a total idle period should be observed before transmission. Its value is chosen randomly from 1 to q , where q has a value from 4 to 32. When a CCA slot is idle, the counter will be cut by one. The equipment can transmit even if the counter reaches zero. In addition, the maximum channel occupancy time is calculated by $(13/32) * q$ ms. Therefore, the maximum channel occupancy time is 13 ms when q equals to 32 which is the best coexistence parameter.

2.1.2 Coexistence between LTE and ZigBee

Coexistence between different communication standards have been studied before in the literature [19]. In [12, 13], coexistence among IEEE 802.15.4, WiFi and Bluetooth has been investigated. Interference suppression technique for coexistence has been presented in [52, 53]. In [54], simultaneous operation of ultra-wideband (UWB) and WiMAX on same frequency band has been proposed using spectrum sensing by detect-and-avoid mechanisms. In [14], coexistence among UWB, WiFi and ZigBee have been studied. In [55, 56], coexistence of LTE and WiFi using different mechanism such as listen before

talk (LBT), silent gap and common database have been proposed. However, for smart grid applications, coexistence of ZigBee and LTE on the same frequency spectrum calls for investigation.

For LTE and ZigBee to coexist in the same spectrum, the main challenge is that LTE uses dynamic scheduling for user equipments (UEs) whereas ZigBee utilizes collision sensed multiple access/ collision avoidance (CSMA/CA) mechanism for accessing network. In case of simultaneous CSMA/CA and LTE operation, several techniques are proposed such as carrier sensing and co-existence gap in transmission frame [41, 57]. In LBT mechanism of LTE operation, Request-to-Send (RTS) and Clear-to-Send (CTS) messages are exchanged prior to LTE transmission [57]. In blank subframe allocation technique [41], LTE will refrain from transmission in certain subframes. Both the techniques have been proposed for coexistence of LTE and WiFi where LTE and WiFi transmission power are greater than 20 dBm. On the other hand, ZigBee transmission power is between -3 dBm to 10 dBm. Therefore, ZigBee will have less interference effect on LTE while LTE will cause significant interference on ZigBee.

2.2 Latency

Low latency is a critical requirement in the next generation- 5G networks [23]. Applications in manufacturing, transportation, robotics and telepresence, virtual reality, health care, gaming, smart grid, and educational purpose may require latency 1ms- 100ms. In some case such virtual reality, tel-surgery, and machine tool operations, latency on the order of 1ms is required. In this regard, drastic changes in the network architecture including core and radio network has been proposed.

In RAN network, various enhancements in packet or frame structure, waveforms/multiple access, modulation and coding, transmission, control channel, and symbol detection have

been proposed to achieve low latency. In case of packet or frame structure, short TTI [58], small packet transmission [59], subcarrier spacing [60], TDD optimization [61–64] are proposed. For advanced waveforms/multiple access, Filtered CP-OFDM, UFMC and FBMC are proposed [65–67]. In case of modulation and coding Polar coding [68], Turbo decoding with combined sliding window algorithm and cross parallel window [69], IFFT design with butterfly operation [70], Sparse code multiple access [71], Balanced truncation [72] are proposed techniques to achieve low latency. For transmitter adaptation, Asymmetric window [73], Transmission power optimization [74], Path-switching method and a packet-recovery method [75], and Diversity implementation [76] are the recommended techniques. Scaled control channel design [77], Control channel sparse encoding [78], Symbol-level frequency hopping and sequence-based sPUCCH [79], Radio bearer and S1 bearer management [80], and Outer-loop link adaptation [81] are proposed approaches in control signaling. In case of symbol detection on the receiver side, Linear MMSE [82], SM-MIMO detection scheme with ZF and MRC-ZF [83], Space-time encoding and widely linear estimator [84], Compressed sensing [7, 85, 86], and Low complexity receiver design [87] are the recommended approaches. In addition to those approaches, carrier aggregation in mmWave [88–91], location aware communication [3, 92, 93], QoS/QoE reinforcement [94–103], and cloud RAN [104–108] are also proposed in RAN.

To achieve low latency, new entities such as SDN, NFV, are MEC/fog network proposed in the core network. SDN based architecture provides larger throughput, massive connectivity and low latency [109–123]. On the other hand, NFV eliminates the hardware-software dependency and resources sharing of RAN [114, 116–118, 123–129]. MEC/Fog networks separate the data planes from control plane and bring the storage and computation near to the users [34, 109, 118, 124, 126, 127]. In addition to core network entities, various modifications in backhaul have been proposed in latency reduction

process. In case of general backhaul, SDN and cache enable architecture [130], MAC-in-MAC Ethernet based unified packet-based transport network [131], PON-based architecture with a tailored dynamic bandwidth allocation algorithm [132], Modified VLC technology to set up an OW link [133] , and GTP tunnel optimization [109] are several prominent approaches. On the other hand, mmWave based fronthaul and backhaul [134], mmWave based backhaul frame structure [135], digitally-controlled phase-shifter network for mmWave massive MIMO [136], framework supporting of in-band, point-to-multipoint, non-Line-of-sight and mmWave backhaul [137], and Ultra dense wavelength division multiplexing passive optical networks based backhaul [138, 139] are the several solution for mmWave backhaul.

Caching is the very important tool to bring storage and computation near the users. Additionally, it can help in separation of data plane from the control plane. For 5G networks, mainly four types of caching; (1) local caching, (2) device-to-device caching, (3) small base station caching, and (4) Macro base station caching is proposed. Generally, the file delivery can be divided into two parts: (1) Cache placement and (2) content delivery. In case of cache impalement, it can be centralized or distributed based on network size, design and requirement. In [140], a new scheme is proposed in small base station (SBS) with large memory and weak backhaul link which ensures QoS at the user levels. In order to reduce download delay, the authors investigates the cache placement matrix in [141]. It was shown that cache placement can be affected by backhaul prorogation delay. A distributed cache placement is proposed in [142] to reduce download delay with constraint to BS storage capacity. Separate caching and delivery schemes aim to operate in centralized and distributed manner is proposed in [143]. In this approach, trade-off between the spatial reuse and multi casting is manifested. In [144], cooperative content caching and delivery is proposed which reduces content delivery delay compared to existing content caching. To reduce the traffic of backhaul and download, a weighted optimization has

been proposed in [145] considering caching memory and bandwidth constraint. In [146], caching and forwarding strategies are proposed to improve latency experience in cache enabled network. In [147], multi-cast and cooperation based caching approach has been proposed which reduces 13% latency compared to similar multi-cast caching. In [148], content distribution strategy is presented in cooperated caching where different nodes interact cooperatively in a centralized cloud server networks. Packet transmission is investigated in cache enable network along with RAN and wired backhaul in [149]. Following this, peak and off-peak network performance is investigated along with and without cached enable network.

2.3 Data Security in Smart Grid

In the recent years, security issues in AMI attracted a significant attention of different communities (electrical engineers, computer science graduates, IT experts, etc.) due to extensive use of wireless communication. In [150], a security scheme is proposed for smart meter using digital signatures that a trusted third party might sign with a time stamp. Additionally, data hashing using SHA-256 before performing the signature provides an added layer of security. In [151], randomization of the AMI configuration is proposed to make its behavior unpredictable to the hacker, whereas the behavior is predictable to the control center. In [152, 153], authors introduced anonymization of data by randomizing node identity using a TTP. But, communication overhead is increased due to the need for the TTP to communicate with all nodes simultaneously. In [154, 155], homomorphic encryption has been introduced. Though it requires minimal calculation at data retrieval, it may be complicated for a large network.

In [156], an Identity-Based Signcryption (IBS) with zero configuration encryption and authentication has been proposed for an end-to-end communication solution. In [157],

a node-to-node encryption with its own secret key has been proposed. But for a large network, the packet overhead might be increased for both IBS and node-to-node authentication. In [158], DiffieHellman key protocol based message authentication is proposed in addition to Hash based message authentication. This approach provides higher scalability, lower memory utilization and less delay for decryption. In [159], authentication between smart meter and utility server along with low overhead key management has been proposed. The mutual authentication consists of four steps whereas the key management is founded on ID based public/private key pair model with lower transmission overhead for key refreshment. Four broad countermeasures to thwart attacks on smart meters have been proposed in [160]: (1) authentication and strong encryption of communication that deals with HAN and NAN and buses within smart meters, (2) secure key management which form the critical backbone to a secure AMI, (3) securing the firmware to avoid being manipulated by the attackers or mistakenly by authorized personnel, and (4) security-driven firmware development cycle that conducts frequent walkthroughs and security assessments.

In [161], event driven asset centric key management is proposed where key management (i.e. key generation, refreshment, revocation, etc.) is orchestrated automatically based on events from assets or nodes. In [162], public key management has been proposed for smart grid based on elliptic curve public key cryptography and Needham Schrouder symmetric key authentication. Even though, scalability and simplicity are two advantages of this approach, it does not come with experimental proof. In [163], symmetric key establishment mechanism is proposed based on X.1035 standard which reduces data delivery time up to 75%. In [164], group key management with three-tier network model is proposed which requires moderate key storage. To distribute the keys and manage the network, a wireless sensor network based Public Key Management Infrastructure (PKI) has been proposed in [165, 166]. However, it requires to generate a large numbers of

unique keys for a large networks. In [167], a Key Management System (KMS) has been introduced based on DLMS/COSEM standard providing two main information security features: data access security and data transport security. Since DLMS/COSEM is an open standard and allows a number of variations in the protocol implementation, it might increase the complexity in the client side. In [168], information centric KMS has been proposed. But, large number of unique key generation and communication overheads due to unicast, multicast and broadcast might be an issue for large networks. In [169], a two layer security scheme for meter to Data Concentrator (DC) and DC to control center has been proposed. For the meter to DC, IEC 62056 based encryption method, and for DC to the control center, the public KMS has been proposed. But here in each time step, encryption and decryption need to be performed twice.

CHAPTER 3

LTE and ZigBee/WiFi Coexistence

To have LTE coexist with ZigBee/WiFi in the same spectrum, the main challenge is that LTE uses dynamic scheduling for user equipments (UEs), whereas ZigBee/WiFi utilizes collision sensed multiple access/collision avoidance (CSMA/CA) mechanism for accessing network [16, 19]. Unlike ZigBee/WiFi, LTE does not implement the carrier sensing before transmitting the packets. Therefore, in a typical scenario of coexistence, the ZigBee/WiFi transmission is most likely to be blocked by LTE transmission. To facilitate coexistence between LTE and WiFi, mainly three types of techniques is found in the literature: (1) Dynamic channel selection, (2) Listen Before Talk (LBT), and (3) Coexistence gaps. However, for LTE and ZigBee coexistence, there is no significant work found in the literature.

In this chapter, we study two types of coexistence- (1) LTE and ZigBee, and (2) LTE and WiFi. At first, we present a general overview of LTE, ZigBee, WiFi and metering infrastructure of smart grid. Following this, we present the performance of coexistences of LTE and ZigBee, and LTE and WiFi using a smart grid scenario.

3.1 Overview of LTE Systems

LTE is a standard for high speed wireless communication to meet the rapid increase of mobile data usage in the future. The standard is developed by the 3rd Generation Partnership Project (3GPP) and is an upgrade from 3G standards for significantly reducing data transfer latency and increasing capacity of data transfer. The PHY layer of LTE includes the DL and UL features. The requirements of this layer are high peak transmission rates, spectral efficiency and multiple channel bandwidths. Therefore, in order to meet these requirements, Orthogonal Frequency Division Multiplex (OFDM) technology is used due to its robustness against fading and interference. To further improve the performance of

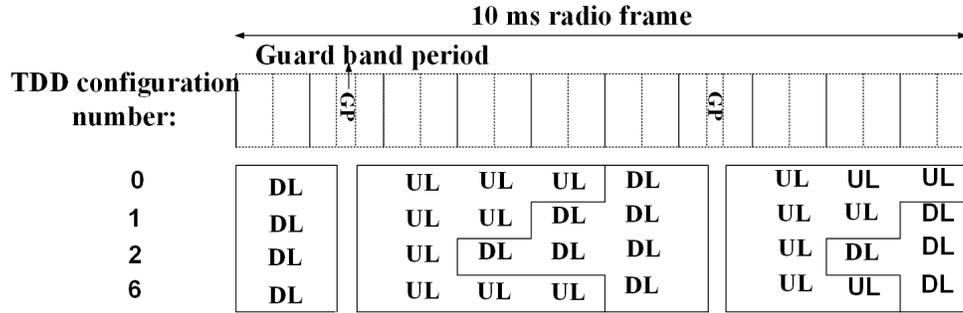


Figure 3.1: 10 ms radio frame of LTE [2].

this standard, multiple antenna techniques are used, which are responsible for increasing channel capacity and increasing the robustness of transmitted signals. The MAC layer, on the other hand, provides an interface between logical channels and physical channels. It is responsible for transport format of the frames, in addition to scheduling techniques and error correction using HARQ.

The LTE DL transmission consists of user-plane and control plane data in the protocol stack multiplexed with physical layer signaling to support the data transmission, facilitated by the Orthogonal Frequency Division Multiple Access (OFDMA). The transmission resource is structured in 3 dimensions- time, frequency, and space. In time domain, the largest radio frame is 10 ms, subdivided into ten 1 ms subframes as illustrated in Fig. 3.1 . Each subframe is further split into two 0.5 ms slots, which is termed as a resource block (RB).

On the other hand, UL utilizes Discrete Fourier transform-Spread OFDM (DFT-S-OFDM). Before transmitting, the signal is frequency-shifted by half a sub carrier frequency in order to overcome the distortion caused by the d.c. subcarrier being concentrated in a single RB.

For LAA of LTE in unlicensed band, two approaches were proposed. In the early deployment stage of LAA in USA, Korea, China and India, adaptive duty cycle- carrier sensed adaptive transmission (CSAT) [170] was considered whereas in Europe listen be-

fore talk (LBT) [171] was used. For fair sharing of spectrum, recently 3GPP Rel. 13 LAA proposes active LTE transmission time 1-10 ms based on load which is similar to CSAT [172].

3.2 Overview of ZigBee Systems

ZigBee is a low-cost, low-power, short-range, low-data rate and energy-efficient wireless technology, and it suits good to be used for applications that involve wireless M2M communications. It gained ratification from IEEE 802.15.4 in 2003. It consists of 16 channels in the 2.4 GHz ISM band worldwide, 10 channels in the 915 MHz band in North America, and one channel in the 868 MHz band in Europe (see Table 3.1). Its operational range is 30-90 m and it can simultaneously support up to 64000 nodes. In the protocol stack, MAC and PHY layers are defined by IEEE 802.15.4 whereas the upper network layers are defined by ZigBee.

The routing protocol is designed to be supported by both ZigBee alliance and IEEE 802.15.4. ZigBee interprets the software command of applications and passes new command to the MAC layer. IEEE 802.15.4 can work on both peer to peer and star networks. A typical example of ZigBee superframe structure has been illustrated in Fig. 3.2 where $aBaseSuperframeDuration = 5.36$ ms and duty cycle is between 10% – 100%. The superframe is divided into inactive portion and active portion, and the latter is further subdivided into contention access period and contention free period. By specifying a duty cycle, we can decide the active transmission time (as for example, at 18% duty cycle, the transmission duration is 1 ms).

IEEE 802.15.4 works at three different frequency bands. The first band is 868 MHz which supports a data rate of 20 Kbps and uses the BPSK modulation technique. The second band is 915 MHz and supports a maximum data rate of 40 Kbps modulated by

BPSK. The global 2.4 GHz band supports 250 Kbps data rate and is modulated by offset quadrature phase-shift keying (OQPSK).

Table 3.1: IEEE 802.15.4 FREQUENCY NOMENCLATURE.

Band	Number of Channels	Channel Number	Channel Center Frequency	Channel Spacing
868 MHz	1	$k = 0$	868.3 MHz	0
915 MHz	10	$k = 1, 2, 3...10$	$906 + 2(k - 1)$ MHz	2 MHz
2.5 GHz	16	$k = 11, 12...26$	$2405 + 5(k - 1)$ MHz	5 MHz

The MAC and PHY layers are based on the CSMA/CA algorithm along with slotted binary exponential backoff (to reduce collision during simultaneous data transfers by multiple channels). It supports two kinds of channel access modalities- beacon enabled slotted CSMA/CA and simple unslotted CSMA/CA without beacon.

ZigBee utilizes CSMA/CA [173] mechanism to access the network. When a packet comes to a node, the MAC layer initiates two variables- the number of backoff tries and the exponential backoff with a minimum value of 3. The MAC layer generates a random

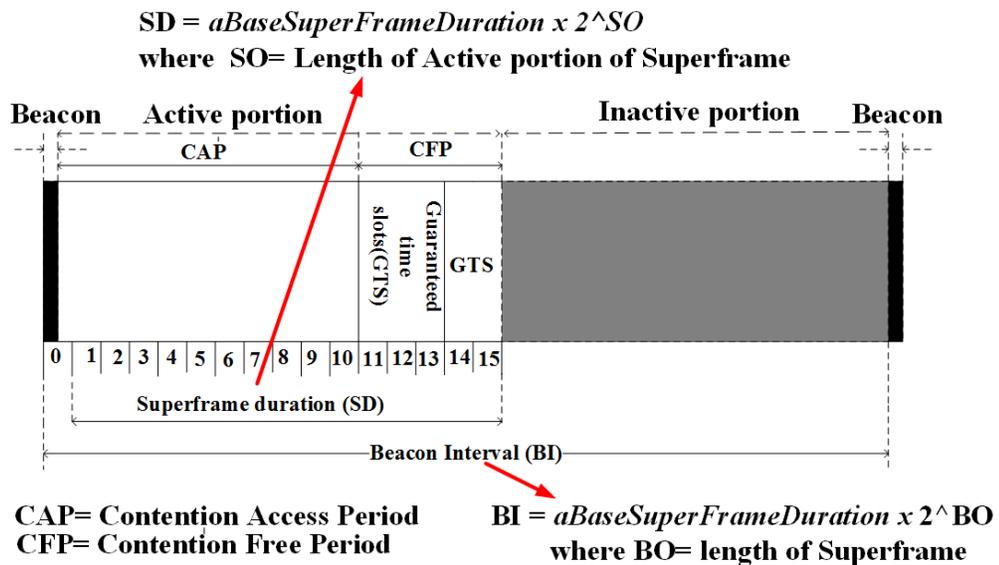


Figure 3.2: A typical example of superframe of ZigBee.

value within the range $[0, 2^{BE} - 1]$ and sets the delay accordingly. When $BE=0$, the MAC senses the channel. If the channel is found idle, packet is transmitted. Otherwise, the value of BE and NB is increased by 1. If the value of NB is less than the maximum number of Backoff (N_{Bmax}), the whole process is repeated for the transmission of the packet. If otherwise, the transmission is discarded.

In the 10 RBs of 10 ms LTE subframe, ZigBee will transmit only during the 2 guard band period (blank RBs). On the other hand, during the remaining 8 RBs, ZigBee will cease transmission due to CSMA/CA mechanism.

3.3 Overview of WiFi Systems

WiFi technology is used for wireless local area networking with devices following the IEEE 802.11 standards. The wireless technologies that come under the WiFi technology umbrella include 802.11n (300Mbps), 802.11b (11Mbps), 802.11g (54Mbps), 802.11a (54Mbps). The most essential aspect of WiFi technology is that it supports diversified range of electronic devices. The WiFi technology is generally referred as a higher layer protocol with IP protocol acting as the dominant one facilitating communication using internet without any protocol translator.

Communication using the WiFi technology is suitable mostly for application with reduced data rate requirement or conditions with low interference. The MAC layer of the WiFi system controls the common channel access of multiple WiFi stations. This control mechanism is fulfilled by using carrier sense channel access with collision avoidance technique (CSMA/CA) technique. In this technique, the STAs can transmit packets during the idle state of channel otherwise transmission is ceased. A better understanding about channel accessing using WiFi can be obtained by studying the various approaches of clear assessment techniques (CCA) of WiFi [15].

One of the CCA techniques for channel access that we have used in this paper is Enhanced Distribution Channel Access (EDCA). The EDCA technique is similar to the DCF technique used in WiFi and has compatibility with four access categories. These are voice, video, best effort and background. In DCF, at the beginning of transmission, a station waits a DIFS before counting down. In EDCA, the station waits for an arbitrated interframe space (AIFS) where AIFS has to be equal to or greater than DIFS. The STAs will cease transmission of packets for arbitrary period during the AIFS time period even if channel is in idle state. This arbitrary wait time for the STAs is decided using the contention window parameters (CW_{min} , CW_{max}) [15], [174].

3.4 Overview of Metering Infrastructure of Smartgrid

Smart grid is the evolution from one-directional power system to modern bi-directional power system which employs innovative communication and distribution to deliver electricity to consumer with enhanced monitoring, control and efficiency [175]. A proclaimed feature of the smart grid is the usage of bidirectional communication for interacting among its entities.

In the arena of smart grid, AMI is the distribution level building block consists of a network of smart meters. AMI is responsible for collecting and sending consumption data from smart meter periodically using wireless communication. It consists of various components which have diverse applications. An example of AMI communication scenario that encompasses communication among home appliances and the control center is shown in Fig. 3.3, which consists of a home area network (HAN), smart meters, neighborhood area network (NAN), and control center.

The network by which home appliances (such as stove, microwave, dishwasher etc) are connected to a meter is termed as the HAN. The most common communication stan-

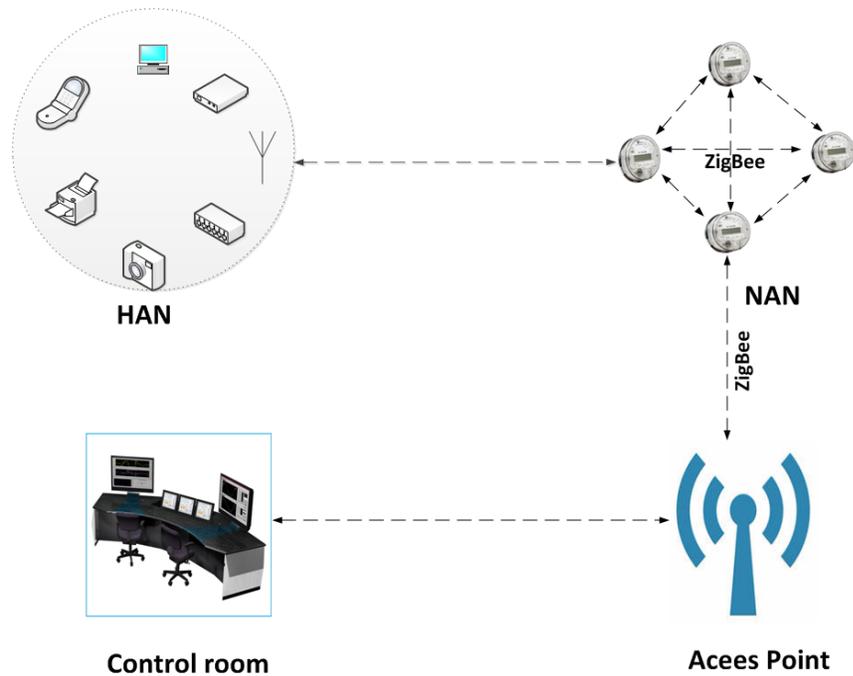


Figure 3.3: AMI communication architecture encompassing Home Area Network (HAN), Neighbor Area Network (NAN), data concentrator, and control center.

standard used for HAN is Bluetooth and WiFi. The energy consumed by home appliances is encapsulated as a consumption unit which is measured, stored and sent by smart meter.

Smart meter is a solid state device responsible for measuring, storing and sending energy consumption data to the back office of the energy service provider. A smart meter sends data every 10 - 30 minutes based on energy service provider's choice. The meter is connected to a network named as the NAN. The popular communication standard for NAN is Zigbee and WiFi. In our study, we use ZigBee for investigation.

The meters are connected among themselves through a mesh connected network. The network may be wired (PLC) or wireless such as LTE, WiFi, ZigBee. The head end of the NAN is the data concentrator or gateway which is connected to a back office by a dedicated wired or wireless connection (optic fiber, cellular network, etc.).

The control center receives consumption data and prepares bills for the consumers. The fine grained data can be used to forecast and optimize the electrical power generation

and distribution. Controlling and monitoring is also performed from remote locations depending on the usage and load requirement.

Since ZigBee is currently used in many utility companies for NAN communication, we will study its coexistence performance in the rest of the paper.

3.5 LTE-ZigBee Coexistence

In this study, we investigate the coexistence of LTE and ZigBee in the unlicensed 2.4 GHz ISM band for AMI communication and usual H2H communication [19]. We consider time division duplexing (TDD)-LTE and ZigBee for simulation on multi layer network layout. In the 10 ms subframe of LTE transmission, ZigBee will transmit during the guard band (blank subframe) time. We evaluate the performance under different LTE traffic arrival rates defined by 3GPP through the FTP traffic model [176] as well as different combination of down link (DL)/ uplink (UL) subframes. The simulation results yield insights about interference effect on each other in the PHY/MAC operating regimes and will help to develop effective coexistence mechanism between LTE and ZigBee.

3.5.1 System Model for LTE-ZigBee Coexistence

To evaluate the coexistence performance of LAA based LTE and ZigBee networks, a suburban building block topology (Manhattan grid) [177] has been considered as shown in Fig. 3.4. In the topology, there are 30×9 blocks in which each block contains 5×5 buildings. Groups of 5 blocks are each served by an STA (ZigBee access point) whereas each 15 block building is served by an eNB. In each building, there are M UEs (LTE subscriber) and a smart meter (ZigBee device) which are uniformly distributed. Each UE/ZigBee device communicates with corresponding eNB or STA. Since a smart meter sends its data to the corresponding STA by hop-to-hop communication, and the STA may

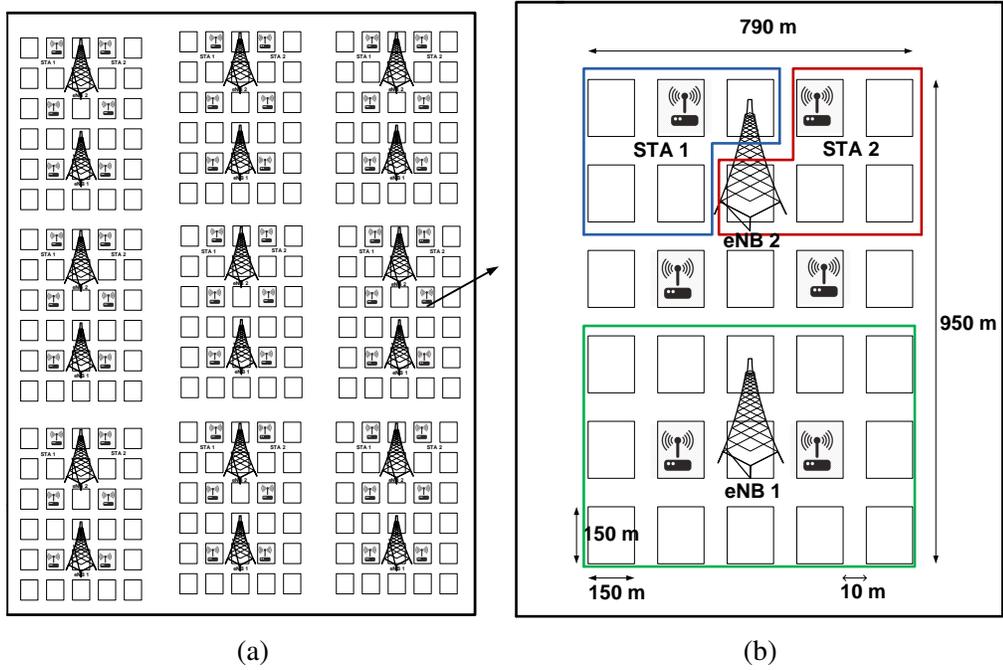


Figure 3.4: (a) Mean square error for different number of nodes and noise variances. (b) Mean square error for different number of nodes and path loss exponent.

also send the data to control center by LTE, its location has been considered to be near an eNB. In each case of 10 ms LTE sub frame transmission, two blank slot (Guard band, 0.5 ms each) has been kept for ZigBee/WiFi transmission and the corresponding ZigBee duty cycle is 18% (i.e. 1 ms). Since LTE and ZigBee are transmitting simultaneously on the same 2.4 GHz spectrum, high power LTE and low power ZigBee transmission cause interference on each other from other cells. To assess the performance, Shannon capacity of both technologies are simulated under interference from each other. For this, PHY layer has been abstracted at the granularity level of a transmission frame of ZigBee and LTE.

The number of received bits (N_B) at a ZigBee/LTE node (i) over a frame duration is given by

	10 ms									
	1	2	3	4	5	6	7	8	9	10
Conf#1	U	G	U	U	U	U	U	U	G	U
Conf#2	D	G	D	D	D	D	D	D	G	D

Figure 3.5: LTE TDD subframe structure for two different configurations.

$$N_B(i) = B \log_2(1 + SINR(i))T, \quad (3.1)$$

where B is the bandwidth, $SINR$ is the signal to interference plus noise ratio and T is the single symbol duration.

Capacity of the ZigBee/LTE node is given by

$$C = \frac{\sum_{i=1}^N N_B(i)}{NT}, \quad (3.2)$$

where N is the number of ZigBee/LTE symbols transmitted in one second.

Two configurations of LTE (as shown in Fig. 3.5) have been employed to obtain an insight into the effect of interference on ZigBee performance owing to the fact that both DL and UL uses different transmission power. The UL fractional power control in LTE is given by the following equation

$$P_{UL} = P_0 + \alpha P_L + 10 \log_{10} R, \quad (3.3)$$

where P_0 is the base power level, P_L is the path loss from BS to UE, α is the path loss compensation factor, and R is the number of RBs allocated for UL transmission of the specific LTE UE.

Since in case of ZigBee, the packet size from smart meter is fixed, a non-Poisson full buffer model has been used. On the hand, for LTE, a non full buffer traffic model has

Table 3.2: LTE MAC/PHY PARAMETERS.

Parameter	Value
Transmission Scheme	OFDM
Central Frequency	2405 MHz
Bandwidth	5 MHz
DL Tx Power	20 dBm
UL Tx Power	PL Based TPC
Frame Duration	10 ms
Scheduling	Round Robin
P0	-106 dBm
TTI	1 ms
Pathloss model	Urban micro (UMi)
Traffic Model	FTP Traffic Model-2 [176]

been considered as given in 3GPP FTP traffic model-2 [176]. For traffic arrival rate λ_L in the LTE transmission, the distribution function of delay between two packets (d) is given by

$$f(d) = \lambda_L e^{-\lambda_L d}. \quad (3.4)$$

For path loss and shadowing parameters, urban micro (UMi) model has been considered in our simulation.

3.5.2 Simulation Results

We used TDD-LTE system with FTP traffic model for simulation in Matlab environment. In case of each building, $M = 1.24$ UEs on the average have been uniformly randomly dropped into the building. Therefore, each block has 31 LTE subscribers. For LTE configuration 1, we used all subframes as UL and in configuration 2, we used all subframes as DL in a 10 ms radio frame duration. In all case, during guard band period (G) of LTE transmission, ZigBee will transmit. The parameters used for the simulation are illustrated in Table 3.2 and Table 3.3.

Table 3.3: ZIGBEE MAC/PHY PARAMETERS.

Parameter	Value
Transmission Scheme	O-OFDM
Central Frequency	2405 MHz
Bandwidth	5 MHz
DL/UL Tx Power	10 dBm
AC	Best Effort
MAC Protocol	EDCA
Slot Time	5.36 ms
CCA-CS Threshold	-82 dBm
CCA-ED Threshold	-62 dBm
macMaxFrame Retries	3
macMaxCSMA Backoffs	4
macMin BE	5
macMax BE	3
Unit Backoff Period	3e-10 seconds
Frame Length	808 bits
CW size	8
Noise Figure	6 [2]
Traffic Model	FTP Traffic Model - 2 [176]

From Fig. 3.6, we found that without ZigBee, the aggregate capacity of LTE DL is 48 Mbps, and with ZigBee, its capacity reduces to 32 Mbps and 28 Mbps on traffic arrival rate $\lambda_L = 1.8$ and $\lambda_L = 2.5$, respectively. On the other hand, with the LTE transmission, ZigBee capacity reduces from 220 Kbps to 40 Kbps at $\lambda_L = 1.8$ and 20 Kbps at $\lambda_L = 2.5$, as shown in Fig. 3.7.

For all LTE UL transmission in a 10 ms radio frame, the Shannon capacity of ZigBee is increased significantly compared to all DL LTE transmission in a 10 ms radio frame, which is reflected in Fig. 3.8. The reason behind this is that the UE uses less power for UL transmission which creates less interference in bit reception at ZigBee node. Contrary to ZigBee, for all LTE DL transmission in a 10 ms radio frame, aggregated capacity of LTE DL is more than all LTE UL transmission while ZigBee transmission is also taking place simultaneously. This is illustrated in Fig. 3.9.

In Fig. 3.10, it is shown that ZigBee has limited interference effect on LTE SINR distribution. For interference caused by ZigBee, the median SINR of LTE is downgraded by 5 dB. On the other hand, LTE DL interference has degraded the ZigBee SINR by over 20 dB, which is illustrated in SINR CDFs Fig. 3.11. This significant SINR downfall causes momentous capacity reduction in ZigBee.

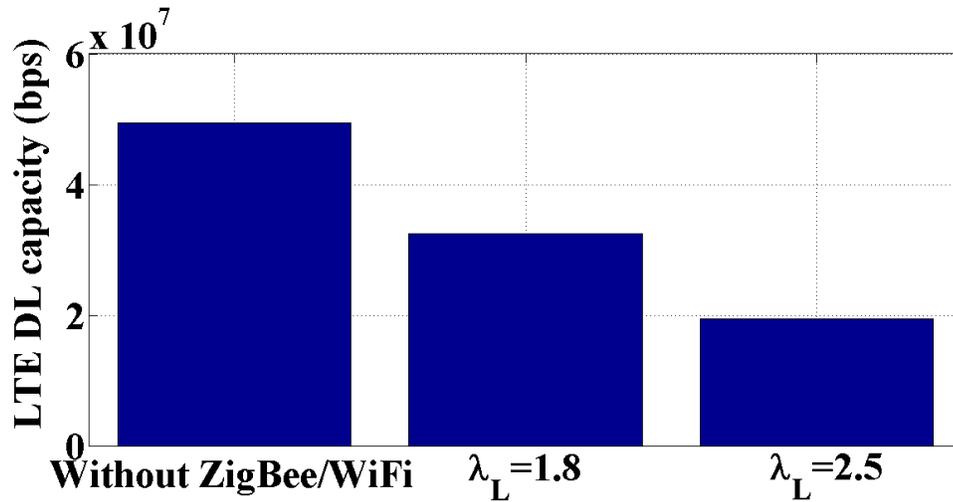


Figure 3.6: LTE DL capacity without ZigBee, and with ZigBee transmission at different traffic rates.

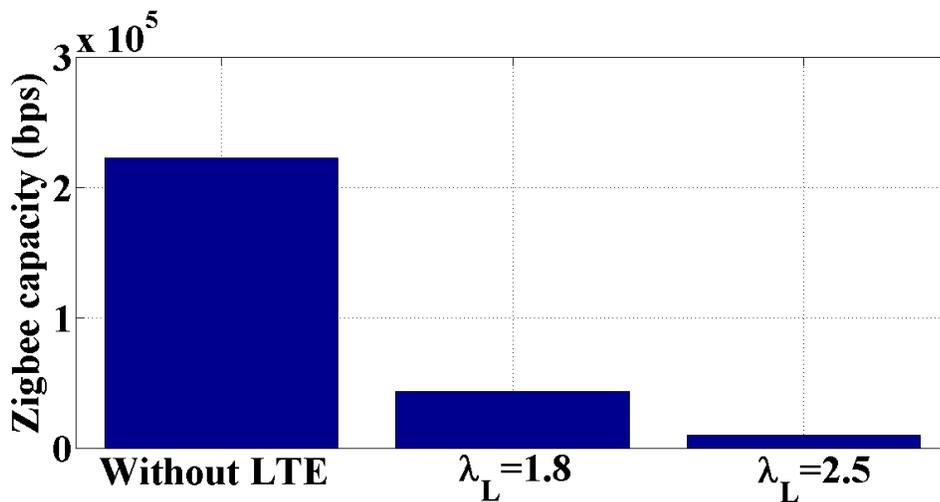


Figure 3.7: ZigBee capacity without LTE, and with LTE DL transmission at different traffic rates.

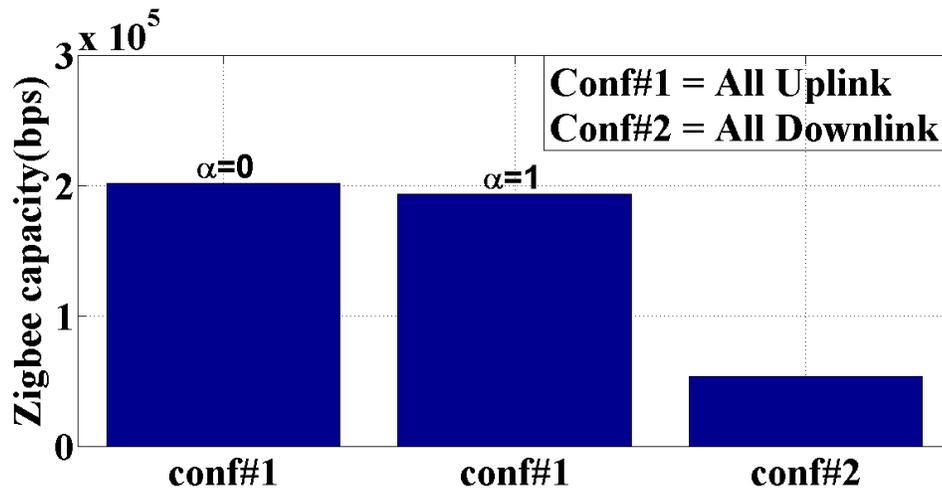


Figure 3.8: ZigBee capacity with all LTE DL or all LTE UL traffic in a 10 ms radio frame at different α , and at $\lambda_L = 1.5$.

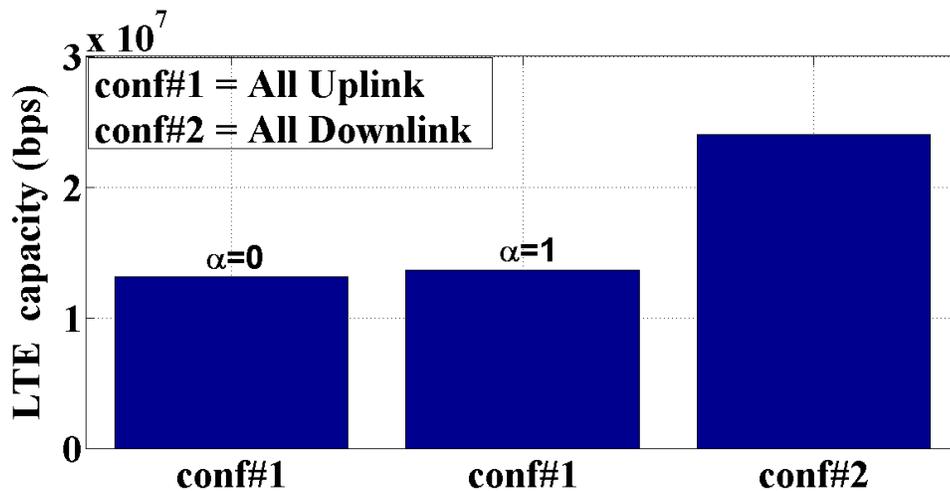


Figure 3.9: LTE capacity with all LTE DL or all LTE UL traffic in a 10 ms radio frame at different α , and at $\lambda_L = 1.5$.

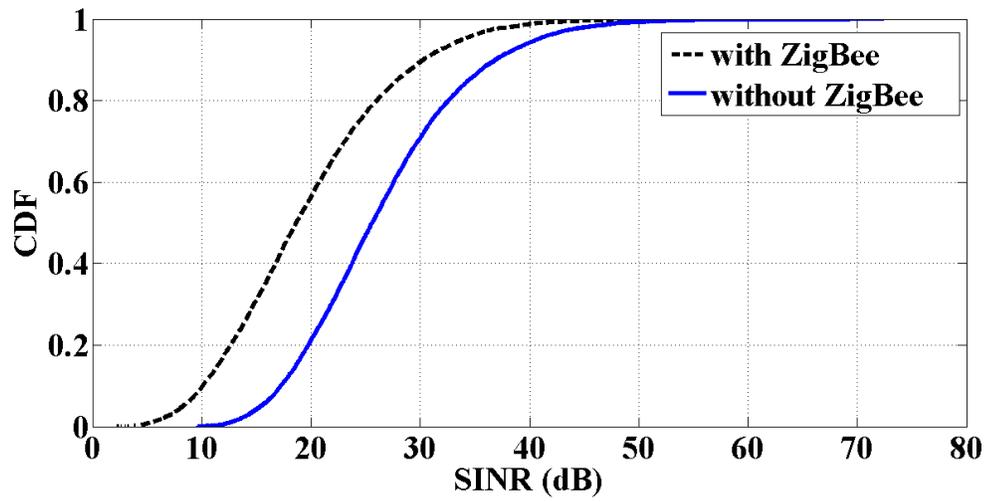


Figure 3.10: SINR distribution of LTE without ZigBee transmission, and with DL ZigBee transmission.

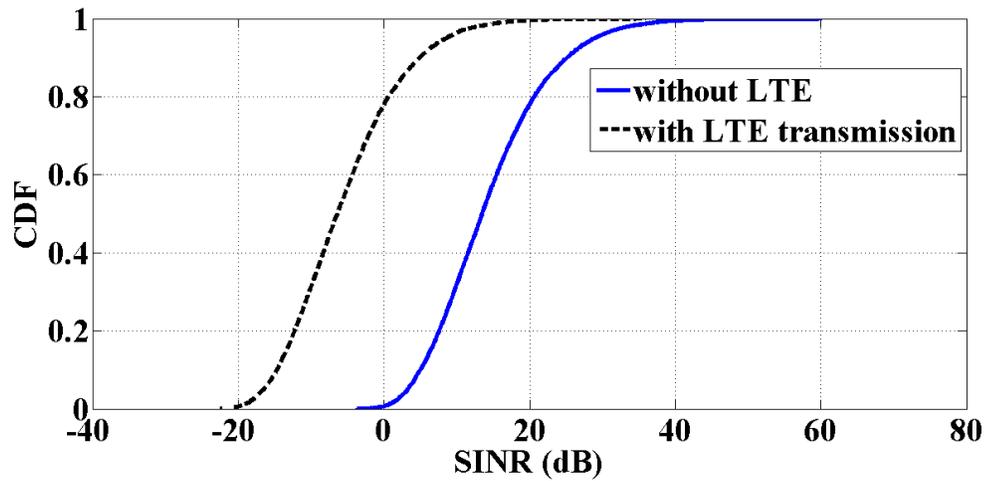


Figure 3.11: SINR distribution of ZigBee without LTE transmission, and with LTE DL transmission.

3.6 LTE and WiFi Coexistence

In this study, we investigate the coexistence of LTE and WiFi with application to smart grid [16]. In our frame work, smart meters use WiFi to transmit data to data collector/Access point (AP). Data collector collects data from a cluster of meters and send the data to control center using LTE. Based on this scenario, we study the performance of coexisted LTE and WiFi in the 3.5 GHz band for AMI communication and usual mobile human-to-human (H2H) communication. We consider a duty cycle based time division duplexing (TDD)-LTE and WiFi for system level simulation on a collocated network layout. LTE system transmits a fixed duty cycle of a period, and on the other hand, WiFi transmits for the rest of the period. The simulation results demonstrate good neighboring coexistence between LTE and WiFi without significantly hampering each other’s performance. Since large amount of clean and free bandwidth is available in CBRS band, coexisted LTE-WiFi based AMI in the CBRS band can be a promising solution for smart grid.

3.6.1 System Model for LTE-WiFi Coexistence

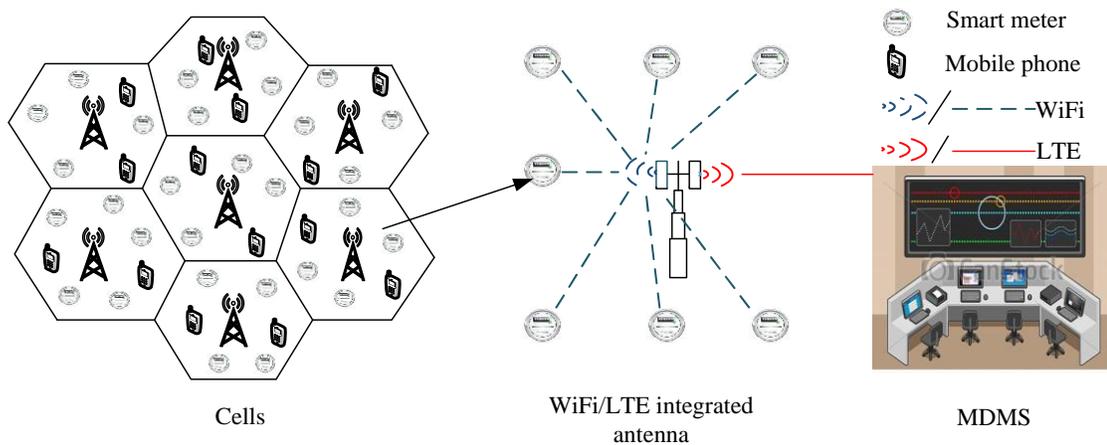


Figure 3.12: Cell layout for AMI of smart grid.

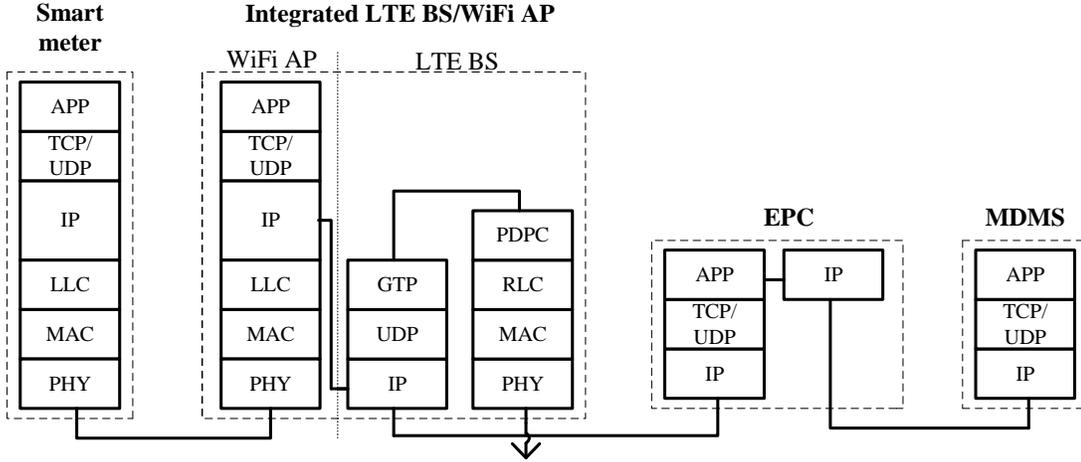


Figure 3.13: Protocol mapping between different entities of WiFi and LTE network. (AP: access point; EPC= Evolved packet core; APP= application; UDP= user datagram protocol; TCP= transmission control protocol; LLC= logic link control; IP= Internet protocol; PHY= physical; MAC= medium access control; GTP= GPRS tunneling protocol; RLC= radio link control; PDPC= packet Data Convergence Protocol)

Let us consider, a collocated LTE-U and WiFi network scenario where LTE-U and WiFi coexists in the 3.5 GHz band as illustrated in Fig. 3.12. In our proposed framework, smart meters use WiFi and APs of smart meters use LTE-U for transferring data. Additionally, collocated WiFi AP and LTE BS are integrated. The data of a cluster of meters is collected by a WiFi AP, and then forwarded to integrated LTE BS. Finally LTE BS transmits data to MDMS. The protocol mapping of various entities of WiFi system and LTE network is illustrated in Fig. 3.13. The PHY layer of smart meter is connected with the PHY layer of WiFi AP. On the other hand, The IP layers of WiFi AP and LTE BS are connected in our model. The communication among LTE BS, EPC and MDMS are based on standard LTE system architecture [178].

Let us consider, the sets of WiFi APs (i.e. data collector), LTE-U BS, WiFi STAs (i.e. smart meter) and LTE-U UE (i.e. MDMS and other UEs) are given by S_w , S_l , U_w^i and U_l^j respectively. The transmission power of WiFi AP i , LTE BS j , meter/WiFi STA l and LTE-U UE/MDMS m are p_r^i , p_r^j , p_r^l and p_r^m .

The channel gain values from WiFi STA/meter x to WiFi AP j , from LTE UE a to WiFi AP j , from LTE-U BS i to WiFi AP j and LTE-U BS $b(i \neq b)$ to WiFi j are $h_{j,r}^x$, $h_{j,r}^a$, $h_{j,r}^i$ and $h_{j,r}^b$ respectively.

The signal-to-noise (SINR) of WiFi AP j during the data reception from meter/WiFi STA x on the resource block r is

$$\text{SINR}_{j,r}^x = \frac{h_{j,r}^x p_r^j}{\sum h_{j,r}^a p_r^a + \sum h_{j,r}^i p_r^i + \sum h_{j,r}^b p_r^b + \sigma^2}, \quad (3.5)$$

where σ^2 is noise variance. The good SINR value ensures high throughput and low SINR results in reduced throughput performance.

The number of successful received bits at WiFi AP j from the WiFi STA x , N_B is

$$N_B^x = BT \sum \log_2(1 + \text{SINR}_{j,r}^x), \quad (3.6)$$

where B is the bandwidth and T is the transmission time such that $T = \sum r$. The number of received bit depends on SINR value.

The up link (UL) capacity of WiFi STA/meter x is

$$C^x = \frac{N_B^x}{T_{tx} + T_{wait}}, \quad (3.7)$$

where T_{tx} and T_{wait} are the transmission and wait time of WiFi, respectively.

For both LTE and WiFi traffic arrival rate λ , the distribution function of delay between two packets (d) is

$$f(d) = \lambda e^{-\lambda d}. \quad (3.8)$$

The higher the value of λ , the more is the number of packet on queue for transmission.

3.6.2 Simulation Results

To evaluate the performance, a collocated 7 cell architecture is considered as shown in Fig. 3.12. We used the Matlab based simulator which was build based on 3GPP standard and was used in [15, 18]. In each cell, for each integrated WiFi AP/LTE-U BS, 10 smart meters (WiFi STAs) and 10 LTE UEs are drooped at random locations. It is noted that one of 10 LTE UEs is to be considered as MDMS. The traffic arrival rates for LTE-U and WiFi are considered as $\lambda_{LTE} = \lambda_{WiFi} = 2.5$. The PHY and MAC layers of LTE and IEEE 802.11n (WiFi) are implemented in the simulation environment. In each transmission time interval (TTI), only one UE is scheduled for the DL transmission and the SINR information is sent to the corresponding BS.

Also based on the number of LTE-U UEs waiting and requesting for the UL transmission during one subframe, bandwidth is equally shared among themselves. The simulation parameter for LTE simulation has been summarized in TABLE 3.4. The parameter value were selected based on 3GPP LTE standard [179].

For WiFi, channel access mechanism CSMA/CA with clear channel assessment assessment (CCA) and enhanced distributed channel access (EDCA) is implemented. WiFi STAs having packets on queue competes for channel access. However, transmission or

Table 3.4: LTE MAC/PHY PARAMETERS.

Parameter	Value
Frequency band	3.5 GHz
Bandwidth	20 MHz
Down link transmission (Tx) Power	20 dBm
LTE-U UE velocity	0 ms
Uplink transmission (Tx) Power	PL Based TPC
Duration of frame	10 ms
Scheduling	Round Robin
P_0	-106 dBm
TTI	1 ms
Packet arrival rate (λ)	2.5

reception is started only after reception of beacon. The WiFi STA sends packets when it sense that the channel is idle. Otherwise, the transmission is ceased and the next transmission will be attempted after a random back off period. The WiFi parameter in our simulation are summarized in the TABLE 3.5. The parameter value were selected based on study presented in [15, 180, 181].

A physical (PHY) layer abstraction is utilized for shannon capacity calculations of WiFi and LTE-U at the $4\mu s$ granularity of WiFi OFDM symbol period of obtaining the number of successfully received bits. FTP Traffic Model-2 [176] is commonly employed for either WiFi and LTE-U. In our simulation, we used 60% and 80% duty cycle of 50 ms transmission time for LTE. Therefore, WiFi will transmit 40% and 20% duty cycle of the 50 ms period.

The throughput performance of coexisted LTE and WiFi in the smart grid scenario is illustrated in Fig. 3.14, and 3.15. Referred to Fig. 3.14, for 60% duty cycle of LTE-U, the capacity of LTE is 36.3 Mbps and the capacity of WiFi is 36.1 Mbps. If we

Table 3.5: WiFi MAC/PHY PARAMETERS.

Parameter	Value
Frequency band	3.5 GHz
Bandwidth	20 MHz
Downlink/Uplink transmission (Tx) power	23 dBm
WiFi STA/meter velocity	0 ms
Access category	Best Effort
MAC protocol	EDCA
Threshold of CCA sensing	-82 dBm
Threshold of CCA Energy detection	-65 dBm
Number of service bits in PPDU	16 bits
Number of tail bits in PPDU	12 bits
Contention window size	$\mathcal{U}(0, 31)$
Noise figure	6
Beacon interval	100 ms
Beacon OFDM symbol detection threshold	10 dB
Beacon error ratio threshold	15
Packet arrival rate (λ)	2.5

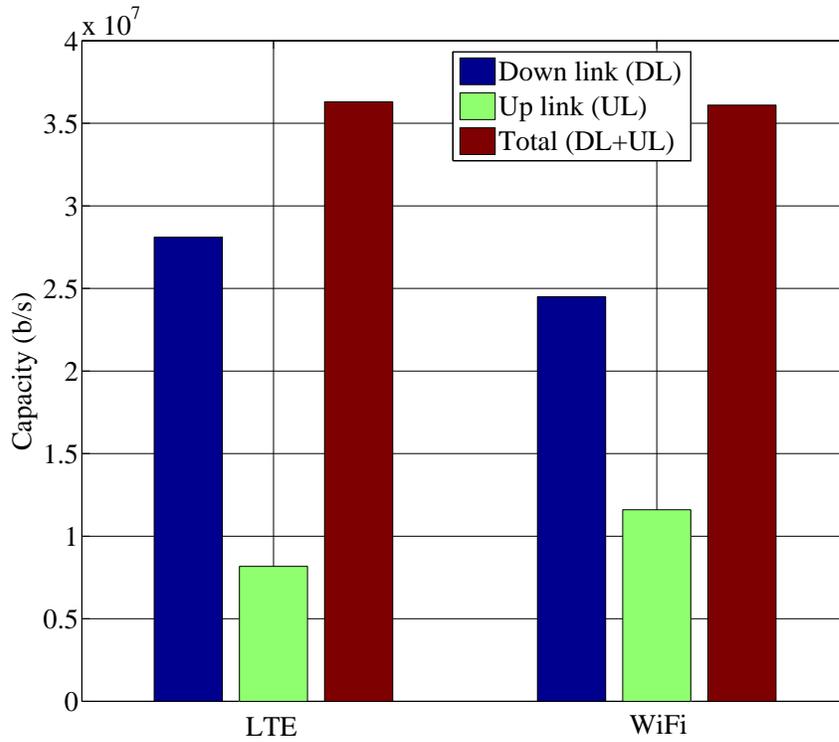


Figure 3.14: Throughput performance of coexisted LTE/WiFi system at 60% duty cycle.

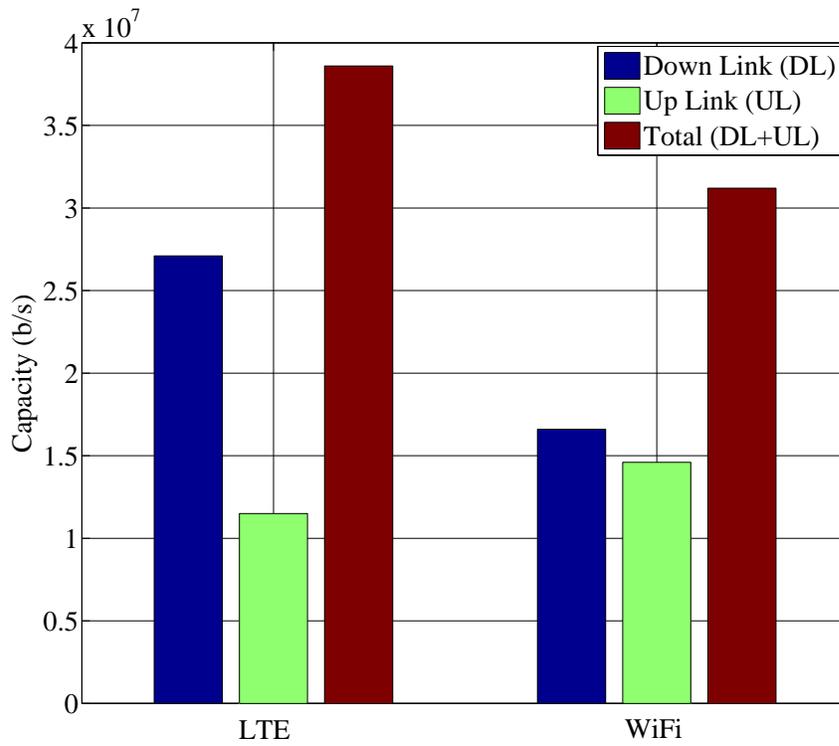


Figure 3.15: Throughput performance of coexisted LTE/WiFi system at 80% duty cycle.

increase the duty cycle of LTE-U to 80%, the LTE capacity is improved to 38.6 Mbps while the capacity of WiFi is decreased to 31.2 Mbps. This is illustrated in Fig. 3.15. The throughput degrade in WiFi is due to the increased transmission back-off on extended transmission time of LTE.

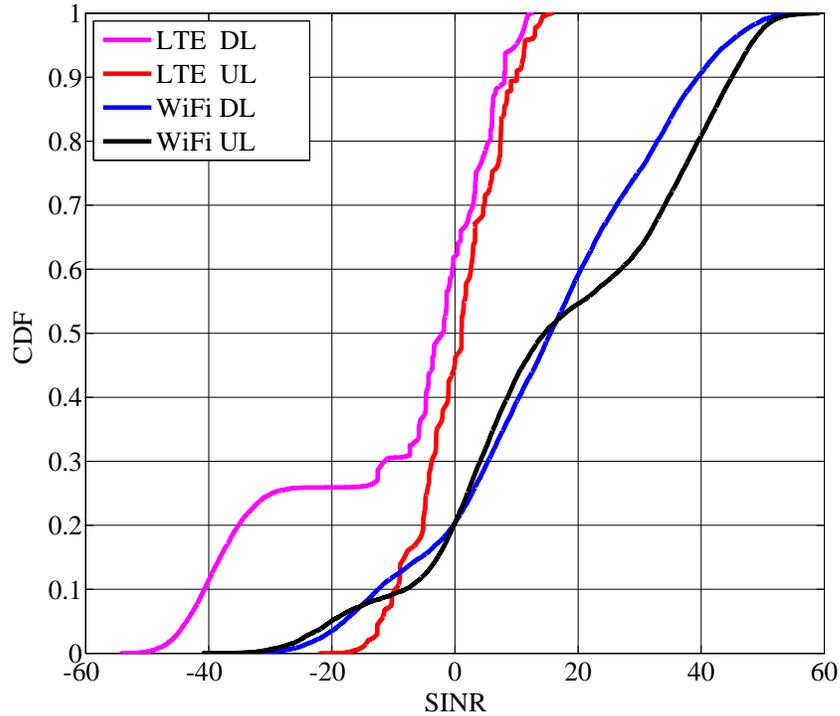


Figure 3.16: SINR distribution of coexisted LTE/WiFi system at 60% duty cycle.

As illustrated in Fig. 3.16, for 60% duty cycle of LTE, the SINR distribution of WiFi is better than that of LTE. However, for 80% duty cycle of LTE, SINR distribution of LTE is improved slightly whereas the SINR distribution of WiFi remained almost same. This is reflected in Fig. 3.17.

The justification of using 60% and 80% duty for LTE is that LTE will be used not only for meter data communication to MDMS, but also it will be used for human-to-human communication (i.e. personal mobile communication). Therefore, we provide more access to LTE transmission. However, more time (i.e. duty cycle) can be allocated for WiFi transmission based on the number of smart meters.

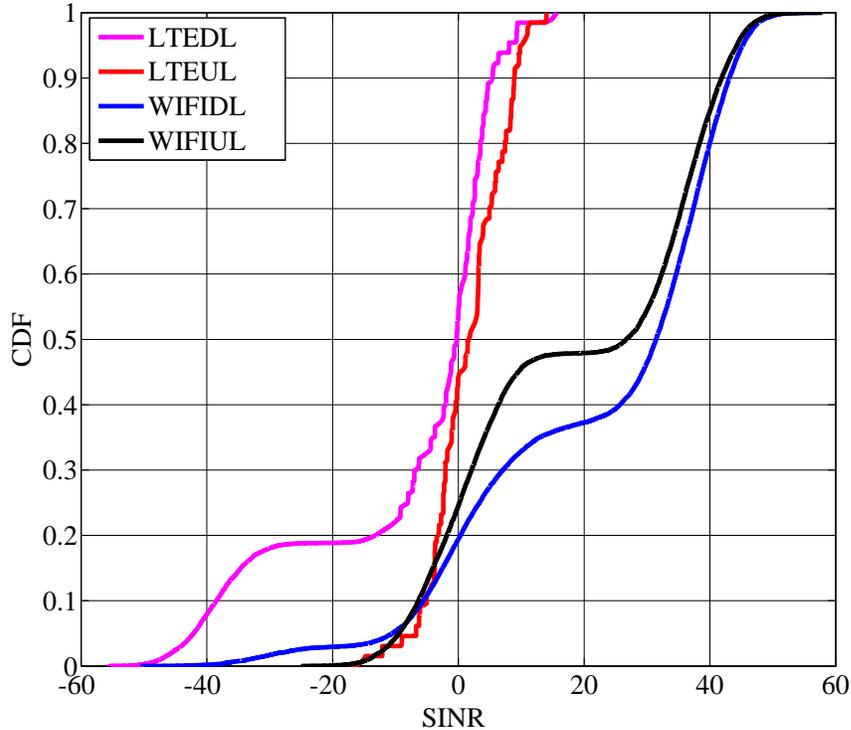


Figure 3.17: SINR distribution of coexisted LTE/WiFi system at 80% duty cycle.

3.7 Conclusion

We study the coexistence of ZigBee and LTE on a multi-layer network. From simulation results, we found that ZigBee capacity is more affected by simultaneous operation of LTE, than LTE capacity being affected by ZigBee transmission. However, by changing the configuration of LTE transmission (DL/UL) and UL power control parameter, we can improve the performance of ZigBee significantly. Since both the ZigBee and WiFi uses CSMA/CA for accessing same channel, SINR of LTE is affected by any of the two. According to US Department of Energy (DOE), the recommended data rate for AMI is 10-100 Kbps/node. Therefore the coexistence of LTE and ZigBee on unlicensed 2.4 GHz band fulfills the requirement for smart meter communication.

In case of LTE and WiFi coexistence, LTE transmits for a fixed duty cycle of a period, whereas WiFi transmits in the rest of the period. However, the duty cycle can be manipu-

lated based on the number of smart meters. The promising simulation results demonstrate that good neighborhood spectrum sharing can be possible without harming each other's performance. Moreover, the 3.5 GHz band has large clean and free bandwidth for data communication. Therefore, 3.5 GHz band sharing by LTE and WiFi can be a promising candidate communication architecture for metering infrastructure of smart grid.

CHAPTER 4

Coexistence of LTE-U and WiFi: A Multi-Armed Bandit Approach

The Third-Generation Partnership Project (3GPP) standardization group has been recently working on standardizing the licensed-assisted access (LAA) technology in the 5 GHz spectrum [21, 22, 182, 183]. The main goal is to develop a global single framework of LAA of LTE in the unlicensed bands, where operation of LTE will not critically affect the performance of WiFi networks in the same carrier. In the initial phase, only downlink (DL) operation LTE-A (LTE Advanced) Carrier Aggregation (CA) in the unlicensed band was considered, while deferring the simultaneous operation of DL and uplink (UL) to the next phase.

Another option for the operation of LTE in the unlicensed spectrum is through a pre-standard approach, referred to LTE-U, where LTE base stations leave transmission gaps for facilitating coexistence with WiFi networks. Development of LTE-U technology is led by the industry consortium known as the LTE-U Forum. LTE-U mainly focuses on the operation of unlicensed LTE in the regions (e.g. USA, China) where listen before talk (LBT) is not mandatory. LTE-U defines the operation of primary cell in a licensed band with one or two secondary cells (SCells), each 20 MHz in the 5 GHz unlicensed band: U-NII-1 and/or U-NII-3 bands, spanning 5150-5250 MHz and 5725-5825 MHz, respectively. However both the LTE-U and LAA need licensed band for control plane. Similar to the 5 GHz band, CBRS band can be utilized for LTE-U operation in the absence of IA users such as radar signal.

In this chapter, we introduce a reinforcement learning (MAB) based adaptive duty cycle section for the coexistence between LTE-U and WiFi [21, 22]. Multi-Armed Bandit (MAB) is a machine learning technique designed to maximize the long-term rewards through learning provided that each agent is rewarded after pulling an arm. Basically

MAB [184, 185] problem resembles a gambler (agent) with a finite number of slot machines in which the gambler wants to maximum his rewards over a time horizon. Upon pulling an arm, a reward is attained with prior unknown distribution. The goal is to pull arms sequentially so that the accumulated rewards over the gambling period is maximized. However, the problem involves the exploration versus exploitation trade-off, i.e., taking actions to yield immediate higher reward on one hand and taking actions that would give rewards in the future, on the other hand.

In our technique, we use a multi arm bandit (MAB) algorithm for selecting appropriate duty cycle. Using a 3GPP compliant Time Division Duplex (TDD)-LTE and Beacon enabled IEEE 802 systems in the 3.5 GHz band, we simulate and evaluate the coexistence performance for different percentage of transmission gaps. We found a significant throughput improvement for both systems ensuring harmonious coexistence. The objectives, subsequently the gains, of this study are not limited to throughput enhancements. The benefits that are achieved in different dimensions with the aid of MAB scheme and the other supporting techniques like PC can be summarized as:

- 1) Proper coexistence is achieved due to the dynamic exploring and exploitation by MAB. So our technique is adaptive.
- 2) The aggregate capacity is improved. Due to the application of MAB algorithm, optimal or suboptimal solutions are achieved.
- 3) Using DL PC higher capacity values are achieved under dense UE and STA configurations.
- 4) Higher energy efficiency is also achieved with PC, which always attempts to reduce the transmission power while increasing the energy efficiency.
- 5) With the use of learning algorithm, a high degree of efficiency is achieved.

To the best of our knowledge, our work is the first study that introduces MAB for improving the coexistence of LTE and WiFi in the unlicensed bands.

4.1 System Model and Problem Formulation

To evaluate the coexistence performance of LTE-U with WiFi in the unlicensed band, a collocated LTE-U and WiFi network scenario is considered. The sets of LTE-U BSs, WiFi APs, LTE-U UEs for BS i and WiFi STAs for AP w are given by \mathcal{B}_L , \mathcal{B}_W , \mathcal{Q}_L^i and \mathcal{Q}_W^w , respectively. $\mathcal{Q}_L = \{ \mathcal{Q}_L^1, \mathcal{Q}_L^2, \dots, \mathcal{Q}_L^i, \dots, \mathcal{Q}_L^{|\mathcal{B}_L|} \}$ and $\mathcal{Q}_W = \{ \mathcal{Q}_W^1, \mathcal{Q}_W^2, \dots, \mathcal{Q}_W^w, \dots, \mathcal{Q}_W^{|\mathcal{B}_W|} \}$ represent the sets of all UEs and STAs. For LTE-U, TDD-LTE is considered. For synchronization of WiFi STAs with the corresponding APs, a periodic beacon transmission is used as in [43].

4.1.1 Interference on DL and UL Transmissions

Interference caused to an LTE-U UE and an LTE-U BS during DL and UL transmissions are shown in Fig. 4.1. A TDD frame structure similar to that in [186, Fig. 6.2] is considered for all the BSs and UEs with synchronous operation. As shown in Fig. 4.1(a), in the simultaneous operation of an LTE-U within a WiFi coverage area, the DL LTE-U radio link experiences interference from other LTE-U DL and WiFi UL transmission. As the same time, WiFi UL suffers from near LTE-U transmission. During an UL transmission subframe, shown in Fig. 4.1(b), LTE-U BS is interfered by the UL transmission of LTE-U UEs, as well as the DL transmissions of WiFi. Similarly, WiFi DL transmission is interfered by other LTE-U ULs where the DL received signal of a WiFi STA is interfered by other LTE-U UL transmissions. In the coexistence scenarios with high density of WiFi users, WiFi transmissions get delayed degrading their capacity performance due to the use of carrier sense multiple access with collision avoidance (CSMA/CA) mechanism [187]. This is an additional degradation other than the performance reduction experienced due

to LTE-U transmissions operated on the same spectrum and this is valid only for WiFi APs and STAs.

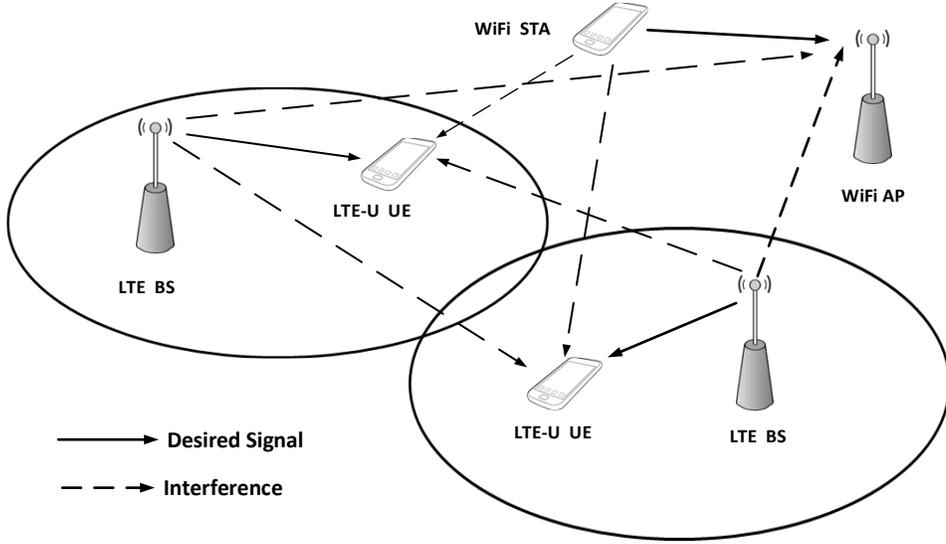
4.1.2 Duty Cycle of LTE-U

In the case of designing a duty cycle for LTE-U, multiple LTE TDD frames are considered. For that purpose, five consecutive LTE frames [186, Fig. 6.2(a)] are used to construct a duty cycle.

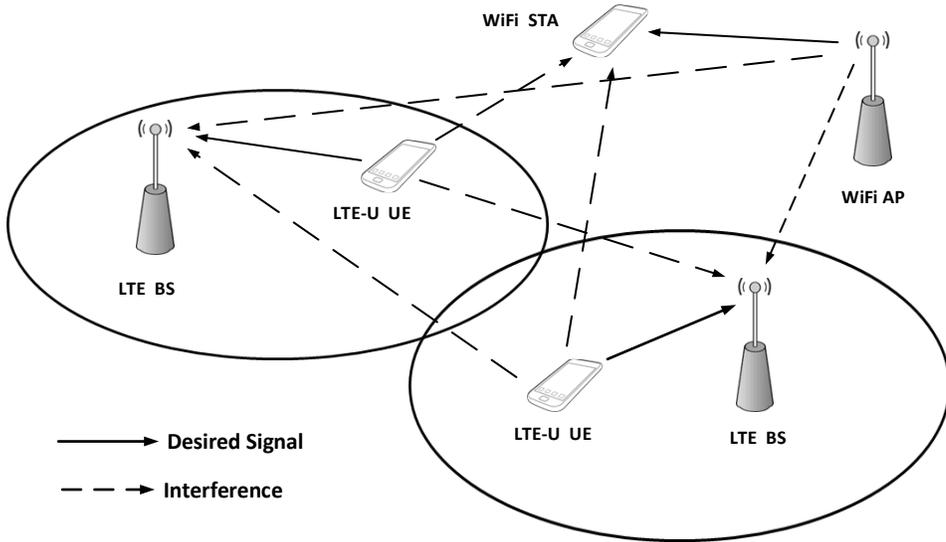
Similar to [43], the LTE-U transmission ON/OFF condition is used to define a duty cycle which is shown in Fig. 4.2 (e.g., 40% duty cycle: during the first two consecutive LTE-U frames, transmission is turned on and it is turned off during the following three frames). One out of these two configurations is used by the UEs and BS in an LTE cell during a duty cycle period. According to this structure, a constant UL:DL duty cycle value is maintained.

4.1.3 Capacity Calculation and Power Control

For any BS $i \in \mathcal{Q}_L$, there are \mathcal{N}^i resource blocks (RBs) for the DL. For a given UE u associated to BS i , n_u^i RBs are allocated where $\mathcal{N}^i = \sum_{u=1}^{|\mathcal{Q}_L^i|} n_u^i$. $p_{s,r}^i, p_{s,r}^b, p_{s,r}^a$ and $p_{s,r}^q$ are transmit power values associated with RB r and the transmit power index s from the LTE-U BS i , LTE-U BS b ($i \neq b$), WiFi AP a and WiFi STA q . The i th BS is considered as the desired BS where the BSs indexed by b are the interference generating BSs. For any AP, UE or STA total transmit power is equally distributed among all RBs. However in every BS, the total transmit power is dynamically changed for every duty cycle according to MAB algorithm. The $h_{u,r}^i, h_{u,r}^b, h_{u,r}^a$ and $h_{u,r}^q$ are the channel gain values from BS i to UE u , from BS b to UE u , from AP a to UE u and from WiFi STA q to UE u , respectively. All channel gain values are calculated considering path losses and shadowing. In that case,



(a) Interference on LTE-U DL and WiFi UL.



(b) Interference on LTE-U UL and WiFi DL.

Figure 4.1: DL and UL interference scenarios for LTE-U/WiFi transmissions.

interference generated to UE u from BSs, APs and STAs are given by I_{BS}^u , I_{AP}^u and I_{STA}^u , respectively. Since a synchronized transmission is considered, there is no interference from the UL transmission of LTE-U UEs. Noise variance is denoted by σ^2 . The Signal-to-Interference-plus-Noise Ratio (SINR) expression for UE u served by BS i on RB r at time interval k is given as

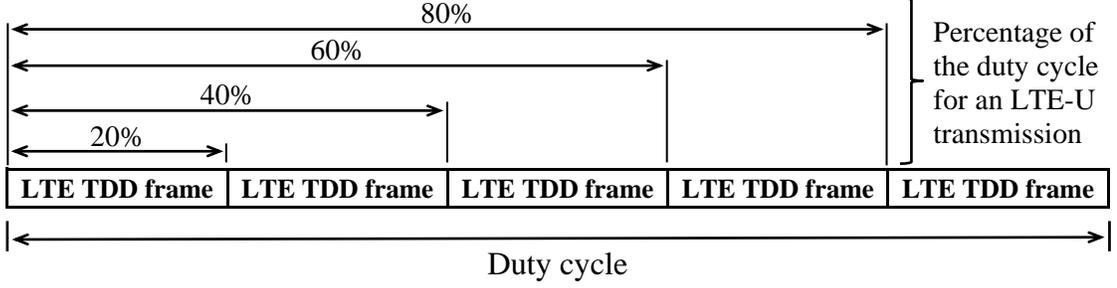


Figure 4.2: Structure of the duty cycle for LTE-U transmissions.

$$\text{SINR}_{u,r}^i [k] = \frac{p_s^i h_{u,r}^i}{\underbrace{\sum_{b \in \mathcal{B}_L \setminus i} p_s^b h_{u,r}^b}_{I_{\text{BS}}^u} + \underbrace{\sum_{a \in \mathcal{B}_W} p_s^a h_{u,r}^a}_{I_{\text{AP}}^u} + \underbrace{\sum_{q \in Q_W} p_s^q h_{u,r}^q}_{I_{\text{STA}}^u} + \sigma^2}, \quad (4.1)$$

where $b, i \in \mathcal{B}_L$.

The amount of successfully transmitted data bits N_B from the i th LTE-U BS during T_{OFDM} time interval k within an active DL subframe/s of a duty cycle is given by

$$N_B^i = \sum_k \sum_{u \in Q_L^i} \sum_r^{R_u} W_{u,r}^i \log_2 (1 + \text{SINR}_{u,r}^i [k]) T_{\text{OFDM}}, \quad (4.2)$$

T_{OFDM} is the orthogonal frequency division multiplexing (OFDM) symbol duration, $T_{\text{Tx}}^i = \mathcal{K}^i T_{\text{OFDM}}$ and \mathcal{K}^i is the total number of transmit T_{OFDM} time intervals for the considered duty cycle. The total allocated bandwidth for RB r for UE u served by BS i is $W_{u,r}^i$. The average capacity over a duty cycle period is used as a performance measure in this study as in [43]. The DL capacity C_i of LTE-U BS i is given as

$$C_i = \frac{N_B^i}{T_{\text{Tx}}^i + T_{\text{Wait}}^i}, \quad (4.3)$$

where T_{Wait}^i is the waiting time due to silent subframe allocation.

The capacity C_i in (4.3) is used as a performance measure for each LTE-U BS. Since the transmit power of one BS contributes to the interference power of the other BS, neighboring BSs are coupled in terms of interference. The goal of every BS is to maximize C_i while minimizing the DL transmit power $p_s^i, \forall i \in \mathcal{B}_L$. By minimizing the transmit power values p_s^i and p_s^b , the goal is to achieve a comparatively higher energy efficiency than the case of constant DL transmit power. In the same time a reduction in interference is also expected while guaranteeing a minimum capacity. Moreover, $P_{\min} \leq p_s^b \leq P_{\max}$ where P_{\min} and P_{\max} are the minimum and maximum transmit power constraints, respectively. The minimum capacity corresponding to a given action is denoted by C_j^{\min} . The objective is to maximize the average capacity while minimizing the transmit power, which can be written as: 1]

$$\text{maximize } \frac{\sum_{i=1}^{|\mathcal{B}_L|} C_i}{|\mathcal{B}_L|} \quad (4.4)$$

$$\text{minimize } p_s^i \quad \forall i \in \mathcal{B}_L \quad (4.5)$$

$$\text{subject to } \{p_s^i, p_s^b\} \leq P_{\max}, \forall i, b \in \mathcal{B}_L, i \neq b, s \in S \quad (4.6)$$

$$\{p_s^i, p_s^b\} \geq P_{\min}, \forall i, b \in \mathcal{B}_L, i \neq b, s \in S \quad (4.7)$$

$$C_i > C_j^{\min}, \forall i \in \mathcal{B}_L, \forall j \in J. \quad (4.8)$$

In the case of energy efficiency, several parameter configurations are considered for (4.8) as

$$\frac{C_i}{p_s^i} > \frac{C_j^{\min}}{p_s^i}, \text{ or } \frac{C_i}{p_s^i} > \frac{C_j^{\min}}{P_{\min}}, \text{ or } \frac{C_i}{p_s^i} > \frac{C_j^{\min}}{P_{\max}}. \quad (4.9)$$

Due to the same denominator, $\frac{C_i}{p_s^i} > \frac{C_j^{\min}}{p_s^i}$ is simplified to (4.8), which can be used as a proportional measure of energy efficiency. The problem is reformulated defining a new objective to maximize energy efficiency as follows:

$$\begin{aligned}
& \text{maximize} && \frac{\sum_{i=1}^{|\mathcal{B}_L|} C_i}{|\mathcal{B}_L|} && (4.10) \\
& \text{subject to} && (4.6), (4.7) \text{ and } (4.9).
\end{aligned}$$

4.2 MAB Techniques for LTE-U WiFi Coexistence

In a MAB problem, an agent selects an action (also known as arm) and observes the corresponding reward. The rewards for given action/arms are random variables with unknown distribution. The goal of MAB is to design action selection strategies to maximize accumulate rewards over a given time horizon. However, the strategies need to achieve a trade-off between exploration (selection of sub-optimal actions to learn their average rewards) and exploitation (selection of actions which have provided maximum rewards so far).

In order to dynamically optimize LTE-U transmission parameters, (i.e., duty cycle and transmit power), a variant of MAB learning techniques, called Thomson sampling [188, 189] algorithm is applied. The scenario is formulated as a multi-agent problem $\mathcal{G} = \left\{ \mathcal{B}_L, \{\mathcal{A}_i\}_{i \in \mathcal{B}_L}, \{C_i\}_{i \in \mathcal{B}_L} \right\}$, considering the BSs as players where \mathcal{A}_i is the action set for player i . During the entire process, each BS needs to strike a balance between exploration and exploitation, where there are M exploration and L exploitation steps, indexed with m , $1 \leq m \leq M$, and l , $1 \leq l \leq L$, respectively.

- **Agents:** LTE-U BSs, \mathcal{B}_L .
- **Action:** The action set of agent i , \mathcal{A}_i is defined as $\mathcal{A}_i = \{d_j^i, p_s^i\}_{j \in J, s \in S}$. $\{d_j^i, p_s^i\}$ is the pair of duty cycle and transmit power elements. Configurations of duty cycles are used as part of the action space \mathcal{D} where \mathcal{D} is common for all players. A given

BS i selects $d_j^i, d_j^i \in \mathcal{D}$ according to the Algorithm 1 where $J = \{1, 2, \dots, |\mathcal{D}|\}$, $j \in J$ and $J \in \mathbb{Z}^+$. Probability spaces of positive integers are denoted by \mathbb{Z}^+ . The set of first elements of the action vector $\mathcal{D}_i = \{d_1^i, d_2^i, \dots, d_{|\mathcal{D}|}^i\}$ of BS i is associated with the duty cycles as $\{20\%, 40\%, \dots, 80\%\}$, respectively. The transmit power values set \mathcal{P} is represented as $S = \{1, 2, \dots, |\mathcal{P}|\}$, $s \in S$ and $S \in \mathbb{Z}^+$. p_s^i is the transmit power of player i where $\mathcal{P}_i = \{p_1^i, p_2^i, \dots, p_{|\mathcal{P}|}^i\}$. For each action \mathcal{A}_i , there a distribution Beta $(\alpha_j^i, \beta_j^i), \forall j \in J$ where α_j^i and β_j^i are the shape parameter. However, in the case of power control (PC), if $C_i > C_j^{\min}$, s is decreased by one ($s \leftarrow s - 1$) reducing the transmit power p_s^i by one level for the next step $m + 1$ and vice versa. Further, when $C_i > C_j^{\min}$ a reward is achieved. As well as, for $C_i > C_j^{\min}$, α_j^i is incremented; otherwise β_j^i is incremented.

- **Decision function:** The DL capacity of a BS i , C_i is used as the utility function. In order to select a duty cycle, a decision function based on the policy UCB1 [190] is used where the accumulated rewards achieved due to values given by C_i are exploited. The decision value for the duty cycle d_j^i related to the exploration step m of BS i , $v_{i,m}(d_j^i)$ is given in (4.11) while d_k^i based on the decision is given in (4.12),

$$v_{i,m}(d_j^i) = \bar{x}_{i,m}(d_j^i) + \sqrt{\frac{2 \ln(m + |\mathcal{D}_i|)}{n_{i,m}(d_j^i)}}, \quad (4.11)$$

$$d_k^i = \arg \max_{d_j^i \in \mathcal{D}_i} (v_{i,m}(d_j^i)), \quad (4.12)$$

where $\bar{x}_{i,m}(d_j^i) = \frac{R_i(d_j^i)}{n_{i,m}(d_j^i)}$. The argument of the maximum value is given by $\arg \max(\cdot)$. $\bar{x}_{i,m}(d_j^i)$, $R_i(d_j^i)$ and $n_{i,m}(d_j^i)$ are the average reward obtained from d_j^i during the exploration step m , total rewards gained from the same d_j^i and the total number of times d_j^i has been played, respectively. Selection of s is totally independent from the decision function.

Algorithm 1 Multi Arm Bandit (Thomson Sampling)

```

1: Initialization:
2: Set the minimum capacity values  $C_j^{\min}, \forall j \in J$ , Exploration steps  $M$ , Beta  $(1, 1)$ ,  $\alpha_j^i$  and  $\beta_j^i$ 
   where  $\forall j : j \in J$ . Select  $d_j^i, \forall j \in J$ , update  $s, n_{i,0}(d_j^i), v_{i,0}(d_j^i)$  and accumulated hypothesis / reward  $R_i(d_j^i)$  based on  $C_i > C_j^{\min}$ 
3: if  $\alpha_j^i(m) = \beta_j^i(m), \forall (l, m) \in M$  then
4:   Exploration:
5:   for  $m = 1, 2, 3, \dots, M$  do
6:     Select  $d_j^i, d_j^i \in \mathcal{D}_i, j \in \{\mathcal{U}(1, |\mathcal{D}_i|) \cap J\}$  and update  $s$ , (4.8)
7:     Execute  $\{d_j^i, p_s^i\}$ , observe  $C_i$  and update  $n_{i,m}(d_j^i)$ 
8:     if  $C_i > C_j^{\min}$  then
9:       Reward,  $R_i(d_j^i) = R_i(d_j^i) + 1$ 
10:      Update  $s$  ( $s \leftarrow s - 1$ ) and  $v_{i,m}(d_j^i)$ , (4.11)
11:      Update  $\alpha_j^i(m) = \alpha_j^i(m) + 1$ 
12:     else
13:       Reward,  $R_i(d_j^i) = R_i(d_j^i) + 0$ 
14:       Update  $s$  ( $s \leftarrow s + 1$ ) and  $v_{i,m}(d_j^i)$ , (4.11)
15:       Update  $\beta_j^i(m) = \beta_j^i(m) + 1$ 
16:     end if
17:     if  $R_i(d_j^i) = R_i(d_a^i), d_j^i, d_a^i \in \mathcal{D}_i, j \neq a, \forall j, a \in J$  then
18:       Select  $d_k^i, d_k^i \in \mathcal{D}_i, k \in \{\mathcal{U}(1, |\mathcal{D}_i|) \cap J\}$ 
19:     else
20:       Select  $d_k^i$ , (4.12)
21:     end if
22:     Exploitation:
23:     for  $l = 1, 2, 3, \dots, L$  do
24:       Execute the action  $\mathcal{A}_i = \{d_k^i, p_s^i\}$ 
25:     end for
26:   end for
27: end if

```

The multi-agent learning problem is addressed using a MAB approach. In the contextual MAB problem handled by the Thomson sampling algorithm [188], current and previous information (i.e., history) are used for the selection of an arm or (i.e., action). Initially $d_j^i, \forall j \in J$ are played once with $p_s^i = p_{|\mathcal{P}|}^i$. Based on the accumulated reward $R_i(d_j^i)$, the parameters $s, n_{i,0}(d_j^i)$ and $v_{i,0}(d_j^i)$ are updated. In the learning process, the accumulated reward is used to play the role of the accumulated hypothesis defined in [191]. Subse-

quently, agents balance between M exploration and L exploitations steps. During the exploration steps, d_j^i is selected randomly where $d_j^i, d_j^i \in \mathcal{D}_i, j \in \{\mathcal{U}(1, |\mathcal{D}_i|) \cap J\}$ where a uniform distribution with the minimum and maximum values x_1 and x_2 are given by $\mathcal{U}(x_1, x_2)$. s is decided based on the last available values of (4.8). Subsequently the same set of parameters is updated. At the end of each exploration step, based on (4.8) and the accumulated rewards an action is selected. Then the same action is repeatedly played for all the L exploitation steps of that particular exploration step as explained in Algorithm 1.

4.3 Simulation Results

For LTE-U, TDD-LTE is considered and it is assumed that all LTE-U UEs are synchronized in both time and frequency domain as in [43] with the serving BSs. A beacon is transmitted periodically for the purpose of synchronization of WiFi STAs with the corresponding APs. To evaluate the performance, an architecture containing two independently operated layers of cellular deployments is considered as shown in Fig. 4.3. Hexagonal cells with omni directional antennas are assumed. LTE-U layer encompasses $|\mathcal{B}_L| = 7$ BSs and $|\mathcal{Q}_L|$ UEs where the WiFi layer includes $|\mathcal{B}_W| = 7$ APs and $|\mathcal{Q}_W|$ WiFi STAs. In each cell, for each AP/BS, STAs/UEs are dropped at random locations. All of them are assumed to be uniformly distributed within the cells of their serving BSs having a mobility speed of 3 km/h and a random walk mobility model. We consider a non-full buffer traffic for both WiFi and LTE networks, where the packet arrivals at the transmitter queues follow a Poisson distribution. The traffic arrival rates for LTE-U and WiFi are $\lambda_{\text{LTE}} = \lambda_{\text{WiFi}} = 2.5$ packet/second.

The LTE and WiFi IEEE 802.11n medium access control (MAC) and physical (PHY) layers are modeled in which a PHY layer abstraction is used for Shannon capacity calculations of WiFi and LTE-U. The time granularity of each WiFi OFDM symbol duration is

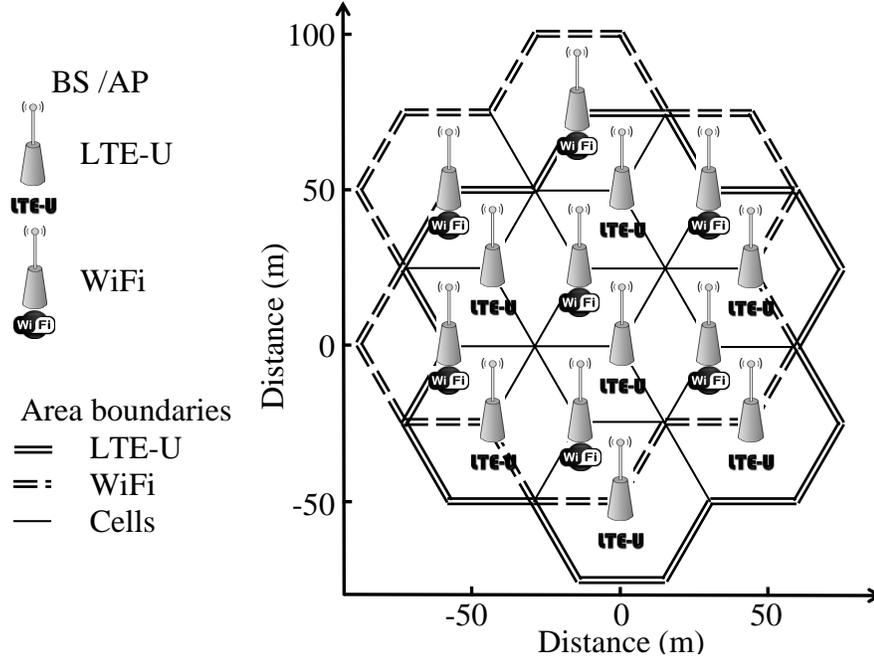


Figure 4.3: Cellular coverage layout used in LTE-U and WiFi coexistence simulations.

4 μs , which we use to periodically capture the number of successfully received bits [43]. For both technologies wireless channel is modeled according to [192], when the systems are operated in the 3.5 GHz band. Indoor Hotspot (InH) scenario is considered with path loss and shadowing parameters. FTP Traffic Model-2 [192] is employed for either WiFi and LTE-U with a noise spectral power density of -95 dBm/Hz.

In each transmission time interval (TTI), DL SINR is reported to the corresponding BS. Based on the number of LTE-U UEs waiting and requesting UL transmission during one sub frame, bandwidth is equally shared among themselves. The simulation parameters for LTE-U transmissions are summarized in TABLE 4.1. TDD configuration 1 [186, Fig. 6.2(a)] is used for the LTE-U frames having a 50 ms total duty cycle period. Minimum required capacity level C_j^{\min} is 10 Mbps and the set of power levels is $\mathcal{P}_i = \{p_1^i, p_2^i, \dots, p_{|\mathcal{P}_i}^i\} = \{8, 13, 18, 23\}$ dBm.

For WiFi, CSMA/CA with enhanced distributed channel access (EDCA) and clear channel assessment (CCA) has been implemented. All WiFi STAs with traffic in their

Table 4.1: LTE MAC/PHY PARAMETERS.

Parameter	Value
Frequency	3.5 GHz
Transmission Scheme	OFDM
Bandwidth	20 MHz
DL Tx Power	23 dBm
UL Tx Power	PL Based TPC
Frame Duration	10 ms
Scheduling	Round Robin
UL Base Power Level P_0	-106 dBm
TTI	1 ms

queue will compete for channel access after receiving a beacon transmission. Without reception of a signal beacon, transmission or reception will not be initiated. The WiFi STA will sense the channel and will transmit if it is idle. Otherwise transmission will be backed off and the next transmission will be initiated after a backoff time. Random backoff time mechanism is used for this study. All the parameters for the WiFi transmission are summarized in TABLE 4.2.

Table 4.2: WiFi MAC/PHY PARAMETERS.

Parameter	Value
Frequency	3.5 GHz
Transmission Scheme	OFDM
Bandwidth	20 MHz
DL/UL Tx Power	23 dBm
Access category	Best Effort
MAC protocol	EDCA
CCA Channel sensing threshold	-82 dBm
CCA Energy detection threshold	-62 dBm
No of service bits in PPDU	16 bits
No of tail bits in PPDU	12 bits
Backoff type	Fixed contention window
Contention window size	$\mathcal{U}(0, 31)$
Noise figure	6 [186]
Beacon interval	100 ms
Beacon OFDM symbol detection threshold	10 dB
Beacon error ratio threshold	15

4.3.1 Aggregate Capacity with MAB

Compared to fixed DCP scenarios, with the use of MAB algorithm, overall capacities are increased. Performance of Epoch-Greedy algorithm with and without PC is presented in the same figure. No significant influence could be observed for LTE-U throughputs due to PC. Even though it is not quantified, both interference caused by LTE-U BSs and the energy consumption at the BSs are reduced due to PC mechanism. When there is a PC mechanism, a degradation in performance particularly in the WiFi activities also can be seen. Capacity reduction of WiFi is caused by the behavior of LTE-U transmissions and CSMA/CS technique. As given by Fig. 4.4, in all the cases a severe performance degradation of LTE-U could be seen in the presence of WiFi transmissions, when compared to the case with only LTE-U transmission. In other words, a very high sensitivity for interference is shown by LTE-U. The same behavior is shown by WiFi as well where there is a considerable throughput reduction compared to the scenario with no LTE-U transmission.

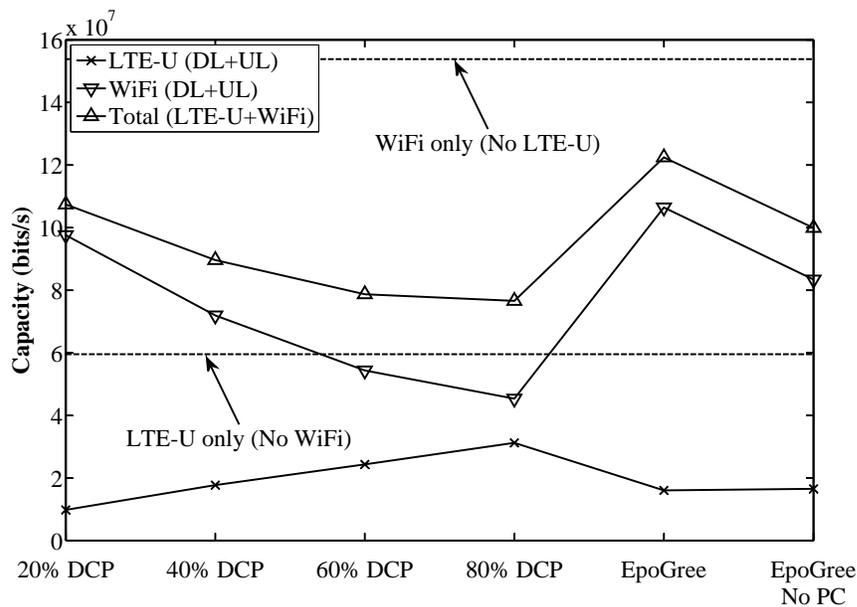


Figure 4.4: LTE-U and WiFi UL/DL capacities.

The aggregate capacity performance of LTE-U UEs and WiFi STAs under different densities is given in Fig. 4.5. There is a capacity improvement with reductions of users in both services. Comparatively high sensitivity could be seen when the density of STAs are changed. When the densities are reduced, particularly the STAs, a significant increase in capacity is achieved under reduced interference conditions. However this reduction is further contributed by the CSMA/CS mechanism as well.

Aggregate capacity of stand-alone WiFi, coexisting LTE-U (80% duty cycle) and WiFi (with no MAB algorithm), and MAB based coexistence of LTE-U and WiFi are presented in Fig. 4.6. The aggregate number of WiFi APs and LTE BSs in all scenarios are kept constant. For the WiFi only deployment, we replace all the LTE BSs in Fig. 4.3 with WiFi APs. It is notable that with the use of MAB, the overall capacity is increased significantly from stand-alone WiFi operation and simultaneous operation of LTE-U and WiFi (without MAB). Also we found that with the increase of inter site distance (ISD) in Fig. 4.3, the capacity decreases. This is because of higher serving area per APs/STA within the ISDs.

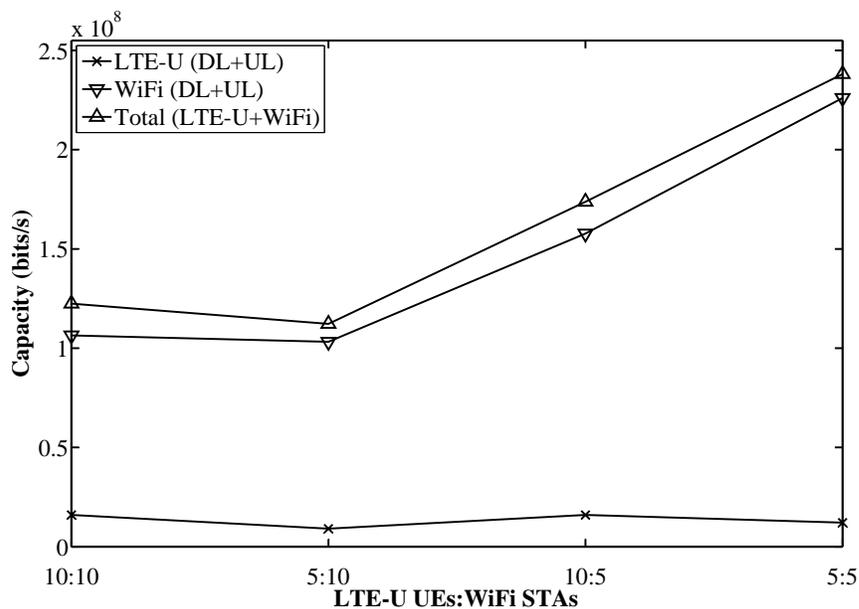


Figure 4.5: Comparison of different UE and STA configurations.

The WiFi throughput performance with and without MAB algorithm is shown in Fig. 4.7, where it is noted that MAB algorithm improves the WiFi throughput over the two other scenarios. Moreover, with the increase of ISD, capacity degrades for all cases. The effect of LTE packet arrival rate on aggregate capacity is shown in Fig. 4.8. We found that the aggregate throughput of coexisting LTE and WiFi networks is maximized for $\lambda_L = 2.5$, but then it decreases for larger values of λ_L due to increased interference levels. Also for full buffer LTE traffic ($\lambda_L = 0$), the coexisting system with MAB has degraded performance compared to coexisting system without MAB.

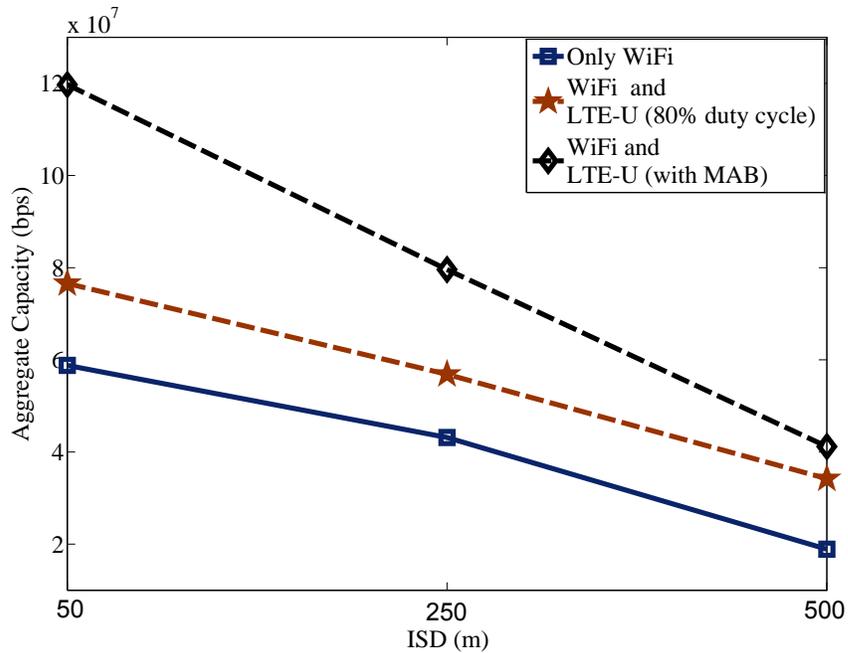


Figure 4.6: Aggregate capacity of coexisting WiFi and LTE-U (80% duty cycle), MAB based coexisting LTE-U and WiFi, and stand-alone WiFi system for different ISDs.

Impact of energy detection threshold on aggregate capacity is shown in Fig. 4.9. It is observed that -62 dBm threshold provides best performance for all scenarios. Sensing threshold less than -62 dBm makes WiFi to back off from transmission in the presence of LTE transmission, and results in lower aggregate capacity. On the other hand, sensing

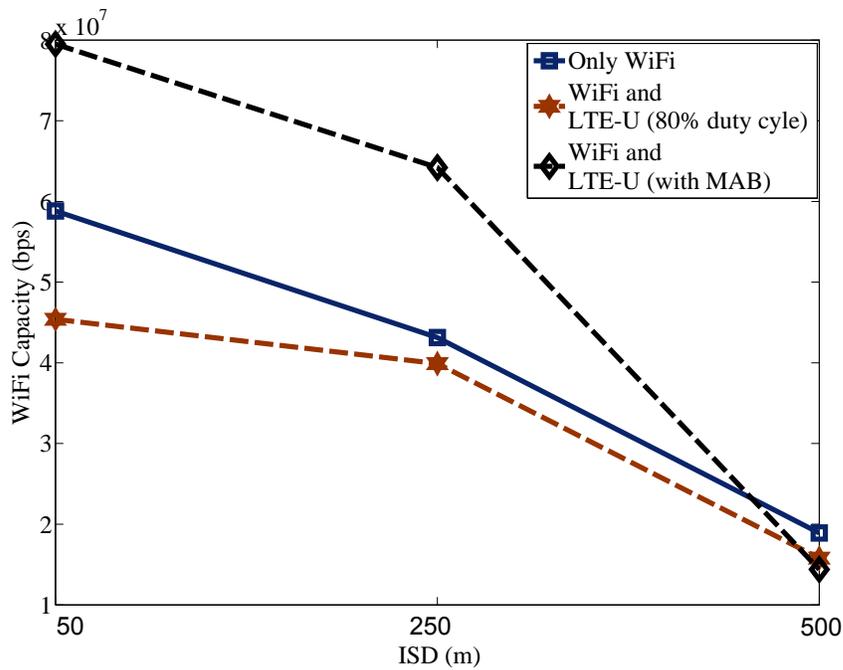


Figure 4.7: WiFi capacity of coexisting WiFi and LTE-U (80% duty cycle), MAB based coexisting LTE-U and WiFi, and stand-alone WiFi system for different ISDs.

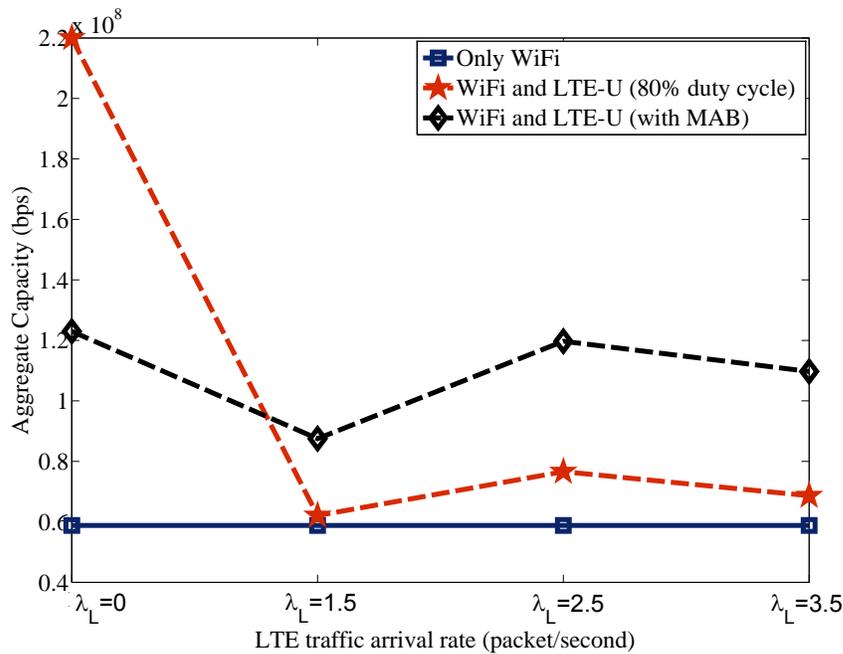


Figure 4.8: Aggregate capacity of coexisting WiFi and LTE-U (80% duty cycle), MAB based coexisting LTE-U and WiFi, and stand-alone WiFi system for different LTE traffic arrival rates.

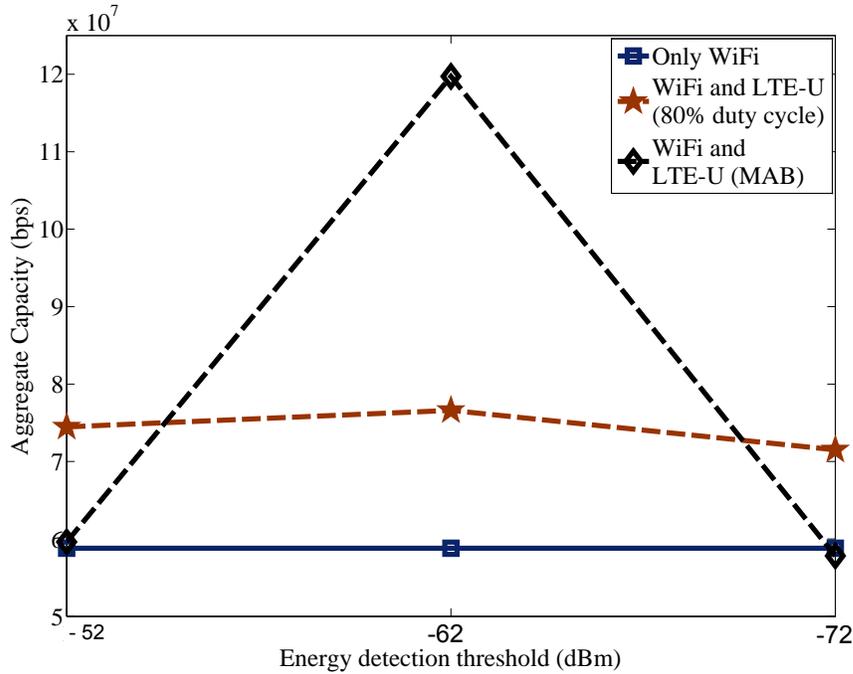


Figure 4.9: Aggregate capacity of coexisting system of WiFi and LTE-U (80% duty cycle), MAB based coexisting LTE-U and WiFi, and stand-alone WiFi system for various energy detection threshold.

threshold more than -62 dBm allows WiFi to transmit in the presence of LTE operation, which reduces aggregate capacity due to higher interference.

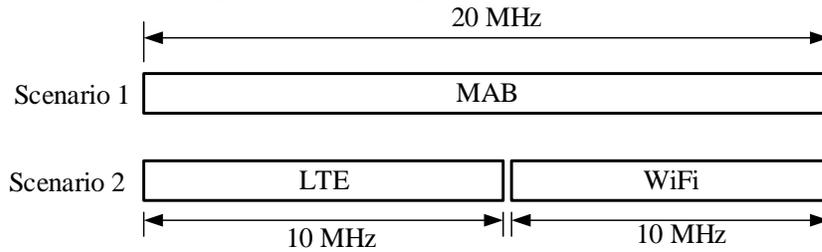


Figure 4.10: Scenario with two cases.

For Fig. 4.11, we consider a scenario with two cases as described in Fig. 4.10. In scenario 1, we consider simultaneous operation of LTE-U and WiFi using MAB on 20 MHz bandwidth. On the other hand, in scenario 2, stand-alone LTE (i.e. 100% duty cycle) and WiFi are operating on separate 10 MHz bandwidth. We find that the overall capacity using MAB is improved significantly when compared with the aggregate capacity of two

stand-alone systems. This reflects how the spectral efficiency can be improved using MAB, and motivates sharing of wireless spectrum among LTE and WiFi networks, rather than deploying them separately.

The impact of LTE-U UEs and WiFi STAs density on aggregate capacity is given in Fig. 4.12. We find that the aggregate capacity improves for the reductions of users in both services. Comparatively high sensitivity could be seen when the density of STAs are changed. When the densities are reduced, particularly the STAs, a significant increase in capacity is achieved under reduced interference conditions. However this reduction is further contributed by the CSMA/CA mechanism as well. Also it is notable that capacity decreases with the increase of ISD.

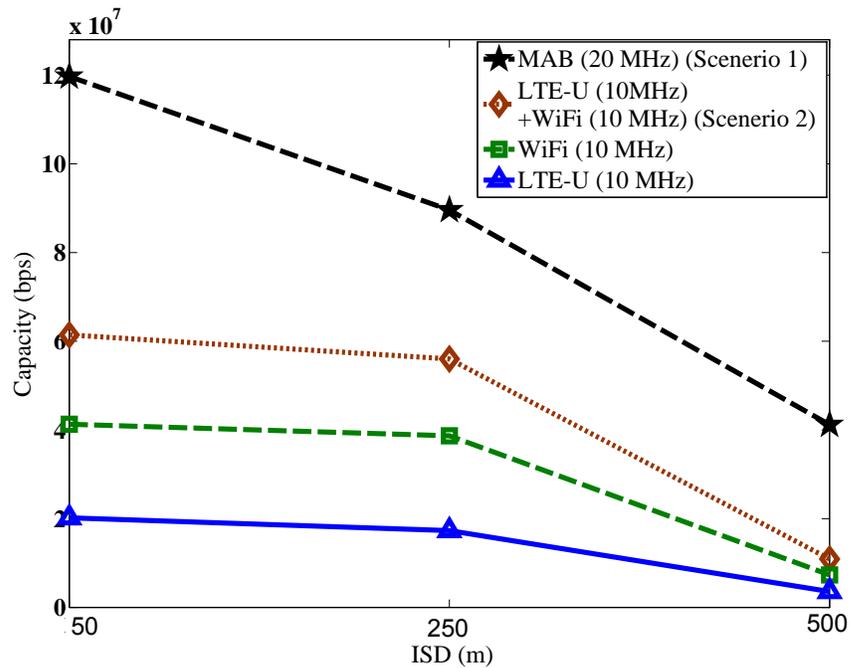


Figure 4.11: Capacity of 10 STAs or/and 10 UEs under stand-alone WiFi, stand-alone LTE, coexisting stand-alone WiFi and LTE-U (scenario 1) and MAB based coexisting LTE-U and WiFi (scenario 2) for different bandwidths and ISDs.

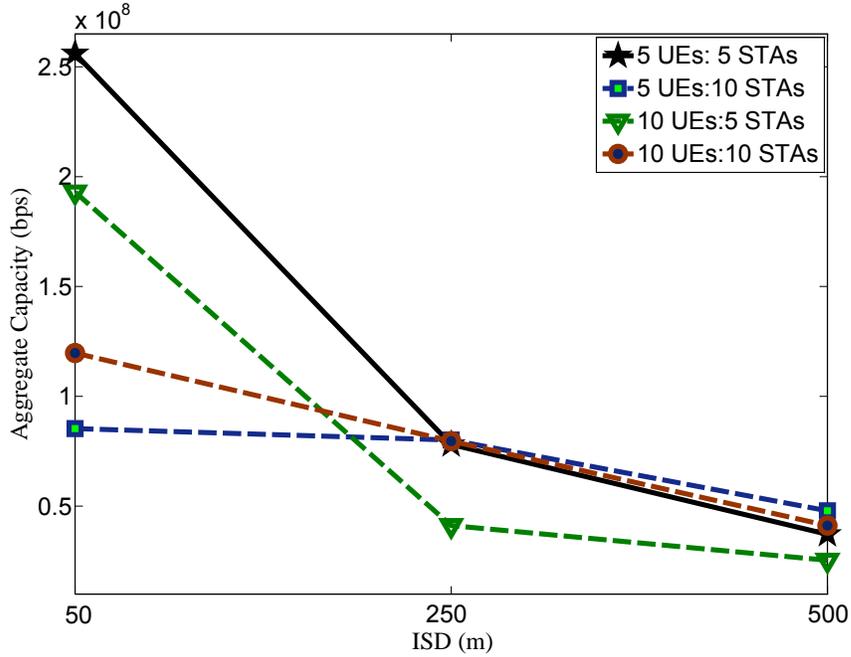


Figure 4.12: Capacity of MAB based coexistence for different UEs and STAs ratios and ISDs.

4.3.2 Cell Edge Performance

In Fig. 4.13, 5th percentile LTE throughput for different user densities of STAs is represented. We found that with the increase of STAs, 5th percentile UE throughput reduces due to more interference caused by STAs. However, with the increment of UEs, the effect of STA density reduces. This means for higher density of UEs and STAs, fewer LTE users will experience higher capacity.

4.3.3 Energy Efficiency Performance

Aggregate capacity of $|Q_L^i| = 10$ and $|Q_W^w| = 10$ is presented in Fig. 4.14 for different power control techniques. Four parameter settings are used for PC. In the first instance, no PC is considered. In the second case, PC is used by replacing the parameters in Step 7 of the Algorithm I with $\frac{C_i}{p_s} > \frac{C_j^{\min}}{P_{\min}}$ where $P_{\min} = 8$ dBm. For the third and fourth cases,

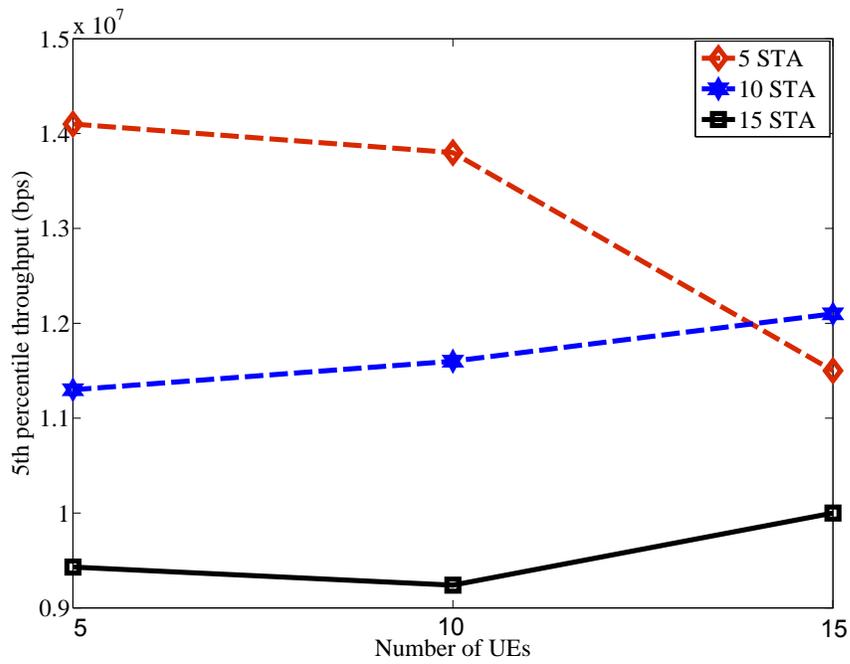


Figure 4.13: 5th percentile throughput of MAB based coexisting LTE-U and WiFi for different UEs and STAs ratios.

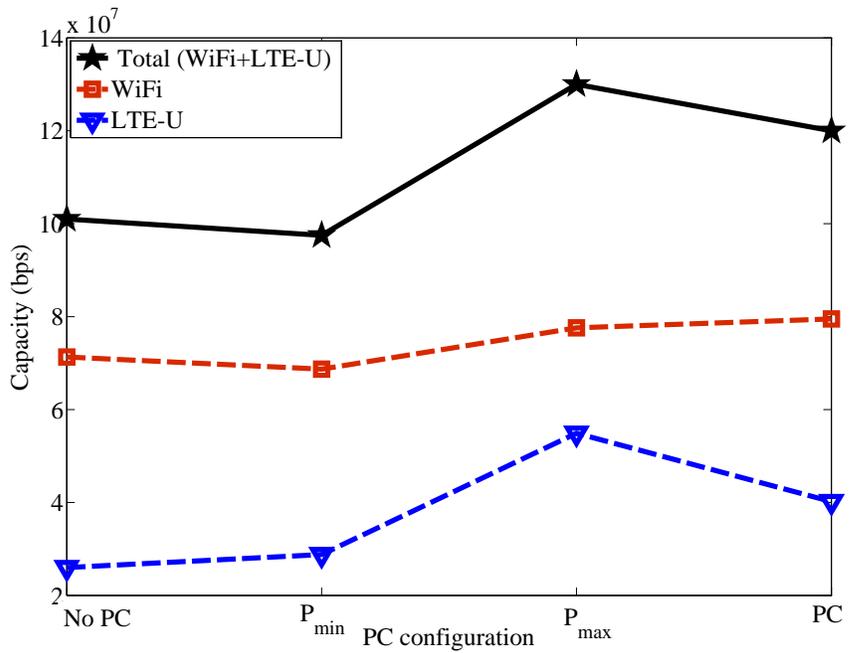


Figure 4.14: Capacity of 10 UEs and 10 STAs under different PC configurations.

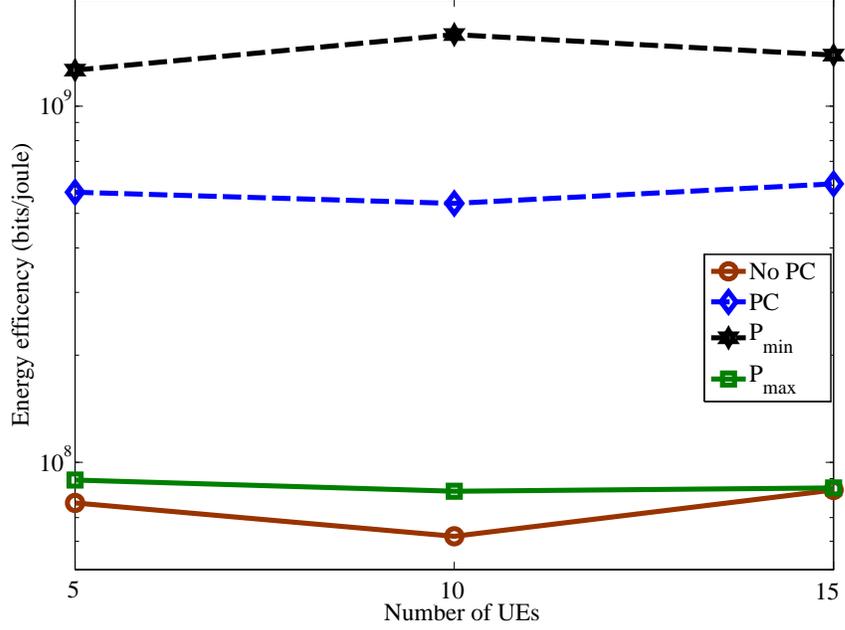


Figure 4.15: Energy efficiency under different PC configurations for various number of UEs (with 10 STAs).

parameters are replaced with $\frac{C_i}{p_s^i} > \frac{C_j^{\min}}{P_{\max}}$ and $C_i > C_j^{\min}$ where $P_{\max} = 23$ dBm. The set of power levels is defined as $\mathcal{P}_i = \{p_1^i, p_2^i, \dots, p_{|\mathcal{P}|}^i\} = \{8, 11, 14, 17, 20, 23\}$ dBm where $P_{\min} = 8$ dBm and $P_{\max} = 23$ dBm. So, in the second and third cases a given level of energy efficiency is aimed at. In the last case, according to the explanation given for (4.9), the level is dynamically adjusted. It is noted that the best and worst performances are found for P_{\max} and P_{\min} . For MAB with PC, optimum result is found.

In Fig. 4.15, different number of UEs are considered to evaluate energy efficiency performance. For all the densities, the least efficiency is achieved with no PC. In the most dense scenario, the best efficiency can be observed under the second configuration, $\frac{C_j^{\min}}{P_{\min}}$ [see (9)]. As it is expected with the reduction of densities, energy efficiency is increased. However after a certain average energy efficiency level, no significant improvements could be observed.

4.4 Conclusion

In this chapter, a MAB based dynamic duty cycle selection method was proposed to facilitate spectrum sharing between WiFi and LTE-U in the same unlicensed band. Performance of the proposed algorithm was further enhanced by using a DL PC technique. Subsequently, the proposed concept was extended to optimize energy efficiency. Considerable gains in overall throughputs could be achieved via the proposed MAB while ensuring a minimum capacity for LTE-U based services in the same band. Significant gains in terms of energy efficiency could be achieved where it is observed that the gains under different parameter settings with PC is much higher than those of with no PC. Our future work includes extending our framework to scenarios with IA and PAL users in the same spectrum.

CHAPTER 5

Low Latency Towards 5G: RAN, Core Network and Caching Solutions

The focus of next generation mobile communication is to provide seamless communication for machines and devices building the Internet-of-Things (IoT) along with personal communication [23]. New applications such as tactile Internet¹, high-resolution video streaming, tele-medicine, tele-surgery, smart transportation, and real-time control dictate new specifications for throughput, reliability, end-to-end (E2E) latency, and network robustness [6]. Additionally, intermittent or always-on type connectivity is required for machine-type communication (MTC) serving diverse applications including sensing and monitoring, autonomous cars, smart homes, moving robots and manufacturing industries.

Several emerging technologies including wearable devices, virtual/augmented reality, and full immersive experience (3D) are shaping the demeanor of human end users, and they have special requirements for user satisfaction. Therefore, these use cases of the next generation network push the specifications of 5G in multiple aspects such as data rate, latency, reliability, device/network energy efficiency, traffic volume density, mobility, and connection density. Current fourth generation (4G) networks are not capable of fulfilling all the technical requirements for these services.

Fifth generation (5G) cellular network is the wireless access solution to fulfill the wireless broadband communication specifications of 2020 and beyond [194, 195]. In ITU, 5G ITU-R working group is working for development of 5G under the term IMT-2020 [196]. The vision of this work is to achieve one thousand times throughput improvement, 100 billion connections, and close to zero latency [6, 194]. In particular, 5G will support enhanced mobile broadband (MBB) with end-user data rates of 100 Mbps in the uniform spatial distribution with peak data rates of 10-20 Gbps [194, 195]. Based on con-

¹A network or network of networks for remotely accessing, perceiving, manipulating or controlling real or virtual objects or processes in perceived real time by humans or machines [193].

sensus, 5G will not only provide personal mobile service, but also massive machine type communications (MTC), and latency/reliability critical services. In mission critical communication (MCC)/ultra reliable low latency communication (uRLLC²), both the *latency* and reliability issues need to be addressed [197–199]. In many cases, the corresponding E2E latency as low as 1 ms needs to be met with reliability as high as 99.99% [62].

To achieve low *latency* for MCC, drastic changes in the network architecture need to be performed. Since the delay is contributed by radio access network (RAN) and core network along with backhaul between RAN and core network, new network topology involving software define network (SDN), network virtualized function (NFV), and mobile edge computing (MEC)/caching can be employed to reduce the latency significantly. This can happen due to the capability of seamless operation and independence from hardware functionality provided by these entities. Moreover, new physical air interface with small time interval transmission, small size packets, new waveforms, new modulation and coding schemes are the areas of investigation for attaining low latency. In addition, optimization of radio resource allocation, massive MIMO, carrier aggregation in millimeter wave, and priority of data transmission need to be addressed. All in all, a robust integration with existing LTE is necessary for 5G networks that will enable industries to deploy 5G quickly and efficiently when it is standardized and available. In summary, 5G wireless access should be an evolution of LTE complemented with revolutionary architecture designs and radio technologies.

In this chapter, we present a comprehensive survey of latency reduction solutions particularly in the context of 5G wireless technology [23]. We first present the sources and fundamental constraints for achieving *low latency* in a cellular network. We also overview an exemplary 5G network architecture with compliance to IMT-2020 vision.

²uRLLC allows E2E latency of less than 1 ms on all layers with packet error rates of 10^{-5} to 10^{-9} .

Finally, we provide an extensive review of proposed solutions for achieving low latency towards 5G. The goal of our study is to bring all existing solutions on the same page along with future research directions. We divide the existing solutions into three parts: (1) RAN solutions; (2) Core network solutions; (3) Caching solutions. However, detailed comparison of these solutions are beyond the scope of this work.

5.1 Low Latency Services in 5G

Latency is highly critical in some applications such as automated industrial production, control/robotics, transportation, health-care, entertainment, virtual reality, education, and culture. In particular, IoT is quickly becoming a reality which connects anything to any other thing anytime, and anywhere. Smart wearable devices (smart watches, glasses, bracelets, and fit bit), smart home appliances (smart meters, fridges, televisions, thermostat), sensors, autonomous cars, cognitive mobile devices (drones, robots, etc.) are connected to always-on hyper-connected world to enhance our life style [31, 200, 201]. Even though operators are supporting these IoT applications through existing 3G/LTE, some applications require much more stringent requirements from underlying networks such as low latency, high reliability [202, 203], and security [204, 205]. In some cases, we need *latency* as low as 1 ms with packet loss rate no larger than 10^{-2} . Several latency critical services which need to be supported by 5G are described as follows.

- **Factory Automation:** Factory automation includes real-time control of machine and system for quick production lines and limited human involvement. In these cases, the production lines might be numerous and contiguous. This is highly challenging in terms of latency and reliability. Therefore, the E2E *latency* requirement for factory automation applications is between 0.25 ms to 10 ms with a packet loss rate of 10^{-9} [200, 206]. In factory automation, the latency is measured as E2E in which the sensors measuring data are at one end and transmit the data for processing

to the other end for programmable logic controller (PLC). The proposed values for the latency are based on the KoI (Koordinierte Industriekommunikation) project, in which a detailed questionnaire-based survey is conducted to collect the information from an extensive range of factory automation processes [207].

- **Intelligent Transportation Systems:** Autonomous driving and optimization of road traffic requires ultra reliable low latency communication. According to intelligent transportation systems (ITS), different cases including autonomous driving, road safety, and traffic efficiency services have different requirements [200, 208]. Autonomous vehicles require coordination among themselves for actions such as platooning and overtaking [209]. For automated vehicle overtaking, maximum E2E latency of 10 ms is allowed for each message exchange. For video integrated applications such as *see-through-vehicle* application described in [210] requires to transmit raw video which allows maximum delay of 50 ms [211]. Road safety includes warnings about collisions or dangerous situations. Traffic efficiency services control traffic flow using the information of the status of traffic lights and local traffic situations. For these purposes, *latency* of 10 ms to 100 ms with packet loss rate of 10^{-3} to 10^{-5} is required.
- **Robotics and Telepresence:** In the near future, remote controlled robots will have applications in diverse sectors such as construction and maintenance in dangerous areas. A prerequisite for the utilization of robots and telepresence applications is remote-control with real-time synchronous visual-haptic feedback. In this case, system response times should be less than a few milliseconds including network delays [200, 212, 213]. Communication infrastructure capable of proving this level of real-time capacity, high reliability/availability, and mobility support is to be addressed in 5G networks.

Table 5.1: TYPICAL LATENCY AND DATA RATE REQUIREMENTS FOR DIFFERENT MISSION CRITICAL SERVICES.

Use case	Latency	Data rate	Remarks
Factory Automation	0.25-10 ms [200]	1 Mbps [214]	<ul style="list-style-type: none"> – Generally factory automation applications require small data rates for motion and remote control. – Applications such as machine tools operation may allow latency as low as 0.25 ms.
Intelligent Transport Systems (ITS)	10-100 ms [200]	10-700 Mbps [215]	<ul style="list-style-type: none"> – Road safety of ITS requires latency on the order of 10 ms. – Applications such as virtual mirrors require data rates on the order of 700 Mbps.
Robotics and Telepresence	1 ms [216]	100 Mbps [217]	<ul style="list-style-type: none"> – Touching an object by a palm may require latency down to 1 ms. – VR haptic feedback requires data rates on the order of 100 Mbps.
Virtual Reality (VR)	1 ms [212]	1 Gbps [217]	<ul style="list-style-type: none"> – Hi-resolution 360° VR requires high rates on the order of 1 Gbps while allowing latency of 1 ms.
Health care	1-10 ms [213]	100 Mbps [217]	<ul style="list-style-type: none"> – Tele-diagnosis, tele-surgery and tele-rehabilitation may require latency on the order of 1 ms with data rate of 100 Mbps.
Serious Gaming	1 ms [212]	1 Gbps [217]	<ul style="list-style-type: none"> – Immersive entertainment and humans interaction with the high-quality visualization may require latency of 1 ms and data rates of 1 Gbps for high performance.

Table 5.1: TYPICAL LATENCY AND DATA RATE REQUIREMENTS FOR DIFFERENT MISSION CRITICAL SERVICES.

Use case	Latency	Data rate	Remarks
Smart Grid	1-20 ms [200, 212]	10-1500 Kbps [218]	<ul style="list-style-type: none"> – Dynamic activation and deactivation in smart grid requires latency on the order of 1 ms. – Cases such as wide area situational awareness require data rates on the order of 1500 Kbps.
Education and Culture	5-10 ms [212]	1 Gbps [217]	<ul style="list-style-type: none"> – Tactile Internet enabled multi modal human-machine interface may require latency as low as 5 ms. – Hi-resolution 360° and haptic VR may require data rates as high as 1 Gbps.

- **Virtual Reality (VR):** Several applications such as micro-assembly and tele-surgery require very high levels of sensitivity and precision for object manipulations. VR technology accommodates such services where several users interact via physically coupled VR simulations in a shared haptic environment. Current networked communication does not allow sufficient low latency for stable, seamless coordination of users. Typical update rates of display for haptic information and physical simulation are in the order of 1000 Hz which allows round trip *latency* of 1 ms. Consistent local view of VR can be maintained for all users if and only if the latency of around 1 ms is achieved [212, 213, 219].
- **Augmented Reality (AR):** In AR technology, the augmentation of information into the user’s field of view enables applications such as driver-assistance systems, improved maintenance, city/museum guides, telemedicine, remote education, and as-

sistive technologies for police and firefighters [212]. However, insufficient computational capability of mobile devices and latency of current cellular network hinder the applications. In this case, *latency* as low as a few milliseconds is required.

- **Health care:** Tele-diagnosis, tele-surgery and tele-rehabilitation are a few notable healthcare applications of low latency tactile Internet. These allow for remote physical examination even by palpation, remote surgery by robots, and checking of patients' status remotely. For these purposes, sophisticated control approaches with round trip latency of 1-10 ms and high reliability data transmission is mandatory [212,213].
- **Serious Gaming:** The purpose of serious gaming is not limited to entertainment. Such games include problem-solving challenges, and goal-oriented motivation which can have applications in different areas such as education, training, simulation, and health. Network *latency* of more than 30-50 ms results in a significant degrade in game-quality and game experience ratings. Ideally, a round trip time (RTT) on the order of 1 ms is recommended for perceivable human's interaction with the high-quality visualization [212].
- **Smart Grid:** The smart grid has strict requirements of reliability and latency [30, 220–222]. The dynamic control allows only 100 ms of E2E *latency* for switching suppliers (PV, windmill, etc.) on or off. However, in case of a synchronous co-phasing of power suppliers (i.e. generators), an E2E delay of not more than 1 ms is needed [194,212]. Latency more than 1 ms which is equivalent to a phase shift of about 18° (50 Hertz AC network) or 21.6° (60 Hertz AC network), may have serious consequence in smart grid and devices.
- **Education and Culture:** Low latency tactile Internet will facilitate remote learning/education by haptic overlay of teacher and students. For these identical multi-

modal human-machine interfaces, round trip latency of 5-10 ms is allowed for perceivable visual, auditory, and haptic interaction [212, 213]. Besides that, tactile Internet will allow to play musical instruments from remote locations. In such scenarios, supporting network latency lower than few milliseconds becomes crucial [212].

Based on the applications and use case scenarios above, latency critical services in 5G networks demand an E2E delay of 1 ms to 100 ms. The latency requirements along with estimated data rates for various 5G services are summarized in Table 5.1. Some use cases such as VR and online gaming may require round trip latency on the order of 1 ms with data rates as high as 1 Gbps. On the other hand, use cases such as factory automation and smart grid require lower data rates on order of 1 Mbps with demanding latency of 1 ms. For required data rates on the order of 1 Gbps, [217] reports that bandwidth of 40 MHz is sufficient at 20 node density per square kilometer. For data rates of few Mbps, bandwidth of 20 MHz and lower can be sufficient for most scenarios. This means spectral efficiency supported by 5G is 50 bps/Hz while LTE-A can support upto 30 bps/Hz [223]. For lower bandwidth, spectrum below 6 GHz can be utilized while for high bandwidth requirement, mmWave can be an attractive choice [217].

In the next section, the major sources of latency in a cellular network are discussed.

5.2 Sources of Latency in a Cellular Network

In the LTE system, the *latency* can be divided into two major parts: (1) user plane (U-plane) latency and (2) control plane (C-plane) latency. The U-plane latency is measured by one directional transmit time of a packet to become available in the IP layer between evolved UMTS terrestrial radio access network (E-UTRAN) edge/UE and UE/E-UTRAN node [224]. On the other hand, C-plane latency can be defined as the transition time of a

UE to switch from idle state to active state. At the idle state, an UE is not connected with radio resource control (RRC). After the RRC connection is being setup, the UE switches from idle state into connected state and then enters into active state after moving into dedicated mode. Since the application performance is dependent mainly on the U-plane latency, U-plane is the main focus of interest for low latency communication.

In the U-plane, the delay of a packet transmission in a cellular network can be contributed by the RAN, backhaul, core network, and data center/Internet. As referred in Fig. 5.1, the total one way transmission time [109] of current LTE system can be written as

$$T = T_{\text{Radio}} + T_{\text{Backhaul}} + T_{\text{Core}} + T_{\text{Transport}} \quad (5.1)$$

where

- T_{Radio} is the packet transmission time between eNB and UEs and is mainly due to physical layer communication. It is contributed by eNBs, UEs and environment. It consists of time to transmit, processing time at eNB/UE, retransmissions, and propagation delay. Processing delay at the eNB involves channel coding, rate matching, scrambling, cyclic redundancy check (CRC) attachment, precoding, modula-

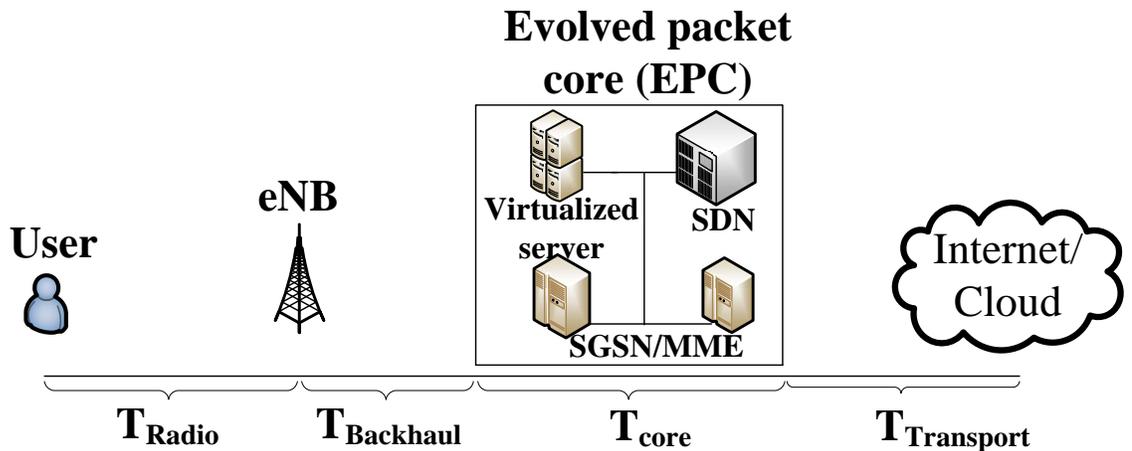


Figure 5.1: Latency contribution in E2E delay of a packet transmission.

tion mapper, layer mapper, resource element mapper, and OFDM signal generation. On the other hand, uplink processing at UE involves CRC attachment, code block segmentation, code block concatenation, channel coding, rate matching, data and control multiplexing, and channel interleaver. Propagation delay depends on obstacles (i.e. building, trees, hills etc.) on the way of propagation and the total distance traveled by the RF signal;

- T_{Backhaul} is the time for building connections between eNB and the core network (i.e. EPC). Generally, the core network and eNB are connected by copper wires or microwave or optical fibers. In general, microwave involves lower latency while optic fibers come with comparatively higher latency. However, spectrum limitation may curb the capacity of microwave [225];
- T_{Core} is the processing time taken by the core network. It is contributed by various core network entities such as mobility management entity (MME), serving GPRS support node (SGSN), and SDN/NFV. The processing steps of core network includes NAS security, EPS bearer control, idle state mobility handling, mobility anchoring, UE IP address allocation, and packet filtering;
- $T_{\text{Transport}}$ is the delay to data communication between the core network and Internet/cloud. Generally, distance between the core network and the server, bandwidth, and communication protocol affect this latency;

The E2E delay, T_{E2E} is then approximately given by $2 \times T$. The T_{Radio} is the sum of transmit time, propagation latency, processing time (channel estimation, encoding and decoding time for first time), and retransmission time (due to packet loss). In particular, the T_{Radio} for a scheduled user [226, 227] can be expressed as:

$$T_{\text{Radio}} = t_{\text{Q}} + t_{\text{FA}} + t_{\text{tx}} + t_{\text{bsp}} + t_{\text{mpt}} \quad (5.2)$$

where

- t_Q is the queuing delay which depends on the number of users that will be multiplexed on same resources;
- t_{FA} is the delay due to frame alignment which depends on the frame structure and duplexing modes (i.e., frequency division duplexing (FDD) and time division duplexing (TDD));
- t_{tx} is the time for transmission processing, and payload transmission which uses at least one TTI depending on radio channel condition, payload size, available resources, transmission errors and retransmission;
- t_{bsp} is the processing delay at the base station;
- t_{mpt} is the processing delay of user terminal. Both the base station and user terminal delay depend on the capabilities of base station and user terminal (i.e., UE), respectively.

In compliance with ITU, T_{Radio} should not be more than 0.5 ms for *low latency* communication [228]. In this regard, radio transmission time should be designed to be on the order of hundreds of microseconds while the current configuration in 4G is 1 ms. For this, enhancement in various areas of RAN such as packet/frame structure, modulation and coding schemes, new waveform designs, transmission techniques, and symbol detection need to be carried out. In order to reduce the delay in $T_{Backhaul}$, approaches such as advanced backhaul techniques, caching/fog enabled networks, and intelligent integration of AS and NAS can provide potential solutions. For T_{Core} , new core network consists of SDN, NFV, and various intelligent approaches can reduce the delay significantly. For $T_{Transport}$, MEC/fog enabled Internet/cloud/caching can provide reduced latency.

In the following section, we discuss the constraints and approaches for achieving low latency.

5.3 Constraints and Approaches for Achieving Low Latency

There are major fundamental trade-offs between capacity, coverage, latency, reliability, and spectral efficiency in a wireless network. Due to these fundamental limits, if one metric is optimized for improvement, this may result in degradation of another metric. In the LTE system, the radio frame is 10 ms with the smallest TTI being 1 ms. This fixed frame structure depends on the modulation and coding schemes for adaptation of the transmission rate with constant control overhead. Since *latency* is associated with control overhead (cyclic prefix, transmission mode, and pilot symbols) which occupies a major portion of transmission time of a packet (approximately 0.3-0.4 ms per packet transmission), it is not wise to consider a packet with radio transmission time less than 1 ms. If we design a packet with time to transmit of 0.5 ms, more than 60% of the resources will be used by control overhead [109]. Moreover, retransmission per packet transmission takes around 8 ms, and removal of retransmission will affect packet error significantly. As a result, we need radical modifications and enhancements in packet/frame structure and transmission strategy. In this regard:

- First, a novel radio frame reinforced by limited control overhead and smaller transmission time is necessary to be designed. For reduction of control overhead, procedures for user scheduling, resource allocation, and channel training can be eliminated or merged.

- Second, packet error probability for first transmission should be reduced with new waveforms and transmission techniques reducing the retransmission delay.
- Third, since latency critical data needs to be dispatched immediately, techniques for priority of data over normal data need to be identified.
- Fourth, synchronization and orthogonality are the indispensable aspects of OFDM that are major barriers for achieving low latency. Even though asynchronous mode of communication is more favorable over synchronized operation in terms of latency, it requires additional spectrum and power resources [229].
- Fifth, since the latency for data transmission also depends on the delay between the core network and the BS, caching networks can be used to reduce latency by storing the popular data at the network edge.

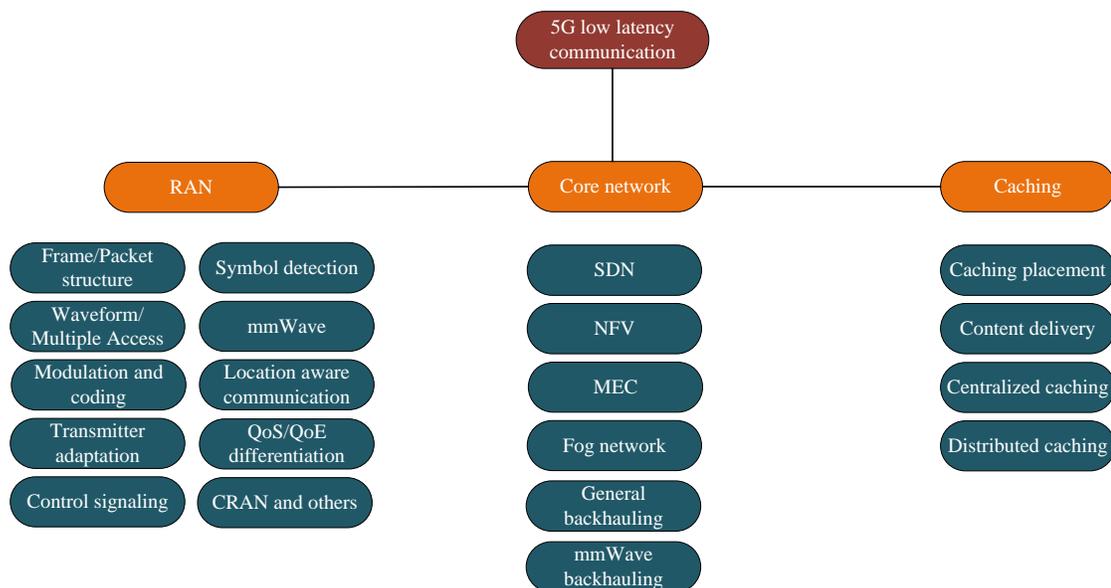


Figure 5.2: Categories of different solutions for achieving low latency in 5G.

Researchers proposed various techniques/approaches for achieving low latency in 5G. As summarized in Fig. 5.2, we divided the existing solutions into three major categories: (1) RAN solutions, (2) core network solutions, and (3) caching solutions. The RAN solutions include new/modified frame or packet structure, waveform designs, multiple access techniques, modulation and coding schemes, transmission schemes, control channels enhancements, low latency symbol detection, mmWave aggregation, cloud RAN, reinforcing QoS and QoE, energy-aware latency minimization, and location aware communication techniques. On the other hand, new entities such as SDN, NFV, MEC and fog network along with new backhaul based solutions have been proposed for the core network. The solutions of caching can be subdivided into caching placement, content delivery, centralized caching, and distributed caching, while backhaul solutions can be divided into general and mmWave backhaul. In the following sections, these solutions are described in further details.

5.4 RAN Solutions for Low Latency

To achieve *low latency*, various enhancements in the RAN have been proposed. Referring to Table 5.2, RAN solutions/enhancements include frame/packet structure, advanced multiple access techniques/waveform designs, modulation and coding scheme, diversity and antenna gain, control channel, symbol detection, energy-aware *latency* minimization, carrier aggregation in mmWave, reinforcing QoS and QoE, cloud RAN and location aware communication. In what follows, the detailed overview for each of these solutions is presented.

Table 5.2: OVERVIEW OF TECHNIQUES IN RAN FOR LOW LATENCY.

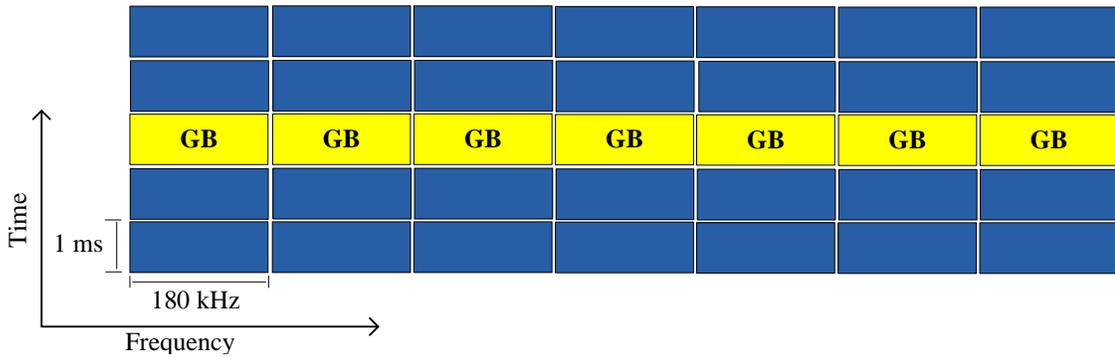
Case	Reference	Approach	Summary
Frame /Packet structure	[58, 59]	Small packets/short TTI	Transmission of small scale data is investigated for packet loss rate of 10^{-9} and latency as low as 1 ms.
	[60]	Subcarrier spacing	Subcarrier spacing is enlarged to shorten the OFDM symbol duration, and the number of OFDM symbols is proposed to keep unchanged in each subframe.
	[61–64]	Flexible OFDMA based TDD sub-frame	TDD numerology is optimized for dense deployment with smaller cell sizes and larger bandwidth in the higher carrier frequencies.
	[230]	Modification of physical subframe	Different control and data part patterns for consecutive subframes, TX and RX control parts are proposed to be separated from each other, and from the data symbols with a GP, leading to total number of 3 GPs per subframe.
	[226, 231–233]	Numerology, flexible sub frame and resource allocation	Numerology and subframe structure are defined considering diverse carrier frequencies and bandwidths to envision 5G including <i>low latency</i> . Cyclic prefix, FFT size, subcarrier spacing, and sampling frequency were expressed as the function of carrier frequency.
Advanced multiple access /Waveform	[65], [66], [67]	Filtered CP-OFDM, UFMC and FBMC	UFMC outperforms over OFDM by about 10% in case of both large and small packets. FBMC demonstrates better performance in case of transmitting long sequences; however, it suffers during the transmission of short bursts/frames.
Modulation and coding	[1, 68]	Polar coding	Based on simulation and field test, polar coding has been proposed for 5G, outperforming over turbo coding in case of small packet transmission.
	[69]	Turbo decoding with combined sliding window algorithm and cross parallel window (CPW) algorithm	A highly-parallel architecture for the latency sensitive turbo decoding is proposed combining two parallel algorithms: the traditional sliding window algorithm and cross parallel window (CPW) algorithm.
	[70]	New IFFT design with butterfly operation	Input signal of IFFT processor corresponding to guard band are assigned as null revealing the existence of numerous zeros (i.e., 0). If the sequence of OFDM symbol data which enter the IFFT is adjusted, the memory depth can be reduced from 1024 to 176.
	[71]	Sparse code multiple access (SCMA)	A dynamic shrunk square searching (DSSS) algorithm is proposed, which cuts off unnecessary communication control port (CCP) calculation along with utilization of both the noise characteristic and state space structure.

Table 5.2: OVERVIEW OF TECHNIQUES IN RAN FOR LOW LATENCY (CONTINUED).

Case	Reference	Approach	Summary
	[234]	Priority to latency critical data	A latency reduction approach by introducing TDM of higher priority ultra-low latency data over other less time critical services is proposed which maps higher priority user data during the beginning of a subframe followed by the normal data.
	[72]	Balanced truncation	Balanced truncation is applied for the model reduction in the linear systems that are being coupled over arbitrary graphs under communication <i>latency</i> constraints.
	[235]	Finite block length bounds and coding	Recent advances in finite-block length information theory are utilized in order to demonstrate optimal design for wireless systems under strict constraints such as low latency and high reliability.
Transmitter adaptation	[73]	Asymmetric window	Asymmetric window is proposed instead of well-known symmetric windows for reduction of cyclic prefix by 30%. This technique suppresses OOB emission but makes the system more susceptible to channel induced ISI and ICI.
	[74]	Transmission power optimization	Transmission power is optimized by steepest descent algorithm considering transmission delay, error probability and queuing delay.
	[75]	Path-switching method and a packet-recovery method	Low latency packet transport system with a quick path-switching method and a packet-recovery method are introduced for a multi-radio-access technology (multi-RAT) environment.
	[214]	Diversity	Diversity could be employed through various approaches such as spatial diversity, time diversity, and frequency diversity.
Control signaling	[78]	Control channel sparse encoding (CCSE)	CCSE is introduced in order to provide the control information using non-orthogonal spreading sequences.
	[77]	Scaled control channel design	A scaled-LTE frame structure is proposed assuming the scaling factor to be 5 with a dedicated UL CCHs for all sporadic-traffic users in each transmission time interval with possible smallest SR size.
	[79]	Symbol-level frequency hopping and sequence-based sPUCCH	A sequence-based sPUCCH (SS-PUCCH) incorporating two SC-FDMA symbols is introduced in order to meet a strict latency requirement. Symbol-level frequency hopping technique is employed to achieve frequency diversity gain and reliability enhancement.
	[80]	Radio bearer and S1 bearer management	Establishment of radio bearer and S1 bearer in parallel are proposed where eNB and mobility management element (MME) manages and controls radio bearer and S1 bearer, respectively. The eNB sends only single control signal in order to configure radio bearers such as SRB1, SRB2 and DRBs, that decreases the signaling interaction rounds between the UE and the eNBs.

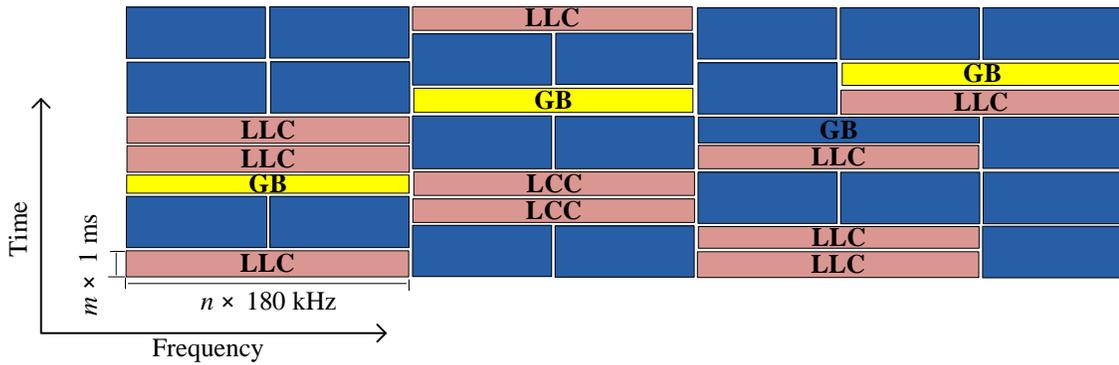
Table 5.2: OVERVIEW OF TECHNIQUES IN RAN FOR LOW LATENCY (CONTINUED).

Case	Reference	Approach	Summary
Symbol detection	[81]	Outer-loop link adaptation (OLLA) scheme	The proposed scheme controls the size of the compensation in the estimated SINR based on the time elapsed after a UE transits from an idle state to an active state, which helps to reduce latency for small packet applications.
	[83]	SM-MIMO detection scheme with ZF and MRC-ZF	A low-complexity and low latency massive SM-MIMO detection scheme is introduced and validated using SDR platforms. The low complexity detection scheme is proposed with a combination of ZF and MRC-ZF.
	[82]	Linear MMSE	A linear MMSE receiver is presented for low latency wireless communications using ultra-small packets.
	[84]	Space-time encoding and widely linear estimator	Space-time encoding is introduced within a GFDM block for maintaining overall low latency in the system. On the other hand, a widely linear estimator is used to decode the GFDM block at the receiver end, which yields significant improvements in gain over earlier works.
	[7, 85, 86]	Compressed sensing	Compressed sensing has been proposed to be effective in reducing latency of networked control systems if the state vector can be assumed to be sparse in some representation.
	[87]	Low complexity receiver design	A low complexity receiver is designed and using this, the performance of an SCMA system is verified via simulations and real-time prototyping. This approach triples the whole system throughput while maintaining low latency similar to flexible orthogonal transmissions.
mmWave	[88–91]	mmWave based air interface	Physical layer air interface is proposed using mmWave aggregation. Large bandwidth along with various approaches such as small frame structure, mmWave backhaul and beamforming can help to achieve low latency.
Location aware communication	[3, 92, 93]	Location information	Issues and research challenges of 5G are discussed followed by the conclusion that 5G networks can exploit the location information and accomplish performance gains in terms of throughput and latency.
QoS/QoE Differentiation	[94–103]	Parameter manipulation	Differentiation of constraints on QoS and QoE can maintain low latency in 5G services including ultra high definition and 3D video content, real time gaming, and neurosurgery.
Cloud RAN (CRAN)	[104–108]	Cloud architecture based RAN	CRANs combine baseband processing units of a group of base stations into a central server while retaining radio front end at the cell sides. Proper optimization of resources can ensure low latency along with capital expenditure reduction.



(a)

m = Fraction of 1 ms
 n = Multiplier of 180 kHz



(b)

Figure 5.3: Physical air interface (a) Conventional LTE radio frame, (b) Exemplary 5G radio frame with flexible time and frequency division for low latency [3] (GB: guard band; LLC: low latency communication).

5.4.1 Frame/packet structure

In the RAN solutions, modification in the physical air interface has been considered as an attractive choice. In particular, most of the proposed solutions are on the physical (PHY) and medium access control (MAC) layers.

In LTE cellular network, the duration of a radio frame is 10 ms. Each frame is partitioned into 10 subframes of size 1 ms which is further divided into 0.5 ms units that are referred as a resource block (RB). Each RB spans 0.5 ms (6 or 7 OFDM symbols) in time domain and 180 KHz (12 consecutive subcarriers, each of which 15 KHz) in frequency domain. Based on this, the subcarrier spacing Δf is 15 KHz, the OFDM symbol duration T_{OFDM} is $\frac{1}{\Delta f} = 66.67 \mu\text{s}$, the FFT size is 2048, the sampling rate f_s is $\Delta f \times N_{\text{FFT}} = 33.72 \text{ MHz}$ and the sampling interval T_s is $1/f_s$.

To reduce TTI for achieving *low latency*, the subcarrier spacing Δf can be changed to 30 KHz [60]. This results the corresponding OFDM symbol duration T_{OFDM} to be $33.33 \mu\text{s}$ and the FFT size N_{FFT} to become 1024 while sampling rate f_s is kept 30.72 MHz similar to LTE systems. The frame duration $T_s=10$ ms can be divided into 40 subframes in which each subframe duration T_{sf} is 0.25 ms and contains 6 or 7 symbols. Two types of cyclix prefixes (CPs) can be employed in this configuration with durations

$$T_{\text{cp1}} = 5/64 \times N_{\text{IFFT}} \times T_s \approx 2.604 \mu\text{s}, \quad (5.3)$$

$$T_{\text{cp2}} = 4/64 \times N_{\text{IFFT}} \times T_s \approx 2.083 \mu\text{s}. \quad (5.4)$$

A conventional LTE radio frame with equal sized RB and an exemplary 5G physical air frame are illustrated in Fig. 8.10(a) and Fig. 8.10(b), respectively.

In [58], an extensive analysis of the theoretical principles that regulates the transmission of small-scale packets with *low latency* and high reliability is presented with metrics to assess their performance. The authors emphasize control overhead optimization for

short packet transmission. In [226], a flexible 5G radio frame structure is introduced in which the TTI size is configurable in accordance with the requirement of specific services. At low offered load, 0.25 ms TTI is an attractive choice for achieving *low latency* due to low control overhead. However, for more load, control overhead increases which affects reliability and packet recovery mechanism resulting in increased latency. This study argues to employ user scheduling with different TTI sizes in the future 5G networks. In [59], the authors try to improve the outage capacity of URLLC and satisfy the low latency requirement of 5G using an efficient HARQ implementation with shortened transmission TTI and RTT. Moreover, some simulations are conducted in order to provide insights on the fundamental trade-off between the outage capacity, system bandwidth, and the latency requirement for URLLC.

In [231], the numerology and subframe structure are defined considering diverse carrier frequencies and bandwidths for low latency 5G networks. Cyclic prefix, FFT size, subcarrier spacing, and sampling frequency were expressed as a function of the carrier frequency. In [64], software defined radio (SDR) platform based 5G system implementation with strict latency requirement is presented. The scalability of the proposed radio frame structure is validated with E2E latency less than 1 ms. In [60], the proposed subcarrier spacing is enlarged to shorten the OFDM symbol duration, and the number of OFDM symbols in each subframe is kept unchanged in the new frame structure for TDD downlink. The subcarrier spacing is changed to 30 KHz resulting the corresponding OFDM symbol duration $T = 33.33 \mu s$. The fast Fourier transform (FFT) size N is 1024, while the sampling rate f_s is kept same as 30.72 MHz. The frame duration T_s is still 10 ms with 40 subframes.

In [230], in order to have fully flexible allocations of different control and data RB in the consecutive subframes, TX and RX control RBs are proposed to be separated from each other and also from the data RB by guard periods (GPs). This leads to total number

of 3 GPs per subframe which separates them. Assuming symmetrical TX and RX control parts with $N_{\text{ctrl},s}$ symbols in each and defining that same subcarrier spacing is used for control and data planes, with $N_{\text{data},s}$ being the number of data symbols and T_{symbol} being the length of an OFDM symbol, the subframe length T_{sf} can be determined as

$$T_{\text{sf}} = (2N_{\text{ctrl},s} + N_{\text{data},s}(T_{\text{symbol}} + T_{\text{CP}})) + 2T_{\text{GP}}. \quad (5.5)$$

In [61], the fundamental limits and enablers for low air interface latency are discussed with a proposed flexible OFDM based TDD physical subframe structure optimized for 5G local area (LA) environment. Furthermore, dense deployment with smaller cell sizes and larger bandwidth in the higher carrier frequencies are argued as notable enablers for air interface latency reduction. In [62], a new configurable 5G TDD frame design is presented, which allows flexible scheduling (resource allocation) for wide area scenarios. The radical trade-offs between capacity, coverage, and latency are discussed further with the goal of deriving a 5G air interface solution capable of providing *low latency*, high reliability, massive connectivity, and enhanced throughput. Since achieving low latency comes at cost of lower spectral efficiency, the proposed solution of the study includes control mechanisms for user requirement, i.e. whether the link should be optimized for low latency or high throughput.

A 5G flexible frame structure in order to facilitate users with highly diversified service requirements is proposed in [63]. Although, in-resource physical layer control signaling is the basis of this proposed radio frame, it allows the corresponding data transmission based on individual user requirements. For this, it incorporates adaptable multiplexing of users on a shared channel with dynamic adjustment of the TTI in accordance with the service requirements per link. This facilitates optimization of the fundamental trade-offs between latency, spectral efficiency, and reliability for each link and service flow. In [232], a scheme that reserves resources for re-transmission for a group of ultra reliable

low latency communication UEs is presented. The optimum dimensioning of groups and block error rate (BLER) target can reduce the probability of contention for the shared retransmission resources. Moreover, the unused resources can be utilized for non-grouped UEs resulting in overall efficiency enhancement.

In [233], fundamental trade-offs among three KPIs (reliability, latency, and throughput) in a 4G network is characterized, and an analytical framework is derived. In cases where the theory can not be extended via mathematical formulations due to complexity of scenario in hand, some guidelines are provided to make the problem tractable. In order to improve the aforementioned trade-offs between these KPIs in future 5G systems different candidate techniques are proposed.

The above approaches of frame/packet structure to achieve *low latency* at the RAN level are tabulated in Table 5.3.

Table 5.3: PHY AND MAC LAYER BASED RADIO INTERFACE SOLUTIONS FOR LOW LATENCY.

References	Approach/Area	PHY layer	MAC layer
[226]	Short TTI	✓	✓
[58, 231]	Numerology and sub frame structure	✓	
[60]	Subcarrier	✓	
[61]	Flexible subframe		✓
[64]	Flexible subframe implementation with SDR platform	✓	✓
[230]	Allocation of control and data RB	✓	✓
[62]	Radio frame and scheduling	✓	✓
[63]	Flexible TTI and multiplexing	✓	✓
[59]	Efficient HARQ implementation with shortened transmission TTI and RTT	✓	
[232]	Reservation of resources		✓
[233]	Calculating the fundamental trade-offs among three KPIs	✓	✓

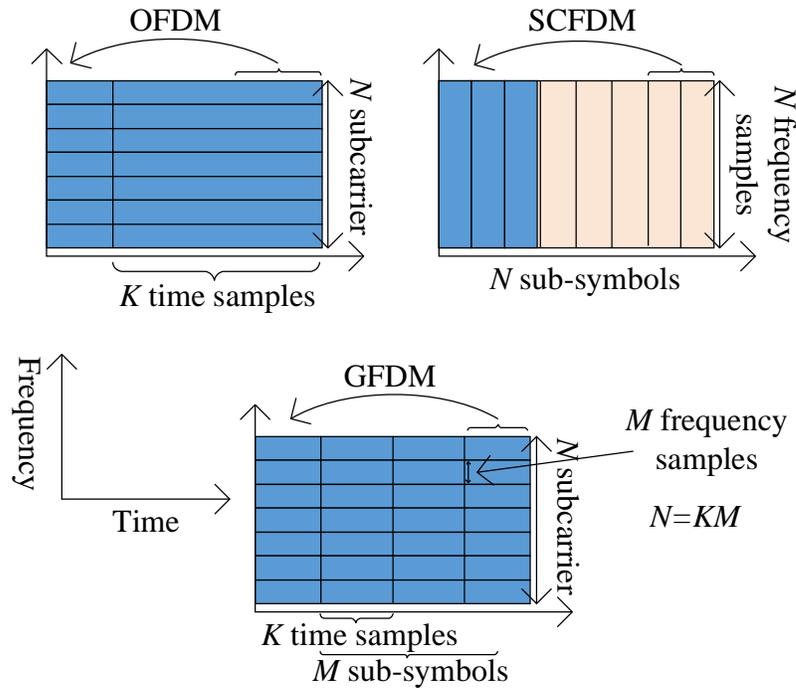


Figure 5.4: Slot placement in GFDM, OFDM and SCFDM.

5.4.2 Advanced Multiple Access Techniques/Waveform

Different kinds of candidate multiple access (MA) techniques and waveforms including orthogonal, non orthogonal and asynchronous access have been proposed for *low latency* communication [65–67, 229]. Since synchronization and orthogonality (integral to OFDM) is a hindrance for achieving *low latency*, asynchronous non orthogonal multiple access techniques have been discussed in [229]. Reduction of symbol duration to $67 \mu\text{s}$ is not a promising solution in critical time budgeting. In this regard, interleave division multiple access (IDMA) has been introduced in [236, 237] for generating signal layers. The IDMA is a variant of the CDMA technique which uses specific interleaving for user segregation in lieu of using a spread sequence to the individual user. Here, channel coding, forward error correction coding and spreading are combined into a single block by a low rate encoder. The spreading can not be considered as a distinct and special task. In-

terleaving usually utilizes a simpler iterative multiuser identification approach. However, this approach needs further rigorous investigation.

Table 5.4: PROPOSED MULTIPLE ACCESS TECHNIQUES FOR 5G.

Cases	IDMA [236,237]	SCMA [238]	GFDM [84,239]
Fundamental concept/features	<ul style="list-style-type: none"> • Specific interleaving • User segregation • Iterative multiuser identification 	<ul style="list-style-type: none"> • Multiple dimensional code word • QAM spreading combination 	<ul style="list-style-type: none"> • Block frame consists of time slots and subcarriers • Non-orthogonal • FFT/IFFT implementation
Low complexity	✓	✓	
Flexibility (in case of covering CP-OFDM and SC-FDE)			✓
Low latency		✓	✓

In order to supply synchronization and orthogonality, sparse code multiple access (SCMA) and non orthogonal multiple access (NOMA) have been presented in [238] for 5G scenarios. In SCMA, symbol mapping and spreading are combined together, and the mapping of multi dimensional codeword over incoming bits is performed directly from SCMA codebook. SCMA is comparatively simpler and has superior performance over low density version of CDMA. Another modulation technique that aims to reduce latency is referred as the generalized frequency division multiplexing (GFDM) is introduced in [84, 239]. The flexibility of covering both the cyclic prefix OFDM (CP-OFDM) and single carrier frequency domain equalization (SC-FDE), and block structure of GFDM help to achieve *low latency*. A typical mapping structure of GFDM, OFDM and SC-FDM

is illustrated in Fig. 5.4. The overall comparison among IDMA, SCMA and GFDM is presented in Table 5.4.

Filter bank multi carrier (FBMC) has been a strong candidate waveform for 5G [194, 240]. FBMC demonstrates better performance in case of transmitting long sequences; however, it suffers during the transmission of short bursts/frames. For usage of cyclic prefix, wide frequency guards and more required coordination, OFDM may be inefficient in case of low latency communication [239]. Universal filtered multi-carrier (UFMC) [239, 241] is upgraded version of FBMC which offsets the disadvantage of FBMC. It outperforms OFDM by about 10% in cases of time frequency efficiency, inter carrier interference (ICI) and transmissions of long or short packets. Additionally, UFMC performs better than FBMC in the case of very short packets while demonstrating similar performance for long sequences. These make UFMC as the one of the best choices for next generation low latency communication.

In case of UFMC, the time domain transmit vector [241] for a user is superposition of sub-band wise filtered components. The time domain transmit vector for a particular multi-carrier symbol of user k with filter length L and FFT length N is

$$X_k_{(N+L-1) \times 1} = \sum_{i=1}^B F_{ik}_{(N+L-1) \times N} V_{ik}_{N \times n_i} S_{ik}_{n_i \times 1}, \quad (5.6)$$

where

- S is the complex QAM symbol vector;
- V is the transformed time domain vector by IDFT matrix; In this case, the relevant columns of the inverse Fourier matrix are incorporated in accordance with the respective subband position within entire available band;
- i is the index of each subband of B ;

Table 5.5: WAVEFORM CONTENDER FOR 5G.

Cases	OFDM	FBMC [194,240]	UFMC [239,241]
Filtering	Generalized filtering to all subcarriers of entire band	Filtering to each subcarriers	Generalized filtering to a group of consecutive subcarriers
Requirement of co-ordination	Higher	Lower	Lower
Time-frequency efficiency (due to CP and guard band)	0.84	1	1
ICI (in case of lower degree of synchronization with UEs and eNBs)	Higher	Lower	Lower
Performance	Performs well for large packets with well coordination	Performs well for large packets with less coordination	Performs well for short packets with less coordination

- \mathbf{F} is a Toeplitz matrix. It is comprised of filter impulse response, and performs the linear convolution.

The symbol duration of $(N + L - 1)$ samples is determined by the filter length and FFT size. Filtering per block per subcarrier allows spectrally broad filters in pass band and shorter in time domain compared to FBMC. The reduced time yields shortened OFDM CP. The filter ramp up and ramp down in shorten time domain ensures symbol shaping in a way that allows protection against ISI and robustness for multiple access users. Furthermore, being orthogonal with respect to complex plain, complex modulation symbol can be transmitted without further complication.

Another advantage of UMFC is the ability of using different subcarrier spacings or filter times for users in different subbands. If a user uses FFT size N_1 and filter length L_1 , and another user uses filter length and FFT size of N_2 and L_2 respectively, then

UFMC symbol duration can be designed such that $N_1 + L_1 - 1 = N_2 + L_2 - 1$. This makes UFMC a remarkable adaptive modulation scheme with capability to be tailored easily under various characteristics of communications, including delay/Doppler spread variations in the radio channel and user QoS needs. The comparative discussion among OFDM, FBMC and UMFC is presented in Table 5.5.

5.4.3 Modulation and Channel Coding

Although use of small packets is a potential approach for achieving *low latency*, appropriate modulation and coding is required for small packet transmission for acceptable reliability. In the literature, mainly three types of coding schemes are proposed for 5G. As presented in [1], low-density parity-check (LDPC) and polar codes outperform turbo codes in terms of small packets while for medium and large packets, the opposite is true. While small packet is a requirement for *low latency*, other aspects such as implemen-

Table 5.6: COMPARISON AMONG CHANNEL CODING SCHEMES FOR LOW LATENCY [1].

Cases	Turbo coding [69]	LDPC-PEG [1]	Convolutional coding [1]	Polar codes [68]
Algorithm complexity for coding 1/3 of 40 bits with respect to turbo codes	100%	98%	66.7%	1.5%
Algorithm complexity for coding 1/3 of 200 bits with respect to turbo codes	100%	98%	66.7%	110.7 %
Performance in short packets		✓		✓
Performance in medium packets	✓		✓	

tation complexity, performance in practical test, and flexibility need to be investigated. In [68], polar code has been tested in field for 5G considering various scenarios: air interface, frame structure, settings for large and small packets, OFDM, and filtered OFDM (f-OFDM) waveforms. In all cases, polar code performed better than turbo codes which makes it a candidate channel coding scheme for 5G. The comparison among the schemes are illustrated in Table 5.6.

In [69], a highly-parallel architecture for the *latency* sensitive turbo decoding is proposed by combining two parallel algorithms: the traditional sliding window algorithm and cross parallel window (CPW) algorithm. New IFFT design with butterfly operation is proposed in [70], which reduces IFFT output data delay through the reduction of IFFT memory size and butterfly operation (e.g. addition/subtraction). Input signal of the IFFT processor corresponding to guard band is assigned as zero (i.e. '0') revealing the existence of numerous zeros. If the sequence of OFDM symbol data which enter the IFFT is adjusted, the memory depth can be reduced from 1024 to 176.

A dynamic shrunk square searching (DSSS) algorithm is proposed in [71], which cuts off unnecessary communication control port (CCP) calculation by utilizing both the noise characteristic and state space structure. In this way, it can maintain close to optimal decoding performance in terms of the block error rate (BLER). This results in reduction of delay in communication. In [234], a *latency* reduction approach by introducing time division multiplexing (TDM) of higher priority ultra-low *latency* data over other less time critical services is proposed, which maps higher priority user data during the beginning of a subframe followed by the normal data. In [72], balanced truncation is applied for the model reduction in the linear systems that are being coupled over arbitrary graphs under communication *latency* constraints. In [235], recent advances in finite-block length information theory are utilized in order to demonstrate optimal design for wireless systems under strict constraints such as low latency and high reliability. For a given set of

constraints such as bandwidth, latency, and reliability the bounds for the number of the bits that can be transmitted for an OFDM system is derived.

5.4.4 Transmitter Adaptation

A representative set of approaches for reducing *latency* using transmission side processing are tabulated in Table 5.7, which will be overviewed in the rest of this subsection.

In [73], an asymmetric window is proposed instead of well-known symmetric windows for reduction of cyclic prefix by 30%, and hence reducing *latency* due to reduced overhead. This technique suppresses out of bound (OOB) emission but makes the system more susceptible to channel induced inter symbol interference (ISI) and inter carrier interference (ICI). Transmission power optimization by the steepest descent algorithm considering transmission delay, error probability and queuing delay is proposed in [74]. In [75], low *latency* packet transport system with a quick path-switching and a packet-recovery method is introduced for a multi-radio-access technology (multi-RAT) environment. In [214], use of diversity gain is proposed as a solution for capacity enhancement and latency reduction. Diversity could be achieved through various approaches such as spatial diversity, time diversity, and frequency diversity.

In [242], a mmWave based switched architecture system is proposed where control signals use low-resolution digital beamforming (to enable multiplexing of small control packets) with analog beamforming in the data plane (to enable higher order modulation). This reduces the overhead significantly due to the control signaling which results in more resources for data transmission. This technique leads to reduction of round trip *latency* in the physical layer.

Recent advancements in full duplex (FD) communication comes forward with feature of doubling the capacity, improving the feedback, and latency mechanism meanwhile

Table 5.7: OVERVIEW OF SOLUTIONS IN TRANSMITTER ADAPTATION FOR LOW LATENCY.

Reference	Techniques	Merits	Demerits
[73]	Asymmetric window	Reduces cyclic prefix by 30% and maintains good OOB suppression along with latency reduction.	It assumes that spectral mask will be stricter in 5G networks.
[74]	Transmission power optimization	Queuing delay is considered in optimization along with transmission delay and packet error.	No uniform cross layer information exchange format is provided. Besides that, cross layer signaling may result extra overhead in the nodes.
[75]	Path-switching and packet-recovery method	Provides fast switching and recovery method in multi-RAT environment.	It depends on availability of good channel for path switch, and packet recovery may affect the resiliency.
[214]	Diversity gain	Option of various diversity gains such as space, time and spatial gain for low latency transmission.	Gain depends on various aspects such as beam forming, beam training and antenna array.
[242]	Beam forming using mmWave	Design and analysis of MAC layer under realistic conditions.	Proper channel model in mmWave is under development.
[243–248]	Full duplex communication in same channel	Improves throughput, reduces latency and upholds PHY layer security.	Crosstalk between the transmitter (Tx) and the receiver (Rx), internal interference, fading, and path loss.

upholding steady physical layer security [243–248]. Various proposed techniques of 5G networks such as massive MIMO and beamforming technology providing reduced spatial domain interference can be contributive for FD realization [244]. Besides that intelligent scheduling of throughput/delay critical packets along with proper rate adaption and power assignment can results in capacity gain and reduction of *latency*. However, this field needs to be extensively investigated for studying capacity and latency trade offs.

5.4.5 Control Signaling

When the packet size is reduced as envisioned in 5G systems, control overhead takes the major portion of the packet. Addressing this, various approaches are proposed in order to reduce the control channel overhead. The potential solutions targeting the control channel enhancements to achieve low *latency* are illustrated in Table 5.8.

In [78], control channel sparse encoding (CCSE) is introduced with vision to transmit the control information by means of non-orthogonal spreading sequences. A scaled-LTE frame structure is proposed in [77] assuming the scaling factor to be 5 with dedicated UL control channels (CCHs) for all sporadic-traffic users in each TTI with possible smallest scheduling request (SR) size. In [249], short TTI based uplink frame has been proposed for achieving E2E *latency* no longer than 1 ms. In the proposed scheme, sub-slot consisting of 2 symbols has been proposed for uplink data and control channel. A sequence-based sPUCCH (SS-PUCCH) incorporating two single carrier-frequency division multiple access (SC-FDMA) symbols is introduced in [79] in order to meet a strict latency requirement. Symbol-level frequency hopping technique is employed to achieve frequency diversity gain and reliability enhancement.

In the proposed procedure of [80], establishment of radio bearer and S1 bearer in parallel are proposed where eNB and mobility management element (MME) manages

Table 5.8: OVERVIEW OF SOLUTIONS IN CONTROL SIGNALING FOR LOW LATENCY.

References	Techniques	Merits	Demerits
[78]	Control channel sparse encoding (CCSE)	Uses non-orthogonal spreading sequences.	Needs field test for further validation.
[77]	Dedicated UL CCHs	Provides CCH for sporadic packets with small size scheduling request (SR).	Requires dedicated CCH in each TTI and well designed scheduling request (SR) detector at BS. Also considered scenario with UL and DL signal space of 10 and 40 bits, spatial diversity of 16, and bandwidth 10 MHz may not be always feasible.
[249]	Sub slotted data and control channel	Two symbols are used in each subslot which is compatible with current LTE.	Reliability issue is not addressed.
[79]	SS-PUCCH consists of SC-FDMA symbol	More robust to channel fading compared to reference signal based PUCCH. Symbol level frequency hopping harnesses frequency diversity gain with enhanced reliability.	Need to be validated by field test.
[80]	Radio bearer and S1 bearer management	Control overhead and latency are decreased for both light and heavy traffic. It ensures 100% accessibility of all UEs.	The technique is more suitable for very large traffic networks such as vehicle networks.
[81]	Outer-loop link adaptation	Besides reduction of latency for small packets, it can boost throughput just after changing from idle state to connected state.	Needs field test for further validation.
[250]	Slotted TTI based radio resource management	It can be implemented as an extension of LTE-A.	Validation through simulation and field test is not presented.

Table 5.8: OVERVIEW OF SOLUTIONS IN CONTROL SIGNALING FOR LOW LATENCY (CONTINUED).

References	Techniques	Merits	Demerits
[251]	Adaptive radio link control (RLC)	Besides latency reduction, it improves throughput and reduces processing power.	Control and data plane need to be separated.
[252]	SDN based control plane optimization	Using bandwidth rebalancing strategy, balance between cost and performance is maintained.	For large number of players (vehicles), the game can be complicated. Also real world field test is required for performance evaluation.
[253]	SDN based X2 signaling management	It reduces signaling overhead and handover latency.	The approach has been investigated for only femtocells.
[254]	Inter BS data forwarding and make-over-handover	This technique reduces X2 communication, processing and reconfiguration delays.	Cases such as packet loss, handover failure, scenario with poor communication link are not considered for performance evaluation.
[255]	Optical connected splitters with dynamic bandwidth allocation and tailored MAC protocol	It ensures X2 latency less than 1 ms.	Needs field test for further validation.

and controls radio bearer and S1 bearer, respectively. The eNB sends only single control signal in order to configure radio bearers such as SRB1, SRB2 and DRBs, that decreases the signaling interaction rounds between the UE and the eNBs. In [81], a new outer-loop link adaptation (OLLA) scheme is proposed. The scheme controls the size of the compensation in the estimated SINR based on the time elapsed after a UE transits from an idle state to an active state, which helps to reduce *latency* for small packet applications. The study [250] proposed a slotted TTI based radio resource management for LTE-A and 5G in order to achieve low latency. The approach can serve low latency services utilizing short TTI and enhance download control channel (ePDCCH).

The study [251] proposed a novel mechanism that introduces an adaptive radio link control (RLC) mode which dynamically alternates between unacknowledgment mode (UM) and acknowledgment mode (AM) according to the real-time analysis of radio conditions. This technique reduces system *latency* and processing power, and improves throughput using UM. On the other hand, it improves data reliability by activating AM during the degraded radio conditions. In [252], SDN based control plane optimizing strategy is presented to balance the *latency* requirement of vehicular ad hoc network (VANET), and the cost on radio networks. The interaction between vehicles and controller is formulated and analyzed as a two-stage Stackelberg game followed by optimal rebating strategy, which provides reduced *latency* compared to other control plane structures.

In [253], SDN-based local mobility management with X2 forwarding is proposed where total handover signaling is minimized by reduction of inter node signaling exchanges and X2 signaling forwarded to centralized SDN system. This approach can reduce handover latency while reducing of signal overhead. In [254], QoS, CQI and other parameter based data utilization is proposed among eNBs to reduce X2 latency, processing and reconfiguration delays. Additionally, make-before-handover is proposed for low

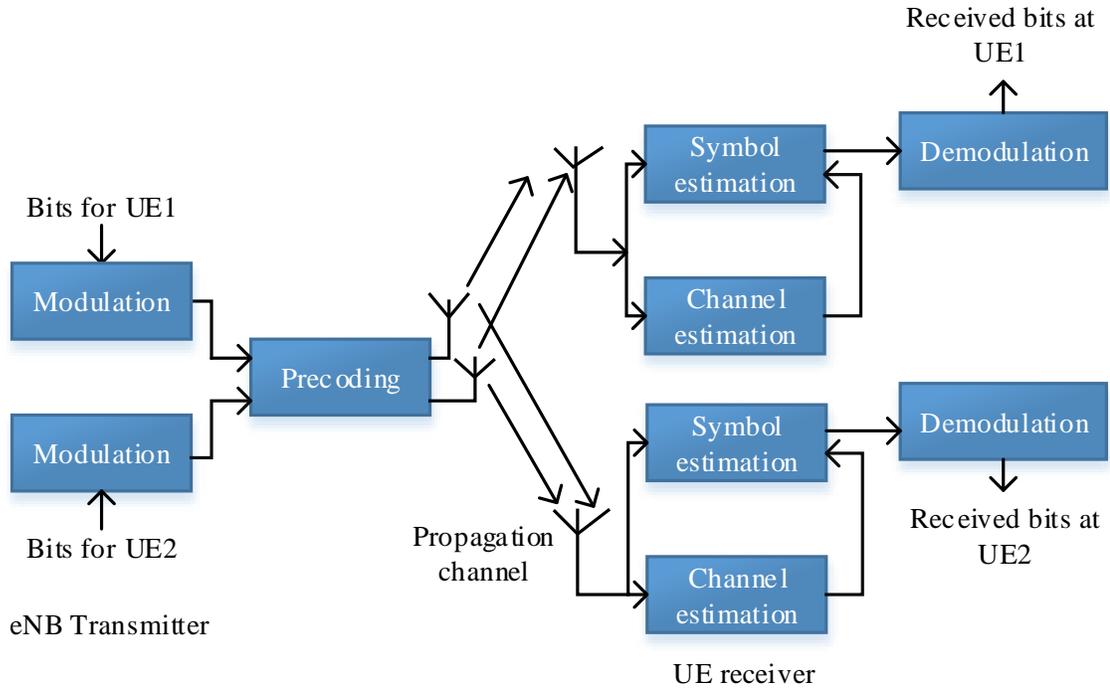


Figure 5.5: Transmission and symbol detection in cellular network.

latency 5G services for no data interruption. In order to meet stringent latency in X2 interface, enhanced passive optical network (PON) based radio network is proposed in [255], where the splitters are connected through optical connections. Following this, dynamic bandwidth allocation algorithm and tailored MAC protocol are introduced for achieving less than 1 ms latency over X2 interface.

5.4.6 Symbol Detection

As illustrated in Fig. 5.5, symbol detection encompasses various processes such as channel estimation and decoding, which can all contribute into the overall *latency*. The related literature in the symbol detection side for latency reduction are tabulated in Table 5.9.

In [83], a low-complexity and low-latency massive SM-MIMO detection scheme is introduced and validated using SDR platforms. The low complexity detection scheme is

Table 5.9: OVERVIEW OF SOLUTIONS IN SYMBOL DETECTION FOR LOW LATENCY.

Reference	Technique/ Approach	Merits	Demerits
[83]	SM-MIMO detection with ZF and MRC-ZF	Significant improvement of SINR and latency is achieved compared to other schemes. Also the method is validated in live environment designed by SDR platform.	Availability of large number of antennas is assumed.
[82]	Linear MMSE receiver	Reduces latency along with throughput gain improvement.	It is not clear how much latency can be reduced in this scheme.
[84]	Space-time encoding and widely linear estimator	Significant improvements in terms of symbol error rate and latency over earlier works.	Complexity at receiver side is increased.
[7, 85, 86]	Compressed sensing	CS algorithm exhibits reduced complexity and increases reliability. It is compatible with the current LTE systems as well, and requires less measurement (resource) to decode control information. It provides sub-Nyquist sampling method for reconstruction of sparse signal efficiently in a linear system.	It is challenging to design CS and sparse recovery system considering diverse wireless conditions and input conditions.
[87]	Low complexity receiver in SCMA system.	The prototype triples capacity while maintaining low latency.	More suitable for MTC. Needs field test for further validation.

proposed with a combination of zero forcing (ZF) and maximum-ratio-combining-zero-forcing (MRC-ZF). In [82], a linear minimum mean square error (MMSE) receiver is presented for low latency wireless communications using ultra-small packets. The estimation of receiver filter using the received samples is proposed during the data transmission period in lieu of interference training period. Additionally, soft decision-directed channel estimation is argued using the data symbols for re-estimation of the channels. In [84], space-time encoding is introduced within a GFDM block in order to achieve transmit diversity for overall low latency in the system. On the other hand, a widely linear estimator is used to decode the GFDM block at the receiver end, which yields significant improvements in terms of symbol error rate and *latency* over earlier works.

In [7, 85, 86], compressed sensing is proposed for latency reduction in networked control systems if the state vector can be modeled as sparse in some representation domain. In [87], a low complexity receiver design is proposed and the superiority of an SCMA system is verified via simulations. In addition, it is demonstrated with a real-time prototype that the whole system throughput triples while maintaining low latency similar to flexible orthogonal transmissions.

5.4.7 mmWave Communications

Carrier aggregation using the mmWave spectrum is widely considered to be a promising candidate technology for 5G, capable of providing massive bandwidth and ultra low *latency*. The mmWave technology is especially critical for VR/AR type of applications which require high throughput and low *latency*. The works in mmWave spectrum for achieving low latency is summarized in Table 5.10.

In [88], a new frame design for mmWave MAC layer is introduced which provides several improvements including adaptable and smaller transmission intervals, dynamic

Table 5.10: OVERVIEW OF SOLUTIONS IN MMWAVE COMMUNICATIONS FOR LOW LATENCY.

Refer-ences	Techniques	PHY layer	MAC layer	NET layer
[88]	mmWave based MAC layer frame structure		✓	
[89]	Low latency core network architecture, flexible MAC layer, and congestion control		✓	✓
[90]	mmWave based physical layer air interface with basic numerology and logical channel arrangement	✓		
[91]	Low latency frame structure with beam tracking	✓		

locations for control signals, and the capability of directional multiplexing for control signals (dynamic HARQ placement). It addresses ultra low latency along with the multiple users, short bursty traffic and beam forming architecture constraints. The study [89] focuses on three critical higher-layer design areas: low latency core network architecture, flexible MAC layer, and congestion control. Possible solutions to achieve improvements in these critical design areas are short symbol periods, flexible TTI, low-power digital beam forming for control, and low latency mmWave MAC, which can all be considered for data channel, downlink control channel, and uplink control channel.

In [90], in order to decrease the latency of the system, two different physical layer numerologies are proposed. The first approach is applicable for indoor or line of sight (LOS) communications, and the second one is suitable for non line of sight (NLOS) communications. This is justified by some channel measurements experiments in 28 – 73 GHz range. In [91], a 5G mmWPoC system is employed to evaluate the throughput functionality in field tests at up to 20 km/h mobile speed in an outdoor LOS environment. Additionally, some improvements for a frame design is obtained which decrease the *latency* in the field tests. In the experiments, it is observed that the new slotted frame design can decrease

the RTT to 3 ms for 70% – 80% of the cases in experiments, alongside the observed throughput up to 1 Gbps.

5.4.8 Location-Aware Communications for 5G Networks

Location knowledge (in particular, the communication link distance) can be considered as a criterion of received power, interference level, and link quality in a wireless network. Therefore, overhead and delays can be reduced with location-aware resource allocation techniques because of the possibility of channel quality prediction beyond traditional time scales. The literature on location-aware communications regarding low latency are tabulated in Table 5.11.

Table 5.11: LOCATION-AWARE COMMUNICATIONS FOR LOW LATENCY.

References	Techniques	Merits	Demerits
[93]	Location information utilization in protocol stack	Latency, scalability and robustness can be improved.	Location accuracy, spatial channel modeling, balancing trade-off between location information and channel quality metric are challenging.
[3]	Physical layer parameters design using FFT, frame duration and local area (LA) physical channel	Improves spectral and energy efficiency along latency reduction.	Needs field test for further validation.
[92]	Utilization of channel quality and traffic statistics from small cell	Coexistence capability with overlay LTE-A network, sleeping modes, contention based data channel, channel quality indicator and interference statistics.	The technique is more feasible for small cells.

In [93], several approaches are presented for monolithic location aware 5G devices in order to identify corresponding signal processing challenges, and describe how location data should be employed across the protocol stack from a big picture perspective. Moreover, this work also presents several open challenges and research directions that should be solved before 5G technologies employ mmWave to achieve the performance gains in terms of latency, connectivity and throughput. In [3], 5G flexible TDD is proposed for local area (5GETLA) radio interface with FFT size of 256 and 512, and short frame structure to achieve *latency* lower than 1 ms. The packets of size of less than 50 kbits can be transmitted with E2E *latency* of 0.25 ms. The main focus in designing physical layer parameters is on FFT, frame duration and physical channel (LA).

In [92], a novel numerology and radio interface architecture is presented for local area system by flexible TDD, and frame design. The proposed framework ensures coexistence with overlay LTE-A network, sleeping modes, contention based data channel, and channel quality indicator and interference statistics. Here, the channel quality and traffic statistics are accumulated from the small cells which can help to gain high throughput and low *latency*. Especially, in order to reduce the *latency*, the delay due to packets containing critical data for the higher layer protocols, for instance transmission control protocol (TCP) acknowledgment (ACK) packets, must be optimized. To do so, one possible approach is to carry out the retransmissions as quick as possible compared to the higher layer timers. Moreover, capability of data transmission to a contention based data channel (CBDCH) can play a key role here. As a result, by introducing CBDCH in small cells that are not highly loaded, the average latency of small packets transmission can be decreased considerably.

5.4.9 QoS/QoE Differentiation

Differentiation of constraints on QoS and QoE can maintain low *latency* in 5G services including ultra high definition and 3D video content, real time gaming, and neurosurgery. The related literature on QoS and QoE control for low latency services are tabulated in Table 5.12.

Table 5.12: LITERATURE OVERVIEW RELATED TO QoS/QoE DIFFERENTIATION.

Reference	Techniques/Approaches	QoS	QoE
[94]	mmWave utilization with beam tracking	✓	✓
[95]	SDN and cloud technology	✓	✓
[96]	QoS-aware multimedia scheduling	✓	
[97]	Client based QoS monitoring architecture	✓	
[98]	Colored conflict graph	✓	
[99]	QoS architecture with heterogeneous statistical delay bound	✓	
[100]	Dynamic energy efficient bandwidth allocation scheme	✓	
[101]	Predictive model based on Internet video download		✓
[102]	Routing using proximity information		✓
[103]	Predictive model based on empirical observations		✓

Abundance of mmWave bandwidth and extensive use of beamforming techniques in 5G will allow high QoS and QoE overcoming the resource and sharing constraints [94]. However, current transmission protocols and technologies cannot be employed simply for addressing technical issues in 5G. The mapping of diverse services including latency critical service to the optimal frequency, SDN and cloud technologies can ensure to achieve the best QoS and QoE, as discussed in [95]. In [96], a QoS-aware multimedia scheduling approach is proposed using propagation analysis and proper countermeasure methods to meet the QoS requirements in the mmWave communications. Mean opinion score (MOS)

which is a criterion for user satisfaction can be employed for functionality evaluation of the newly presented QoS approach and well-known distortion driven scheduling in different frequency ranges.

Client based QoS monitoring architecture is proposed in [97] to address the issue of QoS monitoring from server point of view. Different criteria such as bandwidth, error rate and signal strength are proposed with the well-known RTT delay for maintaining desirable QoS. A colored conflict graph is introduced in [98] to capture multiple interference and QoS aware approaches in order to take the advantage of beamforming antennas. In this case, reduction in call blocking and handoff failure helps to have a better QoS for multi class traffic. Each device can be sensitive to time based on its application. This can be considered as an issue for QoS provisioning. To address this, a novel QoS architecture is presented in [99] with heterogeneous statistical delay bound over a wireless coupling channel. The authors presented the dynamic energy efficient bandwidth allocation schemes in [100], which improve system quality significantly and maintain QoS.

Previous QoS criteria which consist of packet loss rate, network latency, peak signal-to-noise ratio (SNR) and RTT are not sufficient for streaming media on Internet, and therefore, users' perceived satisfaction (i.e. QoE) needs to be addressed [101, 194]. Higher QoS may not ensure the satisfactory QoE. Different routing approaches of video streams in the mobile network operators scenario is discussed in [102] for substantial refinement in QoE considering bit rate streams, low jitter, reduced startup delay and smoother playback. A predictive model from empirical observations is presented in [103] to address interdependency formulated as a machine learning problem. Apart from that, a predictive model of user QoE for Internet video is proposed in [101].

5.4.10 CRAN and Other Aspects

Cloud radio access network (CRAN) (as illustrated in Fig. 5.6) is introduced for 5G in order to reduce the capital expenditure (CAPEX) and simplify the network management [256]. CRANs combine baseband processing units of a group of base stations into a central server retaining radio front end at the cell sides. However, this requires connection links with delay of $250 \mu\text{s}$ to support 5G low latency services. In order to meet strict latency requirements in CRAN, two optimization techniques including (i) fine-tuned real-time kernel for processing latency and (ii) docker with data plane development kit (DPDK) for networking latency have been proposed in [104]. The experimental results clearly demonstrate the effectiveness of the approaches for latency optimization. In [105], split of PHY and MAC layer in a CRAN with Ethernet fronthaul is proposed, and verified through experimental test followed by latency interpretation. It is found that *latency* for packets of size 70 and 982 bytes is $107.32 \mu\text{s}$ and $128.18 \mu\text{s}$ confirming 20% latency increase from small to large packet. The promising results affirm that latency critical services in 5G can be supported by CRAN. In [106], it is demonstrated that based on experimental results from Wi-Fi and 4G LTE networks, offloading the traffic to cloudlets outperforms the response times by 51% in comparison to cloud offloading.

The CRAN can utilize backhaul information to redistribute users for QoE maximization and adaptation of temporal backhaul constraints. Corresponding to this, authors in [107] proposed a centralized optimization scheme to control the cell range extension offset so as to minimize the average network packet delay. In [108], a CRAN on the basis of the optical network (PON) architecture is presented which is called virtualized-CRAN (V-CRAN). The proposed scheme can dynamically affiliate each radio unit (RU) to a digital unit (DU) which results in coordination of multiple RUs with their corresponding DU. Moreover, definition of virtualized BS (V-BS) is brought up which is able to mutually send shared signals from several RUs to a user. V-CRAN can reduce *latency* for joint

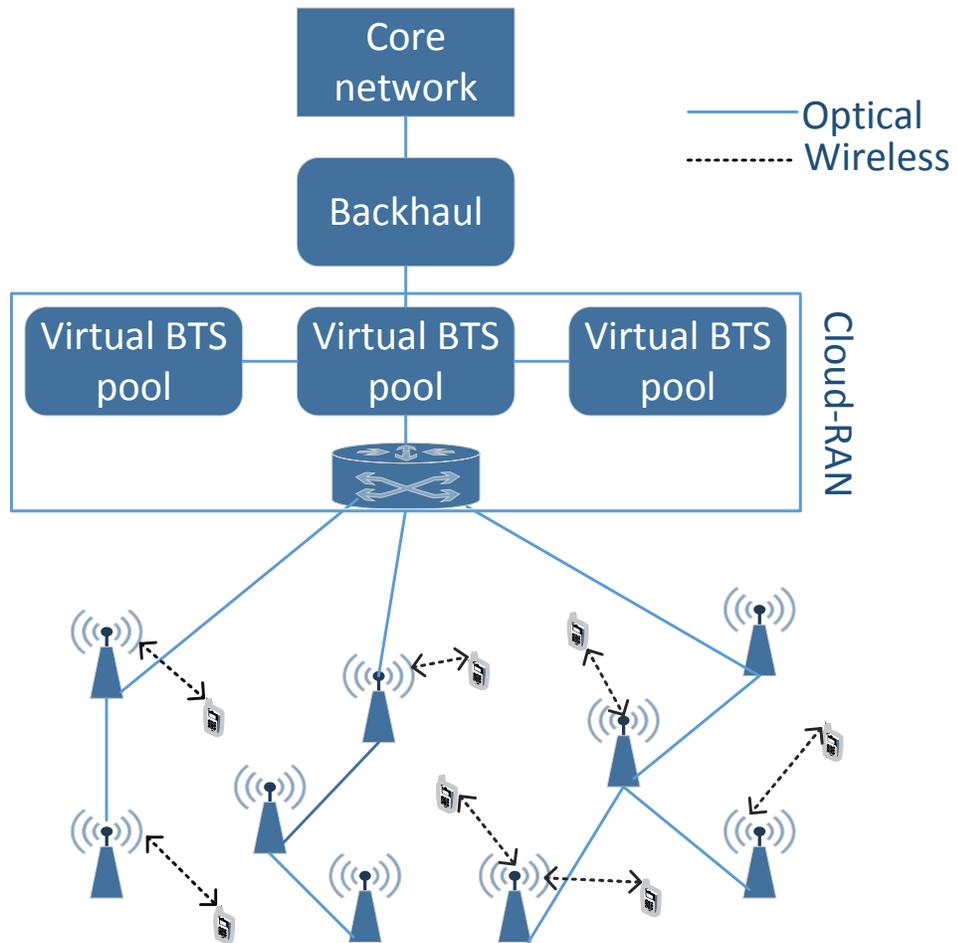


Figure 5.6: Cloud-RAN architecture in 5G networks.

transmission due to the following reasons. First, it can provide more than enough bandwidth for data transmission between RUs and DU. Furthermore, for each DU, a dedicated hardware/software is assigned that can be utilized by joint transmission controller in order to provide data and signaling for RUs. The last but not least, in order to handle the load distribution between DUs, the virtualized PON can connect a DU directly via a linecard.

Though mmWave will be the major contributor in attaining 5G goals, spectrum below 6 GHz is always the primary choice due to less attenuation, supporting long distance, and antenna compatibility. Moreover, conventional cellular networks are usually deployed within the expensive licensed bands and they use reliable core networks that are optimized to provide low-volume delay-sensitive services such as voice. However, with appearance of high-volume delay-insensitive resource-hungry applications including multimedia downloads, such conventional networks may not be cost-effective anymore [257,258]. Such concerns can be tackled or at least partially addressed by spectrum sharing in 5G network improving spectrum and energy efficiency along with QoS/QoE control [16, 18, 181, 259, 260]. In [261], in order to exploit the TV white space for D2D communications underlying existing cellular infrastructure, a framework is proposed. A location-specific TV white space database is proposed in which D2D service can be provided using a look-up table for the D2D link so that it can determine its maximum permitted emission power in the unlicensed digital TV band to avoid interference. In [262], a QoE driven dynamic and intelligent spectrum assignment scheme is proposed which can support both cell and device level spectrum allocation. This technique enhances not only the spectrum utilization, but also can maintain desired QoE including *latency* aspect. The optimization problem of network sum rate and access rate with resource allocation and QoS constraints in D2D communication is presented in [263]. To solve this, a fast heuristic algorithm is proposed to reduce computational complexity resulting desired QoS such as latency. In [264], a game theory and interference graph based optimization prob-

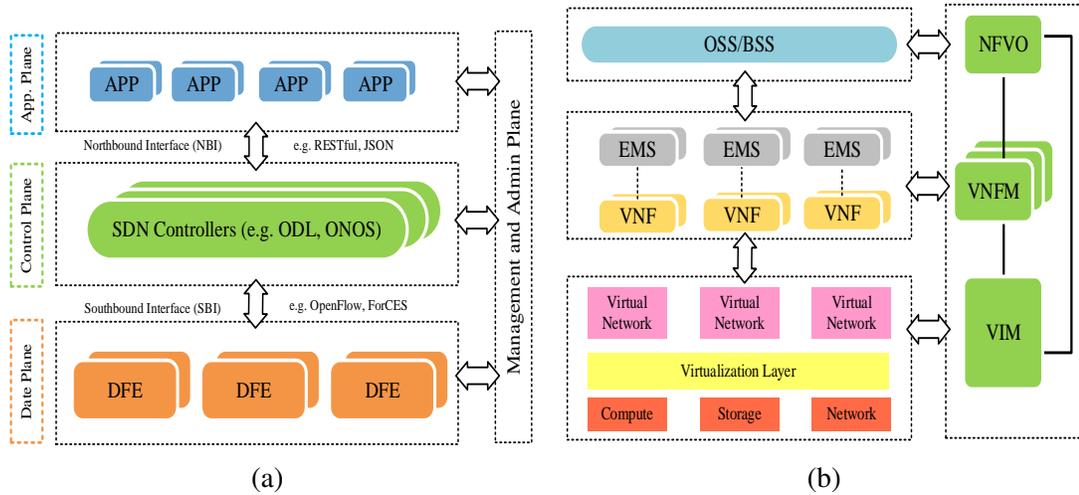


Figure 5.7: Simplified example architectures for core network (a) Architecture of SDN [4], and (b) Architecture of NFV [5] (APP: application; ODL: opendaylight platform; ONOS: open network operating system; DFE: dyna forming engineering; OSS: operations support systems; BSS: base station subsystem; EMS: element management system; NFVO: NFV orchestrator; VNFM: virtual network function manager; VIM: virtualized infrastructure manager).

lem considering user scheduling, power allocation and spectrum access is presented with an aim to maximize user satisfaction across the network. Two algorithms including spatial adaptive play iterative (SAPI) learning are proposed to achieve Nash equilibrium. In [265], a multi agents based cognitive radio framework is proposed for effective utilization of spectrum and meet the goals of 5G including latency. Here, sensing capability of secondary users (SU) is replaced by spectrum agents (SA) where the user can switch between SU mode and SA mode based on available spectrum information.

Before starting the packet transmission in data applications, a tolerable initial delay can be considered. Thus, this short delay can be employed to decrease the energy required to operate the small cell base stations (SBSs). With the goal of average power consumption reduction of SBSs, when the load is low, under-utilized SBSs can be are switched off. As a result, the users and the network can be able to save energy by postponing transmissions. By doing so, the users have time to wait for an SBS with better link quality.

It can be observed that sleeping mode operation for energy efficiency improvements in SBSs will also introduce a source of *latency*. In [266], the energy-efficiency versus delay trade-off is investigated and optimality conditions for UEs transmit power is derived. By postponing the access of the users, energy efficiency of the system can be improved. The optimal threshold distance is derived in order to minimize the average distance between a user and an SBS. Simulation results demonstrate that the energy consumption of SBS can be reduced by about 35% if some of the SBSs are switched off.

In [267], authors introduced an energy efficient, low-complexity technique for load-based sleep mode optimization in densely deployed 5G small cell networks. By defining a new analytic model, the distribution of the traffic load of a small cell is characterized using Gamma distribution. It is shown that the network throughput can be improved significantly while some amount of energy is saved by taking the benefit of the initial delay. In [268], impact of average sleeping time of BS and association radius on the mean delay in an UDN is investigated using a M/G/1/N queuing model. An explicit equation of delay is derived and the effect of average sleeping time and association radius on the mean delay is analyzed.

In [269], the remote PHY (R-PHY) and the remote PHYMAC (R-PHYMAC) based modular broadcast cable network is proposed for the access network. In this architecture, R-PHYMAC can achieve lower mean upstream packet delay compared to R-PHY for bursty traffic and long distance over 100 miles. In [270], a virtual converged cable access platform (CCAP) system and procedure is proposed for hybrid fiber coaxial (HFC) cable network. In this method, a new digital optical configuration is introduced to receive data packets with capability to convert them into RF waveforms. This method improves space and power requirements while enhancing operational flexibility. In [271], a novel remote FFT (R-FFT) module is proposed which can perform physical layer processing of FFT module towards RF transmission. This module reduces fronthaul bit-rate requirement for

CRAN while providing solution for unified data over cable service interface specification (DOCSIS) and LTE service over HFC cable network.

5.5 Core Network Solutions for low latency

To meet the vision of 5G encompassing ultra low *latency* in addition to enhancements in the RAN, drastic changes are also proposed in the core network. The new core network includes some new entities such as SDN, MEC, and NFV as well as new backhaul techniques [138, 272, 273]. These enhancements aim to reduce the processing time, bypass several protocol layers, and ensure seamless operation. The core network solutions for low *latency* are reviewed in further detail in the rest of this section.

5.5.1 Core Network Entities

The SDN and NFV are assumed to be the main candidates for the design of 5G core network [274]. Based on this, in this section, we mainly focus on the role of SDN and NFV technologies in latency reduction in 5G core network. Exemplary architectures of SDN and NFV of the 5G core network are illustrated in Fig. 5.7. The existing literature on the core network entities that can facilitate to achieve low *latency* are summarized in Table 5.13.

The EPC which is developed by 3GPP for the LTE cellular network has some limitations which affect the latency of the overall system. One concern is that the control plane and the data plane in the EPC are not fully separated. There is a level of coupling between Serving Gateway (SGW) and Packet Data Network Gateway (PGW). Decoupling of control plane and data plane seems necessary because they have different network QoS criteria to be met. In particular, the control plane needs low latency to process signaling messages, while the data plane requires high throughput to process the data. Thus,

Table 5.13: OVERVIEW OF TECHNIQUES IN CORE NETWORK FOR LOW LATENCY.

Case	Reference	Approach	Summary
Core Network Architecture	[109], [110–113], [114–123]	SDN-based architecture	The architecture of 5G is proposed based on SDN with 5G vision to meet large throughput, massive connectivity and low latency.
	[114, 116–118, 123–129]	NFV-based architecture	NFV invalidates the dependency on hardware platform and makes easy deployment of EPC functions as well as the sharing of resources in the RAN. This can reduce the E2E <i>latency</i> with improved throughput performance.
	[34,109,118,124, 126, 127]	MEC/fog-based network	MEC/fog provides computation and storage near user end and also separates the data plan from control plan. These reduces the <i>latency</i> .

in order to design such planes efficiently, it is preferable to decouple them completely. Based on the literature, SDN and NFV can be employed in EPC architecture in order to decouple data plane and control plane and have a seamless operation of core network functions [272, 275].

After modification of NFV based EPC in which the whole network elements are implemented using softwares running on Virtual Machines (VM), control plane and user plane can be separated by employing SDN in EPC. An SDN controller can act as an interface between the decoupled planes. In addition to several advantages of SDN/NFV-based user plane and control plane separation, including independent scalability, flexibility of flow distribution, and better user mobility management, such a decoupling can have considerable effect on reducing the *latency* as well. This plane decoupling can facilitate the mobile edge computing technology which decreases the latency. However, adding an

SDN controller to the network can be another source of the latency for the system. On the other hand, the scalability of SDN controller can be addressed by deploying several controllers. Thus, there is a trade off here between the scalability of controllers and latency increase which should be considered in the design process for specific applications [115].

Another limitation is that the data plane of the LTE EPC is implemented in a centralized manner. Even the users that need to communicate locally have to transmit their traffic in a hierarchical system ending with few number of centralized PGWs which increases the E2E latency. Although centralized implementation of network can facilitate the management and monitoring the network by operators, it increases the E2E latency which can not meet the applications requiring low latency including autonomous driving, smart-grid or automated factory. Thus, this kind of implementation leads to inefficient system performance and high latency which can not meet the 5G vision [34].

Recently, by emergence of the new technologies such as cloud computing, fog networks, mobile edge computing, NFV and SDN, the implementation of the network can be more distributed [116, 117]. By employing such technologies, the CAPEX and OPEX of the network can be reduced considerably. Moreover, by bringing the elements of core network closer to the users, the E2E delay can be decreased significantly. In [118], authors proposed SDN/NFV-based MEC networks algorithms that can enable the data plane to create a distributed MEC by placement of network functions at a distributed manner. They demonstrated that the proposed scheme can reduce the redundant data center capacity around 75%, and meet the 5G latency requirement along with considerable backhaul link bandwidth reduction.

Mobility management in core network based on SDN can potentially introduce some delays. In [119], the main contributors for processing delays in an SDN-based mobility management system is discussed. By implementing two proactive and reactive solutions for mobility management using Mininet and OpenFlow, it is observed that with high prob-

ability (almost 95%) in the proactive mobility management system, the overall processing latency is around the median value. By visualizing all of EPC entities as decentralized VMs in different locations, in [128], a carrier cloud architecture is introduced. To improve the E2E latency of the users, the concept of Follow-Me-Cloud is presented. The main point of this concept is that all parts of the network can keep track of the movement of the user which results in seamless connectivity and lower E2E latency. In [120], [121], the authors proposed a decentralized scheme for control plane called SoftMow which is a hierarchical reconfigurable network-wide control plane. The proposed control plane includes geographically distributed controllers where each controller is responsible to serve the network in its particular location. The number of the levels in this hierarchical scheme can be designed based on the available latency budgets.

In [109], authors presented an LTE compliant architecture to decrease *delay* for combination of fog networks, MEC and SDN in which the architecture is supposed to take the advantage of NFV in the evolved packet core (EPC) functions. Following this, optimization of general packet radio service (GPRS) tunneling protocol (GTP) is introduced for supporting low latency services. GTP tunnels management is accomplished by a novel element between the eNB and the mobile network interface with the Internet. In [110, 111], SDN is proposed along with some changes in the existing 4G architecture for moving forward towards 5G. The changes include reduction of number of serving gateways (S-GW) and elimination of some protocol layers. In SDN based system, virtualization is possible, and routes can be optimized. This will allow handling of QoS by setting specific rules in the switches along the data path. Network coding integrated with SDN is proposed in [112] for low latency and reduced packet-retransmission. Network coding can work as network router and can be integrated with SDN, which provides seamless network operation and reduction in *latency*.

The NFV is proposed as a major entity of 5G core network in [114, 124–127]. NFV removes the dependency on the hardware platform and makes flexible deployment of EPC functions as well as sharing of resources in RAN. This can reduce the E2E *latency* with improved throughput performance. SDN and NFV based 5G architecture with enhanced programmability of the network fabric, decoupled network functionalities from hardware, separated control plane from data plane, and centralized network intelligence in the network controller is presented in [114]. In [124], an information centric scheme is presented in order to integrate the wireless network virtualization with information centric network (ICN). In this architecture, key components such as wireless network infrastructure, radio spectrum resource, virtual resources (including content-level slicing, network-level slicing, and flow-level slicing), and information centric wireless virtualization controller have been introduced which can support low latency services.

An NFV-based EPC is introduced in [276] which is an EPC as a service to ease mobile core network (EASE). In this scheme, the elements of the EPC are visualized using VMs. In [123], a simple implementation for EPCaaS is proposed in which one of the drawbacks is the increment in the latency between the EPC and virtual network function components. To address such an issue, in [129], the main idea is to partition the virtual network functions into several subsets/groups based on their interaction and workload in order to reduce the network latency. It should be noted that employing the decentralized control plane and having it closer to the users at the network edge can be helpful for applications with high mobility and low latency. However, it can introduce some issues related to policies and charging enforcements. In [126], authors presented the optimization problem of composing, computing and networking virtual functions to select those nodes along the path that minimizes the overall *latency* (i.e., network and processing latency). The optimization problem is formulated as a resource constrained shortest path problem on an auxiliary layered graph followed by initial evaluation.

In [277], the authors proposed a detailed approach to implement the big data empowered self organizing network (SON) in 5G. The procedure to employ the data based on machine learning and data analytics is demonstrated in order to create E2E visibility of the network for implementation of a more efficient SON. This approach can meet the stringent 5G requirements such as low latency. In [122], the smart gateway (Sm-GW) is employed for scheduling the uplink transmissions of the eNBs. Based on simulations, it is demonstrated that the Sm-GW scheduling can allocate the data rate in uplink transmission to the eNBs in a fair manner along with reducing packet delays. Traffic of heavily loaded eNBs can make the buffer of the queue of a Sm-GW full which results extra latency. This situation happens due to the massive number of connected eNBs to a single Sm-GW. However, by using effective scheduling, connection with Sm-GW can be distributed among eNBs while maintaining QoS. The proposed core network entities for 5G vision are summarized in Table 5.14.

Table 5.14: PROPOSED CORE NETWORK ENTITIES FOR 5G VISIONS.

Reference	SDN	NFV	MEC	Fog networks	SON
[115, 119], [120–122]	✓				
[109]	✓		✓	✓	
[110–113]	✓				
[114], [116, 117, 123]	✓	✓			
[118]	✓	✓	✓		
[124, 128]		✓		✓	
[125, 129]		✓			
[126]		✓	✓		
[127]		✓	✓		
[277]					✓

5.5.2 Backhaul Solutions

Backhaul between base stations and the core network carries the signaling and data from the core and the Internet. Due to the enormous number of small cells and macro cell base stations supporting 1000x capacity, massive connectivity and latency critical services in 5G, the capacity of backhaul is a bottleneck for achieving low *latency*. At current scenario, microwave, copper and optical fiber links are used for backhaul connections based on availability and requirements. 5G backhaul requires higher capacity, lower latency, synchronization, security, and resiliency [278]. Referring to Table 5.15, we divide existing backhaul solutions into 2 parts: (1) General backhaul and (2) mmWave backhaul. The solutions are described as follows.

General backhaul

General backhaul includes a dynamic GPRS tunneling protocol (GTP) termination mechanism that combines cloud based GTP with a quick GTP tunnel proposed in [113]. Based on the user request or other factors, the system can change its mode from a cloud-based GTP tunnel to the quick GTP tunnel. In [109], a 5G vision compliant architecture is presented to reduce *latency* combining with the fog network, MEC and SDN. The optimization of GTP tunnels is accomplished by a novel element acting as an intermediate node between eNB and the mobile network interface accompanied with the Internet. In [133], modified VLC technology is used to set up an optical window (OW) link for low-cost backhaul of small cells to achieve a *latency* of 10 ms. Moreover, using a next generation baseband chipset, E2E latency below 2 ms can be achieved. An efficient PON-based architecture is proposed in [132] that offers ultra-short *latency* for handovers by enhancing connectivity between neighboring cells. Additionally, the authors propose a tailored dynamic bandwidth allocation algorithm for a fast handover between eNBs, which are associated to the same or diverse optical network units.

Table 5.15: OVERVIEW OF LITERATURE IN BACKHAUL SOLUTIONS TO ACHIEVE LOW LATENCY.

Category	Reference	Approach/Techniques
General back-haul	[113]	A dynamic GTP termination scheme combining cloud based GTP with a quick GTP tunnel with a dedicated hardware.
	[109]	GTP tunnel optimization by a new component in 5G complaint network consists of fog networks, MEC and SDN.
	[133]	Modified VLC technology to set up an OW link for low-cost back hauling of small cells.
	[132]	PON-based architecture with a tailored dynamic bandwidth allocation algorithm.
	[131]	MAC-in-MAC Ethernet based unified packet-based transport network.
	[279]	The first architecture is based on over provision transport network whereas the second one is based on dynamic sharing, SDN and NFV controllers.
	[130]	SDN and cache enable architecture for limited backhaul secererio.
mmWave back-haul	[134]	mmWave for fronthaul and backhaul, and split of control and user plane.
	[136]	A digitally-controlled phase-shifter network based hybrid precoding/combining scheme for mmWave massive MIMO.
	[137]	A framework supporting of in-band, point-to-multipoint, non-Line-of-sight and mmWave backhaul.
	[135]	A mmWave based backhaul frame structure in 3 - 10 GHz carrier frequencies.
	[138,139]	Ultra dense wavelength division multiplexing (UD-WDM) passive optical networks (PONs) based backhaul solution for mmWave networks.

In transport networks, latency requirement plays a key role. For instance, the main requirement for machine type communications is the latency that should be kept as low as possible. Therefore, efficient design of 5G transport networks is critical. In [131], a detailed perspective of the 5G crosshaul design is proposed in order to introduce the key goal of transporting the backhaul and fronthaul traffic in a unified packet-based transport network based on MAC-in-MAC Ethernet. Moreover, the SDN/NFV-based 5G-crosshaul control plane architecture is investigated which decouples the logically centralized control plane and data plane. This can contribute in latency reduction aspect. In [279], two candidate technologies for 5G transport networks are presented. One of them is based on the over-provisioning of transport resources while the second architecture is based on dynamic resource sharing and NFV/SDN-based controller to handle the latency requirements.

In [130], SDN and cache enabled heterogeneous network is proposed where C-plane and U-plane are split. The caches of macro and small cells are overlaid and cooperated in a limited backhaul scenario while ensuring seamless user experiences for coverage, low latency, energy efficiency and throughput. In [279], two candidate technologies for 5G transport networks is presented to handle the latency requirements in which one of them is based on the over-provisioning of transport resources, while the second architecture is based on dynamic resource sharing and NFV and SDN-based controller. In [130], SDN and cache enabled erogenous network is proposed where C-plane and U-plane are spitted. The caches of macro and small cells are overlaid and cooperated in a limited backhaul scenario while ensuring seam user experiences such as coverage, energy efficiency and throughput.

mmWave Backhaul

In addition to the presented solutions for backhaul, mmWave employment in backhaul can be considered as a promising solution for *latency* reduction. In order to have the enhanced user experience, the BSs should be in touch with core network and all other BSs via a low *latency* backhaul [280]. In [134], the authors proposed a scheme that employed mmWave links as backhaul, fronthaul and access in which a new separation method between control and user plane is proposed for 5G cellular network. A reasonable split among control and user plane improve the user QoS by providing ubiquitous high data rates in mmWave SBS coverage.

In [136], to implement an ultra-dense network (UDN) for the future 5G network and providing high data rates, the need of a reliable, gigahertz bandwidth, and economical backhaul is emphasized. Since mmWave can be easily integrated with massive MIMO to improve link reliability, and can provide sufficient data rate for wireless backhaul, it is a promising candidate for such a scenario. Considering a massive MIMO scenario, a hybrid precoding approach is considered, in which each BS can cover several SBSs with multiple streams for each SBS at the same time. In [137], the authors proposed a solution framework for supporting an in-band, point to multi-point, NLOS, mmWave backhaul in order to provide a cost-effective and low *latency* solution for wireless backhaul. It is shown that an in-band wireless backhaul for inter BS coordination is feasible while the cell access capacities are not affected considerably.

In [135], a frame design for mmWave communications is proposed for 5G SBS network radio interface in 3-10 GHz. For both of LOS and NLOS scenarios different frame designs are proposed, which have a frame duration of 0.1 ms and 0.05 ms, respectively, to achieve low *latency*. The proposed LOS structure can be assumed as a suitable solution for short distance indoor wireless access or in-band backhaul. In order to obtain high capacity and low latency backhauling, the EU research project 5G STEP-FWD introduced

a novel design that deploys ultra dense wavelength division multiplexing (UDWDM) passive optical networks (PONs) as the backhaul of mmWave networks. The proposed scheme is based on the ultra-narrow wavelength spacing of the UDWDM technology to provide seamless connectivity for dense small-cell networks [139, 281].

5.6 Caching Solutions for Low Latency

In addition to the shortage of the radio spectrum, the insufficient capacity of backhaul links can be considered as a bottleneck for low latency communication. The long delay can be due to the requests of too many users in peak-traffic hours. Thus, latency reduction is crucial for users QoS and QoE in the 5G networks. Caching and in a more general category, information centric networking, can be assumed as one of the promising candidate technologies to design a paradigm shift for latency reduction in next generation communication systems [282, 283].

In this section, referring to Table 5.16, we present a detailed overview of caching concepts for cellular network followed by fundamental limits and existing solutions.

5.6.1 Caching for cellular network

Let us consider, a scenario that a user requests content from a content library $F = \{f_1, f_2, \dots, f_k\}$, where k is the number of files. The files are sorted with popularity where f_1 and f_k are the most and least popular files, respectively. The popularity of a requested file l can be written [304] as

$$\phi_l = \frac{l^{-\gamma}}{\sum_{i=1}^k l^{-\gamma}}, \quad (5.7)$$

Table 5.16: OVERVIEW OF LITERATURE IN CACHING.

Aspect	Reference	Summary
Content caching	[140–142, 145, 146, 284–298]	Filling of appropriate data is investigated by diverse techniques employing time intervals in which the network is not congested.
Content delivery	[145, 146, 284–301]	Content delivery to requested users is presented by different approaches for reduction of latency.
Centralized caching	[140, 141, 145, 284, 286, 288–294, 296–300]	Various centralized caching is investigated with assumption that a coordinator with access to almost all the information about the storage capacities of different BSs, the connectivity of the users and BSs, and etc.
Distributed caching	[142, 146, 285, 287, 295, 298, 301]	Various aspects of distributed caching has been investigated in order to minimize the communication overhead among SBSs and the central scheduler.
Latency-Storage trade-off	[288–293, 302, 303]	Fundamental trade off between storage and latency is investigated in radio networks complemented with cache-enabled nodes.

where $l \in \{1, 2, \dots, k\}$ and γ is the parameter for uneven distribution of popularity in F which follows Zipf distribution. For N eNBs $\mathcal{B} = \{BS_1, BS_2, \dots, BS_N\}$ with each eNB having capacity C , the probability of caching of file f_l by an eNB can be obtained as

$$P_l^{\mathcal{B}} = 1 - e^{-\rho\sigma\pi R^2}, \quad (5.8)$$

where ρ is the spatial density of eNBs following a Poisson point process [305–307], and σ is the probability that file f_l is cached within \mathcal{B} . Then, the total probability of getting content from the eNB can be written as

$$P^{\mathcal{B}} = \sum_{i=1}^N \phi_i P_i^{\mathcal{B}}. \quad (5.9)$$

The probability of getting the content as in (5.9) is directly associated with the latency of downloading it, and hence, effective caching strategies can help in significantly reducing latency in 5G networks

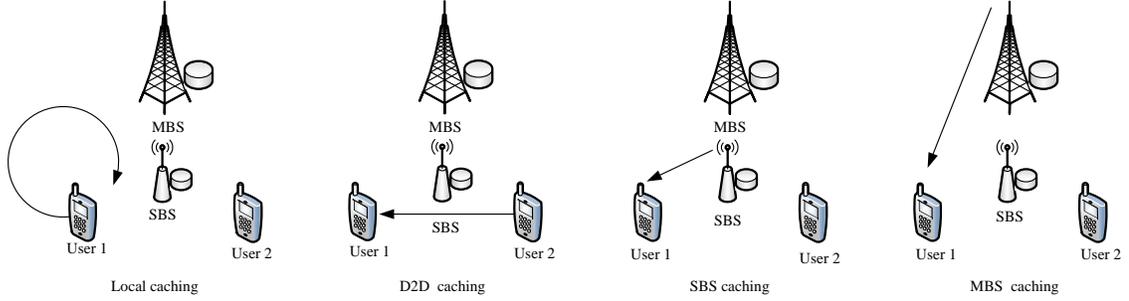


Figure 5.8: Different types of caching in 5G.

The proposed caching schemes for mobile networks can be divided into 4 categories: (1) Local caching, (2) Device to device (D2D) caching, (3) SBS caching, and (4) Macro base station (MBS) caching. Each of these caching types can reduce the *latency* by providing the requested content for the users using a way other than bringing it from the core network using backhaul links. In fact, each user starts from the nearest source to look for its desired content and proceed until finding it in any of the proposed sources. The different types of caching for cellular network are illustrated in Fig. 5.8 followed by the summarized descriptions presented in Table 5.17.

5.6.2 Fundamental Latency-storage trade-off in Caching

There are several fundamental limits for caching in mobile networks including latency versus storage, memory versus rate [294], memory versus CSIT [295], storage versus maximum link load [296], and caching capacity versus delivery rate [297]. As defined in Table 5.18, from an information theoretic point of view, authors employed the metrics such as normalized delivery time (NDT), fractional delivery time (FDT), and delivery time per bit (DTB) for investigation of the latency storage trade-off in caching networks.

Table 5.17: DIFFERENT TYPES OF CACHING SCHEMES FOR THE CELLULAR NETWORK.

Cases	Reference	Description
Local caching	[308]	When a UE wants to access a content, it first checks in itself. Once the content is confirmed in the local caching storage, it is accessed by the UE without any delay.
D2D caching	[304]	If the requested content is not found locally, user will seek it within the range of its D2D communication. If it is found in nearby devices, it is delivered to the requester UE by D2D communication.
SBS caching	[309]	If the requested content is available in the local SBS, the content is delivered to the UE by the local SBS.
MBS caching	[304]	If the content is not found in local caching storage, nearby devices or SBS caching, the content is delivered by MBS caching.

In most of these works [142, 146, 285, 287, 295, 298, 301], for a given scenario, an upper bound or lower bound for the considered metric is derived in order to get useful insights of this trade-off. The summary of *latency* storage trade-off works is presented in Table 5.19.

The authors in [288] investigated the storage latency trade-off using a new metric called NDT. This metric measures the worst-case *latency* that can happen in a cache-

Table 5.18: DEFINITION OF METRICS USED FOR LATENCY EVALUATION IN CACHING SCHEMES.

Cases	Reference	Definition
NDT	[288]	Defined as asymptotic delivery delay per bit in the high-power, long-blocklength case.
DTB	[293]	Defined as the ratio between the duration of transmission in channel to the file size in bits for the very large file size regime.
FDT	[292]	Defined as the worst-case delivery latency for the real load at a rate described by the DoF of the channel.

Table 5.19: SUMMARY OF THE WORKS ON LATENCY-STORAGE TRADE-OFF IN THE CACHING.

Ref.	Used metric	Description
[288]	NDT	Lower bound for NDT is derived for a general cache-enabled network for both perfect and imperfect CSI.
[289]	NDT	The trade-off between NDT and front-haul and caching resources is characterized and optimal caching front-haul transmission is obtained.
[290]	FDT	For a 3×3 wireless interference network the storage-latency trade-off is investigated while all transmitters and receivers are equipped with caches.
[292]	FDT	For a scenario with a 3×3 MIMO system in which the nodes are enabled with several antennas, the trade off between storage and latency is investigated.
[293]	DTB	DTB is used to characterize the system performance as a function of cache storage and capacity of backhaul links connected to SBS.
[302]	NDT	The trade-off between storage and latency for a distributed caching scenario in fog radio access networks is characterized.
[303]	NDT	Cloud-based compressed precoding and edge-based interference management introduced as two major techniques for optimal performance of cloud and caching resources in different cases.
[310]	NDT	Considering heterogeneous timeliness requests depending on application, the fundamental trade-off between the delivery latencies of different users requests is characterized using NDT.
[311]	NDT	Upper and lower bounds for minimum delivery latency as a function of cache and fronthaul resources is obtained over fronthaul and wireless link in a F-RAN with a wireless multicast fronthaul.
[312]	NDT	Assuming an F-RAN when pipelined fronthaul edge transmission is used, lower and upper bounds on the NDT is presented.
[313]	NDT	Assuming both content placement and delivery phases in one time slot, it is shown that the proposed approach outperforms the offline caching.

aided wireless network divided by that of an ideal system with unlimited caching capability. Considering a general cache-aided wireless network, the lower bound for NDT is presented in terms of the ratio of the existing file memory at the edge node and the total size of files for both perfect channel state information (CSI) and imperfect CSI.

Authors in [289] employed NDT as well in order to characterize the trade-off between NDT and fronthaul/caching resources. Using this information-theoretic analysis of fog radio access networks, optimal caching front-haul transmission is obtained. In [290], the *latency* storage trade-off in a 3×3 wireless interference network is investigated while all transmitters and receivers are equipped with caches. Another metric called (FDT) is proposed in order to characterize the trade-off between latency and storage. This information theoretic performance metric is actually a refined version of the metric originally proposed in [288]. The FDT can reflect the load reduction as well. In a similar work [291] the well-known DoF metric is used which does not reflect the load reduction. Moreover, the proposed approach in [291] just considers the one-shot linear processing, but interference alignment scheme in [288] may require infinite symbol extension.

In [291], an optimization problem is designed to minimize the number of required communication blocks for content delivery. Then, a lower bound is proposed on the value of the objective function. Using the same metric, the authors in [292] investigated the fundamental trade-off for a cache-enabled MIMO system. Considering a scenario with a 3×3 MIMO system in which the nodes are enabled with several antennas, the trade off between storage and latency is investigated. In addition to FDT and showing its optimality for some ranges of cache size, the model can consider the effect of real traffic load at a rate specified by the DoF of the channel.

In [293], a cellular network is considered with multiple SBSs with limited cache capacity in which there is interference among them. Here, another information theoretic metric based on delivery latency is defined as well in order to characterize the system

performance as a function of SBS cache memory and capacity of backhaul links connected to SBS. Using this metric which is called DTB, the trade off between latency and system resources is investigated. In [302], using NDT trade-off between storage and latency a distributed caching scenario in fog radio access networks is characterized. In the presented approach, a coded delivery scheme is proposed to minimize the *latency* for delivering user demands for two edge-nodes and arbitrary number of users. It is shown that using decentralized placement, the presented delivery approach can obtain a considerable performance improvement in comparison to the derived lower bound.

In [303], again NDT is employed to characterize the fundamental trade off between delivery latency and system architecture. Considering NDT as the criterion for latency evaluation of the system, some bounds on its value are proposed. In the light of such bounds, useful insights on the latency and storage trade off are obtained. It is demonstrated that in order to obtain the lowest delivery *latency*, cloud-based compressed pre-coding and edge-based interference management should be considered as two major techniques for optimal performance of cloud and caching resources in different cases. In [310], for a fog radio access network, the heterogeneous timeliness requests depending on application is considered while in the existing works the assumption is that all requests have identical *latency* for all files in the content library. The fundamental trade-off between the delivery latencies of different users' requests is characterized using NDT. The minimization of the average delivery *latency* as a function of the content popularity profile remains as a future research direction.

In [311], the total delivery latency over fronthaul and wireless link in a fog radio access network with a wireless multicast fronthaul is investigated. Again using NDT, the optimal delivery *latency* based on cache storage and fronthaul resources is formulated and upper and lower bounds are obtained. It is shown that in contrast to the receiver-side caching, coded multi-casting can not help in decreasing the NDT when two users and two

edge nodes (ENs) are available. In [312], for a F-RAN, the NDT is used to characterize the performance of a fog radio access network when pipelined fronthaul edge transmission is used.

In [313], a F-RAN system is considered in which ENs are cache-enabled with limited storage. On the contrary to the existing works that focused on offline caching where both caching phases are considered separately, the proposed method can integrate both of them in each time slot. The performance is characterized using NDT and compared to that of optimal offline caching schemes. It is shown that the proposed approach outperforms the offline caching.

5.6.3 Existing Caching Solutions for 5G

In general, the file delivery service in mobile networks can be classified into two parts: *cache placement*, and *content delivery* [314]. In cache placement, the cached content on the BSs is determined, which is usually based on the amounts of requests from users. Cache placement can be done using a centralized or distributed approach. In centralized approach, a coordinator is assumed with access to almost all the information about the memory size of BSs, the connectivity of the users, and the BSs. However, in some scenarios that there is no central controller, these schemes can not be applicable and a distributed cache placement is required [142]. In each of the caching phases whether centralized or distributed, the design can reduce the delivery time of the requested content and latency of the system.

There are some efforts in the literature on investigation of different challenges raised up in centralized cache placement problems. In [140], the cache placement problem was investigated in a scenario including SBSs, called helpers with weak backhaul links but large memory size. In experimental evaluation, it is shown that the proposed scheme can

achieve a considerable performance improvement for the users at reasonable QoS levels. In [141], authors aim at minimizing the average download delay of wireless caching networks with respect to caching placement matrix. It is demonstrated that the backhaul propagation delay can affect the caching placement.

In some scenarios, it is more desirable to design the caching problems in a distributed approach. In [142], a distributed cache placement approach is proposed in order to minimize the average download delay while some constraint for BSs storage capacities are met. The formulated optimization problem which is NP-hard is solved using a belief propagation based distributed algorithm with low complexity. This optimization problem makes sense because there is a trade-off between latency and storage capacity in caching networks. The comparison between the performance of the proposed distributed scheme with that of the centralized algorithm in [140] is presented as well. In [285], a decentralized content placement caching scheme is presented. Although there is no coordination, the proposed scheme can attain a rate as good as the optimal centralized scheme proposed in [140]. In [143], two caching and delivery schemes are considered. The first one operates in a centralized manner, while the second one is based on decentralized caching. For both cases, the trade-off between coded multi-casting and spatial reuse is reflected by the code length.

In addition to the aforementioned literature, in [144], content caching, and content delivery schemes are proposed considering cooperation to address the explosive enhancements of demand for mobile network applications. Defining the objective as minimizing the average downloading *latency*, it is demonstrated that the proposed content assignment and delivery policy scheme has a better performance in comparison to the previous known content caching schemes in terms of average downloading *latency*. A weighted optimization problem is formulated in [145] to minimize the traffic of backhaul and downlink while the constraints for cache memory size and bandwidth limitation for D2D communi-

cation is considered. It is shown in [301] that if latency awareness is considered in caching management, it is an effective approach to reduce the delivery time of *latency* sensitive applications, and the global delivery time of users. In the proposed model, two main advantages is claimed. First, it not only has a better performance in term of delivery time at the end-user, but also affects the link load reduction. Second, a faster convergence with respect to probabilistic caching is achieved. In [146], the effect of joint latency awareness and forwarding is investigated in a cache-enabled network. Authors proposed a scheme which is based on caching and forwarding strategies in order to improve E2E experienced latency by the UEs while there is no coordination among them.

In [286], a cooperative content caching approach between BSs in cache-enabled multi-cell network is considered. Due to the trade-off between storage and latency, cooperative caching optimization problem is designed in order to minimize the average delay while a constraint on the finite cache size at BSs is met. It is shown that cooperation among cells can considerably reduce delay in comparison to that of non-cooperative case. Moreover, the gains of the proposed scheme will be increased in more diverse and heavier load traffics. In [287], the aim of the work is to minimize the data transmission delay for the P2P caching system while considering the effect of cache size, all mobiles in a cell are considered as several P2P caching groups. Then, the problem is formulated as a stochastic optimization problem and solved using Markov decision process (MDP) to obtain the optimal solution.

In [147], a cooperative multicast-aware caching strategy is proposed for the BSs to decrease the average latency of content delivery in 5G cellular networks. The proposed scheme is carefully designed in order to take into account the benefit of multicast and cooperation while in the existing caching schemes, the popular content simply is brought close to the users. The optimization problem is formulated in order to minimize average latency for all the content requests. It is demonstrated that via various trace-driven sim-

ulations that the proposed cooperative multicast-aware caching scheme can provide up to 13% decrease in the average content-access latency in comparison to multicast-aware caching scheme with the same total cache capacity.

In [148], the authors presented a cooperative caching architecture in which multiple locally cache-enabled nodes of cloudlets interact cooperatively in a decentralized cloud service networks. By proposing a content distribution strategy, the problem is formulated so that the mean total content delivery delay for all users in the proposed scheme is minimized. It is confirmed that the approach can enhance the cache hit rate, and also reduce the the content delivery latency in comparison to existing solutions. In [149], the E2E packet transmission in a cache-enabled network is modeled in which both the wired backhaul and the RAN are jointly considered. The performances of both the on-peak and the off-peak network are investigated while both the wired backhaul and the RAN are considered. The E2E average packet latency is elaborated with the change of the request rate. It is shown that the average packet *latency* reduces in comparison to that of the system without caching ability due to the traffic offloading of the wired backhaul via caching.

5.7 Field Tests, Trials and Experiments

In this section, we present some representative field tests, trials and experiments for 5G low *latency*. The related literature is summarized in Table 5.20, where each of the individual references will be described in further detail below.

The study [315] presents SDR based hardware platform to verify the concept of 5G. This facilitates initial proof-of-concepts (PoC) of novel 5G air interface and other concepts by extending hardware-in-the-loop (HIL) experiments to small laboratory experiments and finally trials of outdoor tests. Such an SDR based hardware can demonstrate

Table 5.20: PLATFORM AND PERFORMANCE ON FIELD TESTS, TRIALS AND EXPERIMENTS.

Reference	Evaluation methodology	SDR	DSP	mm Wave	Conventional/Proprietary LTE	Remarks
[315]		✓				Trials of 5G concepts along with a novel air interface
[316]	✓					4 test cases and 15 KPIs is proposed.
[317]				✓		mmWave aggregation
[64]			✓			RTT latency of 1 or 2 ms is achieved. Moreover, to achieve latency on the order of couple of hundreds microseconds over air interface, cross later approach is recommended.
[318]		✓				Low latency VANET
[319]			✓			DSP round-trip latency less than $2\mu s$ is achieved for channel aggregation and de-aggregation for 48 20 MHz LTE signals.
[83]		✓				20 times latency reduction in comparison to existing works
[91]				✓		Minimum latency 3 ms
[64]			✓			Latency ≤ 1 or 2 ms
[87]					✓	
[320]					✓	Latency ≤ 17 ms
[60]	✓					HARQ RTT ≤ 1.5 ms
[321]	✓					RTT latency ≤ 1 ms

high-capacity, low *latency* and coverage capabilities of LTE-A solutions. In [316], evaluation methodology including some novel test environments and certain new key performance indicators are discussed in order to evaluate 5G network. Here, four candidate test environments such as indoor isolated environment and high speed train environment, and fifteen key performance indicators such as latency, throughput, network energy efficiency and device connection density are emphasized for performance evaluation.

In [317], 5G system operating at 15 GHz is presented followed by some experimental results. Here 0.2 ms subframe (14 OFDM symbols) is used for throughput, latency and other performance evaluation. The hardware implementation results of digital signal processing (DSP) and SDR based 5G system for *low latency* is presented in [64]. In this study, both the short TTI (sTTI) frame structures and wider subcarrier spacings are implemented in DSP platform. Based on the configurations of the system, RTT latency as low as 1 ms can be achieved. However, for achieving latency on the order of a few μs , optimization in between controllers and processing machines needs to be performed by cross-layer fashion. Additionally, the tail latency is argued to be considered in strict latency requirements assessment while maintaining required reliability.

An SDR based test bed is presented in [318] for cooperative automated driving with some experimental results from lab measurements. It implements flexible air interface consisting of re-configurable frame structure with fast-feedback, new pulse shaped OFDM (P-OFDM) waveform, low *latency* multiple-access scheme and robust hybrid synchronization, which ensure low latency high reliable communication. Results of the experimental trials are presented in [319], which utilizes DSP techniques for channel aggregation and de-aggregation, adjacent channel leakage ratio reduction, frequency-domain windowing, and synchronous transmission of I/Q waveforms and code words used in control and management function. In the proposed experiment, transmission of 48 chunks of 20 MHz LTE signals using a common public radio interface of capacity 59 Gb/s can achieve

RTT DSP latency of less than $2 \mu\text{s}$ and mean error-vector magnitude of about 2.5% after fronthaul fiber communication. This mobile fronthaul technique shows the path towards ultra low latency integrated fiber/wireless access networks.

In [83], a multi-terminal massive SM-MIMO system is evaluated considering realistic scenarios. The authors developed a massive SM-MIMO OFDM system prototype utilizing multiple off-the-shelf SDR modules which serve as IoT terminals. Two linear detection schemes with diverse complexity levels were tested for detection instead of maximum likelihood detection (MLD) schemes. It demonstrates the similar real-time SINR performance of the MLD techniques along with 20 times *latency* reduction over existing works. The promising results urge the utilization of massive SM-MIMO systems for latency reduction and reliability enhancement in IoT transmissions. In [91], the performance of a lower latency frame structure was evaluated in field tests using a 5G mmWave proof-of-concept (PoC) system. It was found that the slot interleaving frame structure can achieve RTT *latency* of 3 ms in the 70 – 80% of the trial course. Additionally, beam tracking and throughput performance were evaluated in field tests at a speed up to 20 km/h on LOS outdoor environment. It was confirmed that mmWave system can obtain throughput of 1 Gbps in the 38% of the trial time at 20 km/h speed.

In [87], a low complexity receiver design was introduced followed by verification of superiority of an SCMA system via simulations and real-time prototyping. It can provide up to 300% overloading that triples the whole system throughput while still enjoying the link performance close to orthogonal transmissions. In [320], the concept of MEC was introduced first time for 5G followed by promising field tests. The MEC was tested and analyzed on various cases including local breakout and network E2E *latency*. It was concluded that MEC can support low *latency* services of not lower than 17 ms. It also urged that stricter requirement of latency needs to be investigated from the new radio technologies or D2D communication. In [60], a lab trial is presented to study the feasi-

bility of ultra-low latency for 5G. It is shown that 1.5 ms HARQ RTT for TDD downlink in a lab trial is achievable when using the available test equipment in the literature while a novel frame structure and the associated signaling procedure is employed. The proposed scheme has 5 times better *latency* performance in comparison to the existing LTE-Advanced standard.

In [321], a novel frame structure is tested using a proprietary quasi-static system simulator for ultra-dense 5G outdoor RANs. In this regard, a frame structure is designed in order to facilitate low latency and multiuser spatial multiplexing on radio interface along with small-scale packet transmission and mobility support. It is found that performance of the introduced 5G network is better than that of LTE in case of air interface *latency*. In particular, considering UL scheduling requests in the RTT latency, the proposed frame structure can achieve *latency* as low as 0.8365 ms which is reduced by a factor of 5 in comparison to that of LTE. This satisfies 5G *latency* requirement (i.e., 1 ms latency).

5.8 Open Issues, Challenges and Future Research Directions

While there are some existing proposals to reduce latency to 1 ms, there are several open issues and challenges for future research. The area of exploration includes RAN, core network, backhauling, caching and resource management. Also, the existing techniques need to be validated in field tests and should evolve from current LTE systems. In the following section, we discuss some of the open issues and challenges which needed to be explored and addressed by researchers from both academia and industry.

5.8.1 RAN Issues

As discussed in Section IV, most of the fundamental constraints for achieving low latency requires modification in PHY and MAC layer which are at RAN level. Even though

several promising solutions are proposed to date, we believe that the following issues at RAN level need to be investigated further.

- For achieving low latency in 5G networks, mmWave is a promising technology which brings massive new spectrum for communications in the 3-300 GHz band. However, mmWave is dependent on diverse aspects such as transmitter/receiver location and environmental topology [194, 322]. Moreover, channel modeling with delay spread, path loss, NLOS beam forming, and angular spread need to be investigated in indoor and outdoor environment which are still evolving [323]. Additionally, more in-depth knowledge of physics behind mmWave regarding aspects such as Doppler, propagation, atmospheric absorption, reflection, refraction, attenuation, and multi-path should be developed for utilization of mmWave.
- In conventional packet transmission, distortion and thermal noise induced by propagation channel get averaged due to large size of packet [58]. However, in case of small size packet, such averaging is not possible. Thus, proper channel modeling followed by simulations and field tests for small packet in diverse carrier bands need be investigated.
- The challenge of admission control in RAN for spectral and energy efficiency with latency constraint is not well explored [324]. CRAN/HRAN provides spectral and energy efficiency while aspects such as caching can ensure low latency [256]. Researchers can work for performance bounds regarding this issue.
- Orthogonality and synchronization is a major drawback of OFDM modulation for achieving low latency. On the other hand, orthogonality and synchronization are very important for data readability. Recently, different non-orthogonal and asynchronous multiple access schemes such as SCMA, IDMA and GFDM, FBMC and UFMC have been proposed. However, more effective access techniques and wave-

forms which require less coordination, ensure robustness in disperse channel, and provide high spectral efficiency is a potential research area [194,325].

- Low complexity antenna, beam steering large antenna array and efficient symbol detection such as compressed sensing is conducive for low latency communication. For this, heuristic beam forming design at BS level, beam training protocols, and weight calculation and reliable error correction technique need to be studied [194, 326]. On the other hand, at the receiver level, low complexity sensing techniques and receiver design should be the focus of the future research.
- One of the major challenges is that latency critical packets are to be multiplexed with other packets. There are solutions such as instant access for latency critical packets ceasing transmission of other packets, and reservation of resources for latency critical services [198]. We believe that these issues are not well-explored and calls for further study.
- Even though the main objective of CRAN is to reduce costs and to enhance energy and spectral efficiency, it might be combined with heterogeneous networks termed as H-RAN. However, it is very challenging to design 5G network with large CRAN/HRAN [256]. In this case, various trade-offs including energy efficiency versus latency can be investigated.

5.8.2 Core Network Issues

In the core network, several new entities such as SDN and NFV have been introduced for supporting large capacity, massive connectivity and low latency with seamless operation. Since these entities are not part of the legacy LTE system, extensive works need to be carried out for the standardization and development in the context of 5G. Some of the issues necessitating research in the core network level are as follows.

- The main challenge of SDN/NFV-based core network design is the management and orchestration of these heterogeneous resources [327]. Effective resource allocation and implementation of functions in this heterogeneous environment while also maintaining low latency is an area of emerging research need.
- Most of the surveyed works are based on Open Flow protocols and their extension integrating control plane and user plane without a detailed implementation plane. Moreover, the scalability of the user plane is not considered while taking mainly control plane into consideration [272]. Researchers have a great opportunity to explore regarding the standardization and scalability of these core network entities.
- To boost spectral/energy efficiency and reduce latency utilizing CRAN and coordinated multi point (CoMP), mmWave is an attractive choice for front/back hauling because of its low implementation cost and unavailability of fiber everywhere [278]. However, research in mmWave regarding front/back hauling is a popular area of investigation. Dynamic, intelligent and adaptive techniques need to be developed with optimized utilization of the heterogeneous back hauling networks while catering low latency.
- Even though MEC is envisioned to reduce individual computation, inclusion of the caching in MEC can further boost users' QoE. The caching enabled MEC will provide content delivery and memory support for memory hungry applications such as VR, and online gaming along with BS level caching [34]. Researchers are encouraged to study various trade-offs such as capacity versus latency, storage versus link load, and memory versus rate.

5.8.3 Caching Issues

Edge caching can be a critical tool for latency reduction along spectral and energy efficiency improvement. Recently, this issue attracted huge attention from researchers in both academia and industry resulting in many different approaches. However, we believe there are still extensive research problems open and need further exploration.

- Even though several exciting works have been carried out for content placement and content delivery time, further studies could be done regarding the latency aspects such as how latency is impacted by caching size and location, and wireless channel parameters [328].
- Assuming that content delivery and content placement are the main two phases of wireless caching, network architecture including caching storage size, placement, and cooperation for caching are potential areas for further study [329,330]. Besides that, the protocol design for caching redundancy and intra cache communication can be investigated with latency constraint [330]. In this regard, performance limits and bounds of caching can be studied for getting insights on optimum performance.
- The BS makes a tunnel between UE and EPC for content request. However, the contents are packeted through GTP tunnel which creates difficulty in content-aware or object-orientated caching [331]. Proper protocol designs can address such problems.
- In low user density areas, caching capacity may be in surplus for serving UE, while in urban areas, the situation may be opposite. Intelligent and co-operative resource allocation and caching strategies can ensure proper hit ratio along with low latency content delivery [330]. In this regard, relatively few works are available in the literature, and can be further investigated.

- Mobility is an important issue in latency critical applications such as AR. The movement and trajectory boost the location and performance information for local caching which handle the current user's experience. The movement of users among cells will incur interference and pilot contamination along with complication in system configurations and user-server association policies. Moreover, frequent handover will introduce latency degrading user's experience [332]. Thus, handover in diverse caching scenarios with focus on low latency can be a potential area of research.

5.9 Conclusion

Along with very large capacity, massive connection density, and ultra high reliability, 5G networks will need to support ultra low latency. The low latency will enable new services such as VR/AR, tele-medicine and tele-surgery; in some cases, latency not more than 1 ms is critical. To achieve this low latency, drastic changes in multiple network domains need to be addressed. In this chapter, an extensive survey on different approaches in order to achieve low latency in 5G networks is presented. Different approaches are reviewed in the domain of RAN, core network and caching for achieving low latency. In the domain of RAN techniques, we have studied short frame/packets, new waveform designs, multiple access techniques, modulation and coding schemes, control channel approaches, symbol detection methods, transmission techniques, mmWave aggregation, cloud RAN, reinforcing QoS and QoE, and location aware communication as different aspects of facilitating low latency.

On the other hand, SDN, NFV and MEC/fog network architectures along with high speed backhaul are reviewed in the literature for core network with vision to meet the low latency requirements of 5G. The new core network will provide diverse advantages

such as distributed network functionality, independence of software platform from hardware platform, and separation of data plane from software plane, which will all help in latency reduction. In caching, distributed and centralized caching with various trade-offs, cache placement and content delivery have been proposed for latency reduction in content download. Following this, promising results from field tests, trials and experiments have also been presented here. However, more practical and efficient techniques in the presence of existing solutions need to be investigated before the standardization of 5G. In this regard, we discussed the open issues, challenges and future research directions for researchers. The authors believe that this survey will serve as a valuable resource for latency reduction for the emerging 5G cellular networks and beyond.

CHAPTER 6

RSS based Loop-free and Low Latency Compass Routing Protocol

Localized routing decision, loop freedom, minimization of hop count, single path routing, minimization of communication overhead, scalability and delivery rates are the metrics of communication protocol [24,333]. To meet these requirements, in this chapter, we propose compass routing technique [334] for data transfer among smart meters and other components of the AMI [24]. Compass routing was first introduced by Kranakis et al. [334], in which the source node sends message to neighboring nodes, measuring the angle with respect to the destination node.

Though compass routing is not inherently loop free, we used a modified algorithm in the context of AMI which not only avoids looping but also minimizes flooding rate and ensures faster delivery [24]. The touted feature of this routing protocol is that it uses the position information of smart meters derived from its radio signal. Though precise location information can be obtained from hard coded GPS installed inside meter, GPS may not work inside the buildings and in some geographically remote places (like forests). So our communication solution for the AMI is acceptable from a global perspective. The effectiveness of the proposed method has been investigated on stationary node environment and an encouraging result was found.

6.1 Localization and Routing Techniques

6.1.1 Localization of Smart Meter

As referred to Fig. 6.1, let us consider an unknown positioned (new) meter at a location (x, y) accompanied by partially dispersed known position meters (anchor meters) at locations (x_l, y_l) , where $1 \leq l \leq n$. The received signal strength at location (x_l, y_l) can be

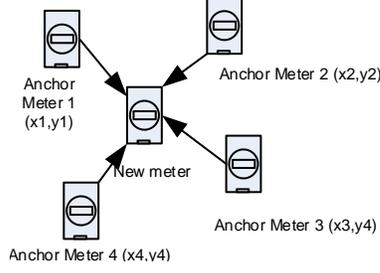


Figure 6.1: Locating the new meter by known position meter (Anchor).

denoted by ψ_l

$$\psi_l = c - 10\gamma \log(d_l) + w_l, \quad (6.1)$$

where c is an unknown constant that depends on transmitted power, frequency etc, and γ is the path loss constant. For a lossy environment, the typical value of γ is 4-6. In our model, $\gamma = 2.93$ has been used considering residential area. The parameter d_l is the euclidean distance between the known and unknown position meter defined as follows:

$$d_l = \sqrt{(x - x_l)^2 + (y - y_l)^2}, \quad (6.2)$$

and w_l is the zero mean random Gaussian noise with standard deviation σ_l . The value of σ_l ranges from 6 to 12 dB.

Let us define, the θ and ψ as $\theta = [x, y, z]^T$ and $\psi = [\psi_1, \psi_2, \dots, \psi_n]^T$.

The likelihood function of θ for a given RSS measurement ψ , $f(\theta|\psi)$ is given by

$$f(\theta|\psi) = c_1 \exp \left\{ - \sum_{l=1}^n \frac{\{\psi_l - c + 10\gamma \log(d_l)\}^2}{2\sigma_l^2} \right\}, \quad (6.3)$$

where c_1 is a constant.

The Maximum Likelihood (ML) estimate of θ , denoted by $\hat{\theta}$, can be found from the following equation

$$\begin{aligned}
\hat{\theta} &= \arg \max f(\theta|\psi) \\
&= \arg \min \left\{ - \sum_{l=1}^n \frac{\{\psi_l - c + 10\gamma \log(d_l)\}^2}{2\sigma_l^2} \right\},
\end{aligned} \tag{6.4}$$

The above equation is an optimization problem. Various optimization techniques such as differential evolution, dynamic relaxation and particle swarp optimization (PSO) can be used to solved equation 6.4. In our problem, we used PSO to solve the non-linear optimization problem. Finally, the ML estimator yields the location (x, y) and reference power z of the unknown positioned meter

$$(x, y, z) = \{\hat{\theta}(1), \hat{\theta}(2), \hat{\theta}(3)\}. \tag{6.5}$$

Optimization Technique- Deterministic Particle Swarm Optimization (D-PSO): The PSO is initialized with a group (population) of random solution of particles. Each particle has two states- its current position x and velocity v . Each particle has the ability to memorize its own best position experienced, $pbest$, and best position swarm experienced, $gbest$. At each iteration, the position and velocity are updated according to formula 6.6 and 6.7 [3]:

$$v_j^{t+1} = wv_j^t + c_1r_1(pbest_j^t - x_j^t) + c_2r_2(gbest_j^t - x_j^t) \tag{6.6}$$

$$x_j^{t+1} = x_j^t + v_j^{t+1} \tag{6.7}$$

where $w \geq 0$ is an inertia weight co-efficient, $c_1 > 0$ and $c_2 > 0$ are the acceleration coefficients, and $r1$ and $r2$ are the diagonal matrices whose components are generated from random number in the region $[0, 1]$.

$x_j^t \in \mathcal{R}^N$ is the position of j th particle in N dimensional space t iteration.

$v_j^t \in \mathcal{R}^N$ is the position of j th particle in N dimensional space t iteration.

$pbest_j^t$ called the best personal position information, indicates the best value of evaluation function of j th particle in t iteration. It is defined as

$$pbest_j^t = \min_{\tau} f(x_j^{\tau}), \tau \leq t \quad (6.8)$$

$gbest_j^t \in \mathcal{R}^N$ also called as the best global position information, indicates the best value of evaluation function of j th particle in swarm in t iteration. It is defined as

$$gbest_j^t = \min_{\tau} pbest_j^{\tau} = \min_{\tau} f(x_j^{\tau}), \tau \leq t \quad (6.9)$$

The particles will fly in the swarm in an N dimensional space to have best co-ordinate positions backed by best current (personal) value of evaluation function.

6.1.2 Compass Routing

In compass routing, the source or intermediate node $\mathbf{A}(x_1, y_1)$ sends a message to the next node based on location information. If the node $\mathbf{B}(x_2, y_2)$, $\mathbf{F}(x_3, y_3)$, and $\mathbf{L}(x_4, y_4)$ are in the transmission range of \mathbf{A} , \mathbf{A} will calculate angles as θ_1 , θ_2 and θ_3 , respectively of \mathbf{AF} , \mathbf{AL} and \mathbf{AB} with references to \mathbf{AD} , where \mathbf{D} is the destination node. The angle is calculated based on formula of Eqn. 6.10:

$$\theta_1 = \tan^{-1}\left(\frac{y_2 - y_1}{x_2 - x_1}\right) \quad (6.10)$$

At the same time, node \mathbf{A} calculates the distance of \mathbf{B} , \mathbf{F} , \mathbf{L} from \mathbf{D} by equation (6.11):

$$d = \sqrt{(x - x_1)^2 + (y - y_1)^2} \quad (6.11)$$

Based on $\theta = \min \theta_1, \theta_2, \theta_3$, node \mathbf{B} is selected.

At the same time, the distance from destination- d_{FD} , d_{LD} , and d_{BD} is measured. For $d = \min\{d_{FD}, d_{LD}, d_{BD}\}$, and $\theta = \min\{\theta_1, \theta_2, \theta_3\}$, the corresponding node is selected. If the node following $\theta = \min\{\theta_1, \theta_2, \theta_3\}$ and $d = \min\{d_{FD}, d_{LD}, d_{BD}\}$ is not same, the node with minimum distance d is selected irrespective value of minimum angle θ . Based on angle and distance, node B (Fig. 6.2) is selected. Now node B acts as the source node and it forwards a message to the next node. In this way, the routing path **ABCED** is selected and the message is transmitted from the source to destination as shown in Fig. 6.2.

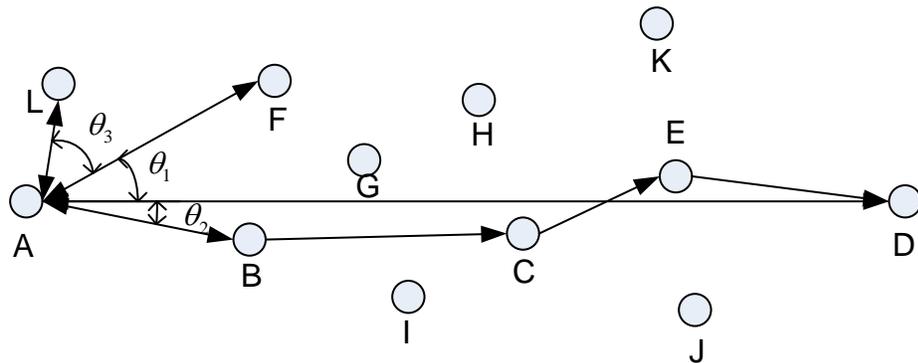


Figure 6.2: Routing diagram.

However, the main drawback of compass algorithm is that it is not loop free. Here in the Fig. 6.3, the message will be forwarded repeatedly in path A-B-C-E-F-A. So, the message is in loop.

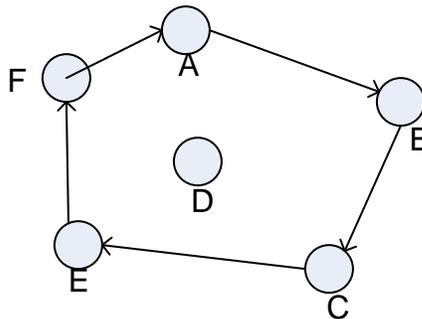


Figure 6.3: Loop in routing path.

To avoid this loop effect, the node which initiated the message looping, we define it as concave node. In the Fig. 6.3, node A is a concave node. We assume, in our algorithm, each message contains the source and destination ID. If a node is concave node, it will return back the message to its sender. We propose that when a node sends back a message, it will return also with the ID of the concave node. Next time, when the sender will send the message, it will send it in the alternate way omitting that path.

6.2 Simulation Result

To evaluate the performance of localization of meters, we use rectangle, hexagon and octagon shaped area of interests (AOIs) (approximate dimensions of 10x10m each). Each edge represents a node (in meters) and the emitter (in unknown node/meter) is the center of the AOI. The estimation of the position of the unknown emitter node (meter) through equation (6.3) is the optimization problem which is defined by equation (6.4). In our simulation, we used Deterministic Particle Swarm Optimization (D-PSO) method. Since, in our optimization problem, the space is 2 dimensional, we used parameters

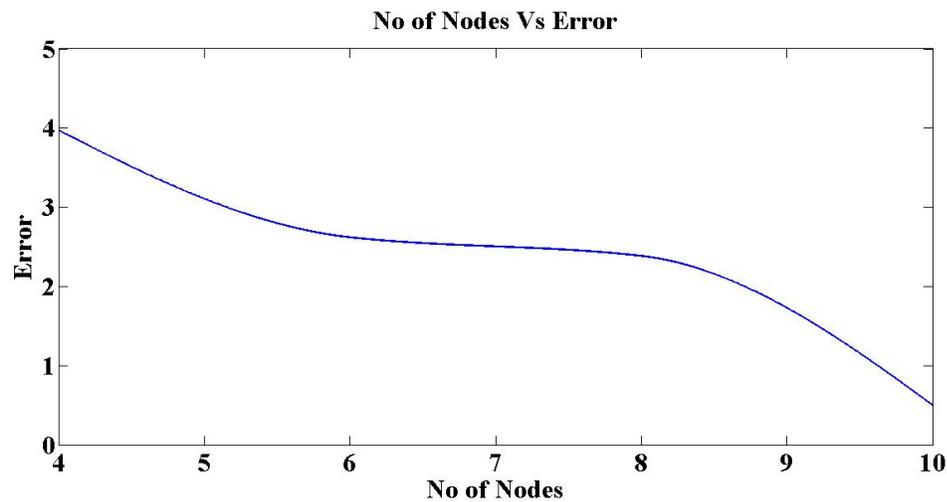


Figure 6.4: The error from exact position for different number of nodes.

Fig. 6.4 shows that the increase of node (in meters) number will decrease the error from exact node position.

In path loss model of radio signal, random noise is added which varies by standard deviation .

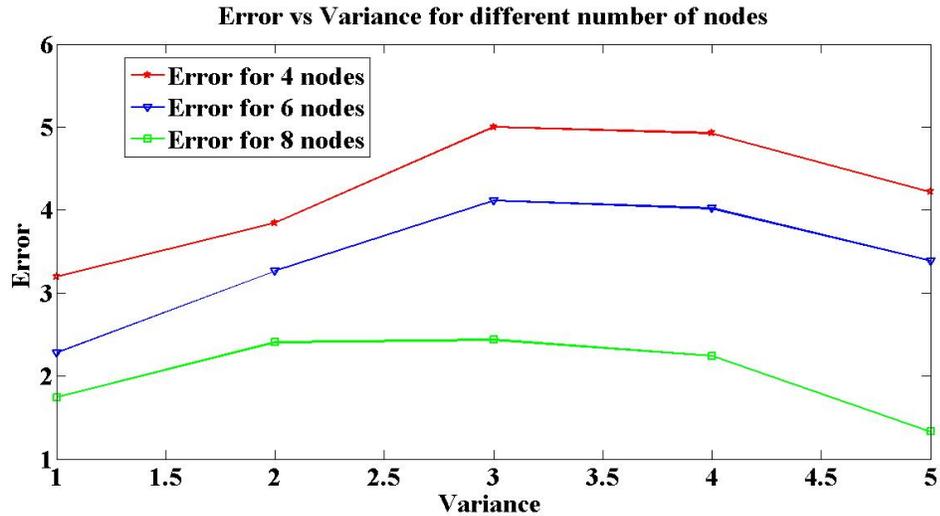


Figure 6.5: The relation between error and variance.

In the Fig. 6.5, the error curve is drawn as a function of variance for different number of nodes. As the variance of the noise increases, the error also goes up.

In the second part of our simulation, in a 10x10m box, we put 20 random nodes. It is observed that an increase in the degree of connectivity will decrease the number of hop as demonstrated in Fig. 6.6.

Now we use CSMA/CA algorithm to transmit data from the source to destination [335–337]. We used the following parameters: 500bit packets, 915MHz frequency and single hop based Markov Chain model for evaluating the performance. It is found in Fig. 6.7 that increasing the node number in the path of communication will decrease the reliability (probability of successful transmission).

Now we vary the packet size. We observe that with the increase of packet size, the reliability also decreases as shown in Fig. 6.7.

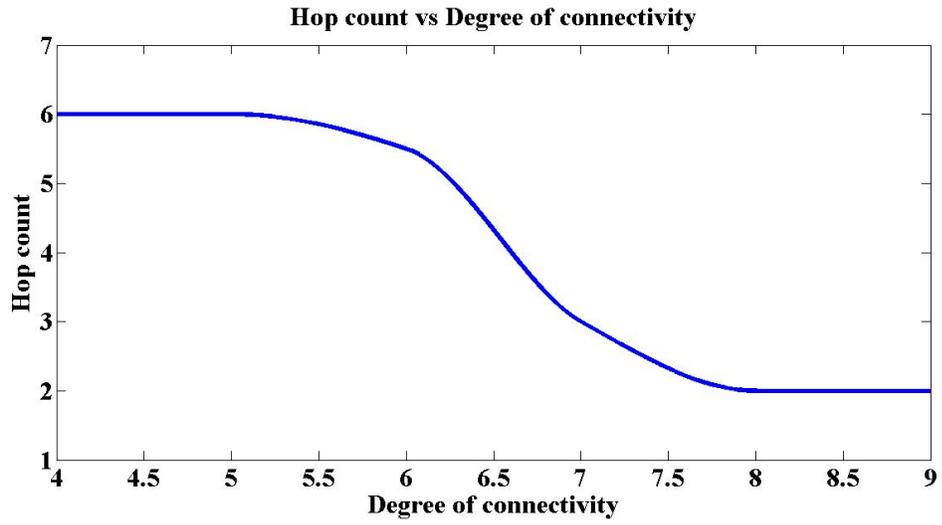


Figure 6.6: Degree of connectivity versus hop count.

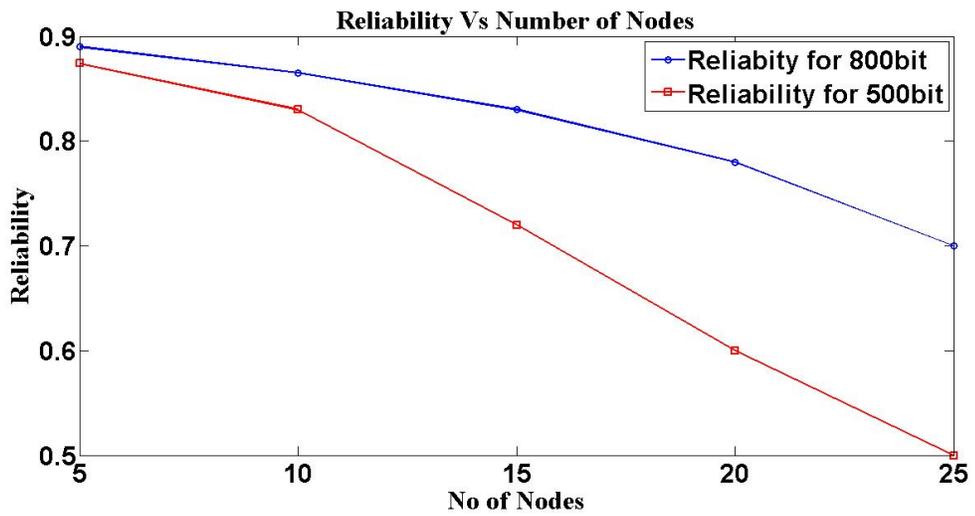


Figure 6.7: Reliability of nodes (meters) at different packet size.

6.3 Conclusion

In compass routing, the position of a smart meter is derived from the electromagnetic signal of neighboring meters. Though the pinpointing the position of a meter comes with error due to various obstacles (trees, buildings etc.), fading etc. occurs, the approximate position is sufficient for routing decision. Besides, in our routing protocol, the source or intermediate node (meter) sends packet to a specific neighboring node. So there is less flooding rate. Furthermore, for choosing only one node in the course of communication, unnecessary routing of data to the neighboring nodes is omitted resulting in a faster data transmission. Besides, to avoid the concave node, the corresponding path is dropped and its information is also updated in the corresponding neighboring nodes.

In the Smart Grid, a voluminous number of smart meters will cover huge geographical area replacing conventional meters. So RSS based compass routing protocol can be a potential candidate for AMI due to its localized routing decision, less flooding rate and faster data transmission, as well as omitting the need of hard coded location information or GPS inside the meter.

CHAPTER 7

Key Management and Authentication based Data Security for Smart Grid

The main obstacle for implementing AMI security scheme is the limited memory and low computational ability of the smart meters which necessitates a lightweight but robust security scheme [338–340]. In the literature, the key management-based encryption method has been accepted as a prominent security scheme for the smart grid, which includes a Trusted Third Party (TTP) [165, 168, 341, 342]. In almost all cases of TTP management systems, it is assumed that the TTP is fully trusted. However, the TTP itself, meters, and communication links among the TTP and meters also could be compromised.

Based on the semi-trusted servers and untrustworthy/unreliable communication links, in this chapter, we present a novel two-layer security scheme. The first layer boosts the security of the transmitted data between the Smart Meter (SM) node and the control center/Metering Data Management Service (MDMS) by data encryption as well as randomized packet transmission. As mentioned in our early work [343], the scheme consists of two independent servers. The master server manages the public-private key to secure the data packet communication between every smart meter and MDMS. On the other hand, the auxiliary server manages the transmitted sequence of the data packet (considering the public key sent by the master server). The private key associated with the public key and generated random sequence are used to retrieve the data at the MDMS. This paper extends [343] by using Received Signal Strength (RSS) and One Class Support Vector Machine (OCSVM) techniques for node-to-node authentications. OCSVM is used to detect malicious packets from unknown sources considering data transmission history like transmission frequency, data packet size, and distance between sender and receiver. RSS algorithm is used for localization of meters using the received signal strength from neighbor meters.

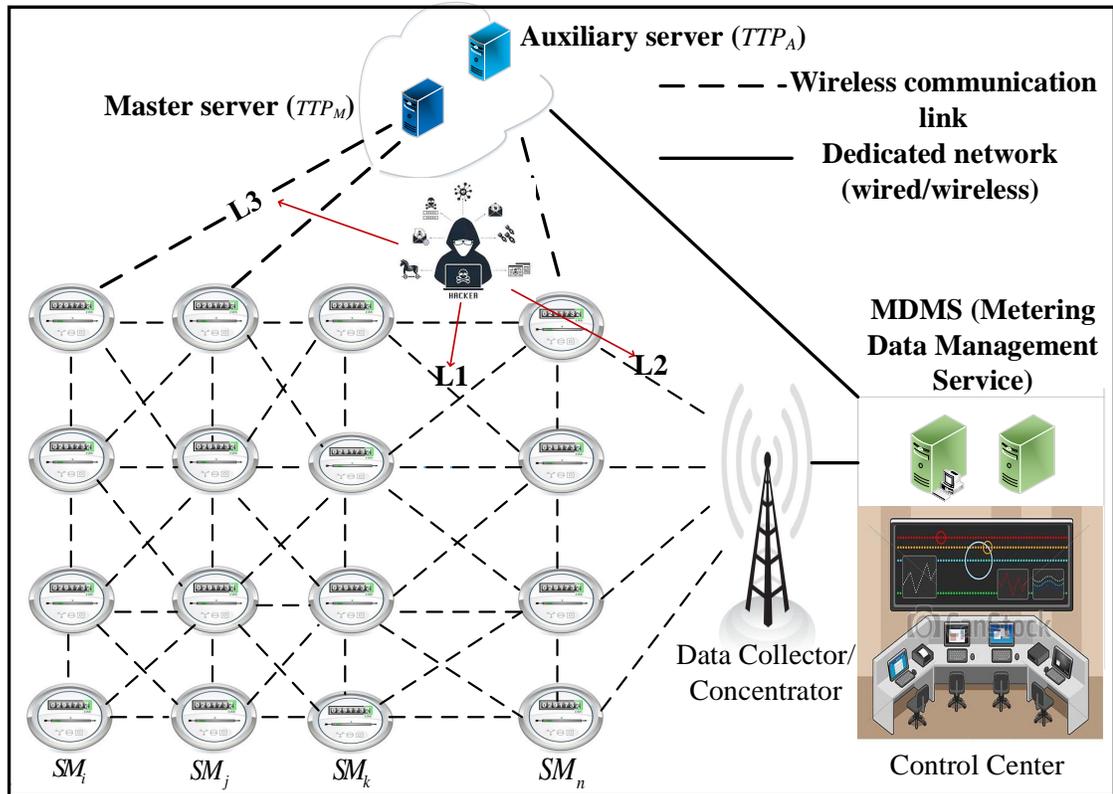


Figure 7.1: AMI architecture consisting of a cluster of mesh/hybrid connected meters, data concentrator, and control center.

7.1 Architecture of the Proposed Secured AMI

The proposed AMI is similar to a typical AMI which is a web-like network with millions of meters except it has two extra servers as shown in Fig. 7.1. Two mechanisms are proposed for the encryption of each data packet and authenticate the communication among meters.

A detailed description of the components of proposed AMI is provided here.

- *Smart meter*: It is a solid state device responsible for collecting, storing and sending data to MDMS using wireless communication in a fixed interval time less than 1 hour.

- *Home Area Network (HAN)*: All home appliances are connected to a smart meter by a network, forming HAN.
- *Neighborhood Area Network (NAN)*: The meters are connected to each other through a mesh or hierarchical or hybrid wireless network termed as NAN. In our architecture, we assume that NAN utilizes ZigBee protocol.
- *Data Collector/concentrator*: The head end of the NAN is the data concentrator or gateway which collects data from NAN and forwards those to MDMS by a dedicated wired or wireless connection (optical fiber, a cellular network, etc.).
- *Metering Data Management Service (MDMS) /control center*: The MDMS receives the consumption data from the AMI network, and calculates bills based on them. Having fine-grained collected data, MDMS also monitors, manages, and optimizes power generation and electricity distribution in the grid.
- *Master server*: Master server generates a pair of public key and private key for a SM for each session. The Master server unicasts the public key for encryption. On the other hand, the private key is sent to Auxiliary server and MDMS for decryption. In the proposed architecture, the connection between the Master server and NAN is via untrusted wireless communication whereas the Master and Auxiliary server are connected to MDMS by reliable communication (such as optical fiber, GPRS).
- *Auxiliary server*: Before the encryption of data with the public key sent by the master server, the smart meter generates a random sequence. This random sequence is sent as encrypted by public key to the Auxiliary server. That server receives random sequence and authenticates it by the private key of the smart meter before forwarding it to MDMS for final decryption.

Since in our model, a cluster of meters is supported by a Master and an Auxiliary server, our scheme can be scalable by prudent design of clusters.

7.1.1 Basic Notifications and Definitions

The notifications and definitions used in the presented algorithm and data flow scheme are stated in Table 7.1.

Table 7.1: SYMBOL NOTATION.

Notation	Description
SM_i	Smart meter node i
TTP_M	Master server
TTP_A	Auxiliary server
\mathcal{AE}_i	Asymmetric encryption scheme for meter i
\mathcal{K}_i	Randomized key generation algorithm for meter i
pk_i	Public key for meter i
sk_i	Secret/private key for meter i
M_i	Cleartext message/data of meter i
C_i	Ciphertext message/data of meter i
$\mathcal{E}(\{u, v\}, w)$	Encrypt the clear text u and v with the key w
$\mathcal{D}(\{y\}, q)$	Deterministically decrypt cipher text y with the key q
t	Time stamp instance
z_i^t	Message packet size for node i at time t
n	Number of packet segments for a given meter i at time t
$(c_1, c_2, \dots, c_n) \in C_i$	Segmented packets of cipher-text for meter i
$(r_1, r_2, \dots, r_n) \in R^t$	Random sequences at time instant t
$(p_1, p_2, \dots, p_n) \in P$	Probability of j th packet transmission
AP	Data concentrator/Access Point
γ	Path loss component
δ_l	Variance of random noise
SM	Set of all smart meters
L_x	Set of ZigBee connections, where $x \in [1, 3]$
Req_i	Key request message sent by SM_i to TTP_M
N	Number of smart meter nodes, i.e. $i \in [1, N]$
PSO	Particle Swarm Optimization
OCSVM	One class support Vector Machine

Let, $\mathbf{SM} = \{SM_i\}_{i=1}^N$ denotes the set of participating smart meters connected as a network in our system. Also, let graph $G = (\mathbf{SM} \cup \{\mathcal{AP}\}, L_1 \cup L_2 \cup L_3)$ represents the network topology of smart meters where:

- AP represents the data concentrator,

- L_1 denotes the set of ZigBee connections connecting neighboring smart meters together,
- L_2 represents the set of ZigBee connections connecting the data collector AP to a few nearby smart meters (see Fig. 1 for illustration)
- L_3 is untrusted communication links (such as ZigBee, WiMax etc.) between the Master sever TTP_M /Auxiliary sever TTP_A and every smart meter in **SM**,
- Both the TTP_M and TTP_A are connected to MDMS via a dedicated network. Additionally, they are connected by a trusted connection.

It is worth mentioning that each SM has a unique ID which is stored in the data packet, TTP_M and TTP_A . TTP_M and MDMS exchange the corresponding private key based on this ID. They both stored the corresponding private keys of the SM in their internal databases to retrieve at the time of packet arrival for decryption purpose.

For a node-to-node authentication of data packets, a built-in machine learning algorithm is proposed to run in each smart meter. The proposed algorithm explores all the incoming data packets in real time and identifies whether it is reliable or spiteful based on the three features of the packet including 1) the distance of the packet sender which is estimated by RSS-based localization algorithm, 2) time intervals that a packet is received in destination, and 3) packet size. Based on the decision of the algorithm, the packet is accepted and forwarded to the next SM in the grid, otherwise the packet will be discarded. Fig. 7.2 illustrates the mechanism.

7.2 RSS Algorithm, OCSVM and Entropy of a Data Packet

As mentioned earlier, packet data is circulated among the meters in the path reaching the destination AP . To point out the malicious packet data from an unauthorized source, the

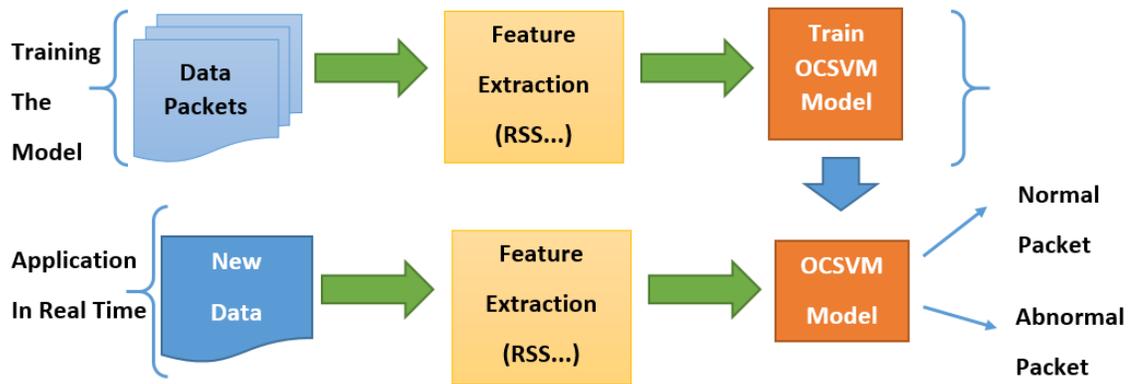


Figure 7.2: Training and application of OCSVM model in smart meters

content of the data packet should be screened carefully before delivery and dispatch to the next node in the grid. One class classification, or concept learning in the absence of counter examples, has the potential to tackle these kinds of problems. Among different implementations of the Support Vector Machine (SVM), one class classification algorithm (OCSVM) is selected in this paper and the performance compared with another anomaly detection algorithm: Isolation Forest [344, 345]. Appropriate features of the packet data should be extracted and fed into the OCSVM for training and later on for testing the new packet. Location (i.e. distance) of the sender, which is an informative feature for detecting the unauthorized source, is not directly defined in the data packet. To extract this feature, RSS algorithm is utilized. RSS can pinpoint the location of neighboring meters based on the received electromagnetic signals. Other features of the data packet such as packet size, and transmission frequency are relatively simple to capture or infer. In the following part, the RSS and OCSVM algorithm are discussed in details.

¹Unless or otherwise specified, node and meter are same thing in the rest of the study.

7.2.1 Received Signal Strength (RSS) based localization

Let us consider, an unknown positioned meter at a location (x, y) accompanied by partially dispersed known position meters at locations (x_l, y_l) , where $1 \leq l \leq n$. The received signal strength at location (x_l, y_l) from the unknown position meter can be denoted by ψ_l [30, 346, 347]

$$\psi_l = c - 10\gamma \log(d_l) + w_l, \quad (7.1)$$

where c is an unknown constant that depends on transmission power, frequency, etc., and γ is the path loss exponent. Path loss exponent defines the decay rate of electromagnetic signal. In our model, $\gamma = 2.93$ has been used considering residential area. The parameter d_l is the euclidean distance between the known and unknown position meter defined as follows:

$$d_l = \sqrt{(x - x_l)^2 + (y - y_l)^2}, \quad (7.2)$$

and w_l is the zero mean random Gaussian noise with standard deviation σ_l . The value of σ_l ranges from 6 to 12 dBm.

Let us define, the θ and ψ as $\theta = [x, y, z]^T$ and $\psi = [\psi_1, \psi_2, \dots, \psi_n]^T$, where z is the reference transmission power.

The likelihood function of θ for a given RSS measurement ψ , $f(\theta|\psi)$ is given by

$$f(\theta|\psi) = c_1 \exp \left\{ - \sum_{l=1}^n \frac{\{\psi_l - c + 10\gamma \log(d_l)\}^2}{2\sigma_l^2} \right\}, \quad (7.3)$$

where c_1 is a constant.

The Maximum Likelihood (ML) estimate of θ , denoted by $\hat{\theta}$, can be found from the following equation

$$\begin{aligned}\hat{\theta} &= \arg \max f(\theta|\psi) \\ &= \arg \min \left\{ - \sum_{l=1}^n \frac{\{\psi_l - c + 10\gamma \log(d_l)\}^2}{2\sigma_l^2} \right\},\end{aligned}\tag{7.4}$$

The above equation is an optimization problem. Various optimization techniques such as differential evolution, dynamic relaxation, and Particle Swarm Optimization (PSO) can be used to solve equation 7.4. In the current problem, PSO is used to solve the non-linear optimization problem. Finally, the ML estimator yields the location (x, y) and reference power z of the unknown positioned meter

$$(x, y, z) = \{\hat{\theta}(1), \hat{\theta}(2), \hat{\theta}(3)\}.\tag{7.5}$$

Now the distance between any two meters SM_i and SM_j is

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}.\tag{7.6}$$

where (x_i, y_i) and (x_j, y_j) are derived positions of meter SM_i and SM_j respectively.

The distance d_{ij} is used as a feature in OCSVM algorithm.

Since GPS doesn't work in some places such as inside the multi-stored building, hilly areas, forests, etc., we used RSS based localization over GPS. GPS may reveal exact position of meters/consumers. On the other hand, RSS technique will build a local map for meters.

7.2.2 OCSVM algorithm

One class classifier is inspired by the SVM classifier [348], [349]. One class classification problem is formulated to find a hyperplane that separates the desired fraction of the training patterns from the origin of the feature space.

OCSVM maps the input vector to feature dimension according to the kernel function, and separates it from the origin with maximum margin.

Let us consider, a set of training data $I = (i_1, i_2, \dots, i_n) \in \mathcal{I}$, and Ω be the feature map $\mathcal{I} \rightarrow \mathcal{H}$ such that the dot product of \mathcal{H} is computed by kernel k

$$k(i, i') = \langle \Omega(i), \Omega(i') \rangle_{\mathcal{H}} \quad (7.7)$$

The regular family for the data set

$$C_{w,\rho}^m = \{i | f_{w,\rho}(i) > 0\} \quad (7.8)$$

where $f_{w,\rho}(i) = \text{sgn}(\langle w, \Phi(i) \rangle - \rho)$ and (w, ρ) is the vector to offset parameterizing a hyperplane in the feature space associated with kernel.

$f_{w,\rho}$ is estimated by minimizing regularization

$$R^{reg}[f_{w,\rho}^m(\cdot)] = R^{emp}[f_{w,\rho}^m(\cdot)] + \frac{1}{2} \|f_{w,\rho}^m(\cdot)\|_{\mathcal{H}} \quad (7.9)$$

It penalizes the outliers by employing slack variables ξ in the objective function and controls carefully the trade off between empirical risk and regularizes the penalty.

The quadratic programming minimization function

$$\min_{w,\xi_i,\rho} \frac{1}{2} \|w\|^2 + \frac{1}{vn} \sum_{i=1}^n n\xi_i - \rho \quad (7.10)$$

$$\text{such that } (w \cdot \Phi(x_i)) \geq \rho - \xi_i, \quad (7.11)$$

$$\text{and } \xi_i \geq 0, \quad i = 1, 2, 3, \dots, n.$$

where Φ is the kernel function for mapping, ξ_i is the slack variables, $v \in (0, 1]$ is a prior fixed constant, and ρ is the decision value that determines whether a given point falls within the estimated high density region. Then the resultant decision function $f_{w,p}^m(x)$ takes the form

$$f(x) = \text{sgn}(w^{*T}\Phi(x) - \rho^*) \quad (7.12)$$

where ρ^* and w^* are the values of w and ρ solving from the equation (10).

In OCSVM, v characterizes the solution instead of c (smoothness operation) [350]:

- It determines an upper bound on the fraction of outliers.
- It is the lower bound on the number of training examples used as support vector.

7.2.3 Entropy of a data packet

Entropy is a metric for analyzing the robustness of an encryption methodology [351]. In other word, entropy demonstrates the feasibility degree of capturing the lock by chance. The more certain about a value, the smaller is the entropy value.

The entropy for a sequence S

$$H(S) = \sum_x P(S = x) \log_2 P(S = x)$$

where $P(S = x)$ is the probability of taking S a value x .

If the size of a random variable or packet generated by a meter is n bit, then the entropy and security strength of the data packet are n and 2^n , respectively. The higher the entropy, the harder the decryption process. For analyzing the performance of the proposed encryption schema, this metric is selected.

7.3 Privacy Scheme implementation and Data Traffic Flow

In this section, the privacy scheme implementation and the data flow process are described in details to clarify how each layer of the schema affects the grid security. In the data flow architecture, the following assumptions exist:

- The Master and Auxiliary servers are independent and semi-trusted. However, the servers might physically be one but virtually divided into two servers.
- The wireless communication links between servers and meters are not fully reliable.
- The meters have limited memory and computational ability.
- The control center/MDMS has the adequate computational ability.
- The meters keep the records of the position of neighboring meters, the frequency of transmission, packet size, and node identity. The frequency of transmission, node identity, and packet size are extracted from the packet header. The node position is derived from electromagnetic signals using RSS based localization as explained in previous section IV(A).
- Every meter transmits data at a constant transmission power.
- The data packet size is constant for every meter, and is 128 KB in the studied grid.
- When a new meter is installed, it starts to record the position of the neighbor meters, frequency of data transmissions, node identity, and packet size.
- The communication links among Master and Auxiliary servers and MDMS are fully trusted.

7.3.1 Attack Model

A simplified attack model for the system model shown in Fig. 7.1 is discussed here. Man-in-the-Middle (MITM) and replay are two of the most common attacks conducted

in the AMI. While MITM leverages the vulnerabilities in the communication medium to intercept, swap, corrupt or steal sensitive data, the replay attacks simply capture a copy of legitimate information and replay it in the future, posing to be the same legitimate user. To this end, it is assumed in this paper that the attacks can happen only in L_1 , L_2 or L_3 , but not between TTP_M , TTP_A and MDMS.

A critical assumption of this attack model is that the attacker's targets are solely on the communication channel between the devices, but not the devices themselves. It is assumed that the smart meter nodes, TTP_M , AP , TTP_A and MDMS are not compromised by the attacker, but only the links L_1 , L_2 or L_3 are. The incentive for the attacker could be offsetting energy consumption to achieve lower bills, stealing sensitive information to monitor energy consumption profile and through that spy on consumer behavior, or corrupting encrypted data packets in L_1 or L_2 . The proposed encryption scheme relies on the strength of the secret key sk , hence falling under the category of Public Key Encryption (PKI).

7.3.2 Data Traffic Flow

The data flow among the meters, servers, and MDMS (as shown in Fig. 7.3) is explained below in details.

STEP 1: Initialization

SM_i sends a request Req_i for a public key pk_i to the Master Server TTP_M . The Master server generates a public key-private key pair, (pk_i, sk_i) , upon receiving the request. While pk_i is unicasted to SM_i for data encryption, sk_i is sent to the Auxiliary server TTP_A and MDMS. Key generation by asymmetric algorithm:

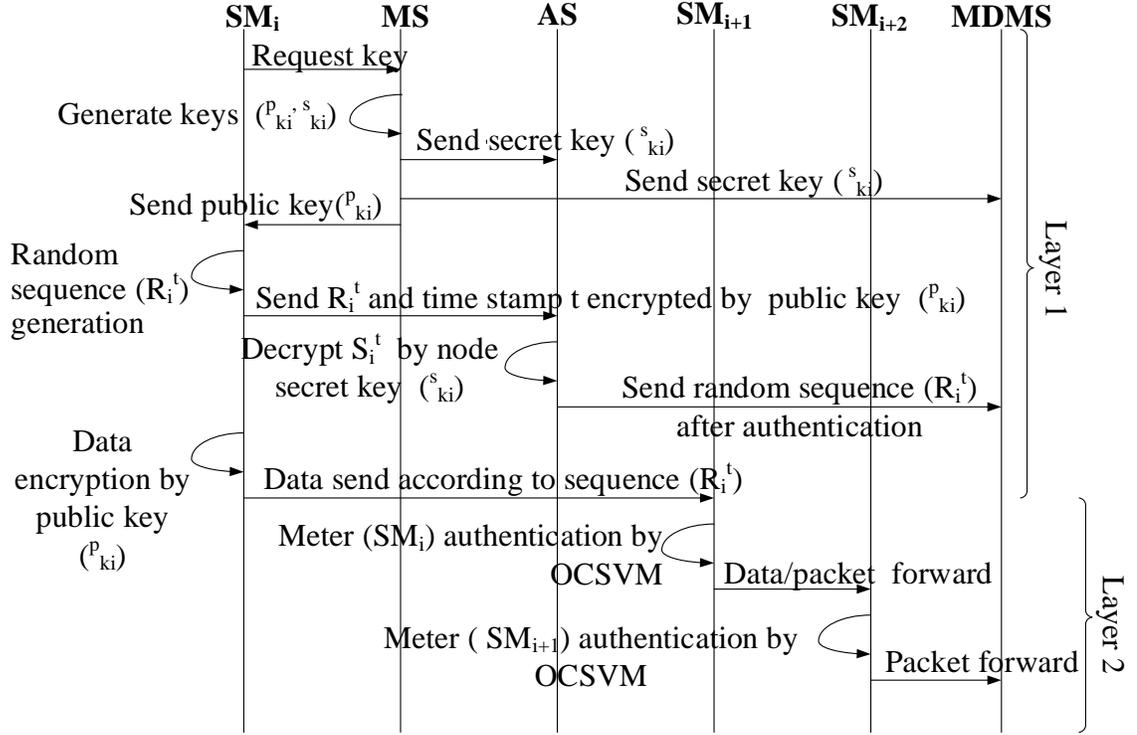


Figure 7.3: Data flow among various components of AMI.

$$\mathcal{AE}_i = (K_i, \mathcal{E}_i, \mathcal{D}_i), i \in [1, n] \quad (7.13)$$

$$K_i \longrightarrow (pk_i, sk_i), i \in [1, n] \quad (7.14)$$

STEP 2: Encryption

The SM_i generates a random sequence (R_i^t) for timestamp instance t . It encrypts both the sequence and the timestamp using pk_i to generate a ciphertext, S_i , and sends it to TTP_A . This server receives S_i and decodes it using sk_i . It then performs two steps to validate the sequence. To guard against the replay attacks, TTP_A ensures if the timestamp instance t it received is within a stipulated limit predefined for "freshness" of data. If it is within that limit, it considers the request, else it rejects the packet. To ensure whether the sequence was sent by a legitimate SM_i , the server re-encrypts the sequence it decoded along with the timestamp instance t using the pk_i it received from TTP_M . It compares the ciphertext

it created with the one it initially received. A discrepancy between the two ciphertexts prompts TTP_A to reject the packet. A match indicates the packet is indeed legitimate. In this way, data transmitted from SM_i is verified. After that, TTP_A sends the sequence R_i^t to MDMS. At the same time, SM_i encrypts its consumption data by the following method:

$$C_i \leftarrow \mathcal{E}(\{M_i, t\}, pk_i) \quad (7.15)$$

STEP 3: Data transmission

The encrypted data C_i is segmented into n packets of equal size, z_i^t .

$$(c_1, c_2, c_3, \dots, c_n) \leftarrow C_i \quad (7.16)$$

Then, the packets are ordered based on sequence R_i^t for transmission.

The transmitting algorithm is explained in Algorithm 2:

STEP 4: Hop to hop data aggregation and forwarding At first, meter SM_i is verified by SM_{i+1} using the node ID that is perceived from packet header. Afterwards, based on previous data receiving records including sender node's distance (d_{xy}), the frequency of data received (f_j) and packet size (z_j)- the node SM_{i+1} verifies source node SM_i and forwards data to the next node SM_{i+2} . OCSVM algorithm is used to authenticate in this process. The data aggregation and forwarding algorithm pseudo code is tabulated in Algorithm 3:

STEP 5: Data retrieval

The MDMS receives the randomized and encrypted packets and decodes them by the secret key pk_i and random sequence R_i^t .

Reordering the data:

$$(c_1, c_2, c_3, \dots, c_n) \xleftarrow{R_i^t} (c_3, c_1, c_4 \dots, c_n) \quad (7.17)$$

Algorithm 2 Transmitting algorithm

- 1: **Initialization:**
 - 2: Get i th meter's data packet C_i , $i \in \{1, 2, \dots, N\}$ and random sequence $R_i^t = \{r_j\}$,
 $j = \{1, 2, 3, \dots, n\}$, $\forall t \in T$ and $|R_i^t| \in \mathbb{N}$
 - 3: **if** $R_i^{t-1} = R_i^t$ **then**
 - 4: Go to step 2
 - 5: **else**
 - 6: Proceed to next step
 - 7: **end if**
 - 8: Segment packet $\mathcal{C} = \{c_j\} \leftarrow C_i$, $j = \{1, 2, 3, \dots, n\}$ and $|\mathcal{C}| = |R_i^t|$
 - 9: Set index set $J_{\mathcal{C}=l}$, $l = 1, 2, 3, \dots, n$ where $f : l \rightarrow \mathcal{C}$ is the particular enumeration
of \mathcal{C}
 - 10: Update $f^{-1}(c_j) = r_j$, $j = \{1, 2, 3, \dots, n\}$.
 - 11: Calculate transmission probability, $P = \{p_j\}$ where $p_j = \frac{1}{r_j} = \frac{1}{f^{-1}(c_j)}$, $j =$
 $\{1, 2, 3, \dots, n\}$.
 - 12: Set index set $K_P = \{k\}$, $k = 1, 2, 3, \dots, n$ where $g : k \rightarrow P$ is the particular
enumeration of P
 - 13: Sort index $l = f^{-1}(c_j) = g^{-1}(\max_{p_j \in P} p_j)$
 - 14: Transmit packet $f(l)$ indexed packet
 - 15: Update $\mathcal{C} = \mathcal{C} - \{f(l)\}$
 - 16: **if** $\mathcal{C} \neq \Phi$ **then**
 - 17: Go to step 9
 - 18: **else**
 - 19: End process
 - 20: **end if**
 - 21: **End**
-

Algorithm 3 Data aggregation and forwarding

- 1: **Training:**
 - 2: For any two meters $\{SM_x, SM_y\} \in \mathbf{SM}$ and time instant l , get regular data $I_{xy}^l = (d_{xy}^l, f_x^l, z_x^l)$, $x \in \{1, 2, \dots, N\}$, $y \in \{1, 2, \dots, N\}$
 - 3: For data set I_{xy}^l , define a family/boundary C_x^m through (8)
 - 4: **Meter/Packet authentication:**
 - 5: For time instant t , get relative distance $d_{i(i+1)}^t$ between source meter SM_i and data receiving meter SM_{i+1} through (6)
 - 6: Calculate data transmission frequency f_i^t and packet size z_i^t for packet from meter i
 - 7: For new data $I_{i(i+1)}^t = (d_{i(i+1)}^t, f_i^t, z_i^t)$, get decision by (12)
 - 8: **if** $I_{i(i+1)}^t \in C_i^m$ **then**
 - 9: The data is within the boundary, forward to next meter SM_{i+2}
 - 10: **else**
 - 11: Reject data from source SM_i which is flagged by algorithm as anomaly
 - 12: **end if**
 - 13: **End**
-

Message unification:

$$C_i \leftarrow (c_1, c_2, c_3, \dots, c_n) \quad (7.18)$$

Decryption:

$$\{M_i, t\} \leftarrow \mathcal{D}(C_i, sk_i) \quad (7.19)$$

7.4 Simulation Results and Performance Analysis

In this section, we present the performance of RSS and OCSVM in our proposed AMI architecture. We also discuss about the security strength of our scheme and compare the performance of the OCSVM algorithm with another state of the art anomaly detection algorithm named Isolation Forest.

7.4.1 Performance of Proposed Algorithm

To get insights into the localization of meters, we consider, Manhattan grid building topology [352] in which the distance between two meters (i.e. house) is 30m. We consider an

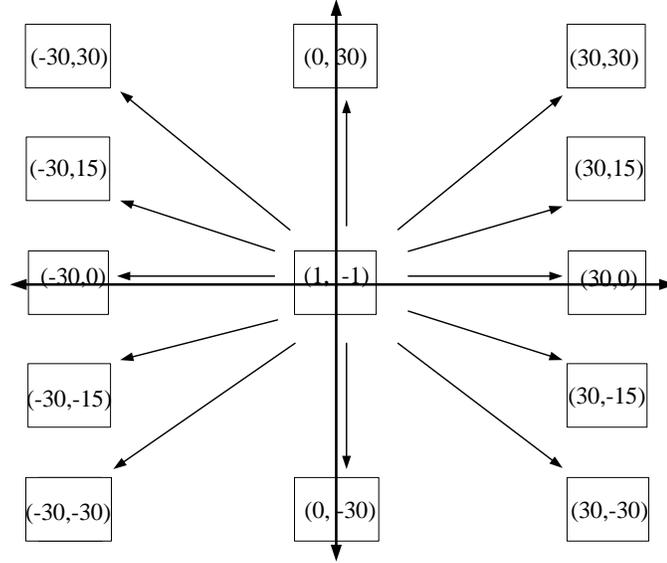


Figure 7.4: Meter positions in Manhattan grid.

area of interest (AOI) of $30\text{m} \times 30\text{m}$ where a new meter is located at $(1, -1)$ position surrounded by known position meters as illustrated in Fig. 7.4. For optimization problem in localization, PSO [353] was used whereas residential path loss model was considered for path loss calculation. The transmission power of each meter is 10 dBm, and iteration number and population size of PSO are 100 and 30 respectively. The simulation results of position and power of an unknown positioned meter surrounded by 4 known positioned meters and environment with path loss constant 3 are tabulated in Table 7.2.

Table 7.2: MEAN SQUARE ERROR (MSE) IN LOCALIZATION FOR DIFFERENT NOISE VARIANCES

Noise Variance	X	Y	Reference Power	MSE
2	-3.3059	-2.2506	10.0005	4.4839
4	-2.8715	-0.2784	9.9997	3.9324
6	7.3873	-1.9575	9.9841	6.4586
8	-4.8517	-5.9137	10.0175	7.6411
10	0.1728	-8.5027	9.9789	7.5482
12	2.3678	-8.8504	9.9818	7.9687

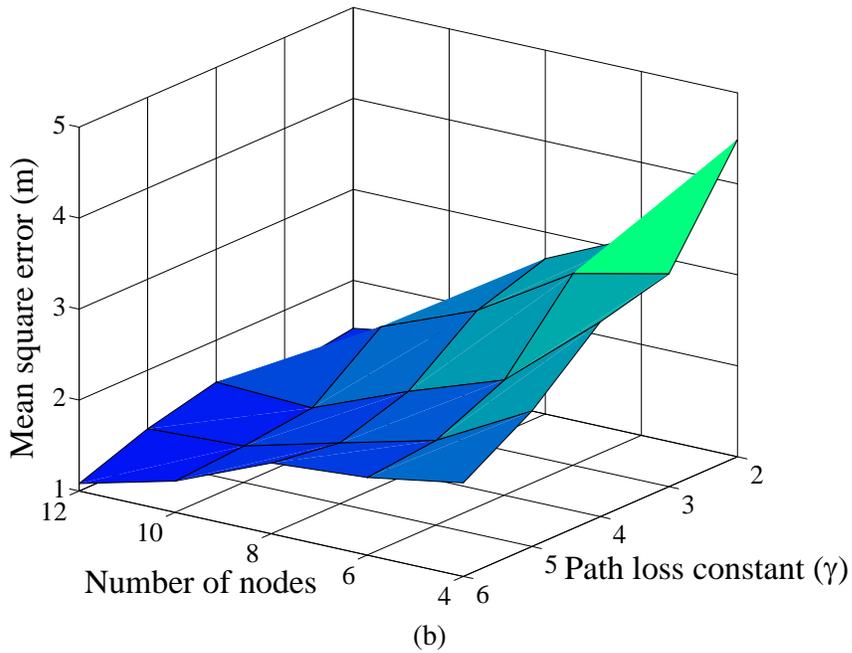
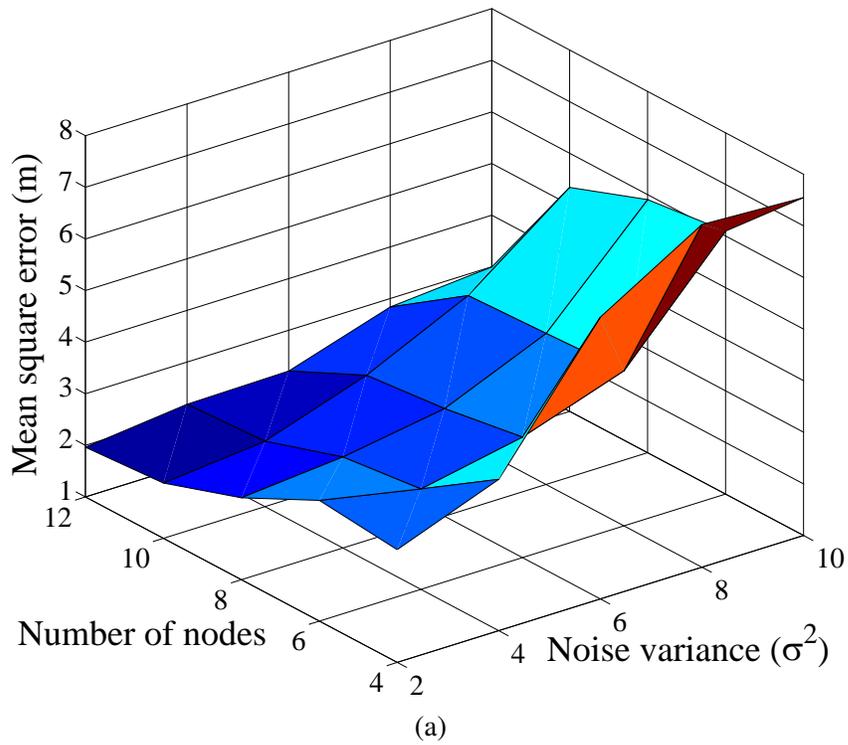


Figure 7.5: (a) Mean square error for different number of nodes and noise variances. (b) Mean square error for different number of nodes and path loss exponent.

With an increase in the number of neighboring meters, the Mean Square Error (MSE) from the exact position of the meter decreases. This means that for the more number of neighboring meters, localization error for a meter will be lower. On the other hand, for the increase of noise variance, MSE also increases. This implies that localization error for a meter is higher for the increased interference or noise. These are illustrated in Fig. 7.5a. For the increase of both neighboring meters and path loss exponent, the MSE decreases from the exact position of meter which is reflected in Fig. 7.5b. Path loss exponent (decaying rate of signal) is associated with the obstacle in the path of electromagnetic signal propagation. Therefore, for the presence of buildings, walls, trees, etc., the error in determining the location of meters will be lower. Furthermore, since the meters are mounted on a stationary wall/pole, and the environment surrounding meters is stable, the calculated error for a meter by RSS based localization method will be almost constant.

In the second part of the simulation, OCSVM is implemented in python using scikit-learn library [354]. While no public dataset in the topic is available and obtaining the real data is not possible, we generated a synthesized dataset simulating only the actual behavior of the normal data packets due to unknown structure of the malicious packets. The structure of the normal packets and their transmission behavior simulated based on the defined standard of the network with a small variation. Regarding the current network standard, each packet data was generated with the following information: meter's position, the frequency of transmission, and the packet size. Some real world data transferring from two smart meters in an AMI network of a local utility company is illustrated in Table 7.3.

Table 7.3: SMART METER REAL WORLD DATA SAMPLE

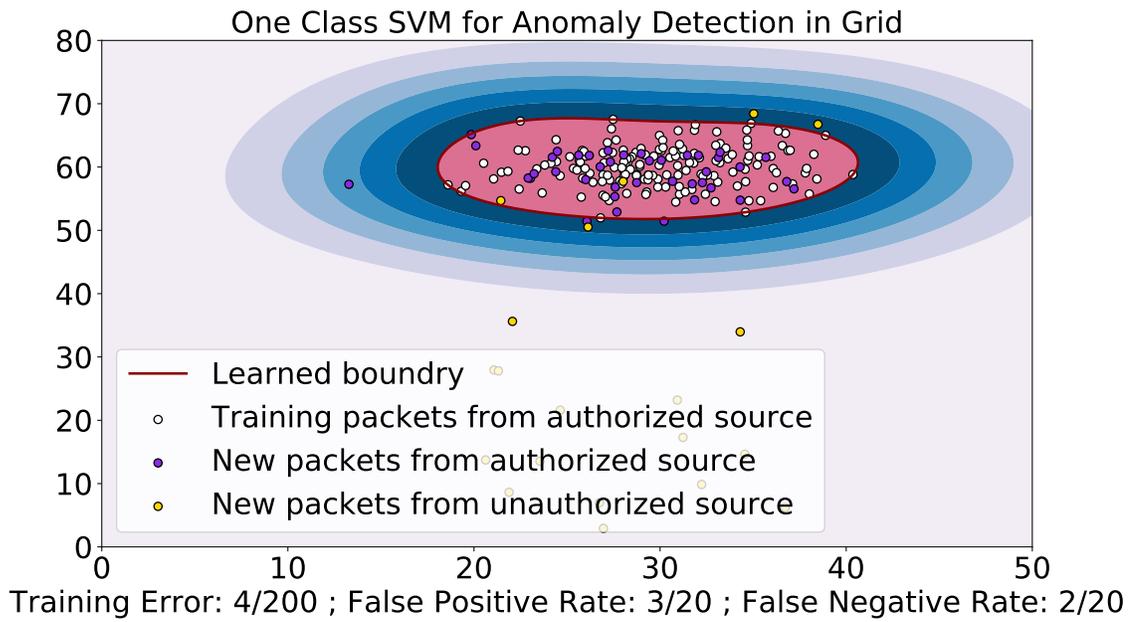
Device-Name	Read-Start-Time	Usage-Unit	Meter ID	Coordinate
G0034501624	5/1/2013 0:00 1:00	0.3218 kWh	98747434	8668675880
G0034501624	5/1/2013 1:00 2:00	0.2757 kWh	98747434	8668675880
G0034501624	5/1/2013 2:00 3:00	0.3561 kWh	98747434	8668675880
G0034501637	5/1/2013 0:00 1:00	0.4587 kWh	98747434	8728604300
G0034501637	5/1/2013 1:00 2:00	0.4101 kWh	98747434	8728604300
G0034501637	5/1/2013 2:00 3:00	0.1346 kWh	98747434	8728604300

Based on the real world utility data (referred to Table 7.3), we defined data transmission frequency equal to one hour. The distance between two meters is assumed to be 30m considering Manhattan Grid. The packet size is directly estimated upon the delivery of packet, and the standard packet size is considered 128KB based on the network topology. The training data is generated with 3 degrees of standard deviation from normal distribution of meter distance, data transmission frequency, and data packet size where the mean of meter distance, data transmission frequency, and packet size are 30m, one hour, and 128KB, respectively.

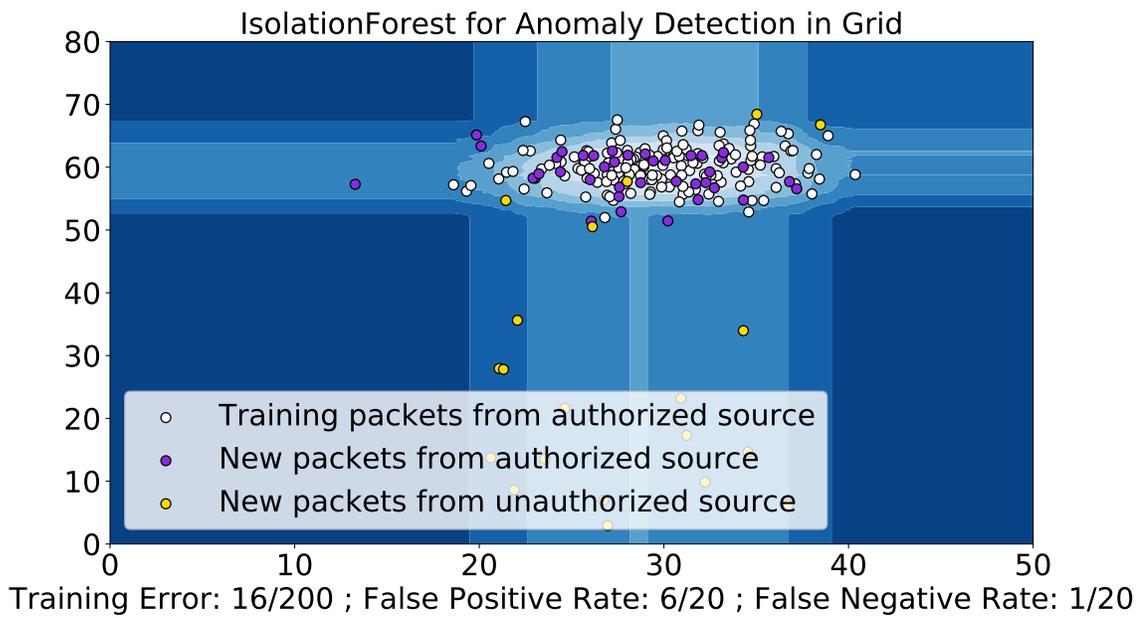
The OCSVM model was trained on the 70% of the data and the model was tuned on the validation set, 10% of the remaining data, with an exhaustive grid search, resulting in the RBF (Guassian) kernel with ν and γ both equal to 0.01. The model was tested by the rest of the data. The mapped decision boundaries of the OCSVM with the best parameter settings is shown in Fig. 7.6. Red lines show the decision boundaries and yellow dots are the packets from unauthorized sources. The training error², false positive rate³ and

²The training error is the ratio between the number of normal data that falls outside of the boundary erroneously and total number of data.

³The false positive rate is the ratio between the number of negative events wrongly categorized as positive (false positives) and the total number of actual negative events (regardless of classification).



(a)



(b)

Figure 7.6: Anomaly Detection with a) One Class SVM and b) Isolation Forest Algorithms

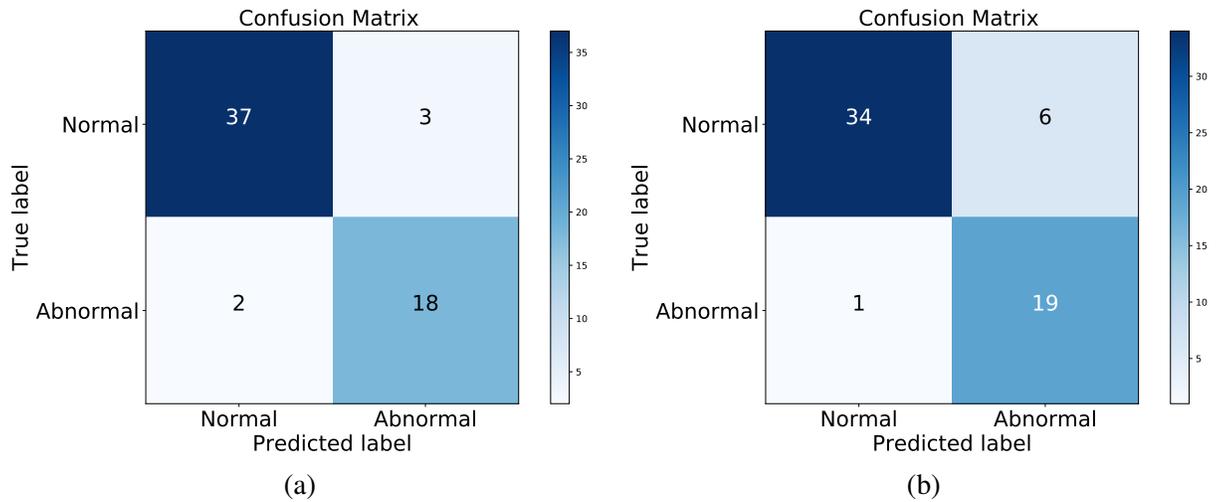


Figure 7.7: Confusion Matrix for a) One Class SVM, and b) Isolation Forest

false negative rate⁴ of OCSVM are $4/200$, $3/20$, and $2/20$, respectively. On the other hand, the training error, false positive rate and false negative rate for Isolation Forest are $16/200$, $6/20$, and $1/20$, respectively. Therefore, it can be conceived that OCSVM can discriminate the authorized and unauthorized (malicious) packets almost precisely in comparison to its counterpart Isolation Forest algorithm.

To evaluate the model's performance, confusion matrix and receiver operating characteristic (ROC) curve [355, 356] are provided. The statistics of true positive, false positive, true negative and false negative of OCSVM and Isolation Forest are illustrated in Fig. 7.7. Referred to the figure, the overall accuracy of OCSVM is $(\frac{37+18}{60}) \times 100 = 91.6\%$. On the other hand, the accuracy of Isolation Forest is $(\frac{34+19}{60}) \times 100 = 88.33\%$.

Receiver Operator Characteristic (ROC) curve plots the percentage of normal samples labeled as abnormal versus the percentage of true positives. ROC for this experiment with different tweaked parameters is shown in Fig. 7.8. It is noted that the area under the curve slightly changes based on the different hyper parameter setting, and for $\gamma = 0.01$ and

⁴The false negative rate is the ratio between the number of positive events wrongly categorized as false (false positives) and the total number of actual negative events (regardless of classification)

$\nu = 0.01$, the maximum area of 0.95 is achieved; therefore the aforementioned parameters are selected as the best setting for the model.

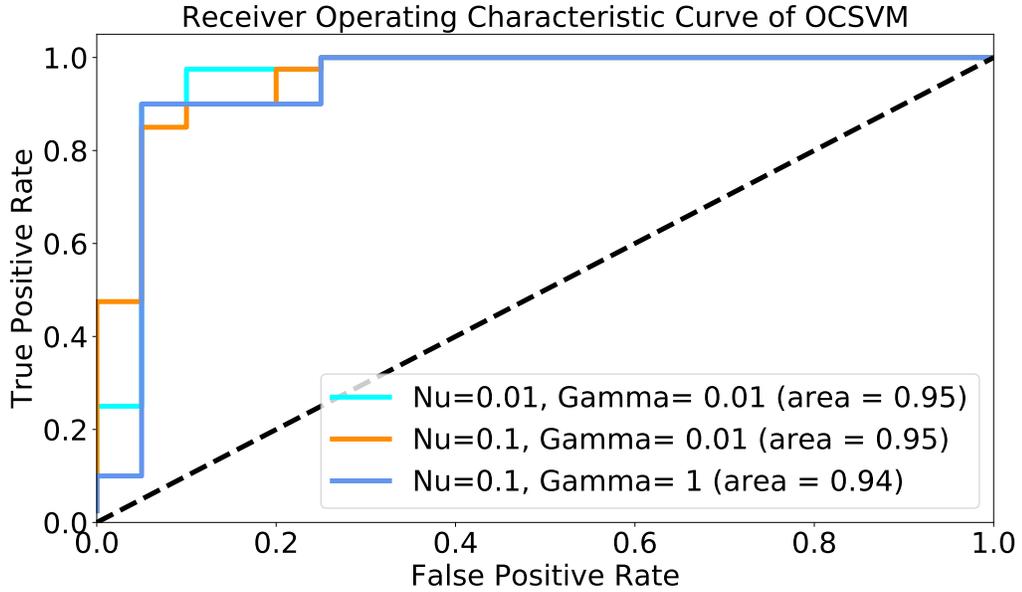


Figure 7.8: ROC curve for OCSVM with different parameter settings. (Nu= ν ; Gamma= γ)

7.4.2 Security Strength Analysis

Let us assume, a meter generates a consumption unit packet of size 128KB (1024×10^3 bit) which is divided into 4000 blocks with each block size of 256 bit. If each block is encrypted by 256 bit public/asymmetric key and transmitted according to a random sequence, then the entropy of each block is 256. The security strength of the data block is 2^{256} .

Furthermore, the security strength of a 256 bit public key is $2^{256/2}$.

So, for 4000 random sequenced packets and 256 bit public key,

Total security strength of the 128KB meter data = $4000 * (2^{256} + 2^{256/2})$

Hence, a hacker needs maximum $4000 \times (2^{256} + 2^{256/2})$ number of iterations (tries) to decrypt a message, which is impractical.

7.5 Conclusion

In our security scheme, a two-level security method has been proposed with data encryption and node authentication. In the data encryption level, encryption by asymmetric keys and randomization of data packets have been proposed. In the conventional key management system, only data encryption is used. On the other hand, in our scheme, randomization of packets along with data encryption ensures enhanced data security. Another contribution of our scheme is the introduction of node-to-node authentication by OCSVM, which utilizes three features- frequency of data reception from a specific meter, packet size, and meter position. The information of data frequency and packet size is easily extractable from the packet header. On the other hand, for derivation of meter's position RSS is used, and it is almost constant due to the stationary position of the meters.

For TTP-to-smart meter communication, we use a bi-directional communication similar to meter data communication of conventional AMI network. Since the communication between meters and servers happens once per every session of sending the meter data to the control center, it doesn't hamper the normal meter data flow from meter to control center. Furthermore, since a random sequence along with the asymmetric key is used to retrieve data in the control center, it also helps to verify data flow from a specific smart meter. Additionally, being a cluster of meters is served by a TTP (a Master and an Auxiliary server), prudent design of clusters can make our approach scalable easily.

CHAPTER 8

A Machine Learning and Geo-location based Data Security Scheme for Smart Grid

The main hindrance of implementing security schemes in advance metering infrastructure (AMI) is the limited memory and computational capability of smart meters [30]. Additionally, AMI is a huge network comprised of thousands of meters. This requires AMI to have light but robust security scheme. In geo-location based encryption scheme [357,358], for localization global positioning system (GPS) was proposed. However, GPS does not work well in some places such as inside the multi-stored building, hilly places as well as coordinate derived by GPS will expose the exact location of consumer house.

In this chapter, we present a key management based security scheme utilizing the location of meters, derived from Received Signal Strength (RSS) of radio signal [30]. The localization of meters by RSS method will create a local positioning map different from geographic coordinate system, in which every meter has its own coordinate. For data encryption, secret keys mapped with the coordinate points of the meters and a random index was proposed in our technique. The keys are distributed among the meters periodically by a trusted third party (TTP) of the key management system. Furthermore, we introduce k-Nearest Neighbors (kNN) algorithm for meter authentication during the transport of data packets. kNN algorithm is a technique used to predict class labels of unknown data [359–361]. kNN classifier is simple, efficient and easy to implement. It is one of the most widely used algorithms in pattern evaluation, text characterization, diagnosis of cancer and many more. In a real world scenario, there are many data sets with little or no prior knowledge about its distribution. kNN is amongst the best choice for classification with data set with little or no prior knowledge. For these reasons, the combination of data encryption by secret key and node authentication using kNN algorithm provides a potential solution for AMI.

8.1 Challenges Faced by AMI Meters

Like any other systems, the AMI needs to fulfill four primary requirements of security viz. confidentiality, integrity, availability, and accountability (non-repudiation). Confidentiality implies that data must be accessible only to the authorized users and all unauthorized attempt must be denied. Since fine grained consumption data of a smart meter conveys consumers' lifestyle patterns, habits and energy usage, they must be concealed. Integrity requires reflecting authentic data correctly without any modification, addition or deletion. Since the hackers as well as the consumers might want to alter the consumption data, integrity is a vital issue in the AMI data.

Availability means that the data must be available on demand at all times for authorized users of the system. Availability follows the concept of authorization which in turn implies the data in the system can be used only by users who are allowed to have access. This involves the concept of access controls wherein not all users have the same degree of freedom and control over the data set of the system. There are restrictions to using specific aspects of data which ensures not everything can be accessed by everyone. Availability takes this one step further by ensuring the accessible data must not be denied to the user by the system at any point of time. Since the adversaries might want to jam the network thereby preventing the system from making the data available, or much worse, incapacitating the system's feature to make the data available, the AMI must comply with this requirement. Accountability (non repudiation) means that an entity doing a specific job must not deny it to do that. In AMI, accountability ensures timely responses to the command and control, and integrity of billing profile etc.

End-user privacy is another challenge of AMI data security. Smart meters are essentially small banks of customer usage snapshots, when aggregated together over a period of time provides an immense wealth of information that if put to the wrong use might

compromise the privacy of customers. Smart meters provide data which is usually granular or fine-grained, and high-frequency type of energy measurements whose illegitimate analysis results in or may result in the invasion of privacy, near real-time surveillance, and behavioral profiling. When the analysis is coupled with an even more threatening hazard such as manipulation of the analyzed data, the attackers get to open a window to observe how many people are at home and at what times, to determine people's sleeping and eating routines, appliance usage patterns, and home vacancy patterns.

Taking it one step further, hackers become capable of wirelessly updating smart meter firmware and remotely disconnect a user or a large section of users. Attackers, armed with different consumer patterns, can stage efficient electricity thefts and frauds, running up bogus charges or swift an electrical appliance to malfunction, shutdown or surge, causing physical damage to life and property.

The AMI meters are inherently susceptible to buffer overflows and the seven state machine flaws as illustrated in [362]. Attacks that exploit its hardware vulnerabilities, such as bus sniffing, clock speed and power glitch are also prevalent [363–366]. An attacker can create abnormal operating conditions by varying the time and voltage levels crucial to the meter performance, consequently gaining access to previously inaccessible parts of the system. Exposing the chip's surface to lasers, micro-probing to inject false signals, capturing or intercepting data and manipulating registers are some of the more advanced methods employed to compromise the meter's integrity in a physical as well as cyber fashion. In the recent times, differential power analysis and other similar techniques have been successfully used to extract the secret keys and circumvent the embedded IC security mechanisms altogether, as shown in [367]. So all these issues need to be addressed in the data security of AMI.

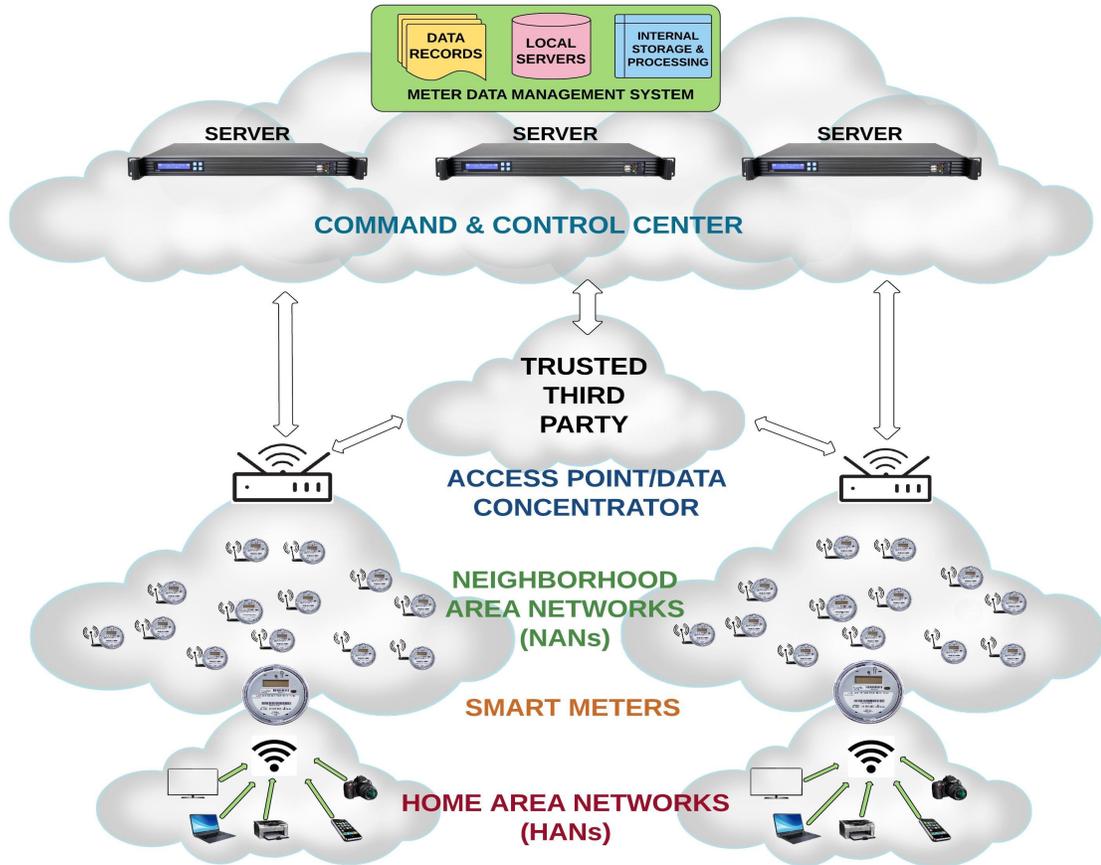


Figure 8.1: Proposed AMI architecture consists of HAN, NAN, Trusted Third Party (TTP), Data concentrator and Control center.

8.2 Architecture of AMI

Millions of smart meters are engineered to communicate with the local utility service provider (SP)/control center using AMI network. The bidirectional networks can be a mesh or a heretical or a hybrid. Periodic collection, storing and transmission of enormous volumes of data packets via the regional access points (APs) form the primary workhorses of AMI. The data packets are transmitted through a gateway or data concentrator, as shown in Fig. 8.1, which then relays the packets eventually to the control center. Each component of the AMI performs its own application. From a broader perspective, the AMI encompasses everything from home appliances to the control center, forming a

comprehensive network. The following are different sub-components of the AMI architecture.

- *Home appliance*: Machines that are employed by consumers for performing everyday chores and activities and those of which consume electric power come under this category. Some examples are washing machines, driers, microwave ovens, and air conditioner. The energy consumed by these machines is calculated per unit of the smart grid system. The unit consumption data is then relayed to the smart meter that measures, calibrates and collects the information reported by these appliances.
- *Smart meter*: They form the backbone of the AMI, being responsible for collecting the consumption unit data from consumers before dispatching the data to the SP. The meters are designed to use the channel in periodic intervals of time by sending short bursts of information. The small network formed between the home appliances and the meter corresponding to that household is termed as Home Area Network (HAN). The smart meters measure, collect and store data before sending the same in the form of encrypted packets.
- *Neighborhood Area Network (NAN)*: Beyond HAN, there is a broader network that is made up of various smart meters within a locality and their corresponding APs. These meters communicate among themselves through a mesh/hierarchical/hybrid connected wired (PLC) /wireless (WiFi, ZigBee, GPRS etc.) network termed as NAN.
- *Trusted third party (TTP)*: The entity administrates security scheme is known as TTP. In our scheme, TTP authenticates meter and conveys the random key index to the control center. Additionally TTP updates the codebook containing encryption keys mapped with meters' coordinate periodically.

- *Control center/Back office/Command and Control Center (CCC)* : The CCC is connected to the NAN by a wired or wireless connection such as fiber optic or cellular network. A bill is issued for the consumer based on the data received by the data center over the period of one month. Each day, the utility receives the data in 15-60 minutes intervals. This data is also used in the optimization of the electric power generation and distribution. Additionally, it also helps in control and monitoring of the load from a remote location.

8.3 Localization Algorithm and Node Authentication Technique

In our proposed scheme, we use RSS technique for localization of meters and kNN algorithm for meter (node) authentication. In the rest of this section, the two algorithms have been explained in further detail.

8.3.1 Localization of Smart Meter by RSS method

Let us assume, there are n partially dispersed known position meters at positions (x_i, y_i) where $i = 1, 2, \dots, n$ and a new meter is at at (x, y) as shown in Fig. 8.2. If the received signal strength of new meter at (x_i, y_i) is \mathcal{U}_i , then it follows the model [368]:

$$\mathcal{U}_i = c - 10\gamma \log_{10}(d_i) + w_i \quad (8.1)$$

where c denotes a constant dependent on the transmitted signal power, frequency etc.

$\gamma > 0$ is the path loss constant. The generic value of γ is 4-6 from which a value of 2.93 has been used here considering the residential area.

d_i defined as the Euclidean distance among new meter and other meters is given by

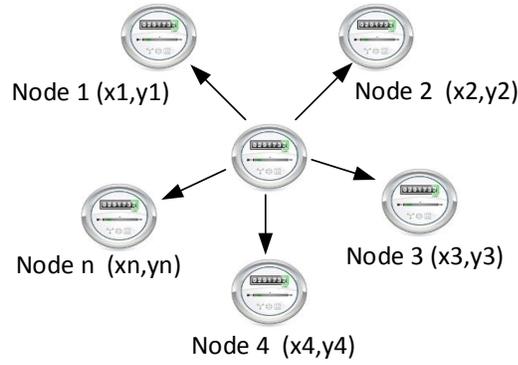


Figure 8.2: Localization of new meter (unknown positioned meter).

$$d_i = \sqrt{(x - x_i)^2 + (y - y_i)^2} \quad (8.2)$$

w_i is the zero mean Random Gaussian Noise with standard deviation σ_i . The typical value of σ is 6 to 12 dB.

We define the θ and \mathcal{U} as

$$\theta = [x, y, z]^T \text{ and } \mathcal{U} = [\mathcal{U}_1, \mathcal{U}_2, \dots, \mathcal{U}_n]^T$$

The Likelihood function of θ given an RSS measurement \mathcal{U}_i , $f(\theta/\mathcal{U})$ can be written as

$$f(\theta/\mathcal{U}) = c_1 \exp\left(-\sum_{i=1}^n \frac{\mathcal{U}_i - c + 10\gamma \log_{10}(d_i)}{2\sigma_i^2}\right) \quad (8.3)$$

where c_1 is a constant.

The Maximum Likelihood Estimation of θ denoted by $\hat{\theta}$ is

$$\begin{aligned} \hat{\theta} &= \arg \max f(\theta/\mathcal{U}) \\ &= \arg \min \left\{ \sum_{l=1}^n \frac{\mathcal{U}_l - c + 10\gamma \log_{10}(d_l)}{2\sigma_l^2} \right\} \end{aligned} \quad (8.4)$$

The ML estimation of a new meter's location (x, y) based on the RSS measurement \mathcal{U}_i , can be written then as:

$$(x_r, y_r) = \{\hat{\theta}(1), \hat{\theta}(2)\} \quad (8.5)$$

The position determination by above RSS technique will create a local coordinate system for HAN, different from the geographical coordinate system.

8.3.2 Meter Authentication by kNN Algorithm

kNN algorithm was proposed by T.M. Cover and P.E. Hart, where k denotes the number of nearest neighbors that are helpful to predict the class of the test sample [361].

Let us consider, a set of meter $M = \{m_i\}, i = \{1, 2, \dots, N\}$ with attribute $a_l^t(x_{m_i}), l = 1, 2, \dots, L, m_i \in M$ at instance t . We define $C_{m_i}^t$ and $c_{m_i}^t$ as the class variable and class value respectively. The standard Euclidean distance between instance t and $t + 1$ is

$$d(x_{m_i}^t, x_{m_i}^{t+1}) = \sqrt{\sum_{l=1}^L (a_l^t(x_{m_i}) - a_l^{t+1}(x_{m_i}))^2} \quad (8.6)$$

When the value of attributes are nominal, the variation of standard Euclidean distance can be written as

$$d(x_{m_i}^t, x_{m_i}^{t+1}) = \sum_{l=1}^L \delta(a_l^t(x_{m_i}) - a_l^{t+1}(x_{m_i}))$$

$$\text{subject to } \delta(a_l^t(x_{m_i}), a_l^{t+1}(x_{m_i})) = \begin{cases} 0, & \text{if } a_l^t(x_{m_i}) = a_l^{t+1}(x_{m_i}) \\ 1, & \text{otherwise} \end{cases} \quad (8.7)$$

The most common class value of $x_{m_i}^t$ at instant t

$$c_{m_i}^t(x_{m_i}) = \arg \max_{c_{m_i} \in C_{m_i}} \delta(c_{m_i}^t, c(x_{m_i}^{t+1})) \quad (8.8)$$

where $x_{m_i}^{t+1}$ is the kNN neighbors of $x_{m_i}^t$ and

$$\delta(c_{m_i}^t, c(x_{m_i}^{t+1})) = \begin{cases} 1 & \text{if } c_{m_i}^t = c(x_{m_i}^{t+1}) \\ 0, & \text{otherwise} \end{cases} \quad (8.9)$$

Since kNN is lazy algorithm and keeps training data for classification, we build a inductive learning classification model at training time to improve its efficiency as shown in [360]. Later, this model is used for classification. An authenticate set of data is used for training. For each meter m_i , parameter - distance between source meter m_j and meter m_i (d), packet size (s) and data transmission frequency (f) are used as data for classification. In the description of algorithm we use terminology "tuple" as data set consists of d, s, f .

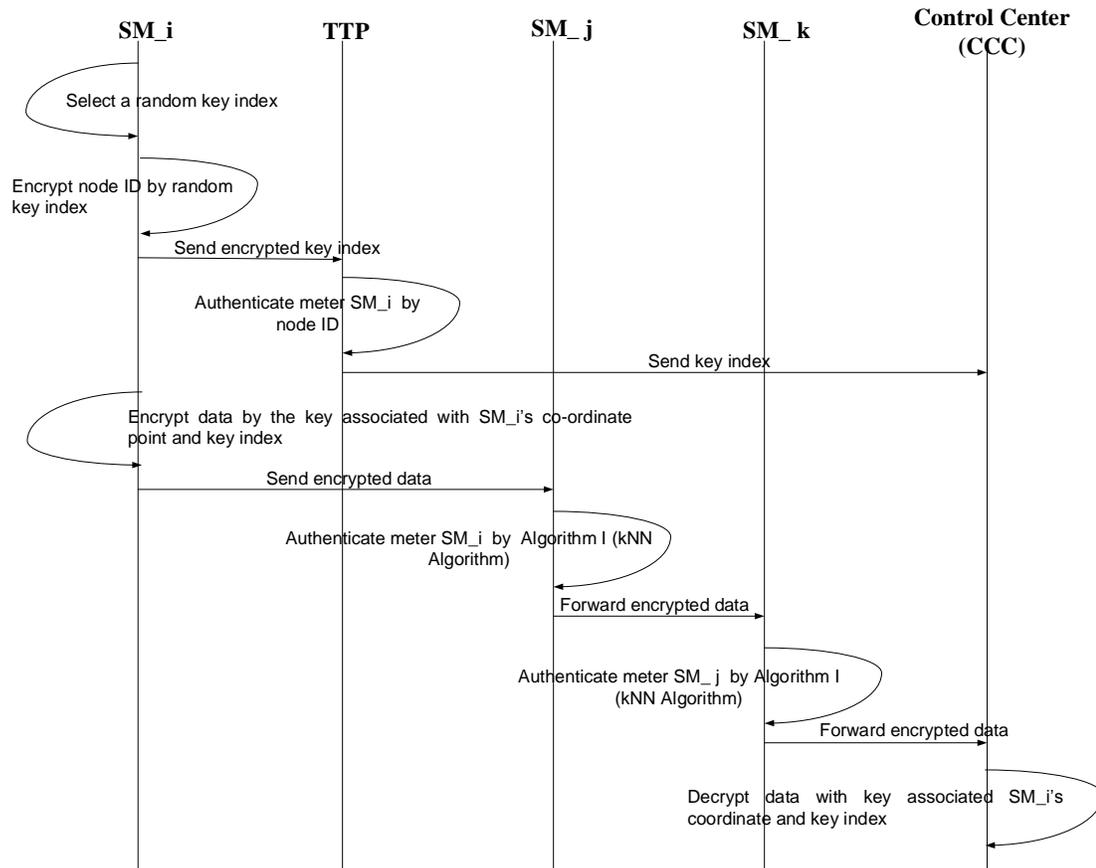


Figure 8.3: Encryption and data transmission process among various components of AMI.

Algorithm 4 kNN algorithm

- 1: For each data tuple $\psi_i \in \Psi, i = \{1, 2, 3, \dots, n\}$, 2 dimensional vector space $\psi_i = \langle W_{i1}, W_{i2}, \dots, W_{in} \rangle$
 - 2: **Training:**
 - 3: Set "ungrouped" tag to all data tuples
 - 4: Calculate euclidean distance $Ed_i = \sum \sum (x_l - x_k)$ among elements of ψ_i where $x_l, x_k \in \psi_i$
 - 5: Build a representative model $\Omega = \langle cls(\psi_i), sim(\psi_i), num(\psi_i) \rangle$ where cls represents class level, sim represents similarity and num of elements
 - 6: **Classification:**
 - 7: For new data tuple ψ_k , calculate representative parameter $\Omega_k = \langle cls(\psi_k), sim(\psi_k), num(\psi_k) \rangle$
 - 8: **if** representative parameter Ω_k belong to Model Ω **then**
 - 9: New data belongs to same group
 - 10: **else**
 - 11: New data belongs to different group
 - 12: **end if**
 - 13: **End**
-

Also we use Euclidean distance for measure of similarity. The algorithm is described in pseudo code in Algorithm 4.

The use of kNN algorithm along with RSS technique will allow data/packet receive from authenticate neighbor meters. So, this will ensure node to node authentication intercepting malicious packets.

8.4 Encryption Process

A detailed description of the entire encryption and data flow process in the AMI is provided in this section. As shown in Fig. 8.3, there is an involvement of a TTP which will perform the authentication of the different smart meters using their node IDs. Once the TTP authenticates a particular smart meter (m_i), it sends the key index to the CCC. At the same time, the encrypted data is sent to the CCC via an intermediary nodes (other smart meters). Finally when the encrypted packet reaches the CCC, the destination will

decrypt the message using the key associated with the random key index and meter (m_i)'s coordinate- latitude and longitude. Before we proceed to the steps of encryption and data flow, the assumptions are outlined below.

Assumptions:

- The meter has a limited memory and computational capability.
- The control center has sufficient computational capability.
- Every meter holds records of the location of its neighboring meters.
- Every meter transmits data at a constant power.
- There is a codebook that has an encryption key associated with each coordinate point of the geo location as shown Fig. 8.4.
- The TTP updates the codebook associated with geo-location /co-ordinate point periodically.

The following are the different steps during the message encoding, transmission and decoding processes:

Initialization:

The meter which is ready to send information, identified as the source meter m_i , performs an initiation process for every session (typically once in every fifteen minutes). This process involves the selection of a random key index. This key index is then encrypted by node ID and sent to the TTP. The TTP has a codebook that also contains information of all smart meters. Hence, it uses the codebook to identify the node by its ID, decrypt the key index and send it to the CCC.

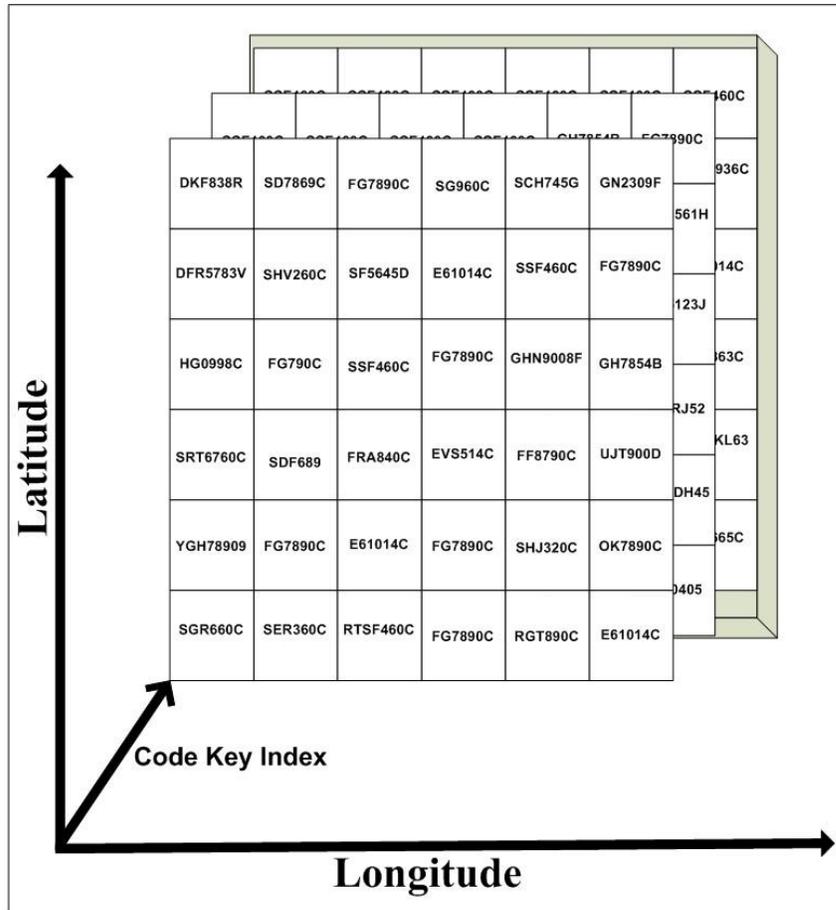


Figure 8.4: Mapping of encryption key based on coordinate point.

Data Encryption:

In the next step, the source meter encrypts the consumption data which is plaintext, with an encryption key associated with its own coordinate points (latitude and longitude) and the key index.

Encryption:

$$k_i \oplus m_i \longrightarrow C_i \tag{8.10}$$

Data forwarding and authentication:

This encrypted message is forwarded to its peers since relaying is the only way the packets can reach the CCC eventually, considering that each meter has very low trans-

mission power. Now, any neighboring meter (m_j) receiving the encrypted packet from a source meter (m_i) determines its authentication by Algorithm 5. Once authenticated, the corresponding packets are forwarded to the next neighboring meters.

Algorithm 5 Transmitting algorithm

- 1: Derive the position of neighbor meter m_i by RSS method
 - 2: Calculate the distance between source meter m_j and neighboring meter m_i , $R_{i,j}$, where $m_i, m_j \in M$
 - 3: Get s and f from packet header
 - 4: Run kNN algorithm (Algorithm 4)
 - 5: **if** New data belongs to same group **then**
 - 6: Forward data to the next meter m_k , where $m_k \in M$
 - 7: **else**
 - 8: Discard the data and report to CCC
 - 9: **end if**
 - 10: **End**
-

Decryption:

The CCC receives the encrypted data and decrypts the same with the help of the key associated with the source meter's location and the key index it received from the TTP.

Decryption:

$$C_i \xrightarrow{k_i} m_i \quad (8.11)$$

8.5 Simulation Results

Rectangle, hexagon and octagonal shapes are employed to evaluate the performance of the proposed localization algorithm. These shapes comprise what is known as the Area of Interest (AOI) with an approximate dimension of 10x10m, as depicted in Fig. 8.5. Each edge represents a known position meter (node), and the emitter (i.e. unknown positioned meter/node) is the center of the AOI. The estimation of position of the unknown positioned meter through Eq.(1) 8.5 is the optimization problem. In our simulation, we use the Particle Swarm Optimization (PSO) for that optimization.

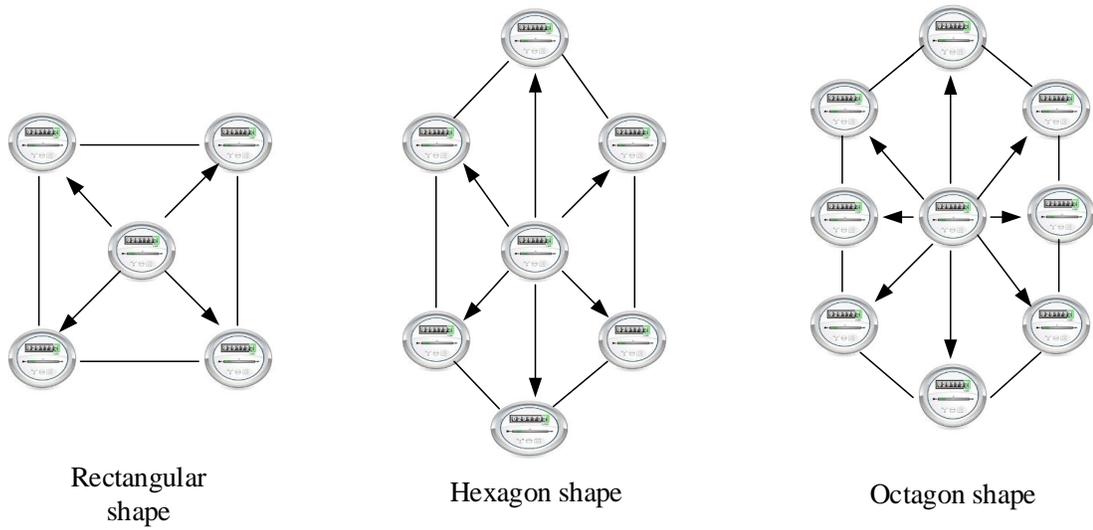


Figure 8.5: Different shapes of AOI.

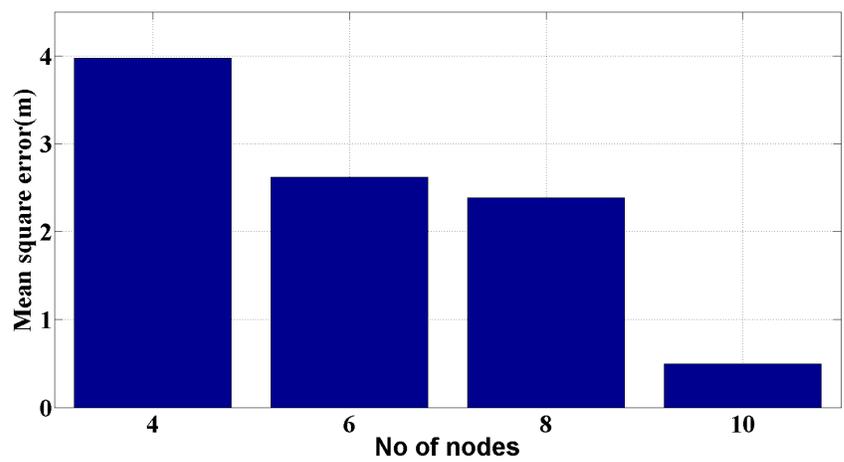


Figure 8.6: Number of nodes vs Mean Square Error.

In Fig. 8.6, we observe that with an increase of the number of the nodes, the mean square error from the exact position will decrease. At the same time, with an increase of the path loss constant, the mean square error will decrease as illustrated in Fig. 8.7. In the path loss model of radio signals, random noise is added which varies by standard deviation. In the Fig. 8.8, the surface diagram is drawn as a function of variance and number of nodes. As the variance of the noise increases, the error also hikes correspondingly.

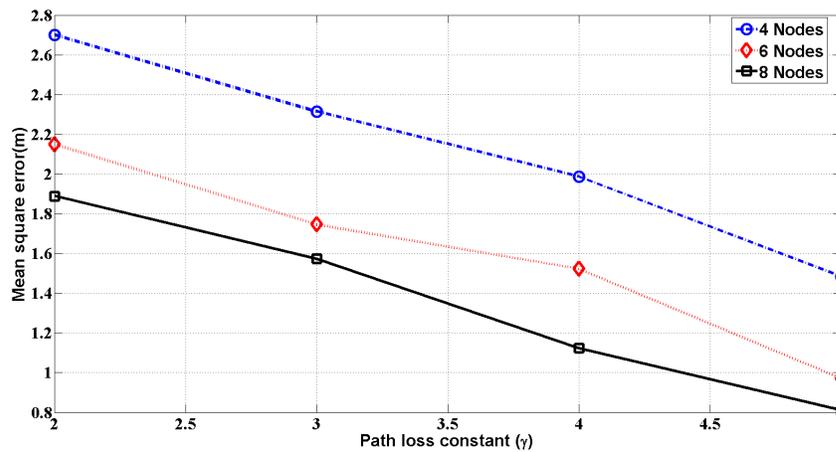


Figure 8.7: Path loss vs Mean Square Error.

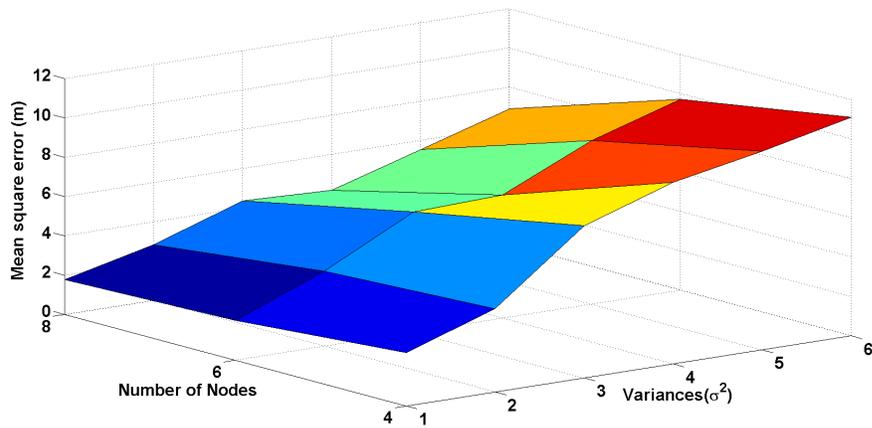
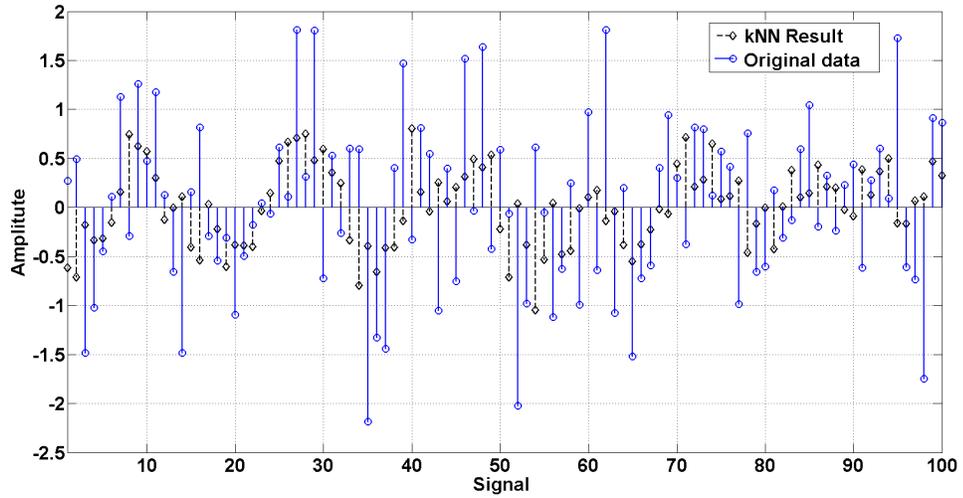


Figure 8.8: Variance vs Mean Square Error.

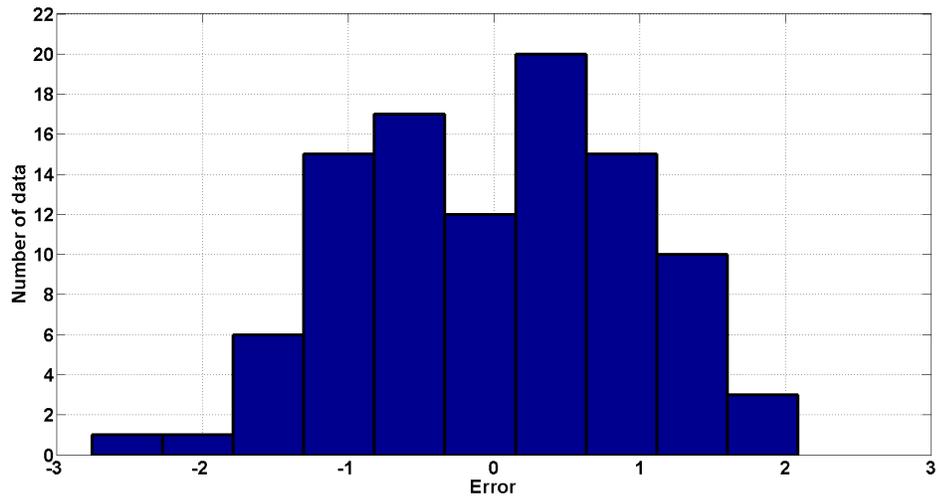
Let the number of neighbor meters $k = 4$, Fig. 8.9a presents a data set of size 100 and the predicted result of the data set derived by the kNN algorithm. Compared with the original data set and the predicted result, Fig. 8.9b shows the error histogram between these two data sets. Similarly, Fig. 8.10a presents a data set of size 200 and the predicted result derived by the kNN algorithm. Fig. 8.10b shows the statistics of the error distribution between original and predicted data set. It is noted that the error is well distributed for both data set of 100 and 200. Performance of kNN algorithm for different data size and neighbors (k) is shown in Fig. 8.11. We found that with the increment of size of data and decrement of k , the mean square error between original data and predicted data increases quite precisely.

Since the position of a smart meter is fixed due to mounted on specific wall or pole and the environment surrounding the meters are also stable, the calculated error in localization method of an unknown/new positioned meter at neighbor meters by RSS method will be almost constant. Also smart meters send the consumption data periodically at a specific interval defined by the utility company, and the data packet size is constant. For these reasons, a meter can authenticate the source meter by kNN algorithm using data of sending frequency, packet size and distance between two meters.

Since the position of a smart meter is fixed due to mounted on specific wall or pole and the environment surrounding the meters are also stable, the calculated error in localization method of an unknown/new positioned meter at neighbor meters by RSS method will be almost constant. Also smart meters send the consumption data periodically at a specific interval defined by the utility company, and the data packet size is constant. For these reasons, a meter can authenticate the source meter by kNN algorithm using data of sending frequency, packet size and distance between two meters.

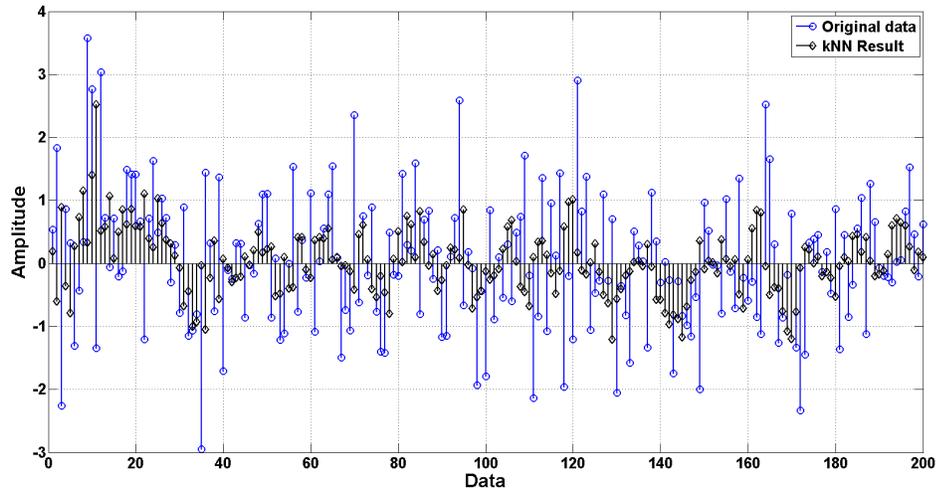


(a)

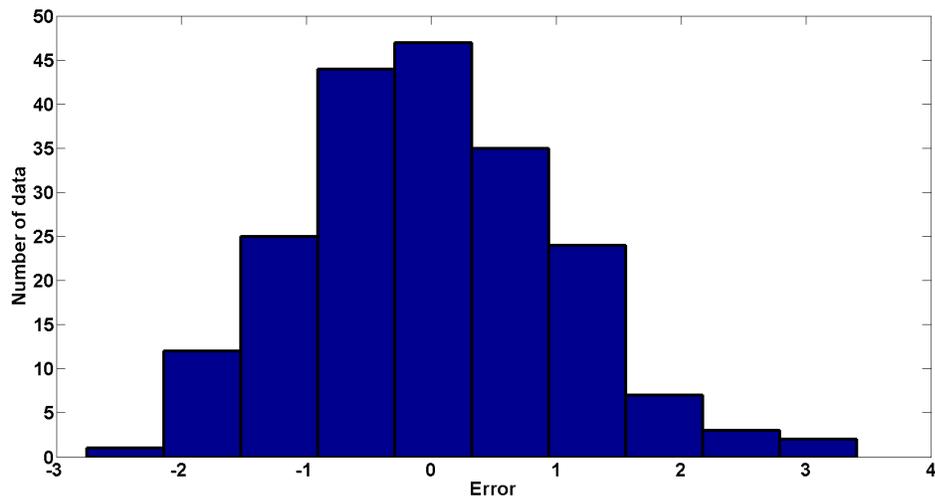


(b)

Figure 8.9: Performance of kNN algorithm for data set of size 100. a) Comparison of data value and their prediction by kNN algorithm. (b) Histogram of error distribution.



(a)



(b)

Figure 8.10: Performance of kNN algorithm for data set of size 200. a) Comparison of data value and their prediction by kNN algorithm. (b) Histogram of error distribution.

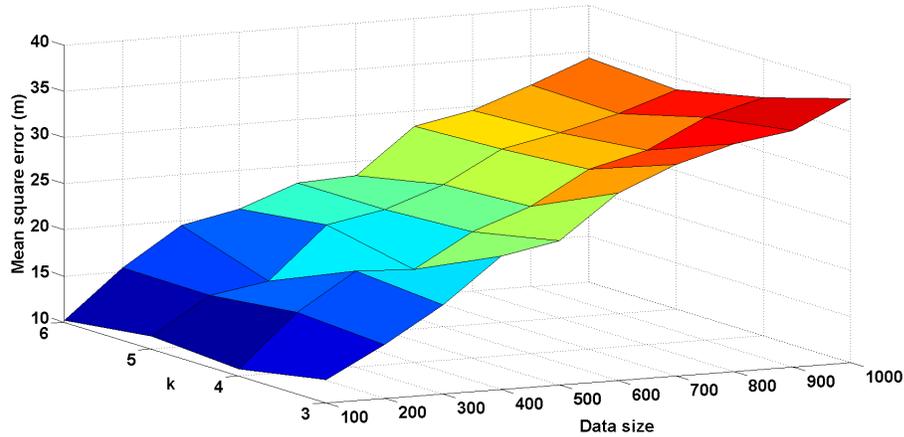


Figure 8.11: The relation of error between number of neighbors and data size.

8.6 Security Strength Analysis

The security strength of a data packet can be measured by Entropy. The value of Entropy reflects the uncertainty of a random variable. The more certain about a value is, the smaller the entropy value.

The entropy for a sequence S [369]

$$H(s) = \sum_S P(S = x) \log_2 P(S = x)$$

where $P(S = x)$ is the probability of taking S value over x .

Let us consider, a smart meter sends a data packet of 128 bit encrypted by 128 bit symmetric key to control center. For a 8 bit random key index, the security strength of the random sequence is 2^8 . On the other hand, for a 128 bit symmetric key algorithm, the security strength is 2^{128} .

So, for a 8 bit random key index and 128 bit symmetric key,

The security strength of the packet= $2^8 + 2^{128}$

So, if a hacker wants to decrypt a data packet of 128 bit, he needs $(2^8 + 2^{128})$ number of tries to decrypt the message unless he is lucky. This is impractical.

8.7 Conclusion

In this chapter, we propose a two level security scheme consisting of data packet encryption and node authentication. For packet encryption, we use an encryption key corresponding to the meter's location and random key index. The codebook relating the encryption code to its geographical location could be updated periodically in a secured way by the TTP. Since the same code book is used for all smart meters, it invalidates the need for using different keys for different meters/nodes each time. Also the inclusion of a random key index increases the uncertainty in packet encryption. In case of node to node authentication, we use kNN classifier which is simple and robust.

The positioning map of meters in HAN is a local map different from geographical coordinate system. So exposing the coordinates will not reveal the exact location of a consumer's house. Besides, the position of smart meters determined by RSS method is nearly constant due to the constant power transmission and stable positions of meters. This ensures the stability and security of the smart grid system without giving opportunity to the potential hackers to manipulate the data.

CHAPTER 9

Conclusion and Future Works

9.1 Concluding Remarks

The 5G cellular networks are to be met the demands of wireless communications of 2020 and beyond. The goals of 5G are not only to provide high speed mobile Internet, but also to support massive connectivity and ultra reliable low latency communication. It will play a critical role in diverse sectors such as industry and manufacturing, transportation, health care, virtual/augmented reality, robotics, smart grid, gaming, and education, along with voice communication. In case of massive connectivity and high throughput requirement, sharing of underutilized free spectrum is a viable solution. On the other hand, to support ultra reliable low latency communication, drastic changes in the core and radio network are required, along with new technologies, such as caching, and fog networks/mobile edge computing. Additionally, data security is a prime concern in wireless communication. In this research, several spectrum sharing techniques, low latency solutions, and data security schemes have been studied with application to smart grid and IoT.

In chapter 2, we present the related works on spectrum sharing, latency, and smart grid security. In case of spectrum sharing, works on coexistence of LTE-ZigBee and LTE-WiFi has been included. For latency, a brief state-of-art work on latency has been discussed. We also present some relevant works on smart grid security.

In chapter 3, we study the coexistence of LTE and ZigBee/WiFi using smart grid as a deployment scenario. We use conventional 10 ms transmission time interval for both LTE and ZigBee/WiFi. It is found that the performance of ZigBee is more affected by LTE than that of LTE by ZigBee in a simultaneous operation. However, uplink power control and manipulating UL/DL configuration of LTE can improve overall performance. For LTE and WiFi coexistence, a fixed duty cycle transmission is introduced. Therefore,

LTE uses different duty cycles, such as 60%, and 80% of transmission time based on its queued data, and WiFi transmits on the rest of the time. It improves the overall efficiency of spectrum. If smart meters utilize WiFi/ZigBee and access point uses LTE for long range communication, coexistence of LTE and WiFi in 2.4/3.5 GHz is definitely a potential communication solution for smart grid. However, interference contributed by other sources such as cellular networks, and public radio communication need to be considered for further study.

In chapter 4, a multi arm bandit (MAB) based duty cycle selection is proposed for the LTE and WiFi coexistence. We investigate the performance in the 3.5 GHz band. Simulation results demonstrate that the aggregated throughput is improved by 33% while maintaining a minimum throughput of LTE. Along with the duty cycle, we incorporate DL link power control which optimizes transmit power. This approach enhances energy efficiency and cell edge performance along with aggregated throughput improvement. However, balancing between the exploration and exploitation is the matter of challenge.

In chapter 5, we summarize the existing works related to latency reduction from the literature. In particular, we divide the solutions in three major divisions: (1) radio network, (2) core network, and (3) caching. We also provide a general overview of 5G network including software defined radio, network function visualization, and caching/mobile edge computing. Following this, we provide future directions for achieving low latency. However, more practical and efficient techniques in the presence of existing solutions need to be investigated before the standardization of 5G.

In chapter 6, we present a compass algorithm based locally decided and loop free routing protocol for stationary nodes. We propose received signal strength for localization of stationary nodes by maximum likelihood and particle swarm optimization. We simulated our approach for a scenario of metering infrastructure of smart grid. The simulation result demonstrates that our routing protocol reduces hop count resulting in reduction of

latency. But, localization error in the localization method will limit performance of the routing protocol.

In chapter 7, we present a geo-location and machine learning based data encryption method for smart grid applications. The meter is encrypted by a key associated with meter's position and a random key index. The position of meter is derived using received signal strength method solved by maximum likelihood and particle swarm optimization. Along with the data encryption, we propose k nearest neighbors algorithm for packet/node authentication. We also evaluate the security strength of our proposed approach. However, compromised nodes, impersonation attack, and interruption in electromagnetic communication are not considered our approach.

In chapter 8, we propose a two layer data security method for packet encryption and node authentication with an application to smart grid. In the encryption layer, data encryption by asymmetric keys and randomization of data packets have been proposed. For node-to-node authentication, one class support vector machine is proposed which utilizes three features- frequency of data reception from a specific meter, packet size, and meter position. We also compare our scheme with state-of-art approach isolation forest. However, impersonation attack, compromised nodes, and electromagnetic signal interruption will affect our approach.

9.2 Future Works

The topics reported in this dissertation addressed some issues of the spectrum sharing, massive connectivity, latency, and data security of wireless communication with application to IoT and smart grid. However, there are many open issues, and challenges that need to be explored. The following issues are the several research directions for further exploration:

- There are abundant mmWave spectrum in 30-300 GHz band which can be shared. However, mmWave has some limitations such as low penetration, and small range. In this regards, the coexistence of LTE and WiFi/ZigBee in mmWave can be investigated for effective techniques.
- The unlicensed spectrum will be shared not only by variant technologies such as, LTE-WiFi and LTE-ZigBee, but also same technology LTE-LTE. Therefore, scenarios including inter technology and diverse technologies can be also an area of exploration.
- For latency reduction in PHY and MAC layer, various learning techniques can be used while maintaining the desired QoS (latency and reliability). In this case, different algorithms can be explored.
- Caching is an important tool to meet the goals of 5G networks. For content delivery, cache placement and hit ratio improvement, exploration can be made with diverse and holistic machine learning techniques.
- For the derivation of routing protocol, we considered only stationary nodes. We look forward to studying our technique in the mobile nodes.

BIBLIOGRAPHY

- [1] M. Sybis, K. Wesolowski, K. Jayasinghe, V. Venkatasubramanian, and V. Vukadinovic, "Channel Coding for Ultra-Reliable Low-Latency Communication in 5G Systems," in *Proc. IEEE Veh. Technol. Conf. (VTC Fall)*, Sep. 2016, pp. 1–5.
- [2] S. Sesia, I. Toufik, and M. Baker (Editors), *LTE, The UMTS Long Term Evolution: From Theory to Practice*. Wiley, 2nd Edition, 2011.
- [3] T. Levanen, J. Pirskanen, T. Koskela, J. Talvitie, and M. Valkama, "Low latency radio interface for 5G flexible TDD local area communications," in *Proc. IEEE Int. Conf. Commun. Workshop (ICCW)*, Jun. 2014, pp. 7–13.
- [4] SDN architecture overview. [Online]. Available: <https://www.opennetworking.org/images/stories/downloads/sdn-resources/technical-reports/SDN-architecture-overview-1.0.pdf>, [Online; accessed 5-May-2017]
- [5] E. N. ISG, "Network Functions Virtualization, white paper," in <https://portal.etsi.org/nfv/nfv-white-paper.pdf>, 2012. [Online; accessed 6-July-2016], Nov. 2016.
- [6] S. Zhang, X. Xu, Y. Wu, and L. Lu, "5G: Towards energy-efficient, low-latency and high-reliable communications networks," in *Proc. IEEE Int. Conf. on Commun. Syst. (ICCS)*, Nov 2014, pp. 197–201.
- [7] C. Bockelmann, N. Pratas, H. Nikopour, K. Au, T. Svensson, C. Stefanovic, P. Popovski, and A. Dekorsy, "Massive machine-type communications in 5G: physical and MAC-layer solutions," *IEEE Commun. Mag.*, vol. 54, no. 9, pp. 59–65, Sep. 2016.
- [8] A. H. Jalal, Y. Umasankar, and S. Bhansali, "Development and characterization of fuel cell sensor for potential transdermal ethanol sensing," *ECS Transactions*, vol. 72, no. 31, pp. 25–31, 2016.
- [9] NSN Whitepaper, "Enhance mobile networks to deliver 1000 times more capacity by 2020," Tech. Rep., Sep. 2013.
- [10] M. Simsek, M. Bennis, and I. Guvenc, "Enhanced intercell interference coordination in HetNets: single vs. multiflow approach," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Atlanta, GA, Dec. 2013, pp. 725–729.

- [11] ———, “Learning based frequency-and time-domain inter-cell interference coordination in HetNets,” *IEEE Trans. Vehic. Technol.*, vol. 64, no. 10, pp. 4589–4602, 2015.
- [12] S. Zacharias, T. Newe, S. O’Keeffe, and E. Lewis, “Coexistence measurements and analysis of IEEE 802.15.4 with WiFi and bluetooth for vehicle networks,” in *Proc. Int. Conf. on ITS Telecom. (ITST)*, Nov. 2012, pp. 785–790.
- [13] W. Yuan, X. Wang, and J.-P. Linnartz, “A Coexistence Model of IEEE 802.15.4 and IEEE 802.11b/g,” in *Proc. IEEE Symp. on Commun. and Vehic. Tech.*, Nov. 2007, pp. 1–5.
- [14] J. Choliz, A. Hernandez-Solana, A. Sierra, and P. Cluzeaud, “Coexistence of MB-OFDM UWB with Impulse Radio UWB and other radio systems,” in *Proc. IEEE Int. Conf. Ultra-Wideband (ICUWB)*, Sept. 2011, pp. 410–414.
- [15] N. Rupasinghe and İ. Güvenç, “Licensed-assisted access for WiFi-LTE coexistence in the unlicensed spectrum,” in *Proc. IEEE Glob. Work.(GC Wkshps)*. IEEE, 2014, pp. 894–899.
- [16] I. Parvez, T. Khan, A. I. Sarwat, and Z. Parvez, “LAA LTE and WiFi based Smart Grid Metering Infrastructure in 3.5 GHz Band,” *CoRR*, vol. abs/1711.05219, 2017. [Online]. Available: <http://arxiv.org/abs/1711.05219>
- [17] J. Cao, M. Ma, H. Li, Y. Zhang, and Z. Luo, “A survey on security aspects for lte and lte-a networks,” *IEEE Communications Surveys Tutorials*, vol. 16, no. 1, pp. 283–302, First 2014.
- [18] Y. Mekonnen, M. Haque, I. Parvez, A. H. Moghadasi, and A. I. Sarwat, “LTE and Wi-Fi Coexistence in Unlicensed Spectrum with Application to Smart Grid: A Review,” *CoRR*, vol. abs/1708.09005, 2017. [Online]. Available: <http://arxiv.org/abs/1708.09005>
- [19] I. Parvez, N. Islam, N. Rupasinghe, A. I. Sarwat, and İ. Güvenç, “LAA-based LTE and ZigBee coexistence for unlicensed-band smart grid communications,” in *Proc. IEEE SoutheastCon*. IEEE, 2016, pp. 1–6.
- [20] I. Parvez and A. Sarwat, “Frequency band for HAN and NAN communication in Smart Grid,” in *Proc. IEEE Conf. Comput. Intellig. Appli. in Smart Grid (CIASG)*. IEEE, 2014.

- [21] M. Sriyananda, I. Parvez, I. Güvenc, M. Bennis, and A. I. Sarwat, "Multi-armed bandit for LTE-U and WiFi coexistence," in *Proc. IEEE Wire. Commun. and Netw. Conf.* IEEE, 2016, pp. 1–6.
- [22] I. Parvez, M. Sriyananda, İ. Güvenç, M. Bennis, and A. Sarwat, "CBRS Spectrum Sharing between LTE-U and WiFi: A Multiarmed Bandit Approach," *Mobile Inform. Systems*, vol. 2016, 2016.
- [23] I. Parvez, A. Rahmati, I. Guvenc, A. I. Sarwat, and H. Dai, "A survey on low latency towards 5G: RAN, core network and caching solutions," *IEEE Commun. Surveys & Tutor.*, 2018.
- [24] I. Parvez, M. Jamei, A. Sundararajan, and A. I. Sarwat, "RSS based loop-free compass routing protocol for data communication in advanced metering infrastructure (AMI) of Smart Grid," in *Proc. IEEE Symp. on Comp. Intell. Appl. in Smart Grid (CIASG)*. IEEE, 2014, pp. 1–6.
- [25] I. Parvez, N. Chotikorn, and A. I. Sarwat, "Average Quantized Consensus Building by Gossip Algorithm using 16 Bit Quantization and Efficient Data Transfer Method," in *Proc. Intern. conf. on Intell. Systems, Data Mining and Inform. Techno. (ICIDIT)*, pp. 1–5.
- [26] I. Parvez, F. Abdul, H. Mohammed, and A. I. Sarwat, "Reliability assessment of access point of advanced metering infrastructure based on Bellcore standards (Telecordia)," in *Proc. IEEE SoutheastCon.* IEEE, 2015, pp. 1–7.
- [27] I. Parvez, A. I. Sarwat, J. Pinto, Z. Parvez, and M. A. Khandaker, "A gossip algorithm based clock synchronization scheme for smart grid applications," *CoRR*, vol. abs/1707.08216, 2017. [Online]. Available: <http://arxiv.org/abs/1707.08216>
- [28] I. Parvez, F. Abdul, and A. I. Sarwat, "A Location Based Key Management System for Advanced Metering Infrastructure of Smart Grid," in *Proc. IEEE Green Techno. Conf.(GreenTech)*, Apr. 2016, pp. 62–67.
- [29] I. Parvez, A. Islam, and F. Kaleem, "A key management-based two-level encryption method for AMI," in *Proc. IEEE Conf. on PES General Meeting — Conf. Expo.*, July 2014, pp. 1–5.
- [30] I. Parvez, A. I. Sarwat, L. Wei, and A. Sundararajan, "Securing metering infrastructure of smart grid: A machine learning and localization based key management approach," *Energies*, vol. 9, no. 9, 2016.

- [31] A. I. Sarwat, A. Sundararajan, I. Parvez, M. Moghaddami, and A. Moghadasi, *Toward a Smart City of Interdependent Critical Infrastructure Networks*. Cham: Springer International Publishing, 2018, pp. 21–45. [Online]. Available: https://doi.org/10.1007/978-3-319-74412-4_3
- [32] I. Parvez, M. Aghili, and A. I. Sarwat, “Key management and learning based two level data security for metering infrastructure of smart grid,” *CoRR*, vol. abs/1709.08505, 2017. [Online]. Available: <http://arxiv.org/abs/1709.08505>
- [33] 3GPP, “Study on licensed-assisted access to unlicensed spectrum,” 3GPP Technical Report (TR 36.889), May 2015. [Online]. Available: http://www.3gpp.org/ftp/specs/archive/36_series/36.889/
- [34] X. Wang, S. Mao, and M. X. Gong, “A survey of lte wi-fi coexistence in unlicensed bands,” *GetMobile: Mobile Computing and Communications*, vol. 20, no. 3, pp. 17–23, 2017.
- [35] G. RP-140808, “Review of regulatory requirements for unlicensed spectrum),” AlcatelLucent, Alcatel-Lucent Shanghai Bell, Ericsson, Huawei, HiSilicon, IAESI, LG, Nokia, NSN, Qualcomm, NTT Docomo, Tech. Rep., June 2014.
- [36] Nokia, “Nokia LTE for unlicensed spectrum,” AlcatelLucent, Alcatel-Lucent Shanghai Bell, Ericsson, Huawei, HiSilicon, IAESI, LG, Nokia, NSN, Qualcomm, NTT Docomo, Tech. Rep., June 2014.
- [37] Qualcomm Research, “Lte in unlicensed spectrum: Harmonic coexistence with wi-fi,” white paper, June 2014.
- [38] Qualcomm, “Qualcomm Research LTE in Unlicensed Spectrum: Harmonious Coexistence with WiFi,” 3GPP RAN1 standard contribution, Tech. Rep., Jun. 2014.
- [39] F. M. Abinader, E. P. L. Almeida, F. S. Chaves, A. M. Cavalcante, R. D. Vieira, R. C. D. Paiva, A. M. Sobrinho, S. Choudhury, E. Tuomaala, K. Doppler, and V. A. Sousa, “Enabling the coexistence of LTE and Wi-Fi in unlicensed bands,” *IEEE Commun. Mag.*, vol. 52, no. 11, pp. 54–61, Nov. 2014.
- [40] Y. Li, F. Baccelli, J. G. Andrews, T. D. Novlan, and J. C. Zhang, “Modeling and Analyzing the Coexistence of Wi-Fi and LTE in Unlicensed Spectrum,” *IEEE Trans. on Wireless Commun.*, vol. 15, no. 9, pp. 6310–6326, Sept 2016.
- [41] E. Almeida, A. M. Cavalcante, R. C. D. Paiva, F. S. Chaves, F. M. Abinader, R. D. Vieira, S. Choudhury, E. Tuomaala, and K. Doppler, “Enabling LTE/WiFi coex-

- istence by LTE blank subframe allocation,” in *Proc. IEEE Int. Conf.*, Budapest, Hungary, Jun. 2013, pp. 5083–5088.
- [42] T. Nihtil, V. Tykhomyrov, O. Alanen, M. A. Uusitalo, A. Sorri, M. Moisio, S. Iraj, R. Ratasuk, and N. Mangalvedhe, “System performance of LTE and IEEE 802.11 coexisting on a shared frequency band,” in *Proc. IEEE Wireless Commun. and Net. Conf. (WCNC)*, Shanghai, China, Apr. 2013, pp. 1038–1043.
- [43] N. Rupasinghe and I. Guven, “Reinforcement learning for licensed-assisted access of LTE in the unlicensed spectrum,” in *Proc. IEEE Wireless Commun. and Net. Conf. (WCNC)*, New Orleans, LA, Mar. 2015, pp. 1279–1284.
- [44] A. K. Sadek, Y. Tokgoz, M. Yavuz, T. A. Kadous, M. Fan, and A. A. Reddy, “Opportunistic supplemental downlink in unlicensed spectrum,” Mar. 5 2015, uS Patent App. 14/476,150.
- [45] N. Rupasinghe and I. Guven, “Licensed-assisted access for WiFi-LTE coexistence in the unlicensed spectrum,” in *Proc. IEEE Glob. Work. (GC Wkshps)*, Dec 2014, pp. 894–899.
- [46] L. Li, A. H. Jafari, X. Chu, and J. Zhang, “Simultaneous transmission opportunities for LTE-LAA smallcells coexisting with WiFi in unlicensed spectrum,” in *Proc. IEEE Intern. Conf. on Commun. (ICC)*, May 2016, pp. 1–7.
- [47] ZTE, “Frame structure design for LAA considering LBT,” 3GPP TSG RAN WG1 Meeting no.80, Feb. 2015.
- [48] R. Zhang, M. Wang, L. X. Cai, X. Shen, L.-L. Xie, and Y. Cheng, “Modeling and analysis of MAC protocol for LTE-U co-existing with Wi-Fi,” in *Proc. IEEE Glob. Commun. Conf. (GLOBECOM)*. IEEE, 2015, pp. 1–6.
- [49] F. Liu, E. Bala, E. Erkip, M. C. Beluri, and R. Yang, “Small-cell traffic balancing over licensed and unlicensed bands,” *IEEE transactions on vehicular technology*, vol. 64, no. 12, pp. 5850–5865, 2015.
- [50] O. El-Samadisy, M. Khedr, and A. El-Helw, “Performance Evaluation of MAC for IEEE 802.11 and LAA LTE,” in *Proc. Intern. Conf. on Comput. Science and Comput. Intell. (CSCI)*, Dec 2016, pp. 923–928.
- [51] A. k. Ajami and H. Artail, “Fairness in future Licensed Assisted Access (LAA) LTE networks: What happens when operators have different channel access pri-

- orities?” in *Proc. IEEE Intern. Conf. on Commun. Works. (ICC Workshops)*, May 2017, pp. 67–72.
- [52] M. Sahin, I. Guvenc, and H. Arslan, “Iterative Interference Cancellation for Co-Channel Multicarrier and Narrowband Systems,” in *Proc. IEEE Wireless Commun. and Networking Conf. (WCNC)*, Apr. 2010, pp. 1–6.
- [53] M. B. Celebi, I. Guvenc, H. Arslan, and K. A. Qaraqe, “Interference suppression for the LTE uplink,” *Proc. Physical Comm.*, vol. 9, no. 0, pp. 23–44, 2013.
- [54] A. Rahim, S. Zeisberg, and A. Finger, “Coexistence Study between UWB and WiMax at 3.5 GHz Band,” in *Proc. IEEE Int. Conf. Ultra-Wideband (ICUWB)*, Sep. 2007, pp. 915–920.
- [55] 3GPP, “3GPP Workshop on LTE in Unlicensed Spectrum,” Tech. Rep., Jun. 2014.
- [56] N. Rupasinghe and I. Guvenc, “Licensed-Assisted Access for WiFi-LTE Coexistence in the Unlicensed Spectrum,” in *Proc. IEEE Emerging Tech. 5G Wireless Cellular Networks Workshop (co-located with IEEE GLOBECOM)*, Dec. 2014, Atlanta, GA.
- [57] 3GPP, “Study on Licensed-Assisted Access using LTE,” 3GPP Study Item - RP-141397, Tech. Rep., Sep. 2014.
- [58] G. Durisi, T. Koch, and P. Popovski, “Toward Massive, Ultrareliable, and Low-Latency Wireless Communication With Short Packets,” *Proc. IEEE*, vol. 104, no. 9, pp. 1711–1726, Sep. 2016.
- [59] C.-P. Li, J. Jiang, W. Chen, T. Ji, and J. Smee, “5G ultra-reliable and low-latency systems design,” in *Proc. European Conf. Netw. Commun. (EuCNC)*, June 2017, pp. 1–5.
- [60] P. Guan, X. Zhang, G. Ren, T. Tian, A. Benjebbour, Y. Saito, and Y. Kishiyama, “Ultra-Low Latency for 5G - A Lab Trial,” *CoRR*, vol. abs/1610.04362, 2016. [Online]. Available: <http://arxiv.org/abs/1610.04362>
- [61] E. Lhetkangas, K. Pajukoski, J. Vihril, G. Berardinelli, M. Lauridsen, E. Tirola, and P. Mogensen, “Achieving low latency and energy consumption by 5G TDD mode optimization,” in *Proc. IEEE Int. Conf. Commun. Workshop (ICCW)*, Jun. 2014, pp. 1–6.

- [62] K. I. Pedersen, F. Frederiksen, G. Berardinelli, and P. E. Mogensen, "The Coverage-Latency-Capacity Dilemma for TDD Wide Area Operation and Related 5G Solutions," in *Proc. IEEE Veh. Technol. Conf. (VTC Spring)*, May 2016, pp. 1–5.
- [63] K. I. Pedersen, G. Berardinelli, F. Frederiksen, P. Mogensen, and A. Szufarska, "A flexible 5G frame structure design for frequency-division duplex cases," *IEEE Commun. Mag.*, vol. 54, no. 3, pp. 53–59, Mar. 2016.
- [64] T. Wirth, M. Mehlhose, J. Pilz, B. Holfeld, and D. Wieruch, "5G new radio and ultra low latency applications: A PHY implementation perspective," in *Proc. Asilomar Conf. Signals, Syst. and Comput.*, Nov. 2016, pp. 1409–1413.
- [65] F. Schaich, T. Wild, and Y. Chen, "Waveform Contenders for 5G - Suitability for Short Packet and Low Latency Transmissions," in *Proc. Veh. Technol. Conf. (VTC Spring)*, May 2014, pp. 1–5.
- [66] M. Kasparick, G. Wunder, Y. Chen, F. Schaich, and T. Wild, "5G Waveform Candidate Selection D3.1," in *5Gnow*, Nov. 2013. [Online].
- [67] B. Farhang-Boroujeny, "OFDM Versus Filter Bank Multicarrier," *IEEE Signal Process. Mag.*, vol. 28, no. 3, pp. 92–112, May 2011.
- [68] B. Zhang, H. Shen, B. Yin, L. Lu, D. Chen, T. Wang, L. Gu, X. Wang, X. Hou, H. Jiang, A. Benjebbour, and Y. Kishiyama, "A 5G Trial of Polar Code," in *Proc. IEEE Glob. Works. (GC Wkshps)*, Dec. 2016, pp. 1–6.
- [69] D. Liu, C. Zuo, and Z. Wu, "Benefit and cost of cross sliding window scheduling for low latency 5G Turbo decoding," in *Proc. IEEE/CIC Int. Conf. Commun. China (ICCC)*, Nov. 2015, pp. 1–4.
- [70] I.-G. Jang and G.-D. Jo, "Study on the latency efficient IFFT design method for low latency communication systems," in *Proc. Int. Symp. Intell. Signal Process. Commun. Syst. (ISPACS)*, Oct. 2016, pp. 1–4.
- [71] Y. Zhou, H. Luo, R. Li, and J. Wang, "A dynamic states reduction message passing algorithm for sparse code multiple access," in *Proc. IEEE Wireless Telecommun. Symp. (WTS)*, Apr. 2016, pp. 1–5.
- [72] D. A. Jaoude and M. Farhood, "Balanced truncation of linear systems interconnected over arbitrary graphs with communication latency," in *Proc. IEEE Conf. Decision and Control (CDC)*, Dec. 2015, pp. 5346–5351.

- [73] T. Taheri, R. Nilsson, and J. van de Beek, "Asymmetric Transmit-Windowing for Low-Latency and Robust OFDM," in *Proc. IEEE Globecom Workshop (GC Wkshps)*, Dec. 2016, pp. 1–6.
- [74] C. She, C. Yang, and T. Q. S. Quek, "Cross-layer Transmission Design for Tactile Internet," *CoRR*, vol. abs/1610.02800, 2016. [Online]. Available: <http://arxiv.org/abs/1610.02800>
- [75] H. Hirai, T. Tojo, and M. M. N. Takaya, "Low Latency packet transport methods for remote-controlled devices in multi-RAT environments," in *Proc. IEEE Int. Symp. on Local and Metropolitan Area Netw. (LANMAN)*, Jun. 2016, pp. 1–2.
- [76] N. A. Johansson, Y. P. E. Wang, E. Eriksson, and M. Hessler, "Radio access for ultra-reliable and low-latency 5G communications," in *Proc. IEEE Int. Conf. Commun. Workshop (ICCW)*, Jun. 2015, pp. 1184–1189.
- [77] S. A. Ashraf, F. Lindqvist, R. Baldemair, and B. Lindoff, "Control Channel Design Trade-Offs for Ultra-Reliable and Low-Latency Communication System," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2015, pp. 1–6.
- [78] H. Ji, S. Park, and B. Shim, "Sparse coding for reliable control channel transmission," *to appear on IEEE Commun. Mag.*, 2017.
- [79] S. Xia, X. Han, X. Yan, Z. Zuo, and F. Bi, "Uplink control channel design for 5G ultra-low latency communication," in *Proc. IEEE Person. Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2016, pp. 1–6.
- [80] R. Jin, X. Zhong, and S. Zhou, "The Access Procedure Design for Low Latency in 5G Cellular Network," in *Proc. IEEE Globecom Workshop (GC Wkshps)*, Dec. 2016, pp. 1–6.
- [81] T. Ohseki and Y. Suegara, "Fast outer-loop link adaptation scheme realizing low-latency transmission in LTE-Advanced and future wireless networks," in *Proc. IEEE Radio and Wireless Symp. (RWS)*, Jan. 2016, pp. 1–3.
- [82] B. Lee, S. Park, D. J. Love, H. Ji, and B. Shim, "Packet Structure and Receiver Design for Low-Latency Communications with Ultra-Small Packets," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2016, pp. 1–6.
- [83] H. Tang, W. Zhang, W. Hardjawana, and B. Vucetic, "Improving latency and reliability in 5G Internet-of-Things networks," in *Proc. IEEE Int. Conf. Smart Grid Commun. (SmartGridComm)*, Nov. 2016, pp. 509–513.

- [84] M. Matth, L. L. Mendes, N. Michailow, D. Zhang, and G. Fettweis, "Widely Linear Estimation for Space-Time-Coded GFDM in Low-Latency Applications," *IEEE Trans. Commun.*, vol. 63, no. 11, pp. 4501–4509, Nov. 2015.
- [85] L. D. Gregorio, "A Note on Compressed Sensing for Low Latency," in *Proc. Mobile Syst. Technol. Workshop (MST)*, May 2015, pp. 8–11.
- [86] J. W. Choi, B. Shim, Y. Ding, B. Rao, and D. I. Kim, "Compressed Sensing for Wireless Communications : Useful Tips and Tricks," *IEEE Commun. Surveys Tuts.*, vol. PP, no. 99, pp. 1–1, 2017.
- [87] L. Lu, Y. Chen, W. Guo, H. Yang, Y. Wu, and S. Xing, "Prototype for 5G new air interface technology SCMA and performance evaluation," *China Commun.*, vol. 12, no. Supplement, pp. 38–48, Dec. 2015.
- [88] S. Dutta, M. Mezzavilla, R. Ford, M. Zhang, S. Rangan, and M. Zorzi, "Frame Structure Design and Analysis for Millimeter Wave Cellular Systems," *CoRR*, vol. abs/1512.05691, 2015. [Online]. Available: <http://arxiv.org/abs/1512.05691>
- [89] R. Ford, M. Zhang, M. Mezzavilla, S. Dutta, S. Rangan, and M. Zorzi, "Achieving Ultra-Low Latency in 5G Millimeter Wave Cellular Networks," *CoRR*, vol. abs/1602.06925, 2016. [Online]. Available: <http://arxiv.org/abs/1602.06925>
- [90] T. Levanen, J. Pirskanen, and M. Valkama, "Radio interface design for ultra-low latency millimeter-wave communications in 5G Era," in *Proc. IEEE Globecom Workshop (GC Wkshps)*, Dec. 2014, pp. 1420–1426.
- [91] S. Yoshioka, Y. Inoue, S. Suyama, Y. Kishiyama, Y. Okumura, J. Kepler, and M. Cudak, "Field experimental evaluation of beamtracking and latency performance for 5G mmWave radio access in outdoor mobile environment," in *Proc. IEEE Pers. Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2016, pp. 1–6.
- [92] T. A. Levanen, J. Pirskanen, T. Koskela, J. Talvitie, and M. Valkama, "Radio Interface Evolution Towards 5G and Enhanced Local Area Communications," *IEEE Access*, vol. 2, pp. 1005–1029, 2014.
- [93] R. D. Taranto, S. Muppisetty, R. Raulefs, D. Slock, T. Svensson, and H. Wymeersch, "Location-Aware Communications for 5G Networks: How location information can improve scalability, latency, and robustness of 5G," *IEEE Signal Process. Mag.*, vol. 31, no. 6, pp. 102–112, Nov. 2014.

- [94] D. Wu, J. Wang, Y. Cai, and M. Guizani, "Millimeter-wave multimedia communications: challenges, methodology, and applications," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 232–238, Jan. 2015.
- [95] Huawei, "5G: A technology vision," *White paper*, 2013.
- [96] Y. N. Lien and Y. C. Wun, "Timeliness and QoS Aware Packet Scheduling for Next Generation Networks," pp. 9–16, Aug 2009.
- [97] A. Kamel, A. Al-Fuqaha, and M. Guizani, "Exploiting Client-Side Collected Measurements to Perform QoS Assessment of IaaS," *IEEE Trans. Mobile Comput.*, vol. 14, no. 9, pp. 1876–1887, Sep. 2015.
- [98] O. Bazan and M. Jaseemuddin, "A Conflict Analysis Framework for QoS-Aware Routing in Contention-Based Wireless Mesh Networks with Beamforming Antennas," *IEEE Trans. Wireless Commun.*, vol. 10, no. 10, pp. 3267–3277, Oct. 2011.
- [99] X. Zhang, W. Cheng, and H. Zhang, "Heterogeneous statistical QoS provisioning over 5G mobile wireless networks," *IEEE Network*, vol. 28, no. 6, pp. 46–53, Nov. 2014.
- [100] W. Wang, X. Wang, and A. A. Nilsson, "Energy-efficient bandwidth allocation in wireless networks: algorithms, analysis, and simulations," *IEEE Trans. Wireless Commun.*, vol. 5, no. 5, pp. 1103–1114, May 2006.
- [101] A. Balachandran, V. Sekar, A. Akella, S. Seshan, I. Stoica, and H. Zhang, "Developing a Predictive Model of Quality of Experience for Internet Video," *SIGCOMM Comput. Commun. Rev.*, vol. 43, no. 4, pp. 339–350, Aug. 2013. [Online]. Available: <http://doi.acm.org/10.1145/2534169.2486025>
- [102] A. Roy, P. De, and N. Saxena, "Location-based social video sharing over next generation cellular networks," *IEEE Commun. Mag.*, vol. 53, no. 10, pp. 136–143, Oct. 2015.
- [103] A. Balachandran, V. Sekar, A. Akella, S. Seshan, I. Stoica, and H. Zhang, "A Quest for an Internet Video Quality-of-experience Metric," in *Proc. ACM Workshop Hot Topics in Netw.*, ser. HotNets-XI, 2012, pp. 97–102. [Online]. Available: <http://doi.acm.org/10.1145/2390231.2390248>
- [104] C. N. Mao, M. H. Huang, S. Padhy, S. T. Wang, W. C. Chung, Y. C. Chung, and C. H. Hsu, "Minimizing Latency of Real-Time Container Cloud for Software Ra-

- dio Access Networks,” in *Proc. IEEE Int. Conf. Cloud Comput. Technol. and Sci. (CloudCom)*, Nov 2015, pp. 611–616.
- [105] G. Mountaser, M. Lema Rosas, T. Mahmoodi, and M. Dohler, *On the feasibility of MAC and PHY split in Cloud RAN*. IEEE, 5 2017.
- [106] Y. Gao, W. Hu, K. Ha, B. Amos, P. Pillai, and M. Satyanarayanan, “Are cloudlets necessary?” 2015.
- [107] G. Zhang, T. Q. S. Quek, A. Huang, M. Kountouris, and H. Shan, “Backhaul-aware base station association in two-tier heterogeneous cellular networks,” in *Proc. IEEE Inter. Works. on Sig. Proces. Adv. in Wire. Commun. (SPAWC)*, June 2015, pp. 390–394.
- [108] X. Wang, C. Cavdar, L. Wang, M. Tornatore, Y. Zhao, H. S. Chung, H. H. Lee, S. Park, and B. Mukherjee, “Joint Allocation of Radio and Optical Resources in Virtualized Cloud RAN with CoMP,” in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2016, pp. 1–6.
- [109] C. A. Garcia-Perez and P. Merino, “Enabling Low Latency Services on LTE Networks,” in *Proc. IEEE Int. Workshop Found. Appl. Self Syst. (FASW)*, Sep. 2016, pp. 248–255.
- [110] J. Pag and J. M. Dricot, “Software-defined networking for low-latency 5G core network,” in *Proc. Int. Conf. Military Commun. and Inform. Syst. (ICMCIS)*, May 2016, pp. 1–7.
- [111] R. Trivisonno, R. Guerzoni, I. Vaishnavi, and D. Soldani, “Towards zero latency Software Defined 5G Networks,” in *Proc. IEEE Conf. Commun. Workshop (ICCW)*, Jun. 2015, pp. 2566–2571.
- [112] D. Szabo, A. Gulyas, F. H. P. Fitzek, and D. E. Lucani, “Towards the Tactile Internet: Decreasing Communication Latency with Network Coding and Software Defined Networking,” in *Proc. Euro. Conf.*, May 2015, pp. 1–6.
- [113] J. Heinonen, T. Partti, M. Kallio, K. Lappalainen, H. Flinck, and J. Hillo, “Dynamic tunnel switching for SDN-based cellular core networks,” in *Proc. ACM 4th workshop All things cellular: operations, applications, and challenges*, 2014, pp. 27–32.

- [114] J. Zhang, W. Xie, and F. Yang, "An architecture for 5G mobile network based on SDN and NFV," in *Proc. Int. Conf. Wireless, Mobile and Multi-Media (ICWMMN)*, Nov. 2015, pp. 87–92.
- [115] M. Y. Arslan, K. Sundaresan, and S. Rangarajan, "Software-defined networking in cellular radio access networks: potential and challenges," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 150–156, January 2015.
- [116] X. Costa-Perez, A. Garcia-Saavedra, X. Li, T. Deiss, A. de la Oliva, A. Di Giglio, P. Iovanna, and A. Moored, "5G-crosshaul: An SDN/NFV integrated fronthaul/backhaul transport network architecture," *IEEE Wireless Commun.*, vol. 24, no. 1, pp. 38–45, 2017.
- [117] G. Wang, G. Feng, S. Qin, and R. Wen, "Efficient traffic engineering for 5G core and backhaul networks," *J. Commun. and Netw.*, vol. 19, no. 1, pp. 80–92, 2017.
- [118] R. Ford, A. Sridharan, R. Margolies, R. Jana, and S. Rangan, "Provisioning low latency, resilient mobile edge clouds for 5G," *arXiv preprint arXiv:1703.10915*, 2017.
- [119] C. C. Marquezan, Z. Despotovic, R. Khalili, D. Perez-Caparros, and A. Hecker, "Understanding processing latency of SDN-based mobility management in mobile core networks," in *Proc. IEEE Int. Personal, Indoor, and Mobile Radio Commun. (PIMRC)*, 2016, pp. 1–7.
- [120] M. Moradi, W. Wu, L. E. Li, and Z. M. Mao, "SoftMoW: recursive and reconfigurable cellular WAN architecture," in *Proc. 10th ACM Int. Conf. emerg. Netw. Experiments and Technol.* ACM, 2014, pp. 377–390.
- [121] M. Moradi, L. E. Li, and Z. M. Mao, "SoftMoW: A dynamic and scalable software defined architecture for cellular WANs," in *Proc. 3rd workshop on Hot topics in software defined networking.* ACM, 2014, pp. 201–202.
- [122] A. S. Thyagaturu, Y. Dashti, and M. Reisslein, "SDN-based smart gateways (SmGWs) for multi-operator small cell network management," *IEEE Trans. on Netw. and Serv. Manag.*, vol. 13, no. 4, pp. 740–753, 2016.
- [123] A. Jain, N. Sadagopan, S. K. Lohani, and M. Vutukuru, "A comparison of SDN and NFV for re-designing the LTE Packet Core," in *Proc. IEEE conf. Netw.Function Virtualization and Software Defined Networks (NFV-SDN)*, 2016, pp. 74–80.

- [124] C. Liang, F. R. Yu, and X. Zhang, "Information-centric network function virtualization over 5G mobile wireless networks," *IEEE Network*, vol. 29, no. 3, pp. 68–74, May 2015.
- [125] E. Cau, M. Corici, P. Bellavista, L. Foschini, G. Carella, A. Edmonds, and T. M. Bohnert, "Efficient exploitation of mobile edge computing for virtualized 5G in EPC architectures," in *Proc. IEEE Mobile Cloud Computing, Services, and Eng. (MobileCloud)*, Mar. 2016, pp. 100–109.
- [126] B. Martini, F. Paganelli, P. Cappanera, S. Turchi, and P. Castoldi, "Latency-aware composition of Virtual Functions in 5G," in *Proc. IEEE Conf. Netw. Softw. (NetSoft)*, Apr. 2015, pp. 1–6.
- [127] J. Heinonen, P. Korja, T. Partti, H. Flinck, and P. Pyhnen, "Mobility management enhancements for 5G low latency services," in *Proc. IEEE Int. Conf. on Commun. Workshop (ICCW)*, May 2016, pp. 68–73.
- [128] T. Taleb, "Toward carrier cloud: Potential, challenges, and solutions," *IEEE Wireless Commun.*, vol. 21, no. 3, pp. 80–91, 2014.
- [129] H. Hawilo, A. Shami, M. Mirahmadi, and R. Asal, "NFV: state of the art, challenges, and implementation in next generation mobile networks (vEPC)," *IEEE Network*, vol. 28, no. 6, pp. 18–26, Nov 2014.
- [130] J. Zhang, X. Zhang, and W. Wang, "Cache-Enabled Software Defined Heterogeneous Networks for Green and Flexible 5G Networks," *IEEE Access*, vol. 4, pp. 3591–3604, 2016.
- [131] S. González, A. Oliva, X. Costa-Pérez, A. Di Giglio, F. Cavaliere, T. Deiß, X. Li, and A. Mourad, "5G-Crosshaul: An DN/NFV control and data plane architecture for the 5G integrated Fronthaul/Backhaul," *Trans Emerg. Telecommun. Technol.*, vol. 27, no. 9, pp. 1196–1205, 2016.
- [132] J. Chen and J. Li, "Efficient mobile backhaul architecture offering ultra-short latency for handovers," in *Proc. Int. Conf. Transparent Opt. Netw. (ICTON)*, Jul. 2016, pp. 1–1.
- [133] D. Schulz, C. Alexakis, J. Hilt, M. Schlosser, K. Habel, V. Jungnickel, and R. Freund, "Low Latency Mobile Backhauling using Optical Wireless Links," in *Proc. Symp. Broadband Coverage*, Apr. 2015, pp. 1–3.

- [134] R. J. Weiler, M. Peter, W. Keusgen, E. Calvanese-Strinati, A. De Domenico, I. Filippini, A. Capone, I. Siaud, A.-M. Ulmer-Moll, A. Maltsev *et al.*, “Enabling 5G backhaul and access with millimeter-waves,” in *Proc. IEEE Euro. Conf. Netw. Commun. (EuCNC)*, 2014, pp. 1–5.
- [135] T. Levanen, J. Pirskanen, and M. Valkama, “Radio interface design for ultra-low latency millimeter-wave communications in 5G era,” in *Proc. IEEE Globecom Workshop (GC Wkshps)*, Dec. 2014, pp. 1420–1426.
- [136] Z. Gao, L. Dai, D. Mi, Z. Wang, M. A. Imran, and M. Z. Shakir, “MmWave massive-MIMO-based wireless backhaul for the 5G ultra-dense network,” *IEEE Wireless Commun.*, vol. 22, no. 5, pp. 13–21, 2015.
- [137] R. Taori and A. Sridharan, “Point-to-multipoint in-band mmwave backhaul for 5G networks,” *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 195–201, 2015.
- [138] A. S. Thyagaturu, A. Mercian, M. P. McGarry, M. Reisslein, and W. Kellerer, “Software Defined Optical Networks (SDONs): A Comprehensive Survey,” *IEEE Commun. Surv. Tutor.*, vol. 18, no. 4, pp. 2738–2786, Fourthquarter 2016.
- [139] J. S. Vardakas, I. T. Monroy, L. Wosinska, G. Agapiou, R. Brenot, N. Pleros, and C. Verikoukis, “Towards high capacity and low latency backhauling in 5G: The 5G STEP-FWD vision,” in *Proc. 19th Int. Conf. Transparent Opt. Netw. (ICTON)*, July 2017, pp. 1–4.
- [140] K. Shanmugam, N. Golrezaei, A. G. Dimakis, A. F. Molisch, and G. Caire, “Femto-Caching: Wireless Content Delivery Through Distributed Caching Helpers,” *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 8402–8413, Dec. 2013.
- [141] X. Peng, J. C. Shen, J. Zhang, and K. B. Letaief, “Backhaul-Aware Caching Placement for Wireless Networks,” in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2015, pp. 1–6.
- [142] J. Liu, B. Bai, J. Zhang, and K. B. Letaief, “Content caching at the wireless network edge: A distributed algorithm via belief propagation,” in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2016, pp. 1–6.
- [143] M. Ji, G. Caire, and A. F. Molisch, “Fundamental Limits of Caching in Wireless D2D Networks,” *IEEE Trans. Inf. Theory*, vol. 62, no. 2, pp. 849–869, Feb. 2016.

- [144] W. Jiang, G. Feng, and S. Qin, "Optimal cooperative content caching and delivery policy for heterogeneous cellular networks," *IEEE Trans. Mobile Comput.*, vol. 16, no. 5, pp. 1382–1393, 2017.
- [145] H. Hsu and K. C. Chen, "A Resource Allocation Perspective on Caching to Achieve Low Latency," *IEEE Commun. Lett.*, vol. 20, no. 1, pp. 145–148, Jan. 2016.
- [146] G. Carofiglio, L. Mekinda, and L. Muscariello, "FOCAL: Forwarding and Caching with Latency Awareness in Information-Centric Networking," in *Proc. IEEE Global Workshop (GC Wkshps)*, Dec. 2015, pp. 1–7.
- [147] X. Huang, Z. Zhao, and H. Zhang, "Latency Analysis of Cooperative Caching with Multicast for 5G Wireless Networks," in *Proc. IEEE/ACM Int. Conf. Utility and Cloud Computing (UCC)*, Dec. 2016, pp. 316–320.
- [148] L. Jiang, G. Feng, and S. Qin, "Cooperative content distribution for 5G systems based on distributed cloud service network," in *Proc. IEEE Int. Conf. Commun. Workshop (ICCW)*, Jun. 2015, pp. 1125–1130.
- [149] Y. Liu, C. Yang, Y. Yao, and B. Xia, "Modeling and analysis on end-to-end performance of cache-enabled networks," in *Proc. Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Oct. 2016, pp. 1–5.
- [150] S. K. E. B. N. K. U. N. M. D. G. M. Nabeel, J. Zage, "Cryptographic Key Management for Smart Power Grids," CERIAS Technical Report, Tech. Rep., February, 2012.
- [151] M. Ali, E. Al-Shaer, and Q. Duan, "Randomizing AMI configuration for proactive defense in smart grid," in *Proc. IEEE Smart Grid Commun. (SmartGridComm)*, Oct 2013, pp. 618–623.
- [152] B. Yuce, M. Mourshed, Y. Rezgui, and O. F. Rana, "Preserving prosumer privacy in a district level smart grid," in *2016 IEEE International Smart Cities Conference (ISC2)*, Sept 2016, pp. 1–6.
- [153] C. Efthymiou and G. Kalogridis, "Smart Grid Privacy via Anonymization of Smart Metering Data," in *Proc. IEEE Intern. Conf. on Smart Grid Commun. (SmartGridComm)*, Oct 2010, pp. 238–243.
- [154] S. Tonyali, K. Akkaya, N. Saputro, and A. S. Uluagac, "A reliable data aggregation mechanism with Homomorphic Encryption in Smart Grid AMI networks," in *2016*

13th IEEE Annual Consumer Communications Networking Conference (CCNC), Jan 2016, pp. 550–555.

- [155] A. Abdallah and X. Shen, “A Lightweight Lattice-based Homomorphic Privacy-Preserving Data Aggregation Scheme for Smart Grid,” *IEEE Trans. on Smart Grid*, vol. PP, no. 99, pp. 1–1, 2016.
- [156] H. So, S. Kwok, E. Lam, and K.-S. Lui, “Zero-Configuration Identity-Based Sign-encryption Scheme for Smart Grid,” in *Proc. First IEEE Int. Conf. on Smart Grid Communications (SmartGridComm)*, Oct 2010, pp. 321–326.
- [157] Y. Yan, Y. Qian, and H. Sharif, “A secure and reliable in-network collaborative communication scheme for advanced metering infrastructure in smart grid,” in *Proc. IEEE Int. Conf. on Wireless Communications and Networking Conference (WCNC)*, March 2011, pp. 909–914.
- [158] Q. Li and G. Cao, “Multicast authentication in the smart grid with one-time signature,” *IEEE Trans. on Smart Grid*, vol. 2, no. 4, pp. 686–696, Dec 2011.
- [159] H. Nicanfar, P. Jokar, and V. C. M. Leung, “Smart grid authentication and key management for unicast and multicast communications,” in *Proc. IEEE PES Innov. Smart Grid Techno.*, Nov 2011, pp. 1–8.
- [160] T. B. Florian Skopika, Zhendong Maa and H. Grneis, “A Survey on Threats and Vulnerabilities in Smart Metering Infrastructures ,” *Intern. Journ. of Smart Grid and Clean Energy*, 2013.
- [161] S. R. Bajekal, J. S. Arun, M. D. Nix, and K. K. Yellepeddy, “Event-driven, asset-centric key management in a smart grid,” Aug. 24 2017, US Patent App. 15/445,087.
- [162] D. Wu and C. Zhou, “Fault-tolerant and scalable key management for smart grid,” *IEEE Trans. on Smart Grid*, vol. 2, no. 2, pp. 375–381, June 2011.
- [163] H. Nicanfar and V. C. M. Leung, “Smart grid multilayer consensus password-authenticated key exchange protocol,” in *Proc. IEEE Intern. Conf. on Commun. (ICC)*, June 2012, pp. 6716–6720.
- [164] Z. Sun and J. Ma, “Efficient key management for advanced distribution automation system,” in *Proc. IEEE Intern. Conf. on Network Infr. and Digital Content*, Sept 2010, pp. 794–798.

- [165] A. V. D. M. Kayem, H. Strauss, S. D. Wolthusen, and C. Meinel, “Key Management for Secure Demand Data Communication in Constrained Micro-Grids,” in *Proc. Intern. Conf. on Advanced Inform. Network. and Appli. Workshops (WAINA)*, March 2016, pp. 585–590.
- [166] R. Bhatia and V. Bodade, “Defining the framework for wireless-AMI security in smart grid,” in *Proc. Intern. Conf. on Green Comput. Commun. and Elect. Eng. (ICGCCEE)*, March 2014, pp. 1–5.
- [167] M. Thomas, I. Ali, and N. Gupta, “A secure way of exchanging the secret keys in advanced metering infrastructure,” in *Proc. IEEE Conf. on Power Sys. Techno. (POWERCON)*, Oct 2012, pp. 1–7.
- [168] K. Yu, M. Arifuzzaman, Z. Wen, D. Zhang, and T. Sato, “A Key Management Scheme for Secure Communications of Information Centric Advanced Metering Infrastructure in Smart Grid,” *IEEE Trans. on Instrum. and Measure.*, vol. 64, no. 8, pp. 2072–2085, Aug 2015.
- [169] P.-H. Hsu, W. Tang, C. Tsai, and B.-C. Cheng, “Two-Layer Security Scheme for AMI System in Taiwan,” in *Proc. IEEE Conf. on Para. and Distri. Process. with Appl. Works. (ISPAW)*, May 2011, pp. 105–110.
- [170] Qualcomm, “Extending the benefits of LTE to unlicensed spectrum,” 3GPP RAN1 standard contribution - RWS-140008, Sophia Antipolis, France, Jun. 2014.
- [171] N. DOCOMO, “Views on LAA for Unlicensed Spectrum - Scenarios and Initial Evaluation Results,” 3GPP RAN1 standard contribution - RWS-140026, Sophia Antipolis, France, Jun. 2014.
- [172] 3GPP, “LTE in unlicensed spectrum,” 3GPP Technical Report, Tech. Rep., Jun. 2014.
- [173] IEEE Standard 802.15.4, IEEE, Jun. 2003.
- [174] E. Perahida and R. Stacey, “Next Generation Wireless LANs: Throughput, Capacity, and Reliability in 802.11n,” 2008.
- [175] I. Parvez, A. Sundararajan, and A. I. Sarwat, “Frequency band for han and nan communication in smart grid,” in *Proc. IEEE Symp. Series on Comp. Intell.*

- [176] 3GPP, “Evolved Universal Terrestrial Radio Access (E-UTRA); Further advancements for E-UTRA physical layer aspects (Release 9),” 3GPP Technical Report, Tech. Rep., Mar. 2010.
- [177] J. Markkula and J. Haapola, “LTE and hybrid sensor-LTE network performances in smart grid demand response scenarios,” in *Proc. IEEE Int. Conf. Smart Grid Commun.*, Oct. 2013, pp. 187–192.
- [178] A. F. Molisch, *3GPP LongTerm Evolution*. Wiley-IEEE Press, 2011, pp. 665–698. [Online]. Available: <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5635449>
- [179] 3GPP, “Study on Licensed-Assisted Access using LTE,” RP-141397, Tech. Rep., Sep. 2014.
- [180] I. Parvez, M. Sriyananda, İ. Güvenç, M. Bennis, and A. Sarwat, “CBRS Spectrum Sharing between LTE-U and WiFi: A Multiarmed Bandit Approach,” *Mobile Inform. Systems*, vol. 2016, 2016.
- [181] M. G. S. Sriyananda, I. Parvez, I. Gvene, M. Bennis, and A. I. Sarwat, “Multi-armed bandit for LTE-U and WiFi coexistence in unlicensed bands,” in *Proc. IEEE Wire. Commun. and Netw. Conf.*, April 2016, pp. 1–6.
- [182] “Study on Licensed-Assisted Access using LTE,” 3GPP Study Item, Edinburgh, Scotland, Tech. Rep. RP-141397, 2014.
- [183] “Study on licensed-assisted access to unlicensed spectrum,” 3GPP, Athens, Greece, Tech. Rep. TR 36.899, Feb. 2015.
- [184] J. C. Gittins, “Bandit Processes and Dynamic Allocation Indices,” *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 41, no. 2, pp. 148–177, 1979. [Online]. Available: <http://www.jstor.org/stable/2985029>
- [185] P. Auer, N. Cesa-Bianchi, and P. Fischer, “Finite-time Analysis of the Multiarmed Bandit Problem,” *Machine Learning*, vol. 47, no. 2, pp. 235–256, 2002. [Online]. Available: <http://dx.doi.org/10.1023/A:1013689704352>
- [186] S. Sesia, I. Toufik, and M. Baker, *LTE: The UMTS Long Term Evolution, From Theory to Practice*. John Wiley & Sons Ltd, Feb. 2009.

- [187] E. Perahia and R. Stacey, *LTE: The UMTS Long Term Evolution, From Theory to Practice*. Cambridge University Press, 2008.
- [188] S. Agrawal and N. Goyal, “Analysis of thompson sampling for the multi-armed bandit problem,” *arXiv preprint arXiv:1111.1797*, 2011.
- [189] N. Gupta, O. C. Granmo, and A. wala, “Thompson Sampling for Dynamic Multi-armed Bandits,” in *Proc. Int. Conf. on Machine Learning and Applications and Workshops*, vol. 1, Honolulu, HI, Dec. 2011, pp. 484–489.
- [190] P. Auer, N. Cesa-Bianchi, and P. Fischer, “Finite-time Analysis of the Multiarmed Bandit Problem,” *Machine Learning*, vol. 47, no. 2, pp. 235–256. [Online]. Available: <http://dx.doi.org/10.1023/A:1013689704352>
- [191] J. Langford and T. Zhang, “The Epoch-Greedy Algorithm for Multi-armed Bandits with Side Information,” in *Advances in Neural Information Processing Systems 20*, J. C. Platt, D. Koller, Y. Singer, and S. T. Roweis, Eds. Curran Associates, Inc., 2008, pp. 817–824.
- [192] “Evolved Universal Terrestrial Radio Access (E-UTRA); Further advancements for E-UTRA physical layer aspects (Release 9),” 3GPP, Tech. Rep. TR36.814, V9.0.0, 2010.
- [193] “Tactile Internet,” Tactile Internet ad-hoc definition group, IEEE P1918.1, 2016.
- [194] M. Agiwal, A. Roy, and N. Saxena, “Next Generation 5G Wireless Networks: A Comprehensive Survey,” *IEEE Commun. Surv. Tutor.*, vol. 18, no. 3, pp. 1617–1655, thirdquarter 2016.
- [195] A. Gupta and R. K. Jha, “A Survey of 5G Network: Architecture and Emerging Technologies,” *IEEE Access*, vol. 3, pp. 1206–1232, 2015.
- [196] ITU-R, “Framework and Overall Objectives of the Future Development of IMT for 2020 and Beyond,” Feb. 2015.
- [197] A. Kumbhar, F. Koohifar, I. Guvenc, and B. Mueller, “A Survey on Legacy and Emerging Technologies for Public Safety Communications,” *IEEE Commun. Surv. Tutor.*, vol. 19, no. 1, pp. 97–124, Firstquarter 2017.

- [198] H. Ji, S. Park, J. Yeo, Y. Kim, J. Lee, and B. Shim, "Introduction to Ultra Reliable and Low Latency Communications in 5G," *arXiv preprint arXiv:1704.05565*, 2017.
- [199] A. H. Jalal, J. Yu, and A. A. Nnanna, "Fabrication and calibration of oxazine-based optic fiber sensor for detection of ammonia in water," *Applied optics*, vol. 51, no. 17, pp. 3768–3775, 2012.
- [200] P. Schulz, M. Matthe, H. Klessig, M. Simsek, G. Fettweis, J. Ansari, S. A. Ashraf, B. Almeroth, J. Voigt, I. Riedel *et al.*, "Latency critical IoT applications in 5G: Perspective on the design of radio interface and network architecture," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 70–78, 2017.
- [201] M. R. Palattella, M. Dohler, A. Grieco, G. Rizzo, J. Torsner, T. Engel, and L. Ladid, "Internet of things in the 5G era: Enablers, architecture, and business models," *Journ. on Select. Areas in Commun.*, vol. 34, no. 3, pp. 510–527, 2016.
- [202] M. Tavana, A. Rahmati, and V. Shah-Mansouri, "Congestion control with adaptive access class barring for LTE M2M overload using kalman filters," *Computer Networks*, 2018.
- [203] I. Parvez, F. Abdul, H. Mohammed, and A. I. Sarwat, "Reliability assessment of access point of advanced metering infrastructure based on Bellcore standards (Telecordia)," in *Proc. IEEE SoutheastCon*, April 2015, pp. 1–7.
- [204] A. Ferdowsi and W. Saad, "Deep learning-based dynamic watermarking for secure signal authentication in the internet of things," in *arXiv preprint arXiv:1711.01306v1*.
- [205] I. Parvez, F. Abdul, and A. I. Sarwat, "A location based key management system for advanced metering infrastructure of smart grid," in *Proc. IEEE Green Techno. Conf. (GreenTech)*, April 2016, pp. 62–67.
- [206] O. N. C. Yilmaz, Y. P. E. Wang, N. A. Johansson, N. Brahmi, S. A. Ashraf, and J. Sachs, "Analysis of ultra-reliable and low-latency 5G communication for a factory automation use case," in *Proc. IEEE Int. Conf. Commun. Workshop (ICCW)*, Jun. 2015, pp. 1190–1195.
- [207] B. Holfeld, D. Wieruch, T. Wirth, L. Thiele, S. A. Ashraf, J. Huschke, I. Aktas, and J. Ansari, "Wireless communication for factory automation: An opportunity for LTE and 5G systems," *IEEE Commun. Mag.*, vol. 54, no. 6, pp. 36–43, 2016.

- [208] A. F. Cattoni, D. Chandramouli, C. Sartori, R. Stademann, and P. Zanier, “Mobile Low Latency Services in 5G,” in *Proc. IEEE Veh. Technol. Conf. (VTC Spring)*, May 2015, pp. 1–6.
- [209] C. Campolo, A. Molinaro, G. Araniti, and A. O. Berthet, “Better Platooning Control Toward Autonomous Driving: An LTE Device-to-Device Communications Strategy That Meets Ultra-low Latency Requirements,” *IEEE Veh. Technol. Mag.*, vol. 12, no. 1, pp. 30–38, March 2017.
- [210] R. Alieiev, A. Kwoczek, and T. Hehn, “Automotive requirements for future mobile networks,” in *Proc. IEEE MTT-S Inter. Conf. on Micro. for Intell. Mobil. (ICMIM)*, Apr. 2015, pp. 1–4.
- [211] M. A. Lema, A. Laya, T. Mahmoodi, M. Cuevas, J. Sachs, J. Markendahl, and M. Dohler, “Business Case and Technology Analysis for 5G Low Latency Applications,” *IEEE Access*, vol. 5, pp. 5917–5935, 2017.
- [212] ITU-Report, “The Tactile Internet,” *ITU-T Technology Watch Report*, Aug. 2014. [Online]. Available: <https://www.itu.int/oth/T2301000023/en>
- [213] M. Simsek, A. Aijaz, M. Dohler, J. Sachs, and G. Fettweis, “The 5G-Enabled Tactile Internet: Applications, requirements, and architecture,” in *Proc. IEEE Wireless. Commun. Netw. Conf. (WCNC)*, Apr. 2016, pp. 1–6.
- [214] N. A. Johansson, Y. P. E. Wang, E. Eriksson, and M. Hessler, “Radio access for ultra-reliable and low-latency 5G communications,” in *Proc. IEEE Intern. Conf. on Comm. Works. (ICCW)*, Jun. 2015, pp. 1184–1189.
- [215] J. Choi, V. Va, N. Gonzalez-Prelcic, R. Daniels, C. R. Bhat, and R. W. Heath, “Millimeter-wave vehicular communication to support massive automotive sensing,” *Comm. Mag.*, vol. 54, no. 12, pp. 160–167, Dec. 2016. [Online]. Available: <https://doi.org/10.1109/MCOM.2016.1600071CM>
- [216] M. A. Lema, K. Antonakoglou, F. Sardis, N. Sornkarn, M. Condoluci, T. Mahmoodi, and M. Dohler, “5G case study of Internet of Skills: Slicing the human senses,” in *Proc. Europ. Conf. on Netw. and Commun. (EuCNC)*, Jun. 2017, pp. 1–6.
- [217] A. Ghosh, “5G mmWave Revolution and New Radio,” https://5g.ieee.org/images/files/pdf/5GmmWave_Webinar_IEEE_Nokia_09_20_2017_final.pdf, Sep., 2017, [Online; accessed 19-Nov-2017].

- [218] “Communications Requirements Of Smart Grid Technologies,” US Department of Energy, Tech. Rep., Oct. 2010. [Online]. Available: <http://www.smartgrid.gov/>
- [219] A. T. Z. Kasgari, W. Saad, and M. Debbah, “Human-in-the-loop wireless communications: Machine learning and brain-aware resource management,” *arXiv preprint arXiv:1804.00209*, 2018.
- [220] I. Parvez, M. Jamei, A. Sundararajan, and A. I. Sarwat, “RSS based loop-free compass routing protocol for data communication in advanced metering infrastructure (AMI) of Smart Grid,” in *Proc. IEEE Symp. on Comp. Intell. Appl. in Smart Grid (CIASG)*, Dec 2014, pp. 1–6.
- [221] I. Parvez, A. Islam, and F. Kaleem, “A key management-based two-level encryption method for ami,” in *Proc. IEEE PES Gen. Meet. Conf. Expos.*, July 2014, pp. 1–5.
- [222] I. Parvez, A. I. Sarwat, J. Pinto, Z. Parvez, and M. A. Khandaker, “A gossip algorithm based clock synchronization scheme for smart grid applications,” in *Proc. North Ameri. Power Sympo. (NAPS)*, Sep. 2017, pp. 1–6.
- [223] C. Xu, J. Cosmas, Y. Zhang, P. Lazaridis, G. Araniti, and Z. D. Zaharis, “3D MIMO radio channel modeling of a weighted linear array system of antennas for 5G cellular systems,” in *Proc. Int. Conf. Telecommun. Multimedia (TEMU)*, July 2016, pp. 1–6.
- [224] Latency Analysis in LTE network. [Online]. Available: <http://www.techmahindra.com/Documents/WhitePaper/WhitePaperLatencyAnalysis.pdf>
- [225] K. Wilner and A. P. Van Den Heuvel, “Fiber-optic delay lines for microwave signal processing,” *Proc. IEEE*, vol. 64, no. 5, pp. 805–807, 1976.
- [226] G. Pocovi, K. I. Pedersen, B. Soret, M. Lauridsen, and P. Mogensen, “On the impact of multi-user traffic dynamics on low latency communications,” in *Proc. IEEE Int. Symp. Wireless Commun. Syst. (ISWCS)*, 2016, pp. 204–208.
- [227] ETSI, “Universal Mobile Telecommunications System (UMTS); Feasibility study for evolved Universal Terrestrial Radio Access (UTRA) and Universal Terrestrial Radio Access Network (UTRAN),” ETSI TR 125 912 V7.1.0, Tech. Rep., 09 2006.
- [228] P. K. Agyapong, M. Iwamura, D. Staehle, W. Kiess, and A. Benjebbour, “Design considerations for a 5G network architecture,” *IEEE Commun. Mag.*, vol. 52, no. 11, pp. 65–75, Nov. 2014.

- [229] G. Wunder, P. Jung, M. Kasparick, T. Wild, F. Schaich, Y. Chen, S. T. Brink, I. Gaspar, N. Michailow, A. Festag, L. Mendes, N. Cassiau, D. Ktenas, M. Dryjanski, S. Pietrzyk, B. Eged, P. Vago, and F. Wiedmann, “5GNOW: non-orthogonal, asynchronous waveforms for future mobile applications,” *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 97–105, Feb. 2014.
- [230] E. Lhetkangas, K. Pajukoski, E. Tirola, G. Berardinelli, I. Harjula, and J. Vihril, “On the TDD subframe structure for beyond 4G radio access network,” in *2013 Future Network Mobile Summit*, Jul. 2013, pp. 1–10.
- [231] J. Vihril, A. A. Zaidi, V. Venkatasubramanian, N. He, E. Tirola, J. Medbo, E. Lhetkangas, K. Werner, K. Pajukoski, A. Cedergren, and R. Baldemair, “Numerology and frame structure for 5G radio access,” in *Proc. IEEE Symp. Pers. Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2016, pp. 1–5.
- [232] R. Abreu, P. Mogensen, and K. I. Pedersen, “Pre-Scheduled Resources for Retransmissions in Ultra-Reliable and Low Latency Communications,” in *Proc. IEEE Wireless Commun. and Netw. Conf. (WCNC)*, Mar. 2017, pp. 1–5.
- [233] B. Soret, P. Mogensen, K. I. Pedersen, and M. C. Aguayo-Torres, “Fundamental tradeoffs among reliability, latency and throughput in cellular networks,” in *Proc. IEEE Globecom Workshop (GC Wkshps)*, 2014, pp. 1391–1396.
- [234] K. Ganesan, T. Soni, S. Nunna, and A. R. Ali, “Poster: A TDM approach for latency reduction of ultra-reliable low-latency data in 5G,” in *Proc. IEEE Veh. Netw. Conf. (VNC)*, Dec. 2016, pp. 1–2.
- [235] J. Östman, G. Durisi, E. G. Ström, J. Li, H. Sahlin, and G. Liva, “Low-latency Ultra Reliable 5G Communications: Finite-Blocklength Bounds and Coding Schemes,” *arXiv preprint arXiv:1611.08714*, 2016.
- [236] P. Niroopan and Y.-h. Chung, “A user-spread interleave division multiple access system,” *Int. J. Adv. Resear. Comput. and Commun. Eng.*, pp. 837–841, 2012.
- [237] J. C. Fricke, H. Schoeneich, and P. A. Hoeher, “An interleave-division multiple access based system proposal for the 4G uplink,” *Proc. IST Mobil. Wire. Commun. Summit*, 2005.
- [238] H. Nikopour and H. Baligh, “Sparse code multiple access,” in *Proc. IEEE Pers. Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2013, pp. 332–336.

- [239] N. Michailow, I. Gaspar, S. Krone, M. Lentmaier, and G. Fettweis, “Generalized frequency division multiplexing: Analysis of an alternative multi-carrier technique for next generation cellular systems,” in *Proc. Int. Symp. Wireless Commun. Syst. (ISWCS)*, Aug. 2012, pp. 171–175.
- [240] A. Sahin, I. Guvenc, and H. Arslan, “A Survey on Multicarrier Communications: Prototype Filters, Lattice Structures, and Implementation Aspects,” *IEEE Commun. Surv. Tutor.*, vol. 16, no. 3, pp. 1312–1338, Third 2014.
- [241] F. Schaich and T. Wild, “Waveform contenders for 5G OFDM vs. FBMC vs. UFMFC,” in *Proc. Int. Symp. on Commun., Control and Signal Process. (ISCCSP)*, May 2014, pp. 457–460.
- [242] S. Dutta, M. Mezzavilla, R. Ford, M. Zhang, S. Rangan, and M. Zorzi, “MAC layer frame design for millimeter wave cellular system,” in *Proc. Euro. Conf. Netw. Commun. (EuCNC)*, Jun. 2016, pp. 117–121.
- [243] S. Goyal, P. Liu, S. S. Panwar, R. A. Difazio, R. Yang, and E. Bala, “Full duplex cellular systems: will doubling interference prevent doubling capacity?” *IEEE Commun. Mag.*, vol. 53, no. 5, pp. 121–127, May 2015.
- [244] G. Zheng, “Joint Beamforming Optimization and Power Control for Full-Duplex MIMO Two-Way Relay Channel,” *IEEE Trans. Signal Process.*, vol. 63, no. 3, pp. 555–566, Feb. 2015.
- [245] E. Ahmed, A. M. Eltawil, and A. Sabharwal, “Rate Gain Region and Design Trade-offs for Full-Duplex Wireless Communications,” *IEEE Trans. Wireless Commun.*, vol. 12, no. 7, pp. 3556–3565, Jul. 2013.
- [246] A. Osseiran, F. Boccardi, V. Braun, K. Kusume, P. Marsch, M. Maternia, O. Queseth, M. Schellmann, H. Schotten, H. Taoka, H. Tullberg, M. A. Uusitalo, B. Timus, and M. Fallgren, “Scenarios for 5G mobile and wireless communications: the vision of the METIS project,” *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 26–35, May 2014.
- [247] A. Rahmati, A. Sadeghi, and V. Shah-Mansouri, “Price-based resource allocation for full duplex self-backhauled small cell networks,” in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2015, pp. 5709–5714.
- [248] A. Rahmati, V. Shah-Mansouri, and M. Safari, “Price-based resource allocation for self-backhauled small cell networks,” *Computer Communications*, vol. 97, pp. 72–80, 2017.

- [249] E. Shin and G. Jo, "Uplink frame structure of short TTI system," in *Proc. Int. Conf. Advanced Commun. Technol. (ICACT)*, Feb. 2017, pp. 827–830.
- [250] Y. Hwang and J. Shin, "Slot based radio resource management for low latency in LTE-Advanced system," in *Proc. Int. Conf. Advanced Commun. Technol. (ICACT)*, Feb. 2017, pp. 244–246.
- [251] R. Shreevastav and R. S. Carbajo, "Dynamic RLC mode based upon link adaptation to reduce latency and improve throughput in cellular networks," in *Proc. IEEE Annual Ubiquitous Comput. Electronics Mobile Commun. Conf. (UEMCON)*, Oct. 2016, pp. 1–6.
- [252] H. Li, M. Dong, and K. Ota, "Control Plane Optimization in Software-Defined Vehicular Ad Hoc Networks," *Proc. IEEE Trans. Veh. Technol.*, vol. 65, no. 10, pp. 7895–7904, Oct 2016.
- [253] T. D. Assefa, R. Hoque, E. Tragos, and X. Dimitropoulos, "SDN-based local mobility management with X2-interface in femtocell networks," in *Proc. IEEE Int. Workshops on Comput. Aided Model and Design of Commun. Links and Netw. (CAMAD)*, June 2017, pp. 1–6.
- [254] L. C. Gimenez, P. H. Michaelsen, K. I. Pedersen, T. E. Kolding, and H. C. Nguyen, "Towards Zero Data Interruption Time with Enhanced Synchronous Handover," in *Proc. IEEE Veh. Technol. Conf.*, June 2017, pp. 1–6.
- [255] J. Li and J. Chen, "Passive optical network based mobile backhaul enabling ultra-low latency for communications among base stations," *IEEE/OSA J. Opt. Commun. and Netw.*, vol. 9, no. 10, pp. 855–863, Oct 2017.
- [256] Z. Guizani and N. Hamdi, "CRAN, H-CRAN, and F-RAN for 5G systems: Key capabilities and recent advances," *Int. J. Netw. Management*, vol. 27, no. 5, pp. e1973–n/a, 2017, e1973 nem.1973. [Online]. Available: <http://dx.doi.org/10.1002/nem.1973>
- [257] X. Hong, J. Wang, C.-X. Wang, and J. Shi, "Cognitive radio in 5G: a perspective on energy-spectral efficiency trade-off," *IEEE Commun. Mag.*, vol. 52, no. 7, pp. 46–53, 2014.
- [258] G. Ding, J. Wang, Q. Wu, Y.-D. Yao, R. Li, H. Zhang, and Y. Zou, "On the limits of predictability in real-world radio spectrum state dynamics: From entropy theory to 5G spectrum sharing," *IEEE Commun. Mag.*, vol. 53, no. 7, pp. 178–183, 2015.

- [259] L. Zhang, M. Xiao, G. Wu, M. Alam, Y. C. Liang, and S. Li, "A Survey of Advanced Techniques for Spectrum Sharing in 5G Networks," *IEEE Wireless Commun.*, vol. 24, no. 5, pp. 44–51, Oct. 2017.
- [260] I. Parvez, M. Sriyananda, I. Güvenc, M. Bennis, and A. Sarwat, "CBRS spectrum sharing between lte-u and wifi: A multiarmed bandit approach," *Mobile Inform. Sys.*, vol. 2016, 2016.
- [261] G. Ding, J. Wang, Q. Wu, Y. D. Yao, F. Song, and T. A. Tsiftsis, "Cellular-Base-Station-Assisted Device-to-Device Communications in TV White Space," *IEEE Journ. on Select. Areas in Commun.*, vol. 34, no. 1, pp. 107–121, Jan 2016.
- [262] K. Lin, W. Wang, X. Wang, W. Ji, and J. Wan, "QoE-driven spectrum assignment for 5G wireless networks using SDR," *IEEE Wireless Commun.*, vol. 22, no. 6, pp. 48–55, December 2015.
- [263] Y. Li, T. Jiang, M. Sheng, and Y. Zhu, "QoS-Aware Admission Control and Resource Allocation in Underlay Device-to-Device Spectrum-Sharing Networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 11, pp. 2874–2886, Nov 2016.
- [264] N. Zhang, S. Zhang, J. Zheng, X. Fang, J. W. Mark, and X. Shen, "QoE Driven Decentralized Spectrum Sharing in 5G Networks: Potential Game Approach," *IEEE Trans. Veh. Technol.*, vol. 66, no. 9, pp. 7797–7808, Sep. 2017.
- [265] Z. Zhang, W. Zhang, S. Zeadally, Y. Wang, and Y. Liu, "Cognitive radio spectrum sensing framework based on multi-agent architecture for 5G networks," *IEEE Wireless Commun.*, vol. 22, no. 6, pp. 34–39, Dec. 2015.
- [266] H. Celebi, N. Maxemchuk, Y. Li, and I. Guvenc, "Energy reduction in small cell networks by a random on/off strategy," in *Proc. IEEE Global Commun. Conf. (GLOBECOM) Workshop*, 2013, pp. 176–181.
- [267] H. Celebi and I. Guvenc, "Load Analysis and Sleep Mode Optimization for Energy-Efficient 5G Small Cell Networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2017.
- [268] L. Pei, J. Huilin, G. Shen, D. Fei, and P. Zhiwen, "Impact of BS sleeping and user association scheme on delay in ultra-dense networks," in *Proc. Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Oct 2016, pp. 1–6.

- [269] Z. Alharbi, A. S. Thyagaturu, M. Reisslein, H. ElBakoury, and R. Zheng, "Performance Comparison of R-PHY and R-MACPHY Modular Cable Access Network Architectures," *IEEE Trans. on Broad.*, vol. PP, no. 99, pp. 1–18, 2017.
- [270] S. S. Rakib, "Virtual converged cable access platforms for HFC cable networks," Jan. 20 2015, US Patent 8,938,769.
- [271] A. S. Thyagaturu, Z. Alharbi, and M. Reisslein, "R-FFT: Function Split at IFFT/FFT in Unified LTE CRAN and Cable Access Network," *CoRR*, vol. abs/1708.08902, 2017. [Online]. Available: <http://arxiv.org/abs/1708.08902>
- [272] V. G. Nguyen, A. Brunstrom, K. J. Grinnemo, and J. Taheri, "SDN/NFV-based Mobile Packet Core Network Architectures: A Survey," *IEEE Commun. Surv. Tutor.*, vol. PP, no. 99, pp. 1–1, 2017.
- [273] T. Taleb, K. Samdanis, B. Mada, H. Flinck, S. Dutta, and D. Sabella, "On Multi-Access Edge Computing: A Survey of the Emerging 5G Network Edge Architecture Orchestration," *IEEE Commun. Surv. Tutor.*, vol. PP, no. 99, pp. 1–1, 2017.
- [274] P. K. Agyapong, M. Iwamura, D. Staehle, W. Kiess, and A. Benjebbour, "Design considerations for a 5G network architecture," *IEEE Commun. Mag.*, vol. 52, no. 11, pp. 65–75, 2014.
- [275] A. T. Z. Kasgari and W. Saad, "Stochastic optimization and control framework for 5G network slicing with effective isolation," *arXiv preprint arXiv:1801.10282*, 2018.
- [276] T. Taleb, M. Corici, C. Parada, A. Jamakovic, S. Ruffino, G. Karagiannis, and T. Magedanz, "EASE: EPC as a service to ease mobile core network deployment over cloud," *IEEE Network*, vol. 29, no. 2, pp. 78–88, 2015.
- [277] A. Imran, A. Zoha, and A. Abu-Dayya, "Challenges in 5G: how to empower SON with big data for enabling 5G," *IEEE Network*, vol. 28, no. 6, pp. 27–33, 2014.
- [278] M. Jaber, M. A. Imran, R. Tafazolli, and A. Tukmanov, "5G Backhaul Challenges and Emerging Research Directions: A Survey," *IEEE Access*, vol. 4, pp. 1743–1766, 2016.
- [279] M. Fiorani, B. Skubic, J. Mårtensson, L. Valcarenghi, P. Castoldi, L. Wosinska, and P. Monti, "On the design of 5G transport networks," *Photonic netw. commun.*, vol. 30, no. 3, pp. 403–415, 2015.

- [280] C. Dehos, J. L. González, A. De Domenico, D. Kténas, and L. Dussopt, “Millimeter-wave access and backhauling: the solution to the exponential data traffic increase in 5G mobile communications systems?” *IEEE Commun. Mag.*, vol. 52, no. 9, pp. 88–95, 2014.
- [281] A. S. Thyagaturu, A. Mercian, M. P. McGarry, M. Reisslein, and W. Kellerer, “Software defined optical networks SDONs: A comprehensive survey,” *IEEE Commun. Surveys Tuts.*, vol. 18, no. 4, pp. 2738–2786.
- [282] A. Ioannou and S. Weber, “A Survey of Caching Policies and Forwarding Mechanisms in Information-Centric Networking,” *IEEE Commun. Surv. Tutor.*, vol. 18, no. 4, pp. 2847–2886, Fourthquarter 2016.
- [283] M. Zhang, H. Luo, and H. Zhang, “A Survey of Caching Mechanisms in Information-Centric Networking,” *IEEE Commun. Surv. Tutor.*, vol. 17, no. 3, pp. 1473–1499, thirdquarter 2015.
- [284] S. Mosleh, L. Liu, H. Hou, and Y. Yi, “Coordinated Data Assignment: A Novel Scheme for Big Data over Cached Cloud-RAN,” in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2016, pp. 1–6.
- [285] M. A. Maddah-Ali and U. Niesen, “Decentralized Coded Caching Attains Order-Optimal Memory-Rate Tradeoff,” *IEEE/ACM Trans. Netw.*, vol. 23, no. 4, pp. 1029–1040, Aug. 2015.
- [286] Y. Sun, Z. Chen, and H. Liu, “Delay Analysis and Optimization in Cache-Enabled Multi-Cell Cooperative Networks,” in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2016, pp. 1–7.
- [287] X. Zhang and Q. Zhu, “P2P Caching Schemes for Jointly Minimizing Memory Cost and Transmission Delay over Information-Centric Networks,” in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2016, pp. 1–6.
- [288] A. Sengupta, R. Tandon, and O. Simeone, “Cache aided wireless networks: Trade-offs between storage and latency,” in *Proc. Annu. Conf. Inf. Sci. and Syst. (CISS)*, Mar. 2016, pp. 320–325.
- [289] R. Tandon and O. Simeone, “Cloud-aided wireless networks with edge caching: Fundamental latency trade-offs in fog Radio Access Networks,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2016, pp. 2029–2033.

- [290] F. Xu, K. Liu, and M. Tao, "Cooperative Tx/Rx caching in interference channels: A storage-latency tradeoff study," in *Proc. IEEE Int. Symp. on Info. Theo. (ISIT)*, July 2016, pp. 2034–2038.
- [291] N. Naderializadeh, M. A. Maddah-Ali, and A. S. Avestimehr, "Fundamental Limits of Cache-Aided Interference Management," *IEEE Trans. Inf. Theory*, vol. 63, no. 5, pp. 3092–3107, May 2017.
- [292] Y. Cao, F. Xu, K. Liu, and M. Tao, "A Storage-Latency Tradeoff Study for Cache-Aided MIMO Interference Networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2016, pp. 1–6.
- [293] S. M. Azimi, O. Simeone, and R. Tandon, "Fundamental Limits on Latency in Small-Cell Caching Systems: An Information-Theoretic Analysis," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2016, pp. 1–6.
- [294] M. A. Maddah-Ali and U. Niesen, "Fundamental Limits of Caching," *CoRR*, vol. abs/1209.5807, 2012. [Online]. Available: <http://arxiv.org/abs/1209.5807>
- [295] J. Zhang and P. Elia, "Fundamental Limits of Cache-Aided Wireless BC: Interplay of Coded-Caching and CSIT Feedback," *IEEE Trans. Inf. Theory*, vol. 63, no. 5, pp. 3142–3160, May 2017.
- [296] M. Ji, M. F. Wong, A. M. Tulino, J. Llorca, G. Caire, M. Effros, and M. Langberg, "On the fundamental limits of caching in combination networks," in *Proc. IEEE Int. Workshop Signal Process. Advances in Wireless Commun. (SPAWC)*, Jun. 2015, pp. 695–699.
- [297] M. M. Amiri and D. Gndz, "Fundamental Limits of Coded Caching: Improved Delivery Rate-Cache Capacity Tradeoff," *IEEE Trans. Commun.*, vol. 65, no. 2, pp. 806–815, Feb. 2017.
- [298] A. Sengupta, R. Tandon, and T. C. Clancy, "Fundamental Limits of Caching With Secure Delivery," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 2, pp. 355–370, Feb. 2015.
- [299] X. Peng, J. C. Shen, J. Zhang, and K. B. Letaief, "Joint data assignment and beamforming for backhaul limited caching networks," in *Proc. IEEE Pers. Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2014, pp. 1370–1374.

- [300] E. Batu, M. Bennis, and M. Debbah, “Cache-enabled small cell networks: Modeling and tradeoffs,” in *Proc. Int. Symp. on Wireless Commun. Syst. (ISWCS)*, Aug. 2014, pp. 649–653.
- [301] G. Carofiglio, L. Mekinda, and L. Muscariello, “LAC: Introducing latency-aware caching in Information-Centric Networks,” in *Proc. IEEE Conf. Local Comp. Netw. (LCN)*, Oct. 2015, pp. 422–425.
- [302] A. M. Girgis, O. Ercetin, M. Nafie, and T. ElBatt, “Decentralized Coded Caching in Wireless Networks: Trade-off between Storage and Latency,” *arXiv preprint arXiv:1701.06673*, 2017.
- [303] A. Sengupta, R. Tandon, and O. Simeone, “Cloud RAN and edge caching: Fundamental performance trade-offs,” in *Proc. IEEE Int. Workshop Signal Process. Advances in Wireless Commun. (SPAWC)*, Jul. 2016, pp. 1–5.
- [304] M. Chen, Y. Hao, M. Qiu, J. Song, D. Wu, and I. Humar, “Mobility-aware caching and computation offloading in 5G ultra-dense cellular networks,” *Sensors*, vol. 16, no. 7, p. 974, 2016.
- [305] S. H. Chae, J. Y. Ryu, T. Q. S. Quek, and W. Choi, “Cooperative Transmission via Caching Helpers,” in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2015, pp. 1–6.
- [306] X. Ge, B. Yang, J. Ye, G. Mao, C. Wang, and T. Han, “Spatial Spectrum and Energy Efficiency of Random Cellular Networks,” *CoRR*, vol. abs/1501.05368, 2015. [Online]. Available: <http://arxiv.org/abs/1501.05368>
- [307] S. Mukherjee and I. Guvenc, “Effects of range expansion and interference coordination on capacity and fairness in heterogeneous networks,” in *Proc. IEEE Asilomar Conf. Signals, Systems and Computers (ASILOMAR)*, Nov. 2011, pp. 1855–1859.
- [308] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, “Five disruptive technology directions for 5G,” *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.
- [309] M. Gregori, J. Gmez-Vilardeb, J. Matamoros, and D. Gndz, “Wireless Content Caching for Small Cell and D2D Networks,” *IEEE J. Sel. Areas Commun.*, vol. 34, no. 5, pp. 1222–1234, May 2016.

- [310] J. Goseling, O. Simeone, and P. Popovski, “Delivery Latency Regions in Fog-RANs with Edge Caching and Cloud Processing,” *arXiv preprint arXiv:1701.06303*, 2017.
- [311] J. Koh, O. Simeone, R. Tandon, and J. Kang, “Cloud-aided edge caching with wireless multicast fronthauling in fog radio access networks,” in *Proc. IEEE Wireless Commun. and Netw. Conf. (WCNC)*, Mar. 2017, pp. 1–6.
- [312] A. Sengupta, R. Tandon, and O. Simeone, “Pipelined Fronthaul-Edge Content Delivery in Fog Radio Access Networks,” in *Proc. IEEE Globecom Workshop (GC Wkshps)*, Dec. 2016, pp. 1–6.
- [313] S. M. Azimi, O. Simeone, A. Sengupta, and R. Tandon, “Online Edge Caching in Fog-Aided Wireless Network,” *arXiv preprint arXiv:1701.06188*, 2017.
- [314] C. Fang, F. R. Yu, T. Huang, J. Liu, and Y. Liu, “A survey of energy-efficient caching in information-centric networking,” *IEEE Commun. Mag.*, vol. 52, no. 11, pp. 122–129, 2014.
- [315] T. Wirth, M. Mehlhose, J. Pilz, R. Lindstedt, D. Wieruch, B. Holfeld, and T. Haustein, “An Advanced Hardware Platform to Verify 5G Wireless Communication Concepts,” in *Proc. IEEE Veh. Technol. Conf. (VTC Spring)*, May 2015, pp. 1–5.
- [316] X. Meng, J. Li, D. Zhou, and D. Yang, “5G technology requirements and related test environments for evaluation,” *China Commun.*, vol. 13, no. Supplement2, pp. 42–51, 2016.
- [317] S. Parkvall, J. Furuskog, Y. Kishiyama, A. Harada, T. Nakamura, P. Naucler, and B. Halvarsson, “5G Wireless Access - Trial Concept and Results,” in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2015, pp. 1–6.
- [318] H. Cao, S. Gangakhedkar, A. R. Ali, M. Gharba, and J. Eichinger, “A 5G V2X testbed for cooperative automated driving,” in *Proc. IEEE Veh. Netw. Conf. (VNC)*, Dec. 2016, pp. 1–4.
- [319] X. Liu, H. Zeng, N. Chand, and F. Effenberger, “Efficient Mobile Fronthaul via DSP-Based Channel Aggregation,” *J. Light. Technol.*, vol. 34, no. 6, pp. 1556–1564, Mar. 2016.

- [320] J. Zhang, W. Xie, F. Yang, and Q. Bi, “Mobile edge computing and field trial results for 5G low latency scenario,” *China Commun.*, vol. 13, no. Supplement2, pp. 174–182, 2016.
- [321] P. Kela, M. Costa, J. Salmi, K. Leppanen, J. Turkka, T. Hiltunen, and M. Hronec, “A novel radio frame structure for 5G dense outdoor radio access networks,” in *Proc. IEEE Veh. Technol. Conf. (VTC Spring)*, May 2015, pp. 1–6.
- [322] T. Rappaport, F. Gutierrez, E. Ben-Dor, J. Murdock, Y. Qiao, and J. Tamir, “Broadband millimeter wave propagation measurements and models using adaptive beam antennas for outdoor urban cellular communications,” *to appear in IEEE Trans. Antennas and Propagation*, 2012.
- [323] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, “Millimeter wave mobile communications for 5g cellular: It will work!” *IEEE Access*, vol. 1, pp. 335–349, 2013.
- [324] B. Maaz, K. Khawam, S. Tohme, S. Lahoud, and J. Nasreddine, “Joint User Association, Power Control and Scheduling in Multi-Cell 5G Networks,” in *Proc. IEEE Wireless Commun. Netw. Conf.(WCNC)*, March 2017, pp. 1–6.
- [325] F. Schaich, T. Wild, and Y. Chen, “Waveform Contenders for 5G - Suitability for Short Packet and Low Latency Transmissions,” in *Proc. IEEE 79th Veh. Technol. Conf. (VTC-Spring)*, May 2014, pp. 1–5.
- [326] W. Roh, J. Y. Seol, J. Park, B. Lee, J. Lee, Y. Kim, J. Cho, K. Cheun, and F. Aryanfar, “Millimeter-wave beamforming as an enabling technology for 5G cellular communications: theoretical feasibility and prototype results,” *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 106–113, February 2014.
- [327] R. Casellas, R. Munoz, R. Vilalta, and R. Martinez, “Orchestration of IT/cloud and networks: From Inter-DC interconnection to SDN/NFV 5G services,” in *Proc. Intern. Conf. on Optical Netw. Design and Model. (ONDM)*, May 2016, pp. 1–6.
- [328] J. Kakar, S. Gherekhloo, and A. Sezgin, “Fundamental limits on latency in transceiver cache-aided HetNets,” in *Proc. IEEE Int. Symp. on Inf. Theory (ISIT)*, June 2017, pp. 2955–2959.
- [329] S. Zhang, P. He, K. Suto, P. Yang, L. Zhao, and X. Shen, “Cooperative Edge Caching in User-Centric Clustered Mobile Networks,” *CoRR*, vol. abs/1710.08582, 2017. [Online]. Available: <http://arxiv.org/abs/1710.08582>

- [330] X. Wang, M. Chen, T. Taleb, A. Ksentini, and V. C. M. Leung, “Cache in the air: exploiting content caching and delivery techniques for 5G systems,” *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 131–139, February 2014.
- [331] C. Wang, Y. He, F. R. Yu, Q. Chen, and L. Tang, “Integration of Networking, Caching and Computing in Wireless Systems: A Survey, Some Research Issues and Challenges,” *IEEE Commun. Surveys Tuts.*, vol. PP, no. 99, pp. 1–1, 2017.
- [332] M. Chen, Y. Hao, L. Hu, K. Huang, and V. K. N. Lau, “Green and Mobility-Aware Caching in 5G Networks,” *IEEE Trans. on Wireless Commun.*, vol. 16, no. 12, pp. 8347–8361, Dec 2017.
- [333] I. Stojmenovic and X. Lin, “Loop-free hybrid single-path/flooding routing algorithms with guaranteed delivery for wireless networks,” *IEEE Trans. on Para. and Distrib. Systems*, vol. 12, no. 10, pp. 1023–1032, 2001.
- [334] E. Kranakis, H. Singh, and J. Urrutia, “Compass routing on geometric networks,” in *Proc. Canadian Conf. on Computational Geometry*, 1999, pp. 51–54.
- [335] I. Parvez, M. Jamei, A. Sundararajan, and A. I. Sarwat, “RSS based loop-free compass routing protocol for data communication in advanced metering infrastructure (AMI) of Smart Grid,” in *Proc. IEEE Symp. on Comp. Intell. Appl. in Smart Grid (CIASG)*. IEEE, 2014, pp. 1–6.
- [336] M. H. Zayani and V. Gauthier, “Usage of IEEE 802.15.4 MAC PHY Model,” in *Laboratory CNRS*.
- [337] A. H. Jalal, Y. Umasankar, M. Chowdhury, and S. Bhansali, “A fuel cell based sensing platform for selective detection of acetone in hyperglycemic patients,” *ECS Transactions*, vol. 80, no. 10, pp. 1369–1378, 2017.
- [338] L. Wei, A. H. Moghadasi, A. Sundararajan, and A. I. Sarwat, “Defending mechanisms for protecting power systems against intelligent attacks,” 2015.
- [339] L. Wei, A. I. Sarwat, W. Saad, and S. Biswas, “Stochastic games for power grid protection against coordinated cyber-physical attacks,” *IEEE Transactions on Smart Grid*, vol. 9, no. 2, pp. 684–694, 2018.
- [340] L. Wei, A. I. Sarwat, and W. Saad, “Risk assessment of coordinated cyber-physical attacks against power grids: A stochastic game approach,” in *Industry Applications Society Annual Meeting, 2016 IEEE*. IEEE, 2016, pp. 1–7.

- [341] Y. Yan, Y. Qian, and H. Sharif, “A secure and reliable in-network collaborative communication scheme for advanced metering infrastructure in smart grid,” in *Proc. IEEE Wire. Commun. and Netw. Conf. (WCNC)*, March 2011, pp. 909–914.
- [342] L. S. Beevi, G. Merlin, and G. MoganaPriya, “Security and privacy for smart grid using scalable key management,” in *Proc. Intern. Conf. on Elect., Electr., and Optim. Techn. (ICEEOT)*, March 2016, pp. 4716–4721.
- [343] I. Parvez, A. Islam, and F. Kaleem, “A key management-based two-level encryption method for AMI,” in *Proc. IEEE Conf. on PES General Meeting — Conf. Expo.*, July 2014, pp. 1–5.
- [344] F. T. Liu, K. M. Ting, and Z.-H. Zhou, “Isolation-based anomaly detection,” *ACM Trans. Knowl. Discov. Data*, vol. 6, no. 1, pp. 3:1–3:39, Mar. 2012. [Online]. Available: <http://doi.acm.org/10.1145/2133360.2133363>
- [345] ———, “Isolation forest,” in *Proc. IEEE Inter. Conf. on Data Mining*. IEEE, 2008, pp. 413–422.
- [346] S. Wang, R. Inkol, and B. R. Jackson, “Relationship between the maximum likelihood emitter location estimators based on received signal strength (RSS) and received signal strength difference (RSSD),” in *2012 26th Biennial Symposium on Communications (QBSC)*, May 2012, pp. 64–69.
- [347] I. Parvez, M. Jamei, A. Sundararajan, and A. Sarwat, “RSS based loop-free compass routing protocol for data communication in advanced metering infrastructure (AMI) of Smart Grid,” in *Proc. on IEEE Sym. on Comput. Intell. Applica. in Smart Grid (CIASG)*, Dec 2014, pp. 1–6.
- [348] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, “Estimating the support of a high-dimensional distribution,” *Neural computation*, vol. 13, no. 7, pp. 1443–1471, 2001.
- [349] V. N. Vapnik and V. Vapnik, *Statistical learning theory*. Wiley Press, New York, 1998, vol. 1.
- [350] B. Schölkopf, R. C. Williamson, A. J. Smola, J. Shawe-Taylor, and J. C. Platt, “Support Vector Method for Novelty Detection,” in *NIPS*, vol. 12, 1999, pp. 582–588.

- [351] I. Parvez, A. Islam, and F. Kaleem, “A key management-based two-level encryption method for AMI,” in *Proc. IEEE Conf. on PES General Meeting — Conf. Expo.*, July 2014, pp. 1–5.
- [352] J. Markkula and J. Haapola, “LTE and hybrid sensor-LTE network performances in smart grid demand response scenarios,” in *Proc. IEEE Conf. on Smart Grid Communications (SmartGridComm)*, Oct 2013, pp. 187–192.
- [353] J. Kennedy and R. Eberhart, “Particle swarm optimization,” 1995.
- [354] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine learning in Python,” *Journ. of Machine Learn. Resear.*, vol. 12, pp. 2825–2830, 2011.
- [355] K. A. Heller, K. M. Svore, A. D. Keromytis, and S. J. Stolfo, “One class support vector machines for detecting anomalous windows registry accesses,” in *Proc. of the works. on Data Mining for Computer Security*, vol. 9, 2003.
- [356] T. Fawcett, “An introduction to ROC analysis,” *Pattern recog. letters*, vol. 27, no. 8, pp. 861–874, 2006.
- [357] D. Scott, L. and Denning, “A Location Based Encryption Technique and Some of Its Applications,” in *National Technical Meeting of The Institute of Navigation*, vol. 4, Jan. 2003, pp. 734–740.
- [358] M. Firoozjaei and J. Vahidi, “Implementing geo-encryption in GSM cellular network,” in *Proc. Intern. Conf. on Communications (COMM)*, June 2012, pp. 299–302.
- [359] J. Xiong, Q. Zhang, G. Sun, X. Zhu, M. Liu, and Z. Li, “An information fusion fault diagnosis method based on dimensionless indicators with static discounting factor and knn,” *IEEE Sensors Journ.*, vol. 16, no. 7, pp. 2060–2069, April 2016.
- [360] G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, “KNN Model-Based Approach in Classification,” in *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE*, ser. Lecture Notes in Computer Science, R. Meersman, Z. Tari, and D. Schmidt, Eds. Springer Berlin Heidelberg, 2003, vol. 2888, pp. 986–996. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-39964-3_62

- [361] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. on Inform. Theory*, vol. 13, no. 1, pp. 21–27, January 1967.
- [362] L. Sustek, "Understanding Smart Meters to Design Intelligent, Secure Systems," *EE Times*, 2013.
- [363] *Advanced Metering Infrastructure Attack Methodology*.
- [364] A. H. Jalal, Y. Umasankar, F. Christopher, E. A. Pretto, and S. Bhansali, "A model for safe transport of critical patients in unmanned drones with a watchstyle continuous anesthesia sensor," *Journal of The Electrochemical Society*, vol. 165, no. 8, pp. B3071–B3077, 2018.
- [365] A. H. Jalal, F. Alam, A. Ahmed, and M. A. Ahad, "Precise calibration of optical fiber sensor for ammonia sensing using multivariate analysis," in *Fiber Optic Sensors and Applications XV*, vol. 10654. International Society for Optics and Photonics, 2018, p. 106540W.
- [366] A. H. Jalal, Y. Umasankar, P. J. Gonzalez, A. Alfonso, and S. Bhansali, "Multimodal technique to eliminate humidity interference for specific detection of ethanol," *Biosensors and Bioelectronics*, vol. 87, pp. 522–530, 2017.
- [367] K. Meritt, "Differential Power Analysis attacks on AES," 2013.
- [368] I. Parvez, M. Jamei, A. Sundararajan, and A. Sarwat, "RSS based loop-free compass routing protocol for data communication in advanced metering infrastructure (AMI) of Smart Grid," in *Proc. IEEE Symp. on Comp. Intell. Appl. in Smart Grid (CIASG)*, Dec 2014, pp. 1–6.
- [369] I. Parvez, A. Islam, and F. Kaleem, "A key management-based two-level encryption method for AMI," in *Proc. IEEE PES Gen. Meet. Conf. Expos.*, July 2014, pp. 1–5.

VITA

IMTIAZ PARVEZ

Born, Tangail, Bangladesh.

- 2003- 2008 B.S., Electrical and Electronics Engineering
Bangladesh University of Engineering and Technology
Dhaka, Bangladesh
- 2008- 2011 Specialist, Core Network Planning & Implementation
Robi Axiata Ltd. (A subsidy of NTT DOCOMO and Axiata Group)
Dhaka, Bangladesh
- 2011- 2013 M.S., Electrical Engineering
University of North Texas
Denton, Texas, USA
- 2013 Associate RF Engineer
3S Network Inc.
Raleigh, North Carolina, USA
- 2013-2016 PhD Student, Electrical and Computer Engineering
Florida International University
Miami, Florida, USA
- Presidential Fellowship
Florida International University
Miami, Florida, USA
- 2017-2018 Doctoral Candidate, Electrical and Computer Engineering
Florida International University
Miami, Florida, USA
- 2018 Dissertation Year Fellowship
Florida International University
Miami, Florida, USA

SELECTED PUBLICATIONS AND PRESENTATIONS

[1] Imtiaz Parvez, Ali Rahmati, Ismail Guvenc, Arif I Sarwat, Huaiyu Dai, "A Survey on Low Latency Towards 5G: RAN, Core Network and Caching Solutions," IEEE Communications Surveys and Tutorials, 2018.

- [2] Imtiaz Parvez, M. G. S. Sriyananda, Ismail Guven, Mehdi Bennis, and Arif Sarwat, "CBRS Spectrum Sharing between LTE-U and WiFi: A Multiarmed Bandit Approach," *Mobile Information Systems*, vol. 2016, Article ID 5909801, 12 pages, 2016.
- [3] Parvez I, Sarwat AI, Wei L, Sundararajan A. "Securing Metering Infrastructure of Smart Grid: A Machine Learning and Localization Based Key Management Approach," *Energies*. 2016 Aug 29;9(9):691.
- [4] Imtiaz Parvez, Maryamossadat Aghili, Arif Sarwat, "Online Power Quality Disturbance Detection By OCSVM in Smart Meter," *Journal of Modern Power Systems and Clean Energy*, Springer, 2018. (Accepted)
- [5] I. Parvez, A. Sundararajan and A.I. Sarwat, "Frequency band for HAN and NAN communication in Smart Grid, IEEE Symposium Series on Computational Intelligence (SSCI), Orlando, FL, Dec 9-12, 2014.
- [6] I. Parvez, M. Jamie, A. Sundararajan and A.I. Sarwat, "RSS based Loop-free Compass Routing Protocol for Data Communication in Advanced Metering Infrastructure (AMI) of Smart Grid," IEEE Symposium Series on Computational Intelligence (SSCI), Orlando, FL, Dec 9-12, 2014.
- [7] Parvez, I., Abdul, F., Mohammed, H.; A.I. Sarwat., "Reliability assessment of access point of advanced metering infrastructure based on Bellcore standards (Telecordia)," IEEE SoutheastCon 2015, vol., no., pp.1-7, 9-12 April 2015.
- [8] I. Parvez, N. Islam, N. Rupasinghe, A. I. Sarwat and I. Gven, "LAA-based LTE and ZigBee coexistence for unlicensed-band smart grid communications," IEEE Southeast-Con 2016, Norfolk, VA, 2016, pp. 1-6.
- [9] I. Parvez, F. Abdul and A. I. Sarwat, "A Location Based Key Management System for Advanced Metering Infrastructure of Smart Grid," 2016 IEEE Green Technologies Conference (GreenTech), Kansas City, MO, 2016, pp. 62-67.
- [10] I. Parvez, A.I. Sarwat, J. Pinto, Z. Parvez, M.A. Khandaker, "A Gossip Algorithm based Clock Synchronization Scheme for Smart Grid Applications," IEEE 49th North American Power Symposium, MorganTown, WV.
- [11] Imtiaz Parvez, Tanvir Khan, Arif Sarwat, "LAA-LTE and WiFi based Smart Grid Metering Infrastructure in 3.5 GHz Band," IEEE R10HTC, Dhaka, Dec 2017.