AN EFFICIENT CONGESTION CONTROL SCHEME FOR A LIGHTWEIGHT CONSTRAINED APPLICATION PROTOCOL IN THE INTERNET OF THINGS

by

Godfrey Akpakwu

Submitted in partial fulfillment of the requirements for the degree
Philosophiae Doctor (Electronic Engineering)

in the

Department of Electrical, Electronic and Computer Engineering
Faculty of Engineering, Built Environment and Information Technology

UNIVERSITY OF PRETORIA

June 2020

SUMMARY

AN EFFICIENT CONGESTION CONTROL SCHEME FOR A LIGHTWEIGHT CONSTRAINED APPLICATION PROTOCOL IN THE INTERNET OF THINGS

by

Godfrey Akpakwu

Supervisor:	Dr. G.P. Hancke
Co-Supervisor:	Dr. A.M. Abu-Mahfouz
Department:	Electrical, Electronic and Computer Engineering
University:	University of Pretoria
Degree:	Philosophiae Doctor (Electronic Engineering)
Keywords:	constrained application protocol, context-aware congestion control, Internet of Things, long-term evolution, machine-type communications, particle swarm optimisation, round-trip time, retransmission timeout, 5G new radio.

The Internet of Things is a promising technology which tends to revolutionize and connect the global world via heterogeneous smart devices through seamless connectivity. The current demand for machine-type communications has resulted in a variety of communication technologies with diverse service requirements to achieve the modern Internet of Things vision. More recent cellular standards like long-term evolution have been introduced for mobile devices but are not well suited for low-power and low data rate devices such as the Internet of Things devices. To address this, there is a number of emerging Internet of Things standards. Fifth-generation mobile networks, in particular, aim to address the limitations of previous cellular standards and be a potential key enabler for the future Internet of Things. Additionally, the third-generation partnership project has introduced low-power wide-area cellular-based networks such as extended coverage global systems for mobile communications for the Internet of Things, enhanced machine-type communications and narrowband-Internet of Things

as enabling solutions to support the new service requirements for massive to critical Internet of Things use cases. Therefore, in a comprehensive literature review, this study highlights the state-of-the-art application requirements of the Internet of Things, along with their associated emerging and enabling communication technologies, with the main focus on fifth-generation mobile networks that are envisaged to support the exponential traffic growth for enabling the Internet of Things. The study further investigates the challenges and open research directions pertinent to the deployment of a massive-critical Internet of Things in coming up with a context-aware congestion control in a Constrained Application Protocol for resource-constrained devices.

A profound open research challenge from the literature review is the need for a context-aware congestion control (CACC) approach for a lightweight CoAP/UDP-based Internet of Things traffic. The CACC proposes mechanisms that include retransmission timeout (RTO) estimator, retransmission count based smoothed round-trip-time (RTT) observation, lower-bound RTT restriction approach, and aging concept. The proposed RTO estimators utilise the strong, weak, and failed RTT to identify the exact network status and provide adaptive congestion control. The CACC incorporates a variable of the retransmission count in the request-response interaction model to mitigate the negative variation in RTT due to the fluctuations in the Internet of Things environment. Moreover, with a lower-bound RTO restriction approach, the unnecessary spurious retransmissions are avoided, and the aging mechanism limits the validity of the RTO value to improve the efficiency of the proposed scheme. The proposed CACC model is validated against baseline CoAP and CoCoA+ using Contiki OS and the Cooja simulator. The results obtained are impressive under different network scenarios.

Managing congestion control in a resource-constrained lossy network with a high bit error rate is a challenging task that needs to be given due consideration if the ever-growing promises of the Internet of Things are to be actualised. The primary congestion control mechanism defined by the core CoAP specification is not capable of adapting to the bursty traffic conditions. This calls for and motivates the need for further research in congestion control mechanisms. The study proposes a congestion control scheme that utilises a Particle Swarm Optimisation (PSO)-based Adaptive Congestion control Technique (PACT). PACT applies random and optimal parameter-driven simulations to optimise the default CoAP parameters in order to adapt to the traffic conditions. The PSO-based algorithm varies the retransmission and max-age values for different traffic scenarios. The proposed PACT exhibits significant performance in terms of packet loss, delay, and normalised overhead in comparison with baseline CoAP with Observe under different network and traffic scenarios.

Dedication

This research work is dedicated to my late father Rtd. SS Pius Obite Akpakwu (Oyame I of Ai-Ochechodo) for providing an opportunity for a great foundation for my education and my late siblings Engr. Samuel Ajene Akpakwu and Master Paul Ahmedu Akpakwu, my only one and immediate younger brother and friend. I appreciate you all. Not to be left out is late Mama Roseline Ugbodu Gbadamosi, for your motherly care and continuous prayer that I should be done and back home to my family. Today, I am done Mama d'Mama as I fondly call you but you are no more to see Anuga Obo'Onyeche.

Thank you all. Even after death, I still feel each and everyone of you.

I will forever be grateful to you all!!!

I Love you all but God knows better. So, continue to rest on in the bosom of our Lord.

ACKNOWLEDGEMENTS

I thank God Almighty for giving me the grace, wisdom, knowledge and intellectual capacity to pursue and complete my PhD study.

I would like to appreciate my supervisors Dr. Gerhard Hancke and Dr. Adnan Abu-Mahfouz for their tireless efforts and supervision in ensuring that this research work is completed in due course. I remain grateful for your continuous consultations and guidance during the course of this research work. Personally, I would want to thank Dr. Abu-Mahfouz for the progress meetings he took time to set up every two weeks. Though, it was hectic but indeed the foundation and success of this work. A very special appreciation to Prof. Gerhard Hancke senior, for his overall supervision and fatherly care during the course of this study. Thank you very much Prof.

Also, to acknowledge are the members of staff, department of Electrical, Electronic and Computer Engineering, the head of the advanced sensor network research group, and members of staff of the University of Pretoria for their continuous emails, advice, and support. Also to acknowledge is Prof. Sunil Maharaj, the Dean, faculty of Engineering, Built Environment & Information Technology.

I would also like to acknowledge the Council for Scientific and Industrial Research (CSIR) and department of Research and Innovation, the University of Pretoria for their financial supports, workshops, seminars, and training during the course of this study.

I also acknowledge the Federal University of Agriculture, Makurdi, the Dean, College of Engineering, Prof. Jonathan Adakole Enokela, and members of staff of the Electrical and Electronic Engineering department, Federal University of Agriculture, Makurdi, for their sacrifices during my study fellowship. Also to acknowledge is Dr. Nentawe Yilwatda Goshwe.

I thank all my colleagues in the advanced sensor network research group, the University of Pretoria for their support and suggestions during the course of this study.

Also to acknowledge are my friends both in South Africa and Nigeria for their immense sacrifices. To mention but a few are Engr. PO Omalaye, Engr. Jeffrey Smith Eiyike, Tolulope Smith Babawarun, Dr.

DRE Ewim and family, and finaly my brother and friend, Owoicho Ijiga Emmanuel.

I would also want to acknowledge Mr Vincent Afolabi Adunmo for your fatherly care, continuous advice, support, prayers, and my in-laws. Also to acknowledge is Hon. Adole OJ Gabriel, for your brotherly care and the Akpakwus' extended family too numerous to mention.

I would also want to specially thank and appreciate my family especially my mother, Mrs Mary Enuwa Akpakwu, for her continuous prayers and guidance. Mama, you will live to eat the fruit of your labour. My siblings Engr. Akpakwu John Ahmadu , Mrs Adunmo Victoria Amina, Consul Akpakwu Emmanuel Oloja, Hon. Akpakwu Simon Pius, and Mr Akpakwu Christian Owoicho, and Wives of this great family Mrs Akpakwu Florence Onyemowo, Mrs Akpakwu Daniella Salome, Mrs Akpakwu Alice Onyeche, and Mrs Akpakwu Mabel Ene. Also to appreciate is Master Akpakwu David Oche. Thank you all for your love, sacrifices, financial support and prayers.

I would like to personally thank and appreciate my beautiful wife, Mrs Akpakwu Lora Onyeche for your unconditional love, care, support, prayers and above all for being there for the entire family when I was not close by. Thank you so much "Ihotukum", for you are my stronghold. To my princesses, Miss Akpakwu Daniella Enuwa, Miss Akpakwu Elizabeth Ehi-K'owoicho, and Miss Akpakwu Debora Ene-Ihotu. Thank you all for your understanding.

And finally, I give God Almighty the grace for life, favour, and His divine protection.

LIST OF ABBREVIATIONS

AAA	Authentication, Authorisation, and Accounting
ACK	Acknowledgment
AMQP	Advanced Message Queuing Protocol
API	Application Programming Interface
ARPU	Average Revenue Per User
BEB	Binary Exponential Backoff
BER	Bit Error Rate
BPSK	Binary Phase Shift Keying
CACC	Context-Aware Congestion Control
CAPEX	Capital Expenditure
Cat	Category
CC	Congestion Control
CoAP	Constrained Application Protocol
CoCoA+	Advanced Congestion Control for CoAP
CON	Confirmable
CoRE	Constrained RESTful Environment
CSIR	Council for Scientifc and Industrial Research
CSS	Chirp Spread Spectrum
CSMA/CA	Carrier Sense Multiple Access with Collision Avoidance
CRN	Cognitive Radio Network
CDMA	Code Division Multiple Access
DBPSK	Differential Binary Phase Shift Keying
DSSS	Direct Sequence Spread Spectrum
DSS	Data Distribution Service
DL	Downlink
DLL	Data Link Layers
DTLS	Datagram Transport Layer Security
EC-GSM-IoT	Extended Coverage Global System for Mobile Communications for the Internet of Things
eMTC	Enhanced Machine-Type Communications
eMBB	Enhanced Mobile Broadband

eDRX	Extended Discontinuous Reception
eV2X	Enhancement of Vehicle-to-Everything
eGPRS	Enhanced General Packet Radio Service
ETSI	European Telecommunications Standard Institute
FCC	Federal Communications Commission
FDMA	Frequency Division Multiple Access
FHSS	Frequency Hopping Spread Spectrum
FR	Frequency of Failed RTT
GMSK	Gaussian Minimum Shift Keying
GSM	Global System for Mobile Communications
GPS	Global Position System
GPRS	General Packet Radio Service
GFSK	Gaussian Frequency Shift Keying
Gbest	Global best Position
H2H	Human-to-Human
HTTP	Hypertext Transfer Protocol
IEEE	Institute of Electrical and Electronics Engineers
IETF	Internet Engineering Task Force
IoT	Internet of Things
IIoT	Industrial Internet of Things
IP	Internet Protocol
ISM	Industrial, Scientific, and Medical band
IMT	International Mobile Telecommunications
ITU	International Telecommunication Union
ITS	Intelligent Transportation System
KPI	Key Performance Indicator
LPWA	Low Power Wide Area
LPWAN	Low Power Wide Area Network
LR-WPAN	Low-Rate Wireless Personal Area Network
LTE	Long-Term Evolution
LTE-A	Long-Term Evolution-Advanced
LLN	Lossy and Low-Power Networks

LDR	Static Link Delivery Ratio
M2M	Machine-to-Machine
MTC	Machine-Type Communications
MAC	Medium Access Control
MCL	Maximum Coupling Loss
MEMS	Micro Electronics, Microelectromechanical Systems
MIDs	Message Identifiers
MQTT	Message Queuing Telemetry Transport
MQTT-SN	Message Queuing Telemetry Transport for Sensor Networks
mMTC	Massive Machine-Type Communications
NB-IoT	Narrowband-Internet of Things
NFV	Network Function Virtualization
NON	Non-Confirmable
OPEX	Operational Expenditure
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
PA	Power Amplifier
PACT	Particle Swarm Optimisation-based Adaptive Congestion Control Technique
PSM	Power Saving Mode
PSO	Particle Swarm Optimisation
PRB	Physical Resource Block
Pbest	Local best Position
QoE	Quality of Experience
QoS	Quality of Service
RAN	Radio Access Network
RC	Retransmission Count
RDC	Radio Duty Cycling
RFC	Requests for Comments
RFID	Radio Frequency Identification
RST	Reset
RTO	Retransmission Timeout
RTT	Round Trip Time
RTTVAR	Round Trip Time Variation

SC-FDMA	Single-Carrier Frequency Division Multiple Access
SDN	Software-Defined Networking
SDWSN	Software-Defined Wireless Sensor Network
SIG	Special Interest Group
SR	Frequency of Strong RTT
SRTT	Smoothed Round Trip Time
TDMA	Time Division Multiple Access
TBS	Transport Block Size
TCP	Transmission Control Protocol
TLS	Transport Layer Security
UMTS	Universal Mobile Telecommunication System
UNB	Ultra-Narrowband
UDP	User Datagram Protocol
UDGM	Unit Disk Graph Medium
UE	User Equipment
URLLC	Ultra-Reliable and Low Latency Communications
VBF	Variable Backoff Factor
VM	Virtual Machines
V2V	Vehicle-to-Vehicle
V2I	Vehicle-to-Infrastucture
WR	Frequency of Weak RTT
WSN	Wireless Sensor Network
2G	Second Generation
3G	Third Generation
3GPP	Third Generation Partnership Project
4G	Fourth Generation
5G	Fifth Generation Mobile Network

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 INTRODUCTION

1.1 BACKGROUND

The Internet of Things (IoT) is an emerging and promising technology which tends to revolutionize the global world through connected physical objects. The IoT deals with low-power devices which interact with one another through the Internet. The concept of the IoT [3–8] has attracted the attention of the research community with the end goal of ensuring that wearables, sensors, smart appliances, washing machines, tablets, smart-phones, smart transportation systems, etc., and other entities are connected to a common interface with the ability to communicate with one another. The IoT interconnect things enables machine-to-machine (M2M) communication as a means of data communication between heterogeneous devices without human intervention [9]. According to [10], this can be achieved through a seamless communication medium. The IoT is expected to enable a conducive environment that will impact and influence several aspects of everyday-life and business applications and contribute towards growing the world economy through massive and critical IoT, depending on the nature of the applications to be deployed. Massive IoT applications require an enormous number of smart devices to be connected. These could be deployed in shipping environments, smart-homes (buildings) and smart-cities, smart power systems, and agricultural monitoring environments, etc., which require frequent updates to the cloud with low end-to-end cost. Applications in this domain require low-cost user equipment (UE) with low energy consumption, extended coverage areas, and high scalability for effective deployment of massive IoT. On the other hand, critical IoT applications, including remote healthcare systems (for remote clinical monitoring and assisted living), traffic control and industrial control (Drone/Robot/Vehicle) and tactile Internet etc., require higher availability, higher reliability, safety and lower latency to guarantee the end user experience because failure in such applications would have severe consequences. In general, the various application opportunities enabled by the IoT are countless and their full potential will only be realised by ensuring that more smart devices are

connected through the Internet.

The IoT vision can only be realized through the integration of various enabling telecommunication technologies to provide connectivity solutions for MTC. The majority of the IoT devices were not built or designed to interface with high-bandwidth networks, since these devices were mainly designed with low-power operation in mind. The Long-Term Evolution (LTE) standard, for instance, was conceived mainly for mobile broadband. In this context, the Institute of Electrical and Electronics Engineer (IEEE) working group 802.11ah enhanced communication development to support M2M applications. Among these were Bluetooth Low Energy 4.0, ZigBee, and Wi-Fi/IEEE802.11 to support short-range communication for MTC. LPWA technologies, including Ingenu Random Phase Multiple Access (RPMA), SigFox, and LoRa etc., are promising technologies operating in the unlicensed Industrial, Scientific, and Medical (ISM) spectrum band to provide low-power and long-range communications as proprietary solutions. At the same time, in order to ensure that M2M applications are efficiently supported in 2G, 3G, and LTE Cat-1 and higher networks, the Third-Generation Partnership Project (3GPP) proposed enhancements in its future release for MTC, including Enhanced Machine-Type Communications (eMTC), Extended-Coverage Global System for Mobile Communications for the Internet of Things (EC-GSM-IoT), and Narrowband-Internet of Things (NB-IoT) as cellular-based LPWA technologies for the IoT. It is worth mentioning that enabling modern IoT connectivity in the licensed approved spectrum bands will be a key enabler for massive to critical IoT use cases since it offers diverse applications with different service opportunities within a single network. The challenge however, lies in the way the fifth-generation (5G) mobile network will meet the diverse requirements of the IoT.

Next-generation 5G mobile networks are envisaged to ensure that massive devices and new services such as enhanced mobile broadband (eMBB), massive machine-type communications (mMTC), critical communications and network operations are efficiently supported. It is hoped that basic requirements such as high throughput, low latency in terms of data delivery, high scalability to enable a massive number of devices, efficient energy consumption techniques and the provision of ubiquitous connectivity solutions for end-users will be efficiently supported by using the 5G mobile network for the IoT. Consequently, considering the security mechanism of existing cellular networks which are based on protecting basic connectivity and privacy of end-users, the 5G cellular system is expected to ensure that enhanced security mechanisms are established on the entire network to address issues of authentication, authorisation, and accounting (AAA) for heterogeneous interconnected IoT devices.

The main objective of this study is to provide a complete scope of MTC use case development and requirements, exploring the available connectivity landscape options and promising a network enabler to meet the 5G new service requirements and devising a context-aware congestion control (CACC) mechanism for lightweight CoAP/UDP-based IoT networks for efficient resource utilisation.

1.2 PROBLEM STATEMENT

Whereas significant and tremendous advancements have been achieved in the area of congestion control for the Internet of Things, there are still several open-research problems that require adequate investigation of congestion control in resource-constrained networks. The primary constraint in IoT networks is resource restriction of the nodes. These nodes have limited storage capacity and processing speed, which result in a high round trip time (RTT) and network congestion. Without considering the RTT value, the proper selection of retransmission timeout (RTO) is impossible, and this leads to spurious retransmission and loss of packets due to congestion and unnecessary packet delay.

The conventional transmission control protocol (TCP) and congestion control (CC) schemes are ill-suited for sending small size packets in IoT environments and fail to consider the context information such as the packet loss due to congestion or bit error rate (BER). The IoT is sensitive to the BER. Although, the random selection of low RTO value tends to fast recovery from congestion, it may increase the BER and mislead the indication of packet losses to congestion. Moreover, this encourages the sender node to incur spurious retransmission. For a large RTO value, the congestion control overestimates the RTO, resulting in unnecessary packet delay. The problem statement would be to identify problems and proffer possible solutions to problems associated with congestion control in resource-constrained IoT networks to meet the diverse needs of IoT applications, for optimal and efficient resource utilisation.

1.2.1 Context of the problem

Irrespective of the Internet of Things applications, quality of service (QoS) delivery is imperative for achieving massive machine-type communications (mMTC). The conventional TCP congestion control mechanism is not an efficient scheme for energy resource-constrained IoT networks that generate small size packets because of their logical connection, acknowledgement, and retransmission of lost

packets. To avoid the congestion in IoT, the design of a Constrained Application Protocol (CoAP) has taken into account the IoT application requirements and resource-constrained devices. Conventionally, the CoAP extends a simple congestion control mechanisms, based on RTO and a binary exponential backoff (BEB) algorithm. The drawback in using the conventional congestion control schemes in IoT traffic is the fact that the binary exponential scheme increases the RTO value exponentially when the acknowledgement is not received. Furthermore, this scheme does not consider the context information, such as the packet loss that has resulted from the congestion or BER since IoT is quite sensitive to the bit error rate. Although, the random selection of low RTO value tends to fast recovery from congestion, it may as well increase the BER and mislead the indication of packet losses to congestion. Moreover, this encourages the sender node to incur spurious retransmissions. On the contrary, for a large RTO value, the congestion control overestimates the RTO, resulting in unnecessary packet delay. Hence, the presence of an efficient context-aware congestion control mechanism for massive IoT traffic is a promising research area. Therefore, the need to research and develop such a system that would be able to identify the cause of packet loss and mitigate the congestion control problem in the IoT environment motivates the need for such system to be developed for efficient resource utilisation. For the future evolution of the IoT, it is therefore very important to ensure that a context-aware congestion control mechanism is developed on a resource-constrained Internet of Things network as a multi-objective function that would support the IoT‘s traffic pattern for the massive number of connected smart things for the IoT applications.

1.2.2 Research gap

While there seem to be considerable amounts of work in the literature on congestion control already, in the course of this research, a comprehensive review has indicated that of the many researchers who have studied the TCP congestion control techniques that have been proposed and studied, only a few have undertaken to extend the conventional congestion techniques to IoT traffic patterns. However, the extension of default congestion control schemes of TCP in CoAP is insensitive to the IoT network conditions. The default CoAP congestion control scheme under-performs for IoT traffic patterns, without adapting its behaviour to the network status information. Consequently, a significant research gap has been identified in this context. Therefore, the extension of TCP congestion control in IoT has to be considered as a multi-objective function to support the IoT traffic pattern for efficient resource-utilisation. The study addresses deficiencies by introducing three RTO estimators, a context-

aware observe RTO estimator algorithm, variable retransmision count, a lower-bound RTO restriction approach, and an aging mechanism to achieve optimum performance of congestion control in resource-constrained IoT networks.

1.3 RESEARCH OBJECTIVE AND QUESTIONS

The following are the objectives of the research study:

- Identifying and investigating the critical constraints associated with congestion control in resource-constrained networks. The constraints identified are indeed factors that limit the actualisation of optimality in congestion control solutions for massive machine-type communications.
- Developing system models for congestion control in resource-constrained IoT environments that incorporate the identified factors, and studying possible solution models to address them. The ability to develop system models that extend the piggyback response feature of base CoAP congestion control specifications to address the problem without degrading the system's performance and reliability, and considerably improve the performance of base CoAP over UDP.
- To demonstrate through appropriate evaluation that the proposed models are feasible and can be used to improve on the timeout parameters of base CoAP congestion control specifications for efficient resource utilisation.

The research questions that this study aims to answer include:

- How can we leverage CoAP over UDP in developing an efficient and comprehensive context-aware congestion control mechanism in resource-constrained environments for a massive Internet of Things?
- How can a particle swarm optimisation (PSO)-based adaptive algorithm be developed and evaluated to achieve an optimal number of retransmissions and max-age values for optimal configurations in baseline CoAP with Observe?
- How can the performance metrics of the developed models be evaluated and validated based on the benchmark protocols for efficient resource utilisation and quality of service (QoS) delivery?

1.4 HYPOTHESIS AND APPROACH

The hypothesis and approach for this research study are stated; hence, the promises of the emerging IoT can be realised better if the potential constraints to achieving optimality in congestion control are identified, and viable solutions provided to address such constraints. The null hypothesis would then be that "identifying and addressing the limiting factors in achieving optimal solutions in congestion control (CC) will not necessarily improve the IoT performance and/or thereby make it better equipped to achieve its promises." If the hypothesis is true, as the results presented in subsequent chapters of the thesis suggest, it is the basis for the proposition that "by proposing a context-aware solution that recognises and addresses the various constraints of CC in CoAP, it can be shown that significant and promising improvement can be achieved". This is certainly a positive contribution to the area of congestion control in a constrained Internet of Things network.

The research study has followed the well-established pattern for conducting technical research in the field of engineering, particularly in electronic engineering. The various stages in which this research work was carried out are highlighted as follows:

- **Literature survey:** The first part of this research study has been dedicated to exploring in depth the concept of the IoT, studying congestion control problems based on CoAP in comparison with other related paradigms, identifying open research gaps and opportunities, and considering the round-trip-time mechanism of CoAP and particle swarm optimisation technique and relevant bodies of knowledge in CC. This output at this early stage, has not only elaborated on a comprehensive foundation of the subject matter but also enabled identification of possible research gaps and definition of a focus and direction for the research study carried out. Based on the findings from these investigations at that early stage, a considerable amount of the information has been gathered that makes up chapter two of this thesis.
- **System modelling:** Finding solutions to engineering problems demands the use of network and system models. In this research work, various system models have been developed based on context-aware CC and PSO-based algorithm to achieve optimal CC in constrained IoT networks. The system models incorporate several identified factors that limit the realisation of optimal solutions in CC for CoAP. The solution models developed employ concepts of three-RTO estimators, variable retransmission count, lower-bound RTO restriction approach, aging mechanism and PSO-based

algorithm to provide optimal solutions in base CoAP CC. These models, that capture and addresses each of these constraints, have been thoroughly analysed and the results obtained are presented and discussed in the related chapters of the thesis.

- **Simulation and numerical analysis:** The system models developed have been simulated by using Contiki-OS, a toolset of the Cooja simulator environment. Detailed numerical analysis of the performance of the proposed CACC model has been carried out and results have been obtained. The concept of the PSO-based technique has been further employed in determining the best fitness value based on retransmissions and max-age value to achieve an optimal CC solutions.
- **Verification and validation of results:** The results obtained through simulations have been verified and validated by obtaining and comparing comparative results from related works in the literature using baseline CoAP CC specification and CoCoA+ as benchmark protocols.
- **Thesis write-up:** The research study has been concluded with detailed documentation of the various findings of the research which are presented in this thesis as a contribution to the body of knowledge.

1.5 RESEARCH GOALS

The research goals of this study are as follows:

- To leverage the base CoAP congestion control specification for improved and optimal congestion control in constrained IoT environments.
- To leverage the default CoAP congestion control specification to develop a context-aware congestion control (CACC) scheme for a lightweight CoAP/UDP-based system and demonstrate the performance and efficiency of the proposed system.
- To leverage existing congestion control to develop a PSO-based adaptive congestion control algorithm and evaluate the efficiency of the scheme with base CoAP with Observe.

1.6 RESEARCH CONTRIBUTION

The following contributions have been made to the body of knowledge of congestion control for the IoT in the course of this research work:

- The emerging applications of the IoT promise global connectivity among connected smart devices to actualise the vision of the IoT. In this regard, a comprehensive literature review related to emerging LPWA IoT solutions, including EC-GSM-IoT, eMTC, NB-IoT, and other existing technologies is presented with a primary focus on 5G mobile networks envisaged to support the exponential traffic growth, and on new service requirements, including mMTC, eMBB, critical communications, and network operations towards enabling efficient IoT use cases. The review provides a complete scope of MTC use cases, development and requirements, and explores the available connectivity landscape solutions and promises a network enabler for the new 5G service requirements and presents a "context-aware congestion control (CACC) approach for lightweight CoAP/UDP-based Internet of Things traffic" for efficient resource utilisation. This survey has been published in the IEEE Access Journal.
- In line with a critical research gap uncovered in the literature review, a context-aware congestion control approach for lightweight CoAP/UDP-based Internet of Things traffic has been proposed. This study designs and implements a context-aware CC scheme in Constrained Application Protocol for efficient resource utilisation in resource-constrained Internet of Things networks. The performance results based on the proposed mechanism are evaluated against baseline CoAP CC specification, and CoCoA+ as benchmark protocols to demonstrate and validate the efficiency of the scheme. This system and its associated performance evaluation of the proposed CACC scheme have been submitted and published as a Journal article in the Transactions on Emerging Telecommunications Technologies.
- This study has also made another novel research contribution that utilised a particle swarm optimisation-based adaptive congestion control technique (PACT). For this evaluation, both the random and optimal parameter-driven simulations have been implemented on default CoAP parameters in order to adapt to the traffic conditions. The goal of this contribution is to leverage the number of retransmissions and max-age value to analyse the impact of these parameters, based on the PSO fitness and velocity function against base CoAP with Observe while keeping the desired QoS in terms of packet delivery, delay, and normalised overhead. This work has been submitted for publication and is currently under review.

1.7 RESEARCH OUTPUTS

The contributions from this research work have been published or are currently under review for publication as full-length articles in peer-reviewed journals. As outputs, a list of the publications from the research works as a contribution to the body of knowledge are presented below:

- G. A. Akpakwu, B. J. Silva, G. P. Hancke, and A. M. Abu-Mahfouz, "A survey on 5G networks for the Internet of Things: Communication technologies and challenges," *IEEE Access*, vol. 6, no. 2, pp. 3619-3647, 2017.
- G. A. Akpakwu, G. P. Hancke, and A. M. Abu-Mahfouz, "CACC: Context-aware congestion control approach for lightweight CoAP/UDP-based Internet of Things traffic," *Transactions on Emerging Telecommunication Techologies*, vol.31, no.2, p. e3822, 2020.
- G. A. Akpakwu, G. P. Hancke, and A. M. Abu-Mahfouz, "PACT: An Optimisation-based adaptive congestion control technique for constrained application protocol," *International Journal of Network Management*, **(under review)**.

1.8 DELINEATION AND LIMITATIONS

The field of the emerging Internet of Things is still fairly new and the research study has encountered a lack of IoT-test-beds for congestion control for experimentation at the time of this research. In this regard, all performance evaluations carried out in the course of this research work have been simulation based, using the Cooja simulator, as part of the Contiki operating system toolset for constrained devices and the IoT to verify and evaluate the proposed work against validated benchmark protocols, baseline CoAP congestion control specification, and the Erbium implementation of CoCoA+ obtained from the author. The study has assumed and the validated congestion control scheme compiled for simulation has been optimal, as expected.

1.9 THESIS OVERVIEW

The rest of this thesis is organised as follows: Chapter 2 presents a comprehensive literature overview of wireless sensor networks as the building blocks for the Internet of Things, the emerging low-power wide-area IoT solutions, including EC-GSM-IoT, eMTC, NB-IoT and other existing technologies, with

the main focus on 5G mobile networks as the future telecommunications system, envisaged to support the exponential traffic growth and new services and market technology enablers (SMARTER) for the next-generation networks to enable efficient IoT use cases. Research gaps have been identified in this area and a context-aware congestion control (CACC) mechanism for lightweight CoAP/UDP-based Internet of Things traffic has been found for efficient resource utilisation. In Chapter 3, a novel CACC scheme is proposed which includes all aspects of the system design and benchmark protocols used for validation. The CACC model, including three RTO estimation algorithms, dynamic SRTT and RTO overall estimation, RTT fluctuation-aware RTO, restricted RTO shrinkage, the context-aware RTO, and analysis of the context-aware algorithms are presented and followed by the mode of implementation strategy, including the simulation setup, presentation of results obtained and discussion in Chapter 4. In Chapter 5, the investigation into particle swarm optimisation (PSO)-based algorithm for congestion control in Constrained Application Protocol is proposed; the system and network model, optimisation strategy for tuning base CoAP parameters, the fitness measurement based on PSO-algorithm are presented, followed by the performance evaluation and discussion of results obtained in Chapter 6. Finally, the thesis conclusion is drawn in Chapter 7 and research opportunities presented for future consideration.

CHAPTER 2 LITERATURE STUDY

2.1 CHAPTER OVERVIEW

The emerging Internet of Things (IoT) paradigm promises to provide a conducive networking environment through connected heterogeneous "things." These things (devices) are systems of embedded smart sensors and actuators which are networked to communicate through a wireless communication medium. The massive heterogeneous devices expected to be connected through enterprise networks, optical networks, as well as mobile wireless communication networks, will play a vital role in driving up the signaling load in the mobile network and presents more challenging issues confronting the IoT in terms of data and service management, including data acquisition and aggregation, service provisioning and control, as well as system performance of resource-constrained networks. This chapter provides an overview of the current state-of-the-art the IoT from various perspectives. In general, this study identifies the essential IoT applications and design requirements, their associated IoT connectivity solutions, 5G enablers for the IoT and the research gaps associated with congestion control for resource-constrained nodes in coming up with a context-aware congestion control (CACC) approach for lightweight CoAP/UDP-based Internet of Things traffic. This work has been published in "A survey on 5G networks for the Internet of Things: Communicaton technologies and challenges" [11], a journal article that reveals an absolute concept of 5G mobile network for the IoT.

The rest of this chapter is outlined as follows. In Section 2.2 an overview of wireless sensor networks (WSNs) as buildings block for IoT is provided and followed by a discussion about WSNs transport protocols in Section 2.3. Section 2.4 presents the performance metrics for WSNs. Section 2.5 presents a general overview of IoT application requirements, including emerging IoT applications and their associated design requirements. In Section 2.6 a discussion about the current IoT communication technology is presented, including long-range networks, short-range networks, and this is followed

by the 3GPP cellular solutions for the IoT in Section 2.7. Section 2.8.1 discusses the Constrained Application Protocol (CoAP) and its fundamental features, followed by a discussion about congestion control in Section 2.9. Section 2.10 narrows down the discussion about 5G New Radio enhancements for the IoT with emphasis on new services and markets technology enablers (SMARTER) for the IoT. This is followed by a discussion of network enablers for the IoT, including software-defined wireless sensor networks, network function virtualization, and cognitive radio networks for providing and deploying connectivity solutions for the IoT applications in Section 2.11. Future research challenges associated with network congestion and overhead are identified and in Section 2.12, a context-aware congestion control scheme is proposed as a multi-objective function to manage and handle the network congestion problem in constrained-IoT networks. The chapter summary is provided in Section 2.13.

2.2 OVERVIEW OF WIRELESS SENSOR NETWORKS

According to forecasts from Ericsson [12], it is estimated that about 28 billion smart devices will be connected across the globe by 2021, with more than 15 billion of these devices to be connected via M2M and consumer-electronic devices, and about 5 billion IoT devices to be connected through cellular 3GPP-based solutions [13]. Research has also shown that roughly 7 billion of these devices will be connected by cellular technologies such as 2G, 3G, and 4G which are currently being used for IoT but not fully optimised for the IoT applications and Low-Power Wide-Area (LPWA) technology [14] and with a revenue of about 4.3 trillion dollars [15] to be generated across the entire IoT sector globally. The emergence of the IoT paradigm will continue to be pervasive in the future Internet and for the Industrial Internet of Things (IIoT) [16].

The rapid development of smart sensors recently has given rise to the concept of connecting wireless sensor networks (WSNs) to the Internet. These WSNs comprise micro-sensors and actuators that are capable of sensing both physical and environmental factors, including humidity, temperature, motions, vibrations, etc., process such information and give feedback to the sink. With the emergence of the Internet of Things as a promising paradigm, WSNs will play a vital role in the IoT since most of the smart sensor nodes will constitute the building blocks for the IoT [17]. The advancement in research for Micro Electronics, Microelectromechanical Systems (MEMS), has made it easier for these sensor nodes to be inexpensive, intelligent and networked through wireless and wired communications for the

emerging IoT paradigm. In this section, we briefly discuss some of the unique features of WSNs as building blocks for the IoT, namely; resource constraints, network topology, applications, message size, and traffic characteristics which are unique to the IoT concept.

2.2.1 Resource Constraints

Most of these sensor nodes are faced with several challenges, including limited computational capability, limited memory space, energy and low communication bandwidth which generally affect the efficiency of the projected IoT paradigm. However, it is paramount that specific applications of layer protocols which will enable typical machine-to-machine applications are considered for resource-constrained environments.

2.2.2 Network Topology

These smart sensor nodes do have characteristics of unique network topology which are generally arranged in a multi-hop star-tree form, most explicitly in a flat or hierarchical network form. The sink node collects data and relays information to external networks at the root of the tree. The dynamic nature of this topology is due to the condition of time-varying link and that of the node variation.

2.2.3 Diverse Application

Smart sensors can support diverse applications ranging from shipping environments, smart-homes and smart-cities, intelligent power systems, agricultural monitoring environments, security surveillance, traffic control and industrial control and tactile Internet etc. Depending on the nature of the applications to be deployed, which require higher availability, higher reliability, safety and lower latency to guarantee the end-user experience, as failure in such applications would result in severe consequences [11]. Care must be taken when considering applications to be deployed for both massive to critical IoT based on their unique requirements.

2.2.4 Message Size

As regards the legacy networks, smart sensor networks are characterised by small message size when compared with the existing systems. In this context, the concept of data segmentation is not an option in most WSNs applications.

2.2.5 Traffic Characteristics

In WSNs, the direction of traffic for smart sensor nodes can be both upstream and downstream traffic. The upstream traffic is the primary traffic from the intelligent sensor nodes to the sink node. This type of traffic can be generalised as a many-to-one mode of communication. However, the sink node can also be involved in communication known as downstream traffic for specific query and control processes.

These unique features present new challenges in the design of WSNs to meet the future IoT application requirements of energy conservation, reliability, extended coverage, and quality of service (QoS) provision as depicted in Figure 2.1.

2.3 FUNCTIONS OF WSN TRANSPORT PROTOCOLS

WSNs transport protocol is designed to run on top of the network layer. These protocols enable and guarantee end-to-end reliability, mitigate congestion and minimise the loss of packets and enable adequate allocation of bandwidth in terms of fairness, etc. With the resource-constrained nature of these nodes and wireless network or connection, congestion remains a significant problem in WSNs. Congestion arises when the traffic load on the IoT network reaches the network capacity, and this may lead to packet loss. Therefore, WSNs transport protocols must be able to cope with or handle the problem of congestion and packet loss. However, the traditional transport protocols which are currently in use for the Internet (i.e., user datagram protocol (UDP) and transmission control protocol (TCP) are not easily implemented for WSNs constrained environments [18–20]. For instance, research has shown that UDP does not support the delivery of reliable communication which is necessary for many sensor applications, nor does it provide flow and congestion control which results in loss of packets and wastage of energy.

On the other hand, TCP has its drawbacks with a high overhead which might not be justifiable for most event-driven applications in sensor networks. Therefore, the need for an application layer protocol to implement a congestion control scheme by itself is of great importance. The rest of this section briefly describes some of the performance metrics and the need for WSNs transport protocols.

2.4 WSNS PERFORMANCE METRICS

Sensor nodes constitute the main building blocks in realising the IoT vision through the integration of various enabling telecommunication technologies to provide connectivity solutions for machine-type communications (MTC). Therefore, WSNs transport protocols must ensure the provision of end-to-end reliability and end-to-end QoS in an energy-efficient manner. To evaluate the performance of WSNs transport protocols, specific metrics such as reliability, fairness, packet-loss ratio, latency, and energy efficiency must be carefully considered to meet the demand of the IoT for an efficient QoS provisioning.

2.4.1 Reliability

According to the authors [21] in WSNs, reliability can be categorized into packet reliability and event reliability. In packet reliability, the applications consider loss-sensitivity and therefore require that all packets are successfully transmitted or meet a specification of successful transmission. On the other hand, event reliability involves applications which demand only successful event detection and not the successful transmission of all packets. Therefore, in application deployment, reliability should be taken into consideration to guarantee provision of efficient quality of service.

2.4.2 Energy Efficiency

Considering the resource-constrained nature of the sensor nodes, the transport protocol to be used must be able to maintain high energy efficiency for effective utilisation of the system lifetime. Most of the packet loss in WSNs occurs as a result of bit error rate and congestion, which affect the system's efficiency. Loss-sensitive applications require the lost packets to be retransmitted, which consumes the battery power of these sensor nodes. Therefore, the number of packet retransmissions, the overhead associated with control messages must be considered carefully for an efficient energy system.

2.4.3 Quality of Service

For efficient system utilisation, it is paramount to take into consideration specific metrics such as packet loss ratio, and bandwidth and latency or delay. These metrics for WSNs can be determined based on the application demand. For instance, specific applications demand high-speed bandwidth for continuous transmission of high-speed data streams when compared to event-based applications. In a delay-sensitive applications, such as the tactile Internet, WSNs must deliver such data within a time frame to provide efficient quality of service.

2.4.4 Fairness in Resource Allocation

Sensor nodes are generally scattered over a geographical location. With the many-to-one communication mode of upstream traffic, data transmission becomes more difficult, specifically for sensor nodes which are at a distance away from the sink node. Therefore, transport protocols must ensure an equitable allocation of bandwidth amongst all sensor nodes as this will enable the sink node to receive a fair amount of data from all the connected sensor nodes.

2.5 OVERVIEW OF INTERNET OF THINGS APPLICATION REQUIREMENTS

The current demand for machine-type communications (MTC) applications such as smart buildings and surveillance [22], smart cities [23], smart grid-monitoring systems [24–29], remote maintenance [30–32], and smart water systems [33] etc., has brought about massive connected devices which pose a major research issue in terms of capacity for currently deployed and future communication networks [34]. In developing applications to implement MTC technologies, there are considerations that need to be taken into account to ensure that each aspect is carefully examined, such as application development protocols, suitability of network connection and available middleware frameworks. Furthermore, the IoT devices are resource-constrained and characterised by low capabilities in terms of both computation and energy capacity. Considering the heterogeneous nature of the IoT resource-constrained devices and use cases for MTC, there are several requirements that need to be addressed. One of the basic and most fundamental issues for the IoT applications is the support for low-power operations, because most IoT devices are battery-powered sensor nodes which could either be installed on bridges or in basements (for indoor applications for instance) in inaccessible regions [35], and

replacing or recharging the batteries of such devices is not feasible. Consequently, such IoT devices are expected to be functional and reliable for a specific number of years. Besides the issue of energy efficiency as a requirement, inter-operability is a major issue for MTC applications because these devices are manufactured by different vendors which lack of standardisation and this makes interaction between heterogeneous devices a challenge. The IoT devices may communicate and disseminate data in various formats, use different application protocols and interfaces for implementation, which presents a challenge for heterogeneous devices to achieve MTC applications. Security and privacy are also very important requirements to be considered for the IoT because of the inherent heterogeneity of Internet-connected smart objects and the ability to ensure that sensitive information which is transmitted and physical objects connected through the communication medium are properly monitored and controlled. This section presents some of the emerging applications and design requirements for the IoT.

2.5.1 Emerging IoT Applications

Given recent advances in ubiquitous computing, there is currently a myriad diverse IoT applications for many different environments that are expected to enhance and improve the quality of everyday life for the end-user community. The variety of these applications dictates that there should be no one-fits-all solution, as each of these applications has different characteristics, and they can be broadly categorized into a number of different fields, since they also have different latency and data rate requirements. Some of these applications are discussed below, with focus on the differences in requirements between application domains.

2.5.1.1 Smart Home

The smart home [36] is a home where devices are connected to the Internet and can make decisions autonomously, based on information originating from sensors, thereby contributing to and improving the personal lifestyle of end-users which makes it easier to monitor and control home appliances and systems. Smart homes are expected to communicate regularly with their environments (internal and external) [26]. The internal environment can be considered as all Internet-connected smart devices and home appliances. On the other hand,the external environment refers to entities that are not in control of the smart home such as smart grid entities. An example is an automatic lighting system that senses the presence of a human being and switches on the lights in a specific area of a house accordingly.

This also includes smart appliances which can optimize their energy consumption based on clever scheduling mechanisms and can be remotely switched ON or OFF over the Internet. From an IoT perspective, smart home is one of the main application domains and there are various applications that have been proposed [37]. ZigBee, based on IEEE 802.15.4, is perhaps the most popular standard used in the smart home domain. Proprietary solutions such as Z-Wave [38] are also used but are not as popular.

2.5.1.2 Intelligent Transportation System

Intelligent transportation systems (ITS) are used to ensure that the transportation network is efficiently monitored and controlled [39, 40]. ITS is designed to make use of the following network components, including vehicle subsystems (which are global position system (GPS), radio frequency identification (RFID) reader, onboard unit (OBU), and communication), ITS monitoring unit, station subsystems (such as road-side equipment) and security subsystems to ensure that system reliability, availability, efficiency and safety of the transportation network are guaranteed. Recent research has shown the on-going potential development in autonomous cars (i.e. self-driving cars). Google is a major pioneer in this project. Google in its recent announcements, introduced their prototype vehicles which drove some miles [41] and other autonomous car projects are in development from Audi (Piloted Driving), Ford (Automated Fusion Hybrid), and Mercedes-Benz (Mercedes-Benz Intelligent Drive), etc. Regulatory bodies are currently working on developments that will enable vehicle-to-vehicle communications in new automobiles for future IoT use cases.

2.5.1.3 Smart City

There is an extensive number of applications (ubiquitous services) that are envisaged to improve and enhance the quality of life and the lifestyle of city residents by gathering information that is relevant to their needs [42, 43]. This will enable smart technologies to be interconnected in order to ensure that basic services required by residents are provided, including (transportation, health, homes and buildings etc.,) which fall under this category, with the most popular being environmental monitoring, smart grid, traffic congestion (which includes vehicle-to-vehicle communication) and waste management systems, amongst other applications [44]. Similar to smart homes, the communication devices in these applications are meant for low-power operation but can also be spread out over very large

areas and require much longer communication ranges than devices in smart homes. Typically, the communications requirements can be considered similar to the smart home case. Meter reading for electricity or water usage [45], for instance, requires much less frequent updates than other applications. LoRa is prominent in smart city applications due to the long ranges it can readily provide.

2.5.1.4 Smart Healthcare

The IoT is expected to impact and influence the medical and healthcare systems strongly. Recent developments in the wearables arena have opened up opportunities for connected healthcare, where advanced sensor devices are attached to patients to collect medical data and vital signs (including blood pressure, body temperature, cholesterol level, heart rate etc.,) from a patient to enable diagnosis of conditions [46], track progress and indicate anomalies direct to the healthcare provider, without significant human involvement. This simplifies the process of collecting patient data and providing a vast quantity of data that can be used to advance scientific studies on cures for diseases, diagnosis, etc. [47], where low-power wearables equipped with sensors serve as data-sourcing platforms for doctors and service providers. For instance, Masimo Radical-7 is a special device that can be used remotely to monitor the patient‘s current health status and report anomalies direct to the clinical staff [48]. In recent research, IBM introduced RFID technology at the Ohio Health‘s hospital to be used for tracking hand washing after patients have been diagnosed [49, 50]. This will definitely reduce the high rate of infections that causes a high rate of death of patients.

2.5.1.5 Industrial Application

Unlike the smart city and smart home counterparts, data reliability in the Industrial Internet of Things (IIoT) (especially in-process monitoring and control) has to be high [51]. For wireless communication in industrial environments, data is usually deterministic as it has strict time constraints, and is characterised by low latency and jitter for applications like motion control for instance [52]. For monitoring and supervision such as vibration or temperature sensing, delays in the second scale are acceptable, but for closed-loop control, latency in the millisecond scale (10-500) ms is required [52]. Therefore, the medium access control (MAC) layer in industrial wireless networks usually makes use of Time Division Multiple Access (TDMA) so that medium access by sensor nodes is deterministic. Two of the major issues that have inhibited the vision of IIoT from being realised include the inability of low-power

wireless networks to fulfil the requirements of high reliability and low energy consumption, and the the fact that IP protocol stack for end-to-end communications has not adapted to the requirements of constrained leaf devices [53]. Moreover, unlike the other application domains previously mentioned, the industrial domain is notorious for proprietary solutions that limit inter-operability, whereas the other application domains are usually more open and use standardized protocols more extensively. Nevertheless, ISA100.11a and Wireless-HART are standards based on IEEE 802.15.4 which have been specifically designed for industrial applications and can be connected to an Internet Protocol (IP) network.

It is evident from the descriptions above that the requirements for these applications are different, therefore a standard that supports them has to cater for this diversity. Table 2.1 highlights and summarizes the differences in requirements for these applications. Although, as highlighted above, IoT comprises a diverse set of applications with varying requirements, on the one hand, there are delay-tolerant applications but there are also applications such as closed-loop control which requires low latency and high reliability. With latencies in the region of 1 to 10 ms, this results in an ecosystem with heterogeneous devices and technologies. Also, irrespective of what the data rate or latency is, different applications require different reporting intervals. For industrial applications or alarms in homes, the update interval might be much higher (i.e. every couple of seconds) than for application domains such as smart cities, where only daily updates might be needed.

2.5.2 Internet of Things Design Requirements

In order to ensure that the cellular LPWA technologies are able to provide an efficient connectivity solution for the different use cases from both Massive IoT to Critical IoT, this section presents some of the key requirements to take into account when considering massive deployment of these services, including low deployment cost, long battery life, low device cost, extended coverage area, support for massive number of connected devices (scalability) and security and privacy. The key requirements for the various use cases are shown in Figure 2.1.

Table 2.1. Overview of typical characteristics/requirements of IoT application, taken from [11], © 2017 IEEE.

Application scheme	Application domain	Tolerable delay	Update frequency	Data rate
Structural health [54, 55]	Smart city	30 min	10 min	Low
Waste management [54, 55]	Smart city	30 min	1 hour	Low
Video surveillance [56]	Smart city	s	Real Time	High
Air quality monitoring [54, 55]	Smart home	5 min	30 min	Low
Monitoring and supervision [52]	Industrial	s or ms	s	Low
Closed-loop control [52]	Industrial	ms	ms	Low
Interlocking and control [52]	Industrial	ms	ms	Low
Patient‘s health delivery and monitoring [56]	Healthcare	Low (s)	1 Report per hour/day	High
Real-time emergency response and remote diagnostics [56]	Healthcare	Low (s)	Requires Ad hoc emergency response	High
Real-time management and accuracy of information across supply chain [56]	ITS	Low (s)	1 Report per hour/day	High

2.5.2.1 Low Device Cost

IoT connectivity is expected to serve very low average revenue per user (ARPU) which is a reduction in revenue generation when compared to mobile broadband subscriptions. This implies that a reduction in the device complexity will at the same time be a key enabler for massive-volume, mass-market applications, which will, therefore, enable most of the IoT use cases. Considering cellular LPWA solutions for business perspective, it is expected that the total cost of production of devices shown, including that of ownership, should be extremely low to promote the massive deployment of IoT use cases.

The summary shown in Table 2.1 presents a diverse class of use cases with a wide range of requirements in regard to tolerable delay, data throughput, and update frequency.

2.5.2.2 Low Deployment Cost

To achieve Massive IoT applications, the entire network of IoT connectivity, including both the Capital expenditure (CAPEX) and the annual Operational expenditure (OPEX), should be kept at a minimum cost. This can be achieved by using software upgrades on existing cellular networks to deploy an LPWA IoT connectivity solution which will reduce the entire cost of new hardware and site planning, thereby maintaining both CAPEX and OPEX to the best minimum in order to deploy massive IoT use cases.

2.5.2.3 Long Battery Life

Energy efficiency is perhaps the most important aspect of IoT, especially because most IoT devices are battery powered and are expected to be operational for a very long time without human intervention. For instance, let us imagine a scenario where a fire alarm system sends data direct to the fire management department. The time interval for changing batteries for such a smart connected device is a major cost factor to be considered. Previous research has shown that most energy expended in IoT devices is for communication [57]. Energy efficiency has to be considered in the design of both hardware and software. There are several medium access control (MAC) protocols that support duty cycling, allowing the radio to be put to sleep (i.e.in low-power mode) for periods when it is not expecting to receive data, therefore extending battery life. Energy management techniques also play an important role in low-power operation through the use of lightweight protocols and scheduling optimisation, for instance [58] as well as energy harvesting, where IoT devices have the capability of harvesting ambient energy from various sources. Moreover, this would also allow the connectivity of new smart device applications that are not currently deployed and a minimum of 10 years battery life span of operation will also be achieved for daily connectivity of these devices.

2.5.2.4 Extended Coverage

Extended coverage is a major design requirement for massive IoT connectivity when applications are considered such as smart metering with very low coverage installed in basements, and other indoor applications such as elevators. The end goal is to ensure that deeper indoor coverage is provided as an equivalent of a signal penetrating a wall or floor, which would at the same time increase the indoor

coverage to support the massive deployment of the IoT use cases. A promising technique for an IoT connectivity link budget for coverage enhancement is being targeted to increase the existing maximum coupling loss (MCL) between the device or user equipment and the base station to a maximum of 164 dB.

2.5.2.5 Support for Massive Number of Devices

It has been envisaged that by 2025, the number of connected heterogeneous smart devices will reach seven billion for cellular IoT technologies. This shows that IoT connectivity will grow faster than legacy mobile broadband connections. This is a clear indication that some cell stations will have more densely connected devices. Therefore, it is hoped that the LPWA IoT connectivity solutions will be able to handle most of these connected smart devices simultaneously.

2.5.2.6 Security and Privacy

Several aspects of security and privacy are major design requirements to be considered in the IoT applications. The mobile IoT user‘s real identity should be well protected from the public but should be traceable by authorities if the need arises and location privacy is of critical importance as this can reveal the physical location of the IoT device. Additionally, forward and backward security should be supported for effective deployment of the IoT use cases [59].

2.6 EXISTING INTERNET OF THINGS COMMUNICATION TECHNOLOGY

Although there is still no unified solution for the IoT at this point, there have been several different communication technologies that have been proposed and are currently in operation, having been deployed in a number of devices worldwide. Both fixed and short-range communication standards will be utilised for most connections to achieve both massive IoT and critical IoT connectivity through either traditional cellular IoT or Low-Power Wide-Area Networks (LPWAN). LPWA technologies are suitable for the IoT applications because of their unique features which include wide-area coverage, high energy efficiency, channel bandwidth, data rate, and low power consumption. This technology is a representation of the various technologies which are currently being used in connecting both sensors and controllers to the Internet without the intervention of existing traditional Wi-Fi or cellular

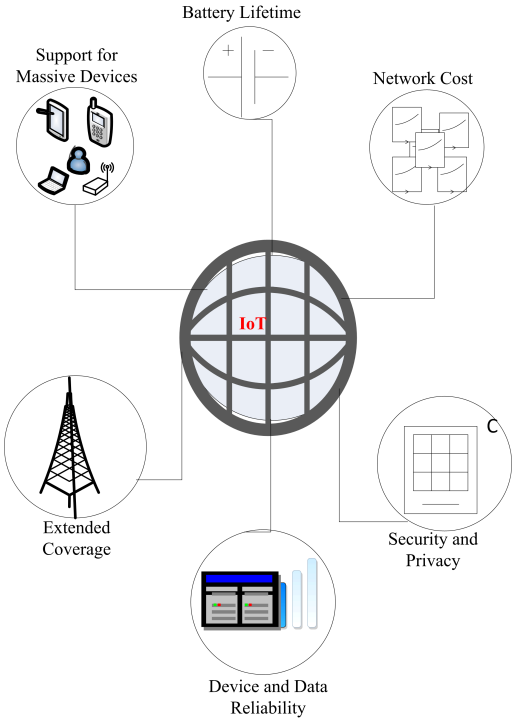

Figure 2.1. IoT design requirements for enabling application scenarios, taken from [11], © 2017 IEEE.

networks. Among these promising technologies are SigFox, Ingenu RPMA, and LoRa. Current and future demand for connectivity to the Internet of Things has motivated cellular technology to introduce their IoT device connectivity landscape solutions such as LTE Cat-M1 (also known as eMTC), EC-GSM-IoT, and NB-IoT (also called LTE Cat-NB1) that will enhance and enable future IoT use cases. LPWA networks are currently being deployed for IoT applications, including smart cities, building management systems, asset monitoring, smart agriculture etc. This section briefly discusses the main features of currently prominent technologies for the IoT and categorizes them into long-range networks, short-range networks and cellular-based technology.

2.6.1 Long-Range Networks

Low-power wide area (LPWA) technologies are among the promising technologies to provide low-power and long-range connectivity solutions for the IoT applications. This section discusses some of the popular LPWAN to support long-range MTC such as SigFox, LoRa, Ingenu RPMA, Weightless, and DASH7 which are relevant to achieve MTC use cases for the IoT.

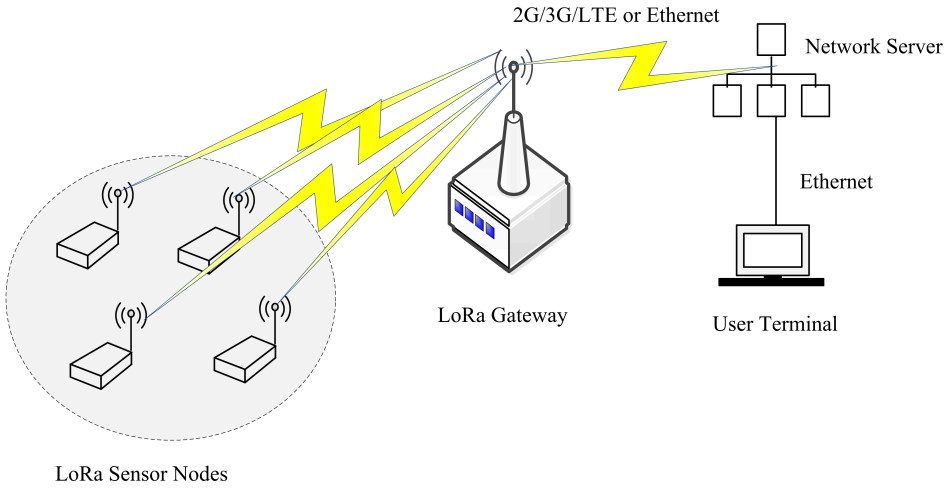

Figure 2.2. Architecture of a LoRa network, taken from [11], © 2017 IEEE.

2.6.1.1 LoRa

LoRa is a physical layer protocol [60] that has emerged as a promising technology for low-cost, low-power and long-range communication. LoRa wireless technology is based on LoRaWAN, a media access control (MAC) layer protocol based on ALOHA [61] for wide coverage area network. LoRa networks are based on a star-to-star network topology where each node (i.e. end device) has a direct single-hop connection to a LoRa gateway. The LoRa architecture consists of end devices (nodes), server, a gateway and a remote terminal as depicted in Figure 2.2. LoRa‘s unique modulation scheme uses a proprietary Chirp Spread Spectrum (CSS) with different bandwidths 7.8 kHz, 10.4 kHz, 15.6 kHz, 31.2 kHz, 41.7 kHz, 62.5 kHz, 125 kHz, 250 kHz, and 500 kHz [62], and provides bi-directional communication. To mitigate the effect of interference, LoRa uses a Frequency-Hopping Spread Spectrum (FHSS) which enables access to available channels. It has been shown that long communication ranges (15 km+) are achievable in urban environments (i.e. with no clear line-of-sight) [63]. LoRaWAN is based on LoRa and adds a network layer to handle network congestion between connected end-devices (i.e. nodes) and central nodes. It uses the 868 and 915 MHz bands for communication at a maximum data rate of 50 kbps, which is sufficient for most IoT applications. LoRa is aimed specifically at IoT applications. Possible data rates with LoRa are dependent on channel bandwidth and spreading factor, where the ALOHA medium access scheme enables multiple devices to communicate by using different spreading factors.

2.6.1.2 SigFox

SigFox [64] low-power wide-area network technology offers a complete end-to-end connectivity solution which is based on their patented technologies by using ultra-narrowband (UNB). Since M2M communications requires a small amount of data to be transferred efficiently on a low bandwidth, SigFox suits such type of communication. This technology is deployed by using proprietary base stations which are configured with cognitive software-defined radios by connecting them to backend servers utilising IP-based network infrastructure. SigFox end devices connect to the network base stations by using a unique modulation scheme called Binary Phase Shift keying (BPSK) [65] in an ultra-narrowband of 100 Hz Sub-GHz Industrial, Scientific and Medical (ISM) band carrier. With UNB, SigFox technology provides higher sensitivity, ultra-low power consumption and long ranges by efficiently utilising its bandwidth at the expense of limited data rates, which is adequate for IoT since most applications do not require high throughputs. SigFox networks use the unlicensed ISM band and as such its frequency of operation varies accordingly between 868 MHz and 915 MHz and enables wide coverage using line-of-sight communication. For instance, in rural areas, a range up to 30-50 km and beyond can be achieved through frequency hopping, but this range is reduced to 3-10 km in urban locations owing to the presence of obstacles.

The SigFox network supports up to 12 bytes of packet size for each message by using typical modulation including Gaussian Frequency-Shift Keying (GFSK) for downlink and Differential Binary Phase Shift Keying (DBPSK) for uplink transmission respectively. Uplink messages are restricted/limited to 140 12-bytes messages per day which conform to the regional regulations which allow no use of license-free spectrum [66], while allowing 4 8-bytes of messages per day for downlink transmission from the base stations to end connected devices. However, ultra-narrowband signals are susceptible to any aggressive bursts exceeding the duration of a bit (i.e. 10 ms), causing devices in a SigFox network to retransmit frames a number of times [67]. This in turn increases the traffic load.

2.6.1.3 INGENU-RPMA

Unlike other LPWA technologies, previously mentioned, that use the 2.4 GHz ISM band for communication, Ingenu-RPMA is a proprietary LPWA technology with more flexible regulations for the use of spectrum across different regions [68], [66]. This means that higher throughput and more capacity

can be achieved when compared to other technologies which are also operating in SUB-GHz band. At the core of the wireless technology, Ingenu uses Random Phase Multiple Access (RPMA) [69] Direct Sequence Spread Spectrum (DSSS), which is used for uplink communication, and allows multiple transmitters to share a single time slot as a variation to Code Division Multiple Access (CDMA). This is achieved by adding a random offset delay to each transmitter within the time slot, consequently reducing overlapping between transmitters [70], and thereby increasing the signal-to-interference ratio for each individual link. Ingenu also provides bi-directional communication. For downlink communication, signals are continuously being spread by base stations to individually connected end devices and such signals are broadcast by using CDMA. Ingenu RPMA is capable of achieving up to -142 dBm receiver sensitivity and a link budget of 168 dB [68]. This technology is made compliant to legacy IEEE 802.15.4k specifications.

2.6.1.4 DASH7

DASH7 is a long-range low-power wireless technology that operates in the 433 MHz ISM band and is an extension of active RFID based on the ISO/IEC 18000 standard [71], where communication can take place direct between devices and they can be used for non-RFID applications. DASH7 employs narrow-band modulation by using two-level GFSK in SUB-GHz bands. This technology is aimed at low-rate applications of a bursty nature, and offers data rates up to 167 kbps. It also supports multi-hopping, albeit limited to 2 hops by default, but it can be extended to more hops. Ranges up to 2 km are possible with DASH7 [72].

2.6.1.5 WEIGHTLESS

Weightless is a new wireless technology which was introduced by the Weightless Special Interest Group (SIP) [73] with three open LPWA standards known as Weightless-W, Weightless-N, and Weightless-P, which operate in both licence-free and licensed spectrum for different ranges and low power consumption. This technology uses cognitive radio and TV white-spaces which enable devices to utilise these bands as opportunistic users without causing interference to the primary user devices as licensed owners. Figure 2.3 depicts the architecture of a Weightless Network.

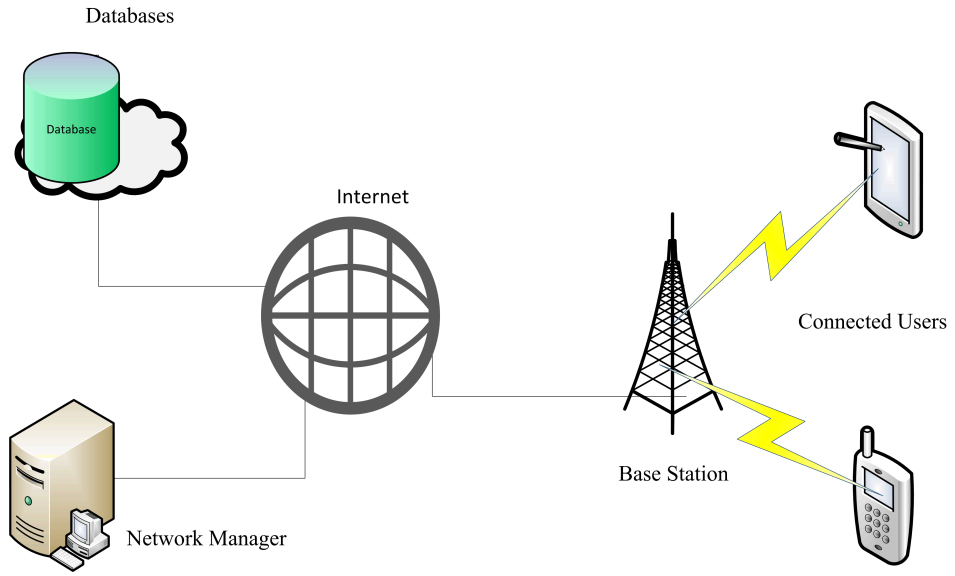

Figure 2.3. Architecture of a Weightless network provider, taken from [11], © 2017 IEEE.

- **WEIGHTLESS-N:** This is an UNB standard which supports only one-way communication (i.e. from end devices to the base station) by using a DBPSK modulation scheme. It exploits TV white-space (SUB-1GHz) in the region of 470 to 790 MHz.
- **WEIGHTLESS-W:** This standard supports different modulation schemes such as Differential-BPSK and 16-Quadrature Amplitude Modulation (16-QAM) with a data rate up to 10 mbps which depends on the link budget. In order to improve energy consumption, end devices are enabled to transmit at a lower power level to the base stations in a narrow band.
- **WEIGHTLESS-P:** This standard, on the other hand, uses GMSK and QPSK modulation and achieves data rates of 100 kbps using narrow channels (12.5 kHz). The main drawback of Weightless-N is the fact that it supports one-way communication, therefore it limits the number of IoT applications for which it can be used. However, the fact that bidirectional communication is not supported extends the battery life for several years more than both Weightless-P and Weightless-W.

2.6.2 Short-Range Networks

This subsection presents some legacy short-range wireless network technologies currently being used to support short-range M2M communication applications, including Bluetooth, ZigBee and low-power Wi-Fi. These technologies are viable and best-fit for consumer use cases of the IoT, but may not be

able to support civic, industrial and other related IoT applications for which the demands are beyond the capacity of their features.

2.6.2.1 Bluetooth

Bluetooth was designed based on the IEEE‘s 802.15.1 wireless personal area communication standard to be used for short-range ad hoc communication (i.e. Master and Slave configuration) between devices operating in the 2.4 GHz ISM bands with achievable data rates in the low mbps. Bluetooth technical specifications and developments are currently being managed by Bluetooth Special Interest Group (SIG) [74]. Bluetooth Low Energy (BLE), which is also called Bluetooth 4.0, was introduced to improve energy consumption. The most recent amendment to the standard uses 40 channels with a width of 2 MHz channel spacing. The modulation scheme used is Gaussian Frequency Shift Key (GFSK) modulation. To make it more robust against interference, and multi-path fading [75], Bluetooth uses a frequency-hopping spread-spectrum (FHSS) scheme where the signal switches carriers over a pre-determined pattern of channels [75].

Although, Bluetooth was originally intended as a replacement for wires in mobile devices, it has evolved to be used in many different applications. However, one of the drawbacks is the restriction of only one-on-one communication between only two devices at a time. The Bluetooth Smart Mesh working group was proposed by the Bluetooth Special Interest Group in order to define and standardize a new architecture for mesh networking for Bluetooth Low Energy which will enhance the communication coverage and enable deployments of Bluetooth Low Energy for IoT. Bluetooth Low Energy is envisaged as a connectivity solution for short-range communication in the IoT applications, including smart energy, healthcare, and smart home applications [76].

2.6.2.2 IEEE 802.15.4 and ZigBee

This standard is currently the de facto standard for low-rate wireless personal area networks (LR-WPAN). Three different frequency bands can be used with IEEE 802.15.4: 868 MHz, 914 MHz and 2.4 GHz supporting 1, 10 and 16 channels, respectively, each with a 2 MHz bandwidth. The maximum supported data rate is 250 kbps [77]. Direct Sequence Spread Spectrum (DSSS) is used as a modulation scheme for IEEE 802.15.4. This standard only defines physical (PHY) and data link layers (DLL).

ZigBee uses the PHY and DLL as defined by IEEE 802.15.4, and builds on it by adding a network layer. A drawback of the original version of IEEE 802.15.4 is the fact that a single static channel is used for communication when the network is established, which is susceptible to interference. It uses Carrier sense multiple access with collision avoidance (CMSA/CA) for channel access. This is particularly important from the IoT perspective, given that a massive number of connected devices might attempt to communicate concurrently. More recent amendments (i.e. 802.15.4e) of the standard have incorporated frequency diversity to counter the impact of interference. The main differences between ZigBee and LoRa are the communication ranges and topology options, as the latter supports star, mesh and cluster tree topologies [78]. Figure 2.4 shows the different topologies - star, peer-to-peer and cluster tree - of a ZigBee network.

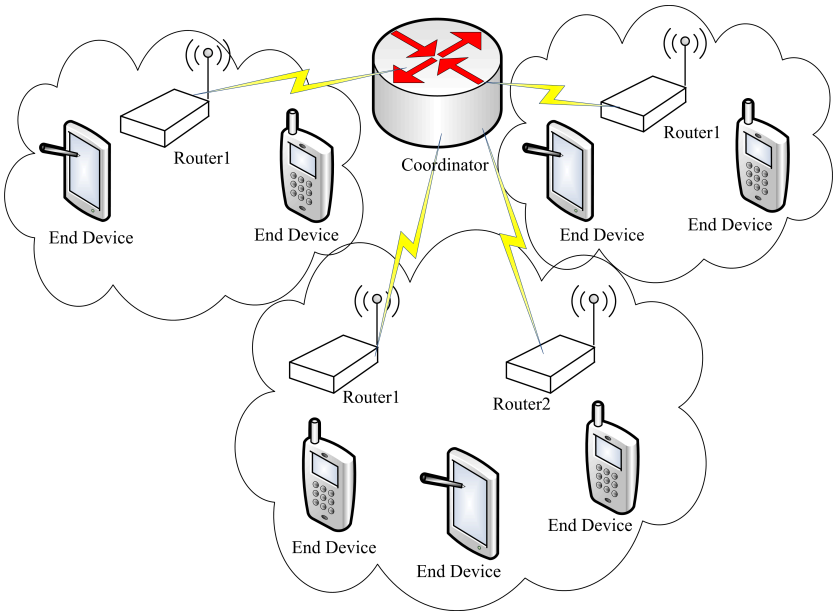

Figure 2.4. ZigBee mesh network, taken from [11], © 2017 IEEE.

2.6.2.3 Wi-Fi

The early version of Wi-Fi was proposed by the IEEE for local area wireless communication which was released without considering the application of modern IoT connectivity. This technology was intentionally designed for high bandwidth communication between devices which are located within a short range with the basic aim of providing high throughput connectivity. This technology is also called wireless local area network (WLAN) and belongs to the IEEE 802.11 standard series [79]. In order to provide Internet connectivity through wireless network access points, the network operates

within the 5 GHz and 2.4 GHz ISM spectrum bands. The access points operate within a coverage area of up to 1 km, which can be increased by using multiple overlapping hotspots. To improve the network performance and throughput while considering dense and congested areas [80, 81] enhanced features have been introduced into the 802.11ac release. In order to extend the Wi-Fi network applications to meet the need and requirements of modern IoT connectivity (basically a large number of smart connected devices, enhanced coverage area, and energy constraints), the IEEE proposed and established low-power Wi-Fi, also called IEEE 802.11ah [82] as an amendment to the legacy standard.

This newly introduced standard aims to achieve low energy consumption down to 100‘s of milliwatts which is suitable for IoT based devices, provides a large coverage area, and can achieve a data rate of up to 347 Mbps. Research has shown that further enhancements are being introduced into low-power Wi-Fi by the Wi-Fi alliance called Wi-Fi HaLow for M2M smart city [83]. Wi-Fi network can provide M2M/IoT applications, including parking metering, autonomous lighting, smart security, smart home thermostats, etc., [84].

Finally, considering the various attributes analysed for low-power wide-area wireless technology for MTC use cases, it is observed that each of these technologies, given the trade-offs, is only able to optimize certain parameters, including battery life-time, data rate, operating frequency band, possible achievable range, scalability and channel bandwidth etc. Based on these attributes, application developers can easily assess which alternative will be a best-fit or viable for deploying IoT use cases for existing alternatives, which are currently available, and wireless network technologies for IoT applications.

2.7 3GPP CELLULAR SOLUTIONS FOR THE INTERNET OF THINGS

Cellular-based technologies such as 3G, 4G and most especially the legacy Third-Generation Partnership Project Long-Term Evolution (3GPP LTE) networks are among the current and promising technologies which are being considered as major landscape connectivity solutions to achieve the modern IoT applications [85]. These promising and appealing technologies are capable of offering wide coverage area, relatively low cost of deployment, high security, dedicated spectrum allocation and efficient management systems. However, having been deployed for optimised broadband networks, they are not suitable for current MTC.

The current IoT landscape comprises different solutions of connectivity which need to be harmonized among the various key industry players in order to ensure that the requirements of the IoT technical key performance indicators (KPIs) are met. 3GPP in its desire to ensure that M2M applications are efficiently supported on 4G broadband networks, including UMTS, and LTE, has been working tremendously to make sure that M2M communications will be efficiently evolved in future and promises 5G New Radio systems which are envisaged for massive IoT applications. 3GPP in its current standardisation of Release-13, recently introduced three main key standards that will enable and enhance the deployment of massive smart connected devices and services such as smart cities, smart grid, wearable devices and connected homes. These introduced features are: EC-GSM-IoT [86], eMTC [87] which are expected to enhance the effective communication of existing cellular-based technologies such as the GSM [88], LTE [89] networks, and the NB-IoT [90]. It is hoped that with these newly introduced LPWA solutions for the IoT, the connectivity profile and basic requirements for the IoT will be achieved when compared to existing cellular networks.

These technologies were introduced in order to provide extensive coverage area, User Equipment (UE) complexity reduction, efficient long battery lifetime, and backward compatibility with existing cellular networks. However, the end goal of the newly emerging standards is to maximize the re-use of legacy cellular network infrastructure which will enhance and support the massive connectivity of IoT applications. In this section, we briefly discuss the emerging and promising technologies which are envisaged as future technology for the Massive deployment of the IoT use cases.

2.7.1 Enhanced Machine-Type Communications

The eMTC, also called LTE Cat-M1, or Cat-M, is a promising cellular LPWA technology introduced in the 3GPP Release-13 standardisation which intends to minimise modem complexity and cost, power consumption, and extend coverage over existing legacy handset modems, such as category 0 user equipment (UEs) from the Release-12 specification for MTC. This technology is an enhancement for LTE networks to support MTC for the IoT.

Cat-M1 UE operates within a limited bandwidth of 1.08 MHz out of the available 1.4 MHz, allowing Cat-M1 UE to use only six physical resource blocks (PRBs) out of the eight available 180 kHz LTE physical resource blocks, which coexist in a broader, general legacy-purpose LTE system (5G Americas,

2016). In order to mitigate the interference level, the two remaining PRBs are used as guard bands. With the support for 1.08 MHz band (narrowband channel) for both radio frequency and baseband, Cat-M1 devices are further reduced in complexity, cost and power over Cat-0 devices. Cat-M1 devices are expected to achieve a maximum throughput of up to 1 Mbps in both uplink and downlink operations for massive IoT. For common control messages, the maximum Transport Block Size (TBS) is further reduced to 1000 bits from the 2216 bits of Cat-0 devices which is an equivalent of unicast data traffic, allowing further processing and memory savings in Cat-M1 devices over the legacy Cat-0 UE.

The eMTC devices have been designed to support either 23 dBm or 20 dBm power classes, unlike the MTC Cat-0 devices which were designed to support a maximum transmission power of 23 dBm, which is approximately 200 mW for uplink (UL). The maximum transmission power of 20 dBm enable the power amplifier (PA) to be integrated as opposed to using a dedicated power amplifier. Consequently, this enables and supports the achievement of lower device cost.

Considering the current LTE numerology, eMTC technology can be deployed within the regular LTE network up to 20 MHz of operation and also to co-exist with other available LTE network services. In view of the reduced bandwidth for Cat-M1 devices, eMTC requires a new set of logical control channels, MTC Physical Downlink Control Channel (MPDCCH) to be used to replace the existing logical control channels such as Physical Downlink Control Channel (PDCCH), Physical Control Format Indicator Channel (PCFICH), and Physical Hybrid Automatic Retransmission Request (ARC) Indicator Channel (PHICH), which are no longer suitable within the new narrower bandwidth for eMTC technology. With the deployment of the eMTC network, series of multiple narrowband regions can be configured. That means that it is possible to configure 6 PRBs each within the LTE carrier for narrowband Physical Downlink Shared Channel (PDSCH) and MPDCCH for data scheduling purposes [56]. It is also important to note that eMTC is designed with an increased link budget of 15 dB with a Maximum Coupling Loss (MCL) of 155.7dB which exceeds the legacy LTE baseline of 140.7dB in order to ensure that coverage is extended for IoT devices which are deployed in remote regions or locations.

eMTC is standardized to ensure that for Massive IoT deployment and coverage, it supports long battery life of about 10 years with a 5 Watt-Hour battery system for effective utilisation. This technology uses power savings management (PSM) and extended discontinuous reception (eDRX) as its power savings mechanisms to achieve long battery life for Cat-M1 devices.

2.7.2 Extended Coverage Global System for Mobile Communications for the Internet of Things

GSM is one of the most dominant and compelling cellular technologies for the deployment of IoT applications because of its extensive and established global and broad ecosystem. 3GPP standardisation in its Release-13 specification introduced EC-GSM-IoT as a standard-based LPWA emerging technology which was designed for high capacity, long-range coverage, low energy and low complexity cellular system based on enhanced General Packet Radio Service (eGPRS) for the IoT [86].

Existing GSM Networks can be upgraded by using a software application in order to ensure that extensive coverage and accelerated time of deployment are determined through optimisation techniques which have been deployed in EC-GSM-IoT and for efficient battery life of about 10 years for a wide range of use cases. EC-GSM-IoT technology is standardized to ensure that its enhancements support extended Discontinuous Reception (eDRX) which improves the power efficiency of devices, minimises idle mode procedures and admission control in terms of QoS. With this technology, GPRS/EGPRS Packet Switched Channels are fully enabled for multiplexing. For effective deployment of Massive IoT applications, new logical channels which were introduced to support extended coverage in EC-GSM-IoT technology are called EC-Channels. These include EC-Shared Channel (EC-SCH), EC-Access Grant Channel (EC-AGCH), EC-Broadcast Control Channel (EC-BCCH), EC-Packet Data Traffic Channel (EC-PDTCH), EC-Paging Channel (EC-PCH), and EC-Packet Associated Control Channel (EC-PACCH). These new logical channels can be incorporated into legacy GPRS spectrum to accommodate EC-GSM devices for IoT services. In order to reach the 20 dB extended coverage which is required when compared to existing legacy GPRS networks, repetitions such as L2 (16 times) and L3 Hybrid Automatic Retransmission reQuest (HARQ) (4 times) are required for effective extended coverage while considering effective utilisation of spectrum, blind repetitions and incremental redundancy (HARQ type II) used for data traffic channels [56]. EC-GSM-IoT is designed with two different modulation schemes which are Eight-Phase Shift Keying (8PSK) and Gaussian Minimum Shift Keying (GMSK) for variable data rates. Figure 2.5 depicts the extended coverage for EC-GSM-IoT technology indicating the various newly introduced logical channels.

Moreover, considering GSM as one of the most widely used wireless standard networks which has been deployed globally, EC-GSM-IoT technology has enhanced legacy GSM networks to ensure that it

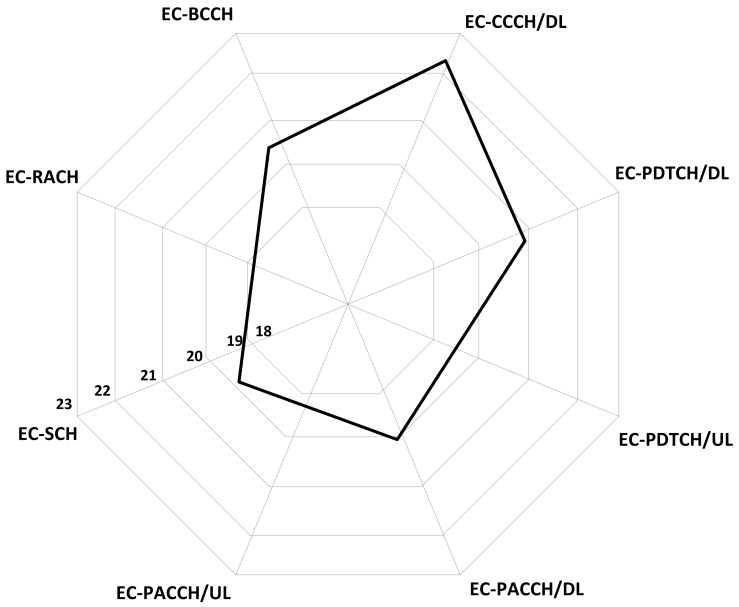

Figure 2.5. Showing EC-GSM-IoT extended coverage, taken from [11], © 2017 IEEE.

supports the global cellular network and deployment of Massive IoT applications into the future which requires low data rates services.

2.7.3 Narrowband-Internet of Things

The NB-IoT, also known as LTE Cat-NB1, is a new and promising cellular low-power wide-area technology introduced in the 3GPP Release-13 specification as an evolution to LTE Cat-M1.

NB-IoT technology is expected to ease the massive deployment of the IoT by allowing an existing operator to introduce NB-IoT within its small portion of legacy network and available spectrum. This technology is designed to ensure that ultra-low end IoT applications, including remote sensors, smart buildings and smart meters, are supported. LTE-Cat-NB1 (NB-IoT) is designed for optimal co-existence performance with legacy GSM, GPRS and with LTE technologies. Cat-NB1 operates within a minimum system bandwidth of 180 kHz for both the downlink and uplink operations, respectively. Because of its choice of operation, it is possible for a GSM operator to replace one GSM carrier of 200 kHz with an NB-IoT application. On the other hand, an operator of the LTE network can as well deploy NB-IoT applications into an LTE carrier by allocating one of its PRBs of 180 kHz to Cat-NB1. The NB-IoT air interface is well optimised to ensure harmonious coexistence with LTE, which means,

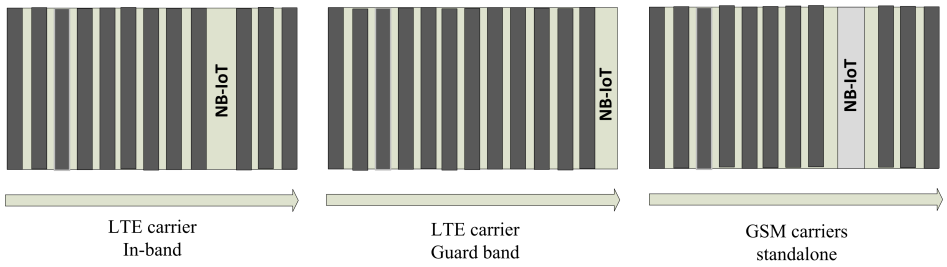

Figure 2.6. Cat-NB1 (NB-IoT) flexible deployment modes, taken from [11], © 2017 IEEE.

when an NB-IoT is deployed inside an LTE carrier, the performance of LTE or Cat-NB1 cannot be compromised.

Therefor, NB-IoT enables flexible deployment of Massive IoT to network providers as:

- In-band, which is integrated as part of the resource regularly used for the eNB communication.
- Guard band, which makes use of the unused frequency band of 180 kHz which is between the last PRB used and the channelization edge.
- Standalone system, which is based on a re-farmed channel (i.e. reusing GSM carrier frequencies) of a legacy GSM/GPRS system which is operated by the service operator. Figure 2.6 shows the flexible deployment options for NB-IoT systems when considering In-band, Guard band and Standalone deployment options.

In order to ensure that the device complexity and cost of NB-IoT technology are limited, the peak data rates for downlink are further reduced considering 32 kbps for in-band scenario, and 34 kbps for standalone deployment, while the uplink peak data rates are limited to 66 kbps and 16.9 kbps for both multi-tone and single-tone transmissions respectively [56].

Cat-NB1 technology is designed to reuse the existing LTE design structure, which includes the numerology, Uplink Single-Carrier Frequency Division Multiple Access (SC-FDMA), downlink Orthogonal Frequency Division Multiple Access (OFDMA), rate matching, Channel Coding, and Interleaving, which reduces the time required to introduce a new system specification for NB-IoT [90]. The first normative phase for introducing NB-IoT in 3GPP started sometime in September 2015 with its core complete specification in June 2016. It has been expected that the commercial launch of Cat-NB1 products and services would commence towards the end of 2016 and in the early part

of 2017. It is forecast that IoT traffic will compound to an annual growth rate of 23 % between 2015 and 2023. Therefore, it is envisioned that the introduction of NB-IoT should have an optimal capacity to accommodate and support such growth at the same time in the near future. Some of its key performance indicators to support this are coverage extension, peak data rates, and high capacity to support Massive IoT, latency, device complexity and battery lifetime. Table 2.2 presents a detailed summary of the high-level complexity differences that have evolved between newly introduced LTE IoT UE Categories.

In conclusion, the NB-IoT (Cat-NB1), is a pioneer technology towards building the 5G New Radio Network which is intended to enable new use cases for the IoT. It is foreseen that NB-IoT will continue to evolve towards meeting future 5G requirements.

In summary, having discussed the 3GPP cellular solutions for IoT aimed at fostering the next generation of 5G new service requirements, and other non-3GPP LPWA technologies that could also be used for deploying massive IoT use cases, a comparison analysis between the 3GPP cellular LPWA solutions (eMTC, NB-IoT, EC-GSM-IoT) and non-cellular technologies such as SigFox, LoRa, and Ingenu RPMA to enable connectivity solutions for MTC applications is summarized. The possibility of deploying LPWAN anywhere around the world depends on the choice and availability of the frequency band to be used without going through any process of modification. One major and noticeable difference is the fact that promising LPWA cellular technologies such as eMTC, NB-IoT, and EC-GSM-IoT are defined by the current 3GPP standardisation and operate within the licensed frequency band, including existing LTE bands and GSM carriers, which offer a high level of security, interference-free, collision-free and quality of service (QoS) guaranteed. LPWA cellular-based connectivity solutions will support massive IoT applications by allocating spectrum resources needed for IoT services with an extended coverage area and reduced complexity and cost as a result of eliminating the complex functional radio by using single antennas and a half-duplex mode of communication. In addition, cellular-based solutions are capable of supporting the trade-offs between network capacity and coverage for LPWAN in terms of lower device and infrastructure costs because of its mature and global ecosystem which supports a massive number of devices, high-throughput use cases, as well as the ability to scale down in order to support low-performance use cases while utilising the same network infrastructure.

On the other hand, LPWA technologies, including SigFox, LoRa, and Ingenu RPMA, are proprietary-based networks that have adopted the universal 2.4 GHz ISM band which operates in unlicensed

Table 2.2. Complexity reduction summary for LTE IoT user equipments (UEs), taken from [11], © 2017 IEEE.

Device category	**LTE-Cat-1**	**EC-GSM-IoT**	**LTE Cat-M1 (eMTC)**	**LTE Cat-NB1 (NB-IoT)**
3GPP release	8	13	13	13
Peak data rate [56]	DL:10 Mbps; UL:5 Mbps	For DL and UL:74 kbps (GMSK), 240 kbps (8PSK)	DL:1 Mbps; UL:1 Mbps	DL:170 kbps; UL:250 kbps
Duplex mode [56]	Supports full duplex FDD/TDD	Supports half duplex FDD only	Supports full duplex FDD/TDD	Supports full duplex FDD only
Bandwidth [56]	20 MHz	0.2 MHz	1.08 MHz (1.4 MHz carrier bandwidth)	180 kHz (200 kHz carrier bandwidth)
MCL [100]	140.7 dB	Support for:164 dB (33dBm); 154 dB (23dBm)	155.7 dB	164 dB
Rx antenna [56]	Supports double Rx	Supports single Rx	Supports single Rx	Supports single Rx
Coverage support	Complementary to Cat-M1 and NB-IoT	20 dB	+15 dB	+15 dB
Battery life	Less than 10 years	Supports within +10 years	Supports within +10 years	Supports within +10 years
Max transmit power [56]	23 dBm	33 dBm or 23 dBm	20 dBm or 23 dBm	20 dBm or 23 dBm
PSM [100]	PSM	PSM, ext. I-DRX	PSM, ext. I-DRX, C-DRX	PSM, ext. I-DRX, C-DRX
Security	It supports 3GPP (128-256bit)	It supports 3GPP (128-256bit)	It supports 3GPP (128-256bit)	It supports 3GPP (128-256bit)
Spectrum	Supports licensed LTE Bands In-band	Supports licensed GSM bands	Supports licensed LTE Bands In-band	Supports licensed LTE In-band, Guard-band and Stand-alone

frequency bands and are susceptible to interference from other networks using the same bandwidth. In order to enhance and provide long-range communication, some of these technologies adopted the sub 1 GHz bands which are highly fragmented. Unlike SigFox and LoRa networks, Ingenu RPMA can support a large-coverage network because of its receiver sensitivity of -145 dBm which is acceptable worldwide without being restricted by the policy regulations for the 2.4 GHz band. Consequently, one issue to be addressed is the scalability of these technologies to support massive capacity. In most of these networks, new base-stations would have to be configured in order to scale up capacity when the original capacity of the network has been exhausted.

Finally, the different attributes presented by LPWA technologies are enormous and there is a need to ensure that the appropriate connectivity solution is considered when deploying IoT use cases. Technologies, including LoRa, SigFox, and Ingenu RPMA, are already deployed in multiple markets, while the promising cellular-based LPWA technologies such as LTE Cat-M1, NB-IoT, and EC-GSM-IoT, are yet to be fully commercialized based on current demand. Therefore, the need for an application layer protocol to support the various connectivity landscape solutions for application scenarios of the IoT is very critical for efficient resource utilisation and for controlling or mitigating congestion due to the massive number of devices that would be connected. There are various related application layer protocols that can be used with different performance features for the IoT, including CoAP, MQTT, AMQP, XMPP etc. For this research work, the CoAP is considered as the application layer protocol over User Datagram Protocol (UDP) as its transport layer protocol for constrained-resource networks to support the massive deployment of IoT applications.

2.8 INTERNET OF THINGS COMMUNICATION STANDARDS

With the ever-growing Internet-connectivity of things, different IoT communication standards are being proposed and standardized to support and facilitate the effort of application programmers' and the network service providers' for efficient QoS delivery for the various and emerging IoT-based applications. Among these are the IETF, IEEE, EPCglobal, ETSI, and W3C. These protocols defined by the various communication standards is a clear indication that there is no singular solution in the deployment of IoT-based use cases. Moreover, base on the nature of the IoT use case to be deployed, it is also critical to take note that some standards may not be needed to support such use cases.

In the IoT, application layer protocols are responsible for orchestrating the network that is expected to be used by resource-constrained environments. These application protocols include Message Queue Telemetry Transport (MQTT), Extensible Messaging and Presence Protocol (XMPP), Advanced Message Queuing Protocol (AMQP), Data Distribution Service (DSS) and CoAP. In this paper, our work is based on the CC approaches with emphasis on CoAP in the UDP-application layer protocols for resource-constrained environments.

2.8.1 Constrained Application Protocol

The IoT as an ubiquitous Internet Protocol (IP) technology will connect billions of devices such as personal mobile gadgets, automation systems, cellular networks, the smart cities, smart grid, and so on which will contribute to changing the world into a global digital system. Most of these Things to be connected are resource-constrained devices with limited capabilities. Consequently, most applications depend on the Web architecture to enable easy access to information and also to perform updates, such as the hypertext transfer protocol (HTTP), as a Representational State Transfer (REST) paradigm. The ever-growing IoT applications require the REST paradigm that would be suitable for devices and networks in the constrained environments, unlike the HTTP. To achieve this, the IETF Constrained RESTful Environment (CoRE) working group introduced the Constrained Application Protocol (CoAP) [91] to meet the demand of the resource-constrained IoT environment and M2M applications as a lightweight protocol.

CoAP is a unique approach to a Web application transfer protocol specifically designed to address the limitations with resource-constrained nodes and memory-constrained networks. The CoAP base specification supports the basic request methods defined in HTTP, including GET, POST, PUT, and DELETE. CoAP operates over the UDP. But UDP being an unreliable transport protocol, CoAP must ensure that it provides reliability and congestion control by itself for reliable message communication. For message reliability, the CoAP makes use of confirmable (CON) message that requires an acknowledgment (ACK) or a reset (RST) message from the destination endpoint. This enables the CoAP to retransmit the sent packet within a timeout interval when the ACK is not received. With the ACK message or CON message, a CoAP response can easily be carried, which on the other hand prevents the duplication of messages based on the message ID used, and the RST message enables the flagging of wrong parameters defined in the message. Another important message of the CoAP is a

non-confirmable (NON) message used for messages that do not require ACK from the destination end node. Fig. 2.7 presents CoAP protocol functionality for inter-connectivity conversion between HTTP and CoAP.

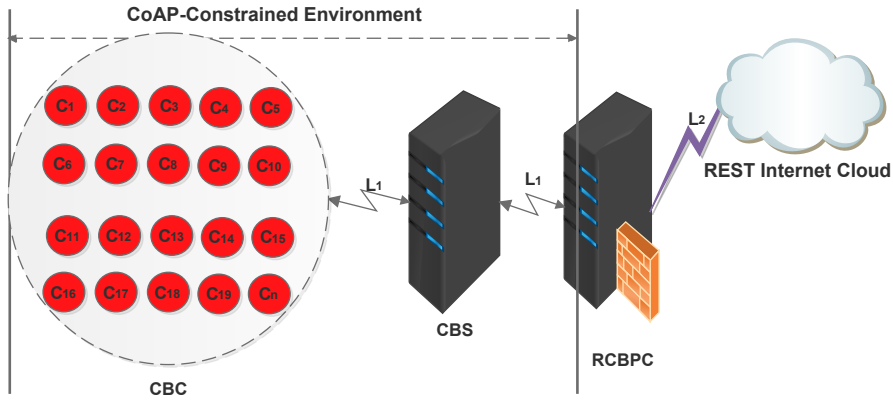

Figure 2.7. Functionality of CoAP protocol showing the REST-CoAP proxies inter-connectivity conversion between HTTP and CoAP. Where CBC, CBS, RCBPC, L_1, and L_2 represent heterogeneous CoAP-based clients, CoAP-based server, REST-CoAP-based proxy, CoAP and HTTP communication links respectively.

The current demand of Internet connectivity of Things motivates the cellular technology to introduce their IoT device connectivity landscape solutions such as LTE Cat-M1 (also known as eMTC), EC-GSM-IoT, and NB-IoT (also called LTE Cat-NB1) that will enhance and enable future IoT use cases. LPWA networks are currently being deployed for the IoT applications, including smart cities, building management systems, asset monitoring, smart agriculture, etc. The Constrained RESTful Environments (CoRE) provides a framework architecture for applications in the constrained IP networks in the IoT base on the progress of Web technology and promises of the REST paradigm. These networks are made up of resource-constrained devices, based on their unique function capability . Accordingly, RFC 7228 defines three classes of terminology for constrained devices [92]. This section briefly discusses the main features of the class-N devices.

- **Class 0** devices are constrained explicitly in terms of memory and processing capabilities; this, however, affects their ability to communicate smoothly over the Internet or IP-based IT systems in a secure manner due to limited resources. Class 0 devices can communicate over the Internet by using application-level devices to act as servers, proxies, or gateways for protocol translation. These devices are very resource-constrained sensor-like Tmote Sky, used for management purposes, such

as basic health indication systems, and keep-alive signals. The characteristic features of this platform in terms of program flash, memory, are minimal and affect their essential capability functions.

- **Class 1** devices remain one of the promising resource-constrained nodes that can easily connect over the Internet because of their security mechanisms. These devices make use of protocol stack for constrained nodes such as CoAP over UDP and communicate with the network without the use of any such application-level gateway. These devices cannot employ full protocol stack such as the HTTP over Transport Layer Security (TLS) because of their limited capability and processing power. Moreover, they require a lightweight protocol with low memory and computational overhead.
- **Class 2** devices are some of the most efficient resource-constrained devices that are capable of supporting most protocol stacks as used on smartphones, notebooks, and/or servers. Furthermore, these devices can benefit from lightweight and energy-efficient protocols for efficient resource utilisation to applications. Class 1 devices most often use IEEE 802.15.4 or BLE as low-power communications technology. The use of recommended protocol stacks which are defined explicitly for constrained devices on Class 2 devices will help minimise the cost of development and improve the inter-operability among connected-node networks.

In summary, the various classes of devices defined by the RFC 7228 are distinct and these devices, based on their capability functions, can be used for different application layers. However, the constraints of M2M solutions and the overheads of HTTP over TCP make them unsuitable for use in such resource-constrained environments, hence, the need for RESTful protocol that is more compactible, with less overhead, easily implemented over lossy communications links, to provide unique features that will enable typical M2M applications.

2.8.2 CoAP Fundamental Features

The RESTful environment motivates the IETF working group to propose and implement CoAP as a Web protocol [93]. CoAP is a client/server interaction model that operates between application-constrained end nodes based on UDP. The performance of UDP in terms of less overhead need, utilisation of smaller header, and enabling connectionless setup gives an edge for use in CoAP over TCP in low-power wireless communication with lossy links [94–96]. Also, in most applications, the need for an efficient QoS is critical as compared to that offered by UDP. To this end, CoAP introduces a thin-control layer for efficient resource utilisation, as shown in Figure 3.1; the request-response

communication model represents the RESTful interaction, and the message sub-layer represents duplication and optional retransmissions. Briefly, this subsection further discusses these unique CoAP protocol features that are of great importance in resource-constrained environments.

2.8.2.1 Message Format

Since CoAP is a binary-encoded model, it is a protocol system that enables message compatibility with low complexity for micro-controller-based systems. The message format for CoAP is a four-byte base header system, followed by a variable-length token, multiple header options, with a payload to enable the CoAP message format citeshelby2014 as shown in Fig. 2.8.

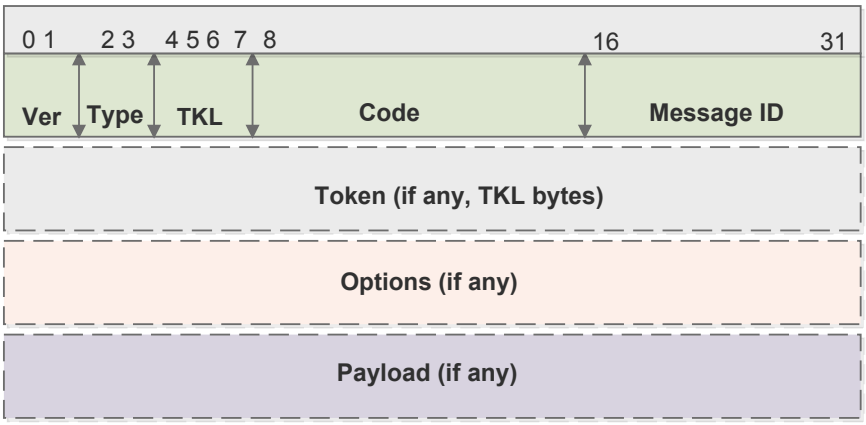

Figure 2.8. CoAP Message Format showing the different fields in the header.

2.8.2.2 Messaging Sub-layer

The messaging sub-layer allows for multiple detection and also serves as an option for reliable message transmission. The reliable transmission of messages in CoAP is based on the simple stop-and-wait mechanism to enable retransmission with truncated binary exponential backoff. For this process to be activated, the sub-layer makes use of 16-bit message identifiers (MIDs) along with unique message types, including Confirmable (CON), Non-confirmable (NON), Acknowledgment (ACK), and Reset (RST).

2.8.2.3 Confirmable (CON)

CoAP uses the CON message type for reliable message transmission. The sender node keeps retransmitting the CON message until the receiver node confirms and acknowledges the message or considers it as timeout at the expiry of the message. CoAP implements a random initial retransmission timeout $(RTO)t_{ini}$ using basic protocol timeout parameters such as $T_0 = 2\ s$ and $C_{rand} = 1.5$ to define the interval bounds in Equation 2.1 [97] which helps to prevent the effect of node synchronization. For each retransmission, CoAP doubles the current timeout until the retransmission counter is said to reach its maximum limits of $R_{max} = 4$. In reliable CoAP transmission, a CON message must be acknowledged with an ACK to indicate successful delivery, otherwise the CON messages will have to be retransmitted a maximum of four times to report loss of the CON or its corresponding ACK. To enable this process of retransmission, the initial retransmission timeout of [2, 3 s] is multiplied after each retransmission event.

$$\begin{aligned} T_0 \leq t_{ini} \leq T_0 \cdot C_{rand} \\ 2s \leq t_{ini} \leq 3s \end{aligned} \tag{2.1}$$

2.8.2.4 Acknowledgment (ACK)

ACK messages are used for reliable message transmission to acknowledge the presence of a specific CON request based on the same MID for unique correlation. An Acknowledgement message cannot be used to signify successful or failed transmission of any request that has been encapsulated in the confirmable message on its own. Also, a piggybacked response may be carried along with the acknowledgement message. Generally, a piggybacked response is rightly added to a CoAP acknowledgement (ACK) message sent out to acknowledge the receipt of the Request for this Response.

2.8.2.5 Non-confirmable (NON)

Specific applications do not require reliability or acknowledgement for messages sent. Messages that are retransmitted often for application requirements such as readings are repeated regularly from a

sensor. For such messages that are routinely triggered, a Non-confirmable (NON) message is used as best-effort delivery.

2.8.2.6 Reset (RST)

RST messages represent feedback to CON and NON messages to indicate the presence of message-processing errors as a result of abnormal context on the receiving node. This unusual missing context can occur during rebooting of nodes and therefore they cannot identify the actual message or are unable to acknowledge the message. RST is often used to indicate failure and to reset the message. Also, this can be ignored for received NON messages since most of these messages can be dropped quickly without further notification.

2.8.2.7 Request-Response Sub-layer

The request-response sub-layer is based on RESTful architecture [93]. The CoAP protocol is similar to the HTTP 1.1 [98], but with some unique features, since it is designed more specifically for the constrained environments and to support the M2M applications. The code field can be either a method or a status code. The method code turns the message into a request, while the status code changes the message into a response. To communicate with resources, CoAP defines four different RESTful syntaxes such as GET, PUT, POST, and DELETE. These RESTful verbs do have similar semantics to those of their counterpart HTTP, which facilitate a stateless and transparent mapping. Moreover, these response codes are based on HTTP 1.1, with some unique features to make the codes more suitable generally for M2M communication. CoAP methods and response codes are currently defined by RFC 7252 [91].

The empty token can then be utilised to minimise the size of messages. Also, it is worth mentioning that the token is independent of the MID, as this can only be used at the messaging sub-layer. In most cases, an ACK used to confirm a CON request can be used to piggyback the response as represented in Figure 2.9. The server defines different MID for separate responses which are either carried in a CON or NON message that cannot be used for response correlation. Table 2.3 presents a summary of the different CoAP message types that can be used for requests and responses.

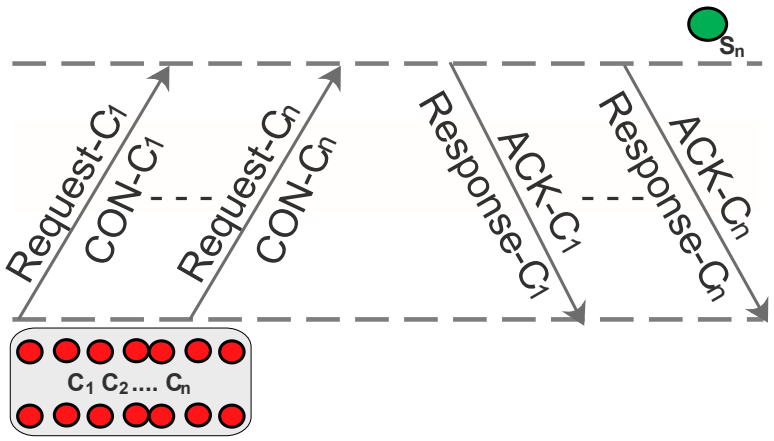

Figure 2.9. A request-response exchange, showing an ACK responding to a CON request with piggyback response. Where C_n, and S_n represent the client nodes and server node respectively.

Table 2.3. CoAP message types for requests and responses: A CoAP ping can be used to represent an empty CON which can be utilised to elicit a RST.

Message	Request	Response	Empty
CON	✓	✓	Ping
NON	✓	✓	×
ACK	×	✓	✓
RST	×	×	✓

2.8.2.8 Group Communication

With a UDP-based transport layer protocol, CoAP makes use of IP multicast to enable group communication. This is achievable by enabling a routing and forwarding protocol on specific gateway devices such as routers, including the Multicast Protocol for Low-power and Lossy Networks (MPL) [99] or Protocol Independent Multicast - Sparse Mode (PIM-SM) [100] which can be used for less constrained networks. To effect requests to a multicast address, it is essential for the use of NON messages and servers that the receiving end should ensure that it randomly delays the corresponding response for a specific time frame, called Leisure, to prevent congestion. A default time of 5 *s* is usually considered;

however, one can adapt the Leisure time T_L to the actual group size G, if the message size S and target data rate R can be estimated [97].

$$T_L = S \cdot G/R \tag{2.2}$$

Moreover, requests can be restricted by using idempotent methods such as (GET), (PUT), and (DELETE), since the client node may not identify quickly which group members have acknowleged the request. Accordingly, to enable RESTful group interaction, additional guidelines and constraints are provided in the group reference documentation [101]. Furthermore, all group manipulation requests are unicast and must be enabled by using a secured DTLS session. Since CoAP itself does not provide secure, nor reliable group communication, it is paramount that multicast requests must operate by using NoSec mode to enable secure group communication [97].

2.8.2.9 Security

The CoAP security model is similar to that of the conventional Web services (i.e. Transport Layer Security (TLS)). In view of the resource-constrained nature and UDP binding, CoAP utilises a UDP-based variant known as Datagram Transport Layer Security (DTLS), also known as a delta to TLS 1.2 [102, 103]. This model is flexible with various cipher suites defining the specific class of cryptographic algorithms to be implemented.

Irrespective of the NoSec mode, CoAP end nodes must ensure that they implement the Raw Public Key mode for a secure system. With this implementation, each end node will have its asymmetric key pair and unique identity. To enable node authentication, each device uses an out-of-band scheme based on the public key instead of full X.509 certificates [104]. Otherwise, CoAP devices can either implement the full X.509 or, preferably, the lightweight Pre-Shared-Key mode which uses the well-known AES block cipher with symmetric keys ($TLS\_PSK\_WITH\_AES\_128\_CCM\_8$) [97]. In resource-constrained nodes, asymmetric public-key cryptography is enabled by using elliptic curve cryptography (ECC) for a secure CoAP system.

2.9 CONGESTION CONTROL

The ever-growing IoT is at the centre of the future Internet, which requires the deployment of smart sensors for intensive IoT applications. Since these smart WSNs constitute the building blocks for the IoT, congestion in WSNs affects the overall energy efficiency and QoS of its application. Congestion occurs when the rate of packet arrival exceeds the packet service rate. Congestion takes place at the sensor end-nodes close to the destination node, and also as a result of the link-level performance, which could occur due to interference, contention, and bit-error-rate. Congestion in resource-constrained nodes results in a buffer overflow which leads to more substantial queuing delays and also higher loss of packets. Moreover, packet loss can also degrade the reliability and QoS of applications, and can cause wastage of the constrained node energy. Besides, congestion can also affect the link utilisation. Therefore, as the demand for resource-intensive applications in the IoT continues to grow, effective and efficient congestion control mechanisms are a critical requirement, to either avoid or mitigate congestion. Congestion control mechanisms for managing congestion in a resource-constrained environment can be categorized into congestion recovery and congestion avoidance.

In congestion recovery, this approach is followed to reinstate an operating state; in other words, a situation when the resource demand has already exceeded the resource capacity. On the other hand, congestion avoidance is a preventive measure for congestion control. This mechanism helps to maintain resource-demand on the network at a point close to its maximum power, to prevent congestion. It is also important to ensure that congestion recovery is active since, without this, the entire network may ultimately cease to be functional. This is known as zero throughput. Also, irrespective of the effectiveness of the congestion avoidance in place, the congestion recovery approach should nevertheless be maintained to sustain the throughput during unforeseen circumstances as a result of the increase in resource-demand, and loss of resources which could also result in network congestion.

2.9.1 Advanced Congestion Control

CoAP modularity makes it easier for resource-constrained application endpoints to enable their specific required features. The CoRE working grouping improves the default protocol specification by providing several extensions as a framework to support RESTful IoT applications in constrained environments.

The advanced congestion control is a protocol extension for congestion control and is described in this subsection.

Since CoAP is an Internet-based protocol, it must provide congestion control by itself to maintain the backbone network, and also prevent running overhead of constrained networks. Therefore, the base congestion control specification defines some conservative default parameter values for the retransmission timers, the allowed number of open requests, and the permitted overall message rate. However, these default parameters can be optimised by powerful CoAP endpoints for optimal and efficient quality of service. The need to enhance the legacy congestion control specification motivates the proposal for a more sophisticated congestion control mechanism called CoAP Congestion Control Advanced (CoCoA) [105]. The primary aim of this scheme is to use RTT measurements to adapts the RTOs for resource-constrained nodes. The CoCoA mechanism utilises two RTO estimators to determine the RTT measurements: the strong RTO estimator is based on the Karn‘s algorithm [106] and also conditions for TCP [107], for the successful transmission of the packet. In contrast, the weak RTO estimator indicates the RTT measurements for messages that require retransmission. These estimators are used to update the overall RTO estimator. CoCaA enhanced its scheme by introducing the exponential backoff factor that can be combined with the initial RTO. CoCoA has shown significant improvement to the base specification of CoAP [108] and since then has acted as a vital reference in the research community for congestion control in constrained-IoT environments. The rationale behind CoCoA is presented in detail in Chapter 3 as a benchmark protocol for evaluation.

2.10 5G NEW RADIO ENHANCEMENTS FOR THE INTERNET OF THINGS

Research has shown that the future 5G mobile networks will have to cater for the massive deployment of IoT with billions of connected smart objects and sensors that will be a global representation of the real world and to support the provision of mission critical IoT use cases, which will require real-time responses and automation of dynamic processes across different fields of operations, including vehicle-to-infrastructure (V2I), high-speed motion, vehicle-to-vehicle (V2V), as well as process control systems [56].

The 5G new radio network which is currently under consideration is expected to cater for both Massive and Critical IoT use cases as the demand for machine communications continues to grow extensively

to connect a massive number of smart devices with the benefits of using cellular networks. In light of this, further enhancements are currently being introduced in M2M and NB-IoT systems as specified in the current 3GPP Release-14 for cellular IoT, which is the first normative phase for 5G standards. Currently, 3GPP standardisation is working towards ensuring that further enhancements of KPIs are introduced into existing 4G networks to ensure that the 5G mobile network is designed from scratch in order to accommodate the growing span of the IoT use cases into the market, and minimise the cost of developing new networks.

In 3GPP Release-14, some of the expected key performance features and enhancements for the M2M and NB-IoT systems highlighted for Massive and Critical IoT applications to be considered for discussion are briefly introduced as follows:

- General Enhancements to MTC
- Enhancements of NB-IoT
- NB-IoT RF requirement for co-existence with CDMA networks
- Release- 14 extensions for Cellular Internet of Things (CIoT)
- New band support for Release- 14 NB-IoT
- New services and Markets Technology Enablers

For this study, the new services and markets technology enablers for the IoT are highlighted..

2.10.1 New Services and Markets Technology Enablers

The 5G mobile network is being considered as the future telecommunications system that promises to provide the opportunity to design a 3GPP network that can be easily optimised to support connected devices and services. 3GPP is currently reviewing Rel-14 for potential 5G service requirements which are expected to cover over 70 use cases under the New Services and Markets Technology Enablers (SMARTER) as promising opportunities for next-generation telecommunications networks [109]. These newly introduced use cases cut across a wide range of new service markets from the IoT to vehicular communications and control, drone control systems, tactile internet, and industrial automation, as well as catering for new services such as device theft prevention and recovery. As some of the applications for the IoT will be supported by current systems, there is a need for improvements in

terms of efficient resource utilisation, adequate support for different access technologies, network flexibility, and network slicing that needs to be implemented into the future 5G radio network which is not readily retrofitted into already functional and existing networks. According to different industry white papers [56], the objective of the future 5G mobile network is a new network system that is expected to ensure that multiple service dimensions are efficiently and effectively supported. These proposed use cases are further categorized as follows:

2.10.1.1 Massive Machine-Type Communications

This proposed Feasibility Study on New Services and Markets Technology Enablers for mIoT covers different applications (use cases such as smart utilities, smart buildings and cities, e-health systems, smart wearables and inventory control systems) with a massive number of connected heterogeneous devices, including wearables, actuators and sensors etc., with various characteristics and demands, which are specifically of importance when these new vertical services are considered [110]. For instance, smart wearables are envisaged to ensure that human clothing is integrated with a number of ultra-light, low-power, waterproof sensors which will be used to evaluate and determine environmental and health conditions (attributes) including temperature, pressure, heartbeats, blood pressure, body temperature, etc. However, it is important that a management system is put in place to control these devices, as well as the data generated, and applications for effective deployment of mMTC.

2.10.1.2 Enhanced Mobile Broadband

This proposed Feasibility Study on New Services and Markets Technology Enablers - Enhanced Mobile Broadband, envisages that users will be provided with accessibility to mobile broadband services anywhere and anytime, including in constrained areas in terms of extended coverage (such as moving from urban to suburban and rural areas). Use cases to be considered in this category are relevant to higher data rates, high density, deployment and coverage (ultra-low cost networks), higher user mobility, and fixed mobile convergence [111]. For instance, network infrastructure and its cost of terminals are not readily deployed by network operators owing to the very low average revenue per user (ARPU) in rural areas (low population density distribution). With this new service requirement, 5G mobile network is envisaged to be more flexible for deployment under ultra-low cost requirements

in order to provide Internet access to such areas, and enable new business models and avenues in underserved regions to be globally connected for efficient IoT applications.

2.10.1.3 Critical Communications

This proposed Feasibility Study on New Services and Markets Technology Enablers for Critical Communications, use cases such as industrial control applications (Drone / Robot / Vehicle), and tactile Internet should be considered. This family of use cases requires a strong demand for real-time interaction with enhancements to be focused on mobility, latency (high throughput), critical reliability and availability which can be achieved through improved radio interface and optimised network architecture [112]. Use cases such as tactile interaction, require a typical tactile control signal and audio or visual feedback system where humans can control real and virtual devices wirelessly. For instance, considering running software applications on the cloud so that the end user interacting with such environments is not aware of the difference between the local and remote contents. However, it is also challenging because of the real-time reaction which is expected to be sub-millisecond in tactile Internet use cases. (For Critical MTC use cases with Ultra-Reliable and Low Latency Communications (URLLC).

2.10.1.4 Network Operation

This proposed Feasibility Study on New Services and Markets Technology Enablers - Network Operation, use case scenario is expected to look into functional system requirements such as network slicing, flexible functions and capabilities, routing, migration and inter-networking, optimisations and enhancements, and security [113] to enable connectivity of heterogeneous networks as a unique feature of 5G mobile networks.

2.10.1.5 Enhancement of Vehicle-to-Everything

These proposed use cases include autonomous driving, safety and non-safety aspects (which are associated with vehicles), requiring provision of ultra-reliable communication based on real-time response in order to prevent the occurrence of road accidents. It is hoped that emerging 5G mobile networks will be able to provide low latency, high reliability, higher accuracy positioning and mission

critical services which are required for future safety applications to reduce the occurrence of road accidents, enhance traffic efficiency and enable the mobility of emergency vehicles, including fire trucks, ambulances etc. Enhancement of Vehicle-to-Everything (eV2X) is foreseen not to be applicable only to vehicle-to-vehicle (V2V) or vehicle to infrastructure communication, but to be also applicable to other vulnerable road users.

The proposed use cases highlighted above as specified in 3GPP, are the basis of normative requirements which are currently under consideration as service requirements for the future 5G next-generation network. Figure 2.10 depicts the proposed new enhancements for service requirements for 5G mobile networks that can efficiently and effectively support multiple service dimensions. Finally, we conclude that with further research on the 3GPP New Radio (NR) (Release-13/14) for the emerging IoT standards, 5G mobile networks aim to enable the basic requirements and KPIs which are required for future 5G new services to enable the IoT use cases. The IoT will continue to be more pervasive in future use cases, which are expected to impact the everyday-life of the end user community positively. However, the massive number of connected things and the heterogeneous nature of communication networks in the IoT pose many research challenges.

Although the IoT paradigm has been the focus of research, there is a need for more intense research work to ensure that the vision of the IoT is globally achieved as projected. The continuous contribution and attention which have been given by academia, industries, and governments have definitely led to great achievements in terms of research projects towards meeting new service requirements for the IoT. Some of the IoT challenges, including security, extended coverage, low device cost, low power consumption and network architecture, have been given some consideration over time while others, including network management, heterogeneity and inter-operability, traffic congestion and control which automatically guarantee the availability of network information, and QoS over a specific time, are still broadly open for more intense research. With the current emerging cellular-based LPWA solutions for the IoT, including eMTC, EC-GSM-IoT, and NB-IoT, there is a need for research into the network positioning of smart connected things and context-aware services in view of the new service requirements for the next-generation mobile telecommunications. Table 2.4 presents a summary analysis of 5G KPIs [114] of the various existing and emerging technologies which have been presented in detail, and considers their modern connectivity characteristics to address the IoT requirements in terms of extended coverage, availability of dedicated spectrum, low deployment cost, battery lifetime, and scalability for the new service requirements.

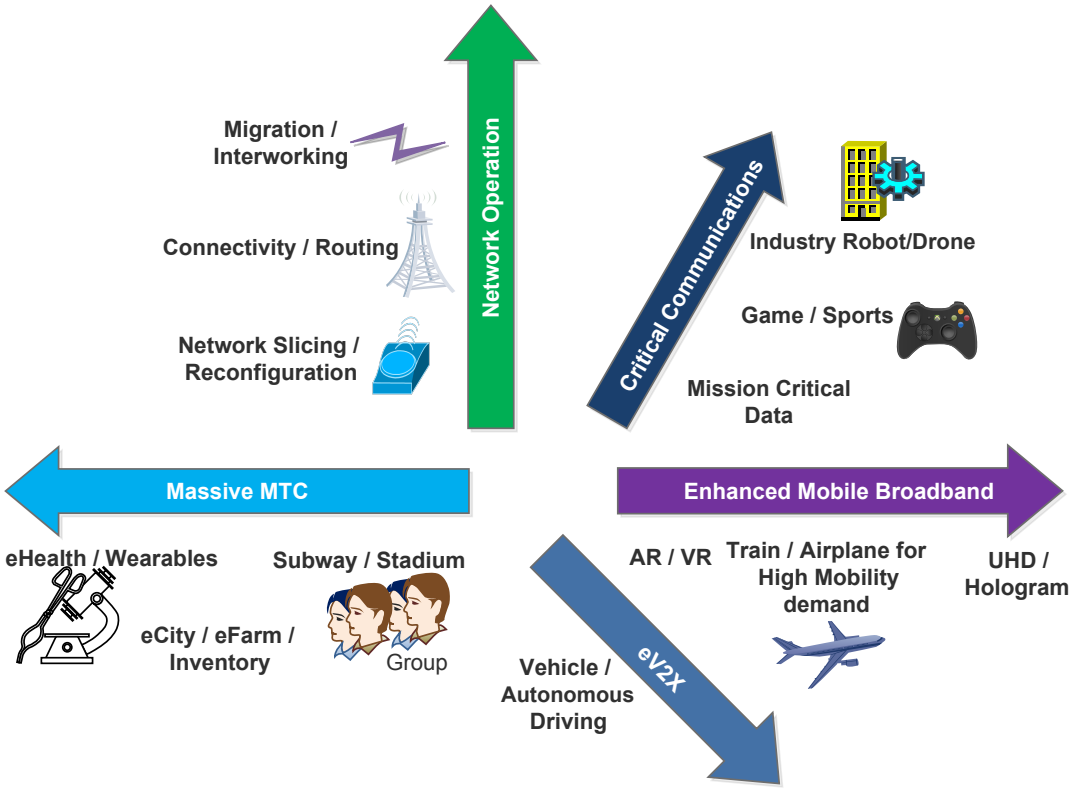

Figure 2.10. SMARTER new service dimension, taken from [11], © 2017 IEEE.

2.11 NETWORK ENABLERS FOR THE INTERNET OF THINGS

5G mobile networks are envisaged as a promising next-generation technology to support the massive deployment of simultaneously connected heterogeneous devices with new service requirements based on wearable things, improved better coverage edge, low latency, high versatility and scalability for efficiently enabling the massive to critical IoT applications. However, it is obvious that the conventional network infrastructure is continuously becoming too outdated to support these features for the IoT architecture with the existing conventional networking system. Consequently, the network management complexity continues to increase because of the manual process which is used in network configurations owing to the limitations of conventional hardware-based networking. Moreover, a network system should be able to enable the ever-evolving networking technologies for future network infrastructure. Moreover, the current traditional networks cannot meet the ever-growing networking technologies' demand for future next-generation networks. To achieve these objectives, emerging technologies such as Software-Defined Wireless Sensor Networking (SDWSN), Network Function Virtualization and

Table 2.4. Key performance indicators (KPIs) analysis for modern IoT connectivity solutions, taken from [11], © 2017 IEEE.

KPI	ZigBee	BLE	Wi-Fi	SigFox	LoRa	EC-GSM-IoT	eMTC	NB-IoT
Extended coverage	×	×	×	✓	✓	✓	✓	✓
Reliability	×	✓	✓	×	×	✓	✓	✓
Low deployment cost	✓	✓	✓	✓	✓	✓	✓	✓
Long battery life	✓	✓	✓	✓	✓	✓	✓	✓
Low latency	×	✓	✓	×	×	✓	✓	✓
Network scalability	×	×	✓	×	×	✓	✓	✓
Mobility support	×	×	×	×	×	✓	✓	✓
SLA support	×	×	×	×	×	✓	✓	✓
Support for Roaming	×	×	×	×	×	✓	✓	✓
Dedicated spectrum	×	×	×	×	×	✓	✓	✓

Cognitive Radio (CR) are among the few network enablers to be discussed briefly in this section to overcome such limitations of the legacy networks by 5G mobile network for the IoT.

2.11.1 Software-Defined Wireless Sensor Network

Cellular technology, which is currently being deployed for wireless communications and most especially for the rapid growth and requirements of the IoT applications, is hardware-based designs and therefore requires that emerging technologies for future next-generation networks are introduced to enhance the flexibility of the network infrastructure to accommodate and process the massive inflow of data for the IoT use cases. SDWSN is a promising new paradigm to achieve Low-Rate Wireless Personal Area Networks (LR-WPAN) [17, 115, 116]. This network paradigm is achieved through the infusion of the Software-Defined Networking (SDN) model into existing Wireless Sensor Networks

(WSNs). SDN was primarily intended to be used in wired communication systems such as data centres and for next-generation Internet connectivity [117–119] but has recently emerged in most wireless communication networks [120], and is envisaged to be a future technology enabler for the next generation of 5G mobile networks [121, 122]. This intelligent network paradigm provides a centralized network abstraction for programmability of the entire network.

The primary purpose of introducing SDN is to decentralise the control logical plane from the network device, e.g., switch, and to enable an external network controller to define the nature and behaviour of the network forwarding infrastructure (i.e. routing and major management processing). SDN also facilitates the introduction and deployment of new application services, and enables the flexibility of network management that would support the exponential traffic growth of the envisaged 5G mobile networks [123] as a major network enabler for the IoT. WSNs is expected to be a vital enabler for IoT systems since most of the sensor nodes are the main entities for this concept [124]. It is hoped that SDWSN will be a key network enabler technology to address the issue of flexibility and inter-operability of future multi-vendor infrastructure in 5G mobile networks for the IoT. With the massive increase in the number of connected devices through Massive to Critical IoT, traditionally deployed network architectures, which are hardware-based, will need to be enhanced to accommodate, manage, and control the large amount of heterogeneous devices and inflow of data into the network.

Therefore, there is a need to introduce SDWSN into next-generation cellular networks that will simplify the entire network infrastructure, manage and control the entire system's requirements and maintain the heterogeneity of the networked environments to enable future IoT use cases.

2.11.2 Network Function Virtualization

Network Function Virtualization (NFV) is highly complementary to SDN, which is envisaged as an enabling network technology for next-generation service requirements for the IoT applications but both are independent of each other. This means that NFV can be implemented successfully without considering SDN and vice versa. In addition, it is possible that both solutions can be combined to achieve optimal performance output.

According to [125], NFV technology can be used to virtualize a set of network functions which can be

further implemented into software packages to be configured in order to provide efficiently for related service requirements such as the existing network infrastructure. The concept of NFV has been realised from the perspective of Virtual Machines (VM) which can be installed to run on various operating systems on the same server machine. With the rapid growth of and increase in the number f connected heterogeneous devices for massive to critical IoT applications towards new service requirements for 5G mobile networks, there is a need for a key technology enabler in 5G technology to enable the massive deployment of MTC devices efficiently.

For network functions (i.e. by relocating the routing decision-making from local hardware and to be implemented into general purpose computing and other storage devices such as servers and cloud), NFV technology would ensure that the deployment of heterogeneous MTC devices to achieve Critical and massive IoT applications is managed and controlled well. According to [126], NFV is being considered in the context of virtualizing core networks, and centralizing base band processing within Radio Access Networks (RAN). In addition, by virtualizing network functions in any deployed network infrastructure, system scalability and flexibility of connected heterogeneous devices, reduction in Capital Expenditure (CAPEX) and Operational Expenditure (OPEX), as well as power consumption, can be achieved efficiently, which will, at the same time, support the market deployment of IoT use cases.

2.11.3 Cognitive Radio Networks

The current demand for MTC to ensure that everything is connected, anywhere and at any time, has resulted in drastic changes which are evolving in currently deployed cellular networks and next-generation networks such as the 5G mobile network. Inter-connectivity of heterogeneous devices with diverse service requirements presents a huge opportunity and challenge to cellular network operators to deploy future IoT applications massively. The current service requirements for IoT applications from both Massive to Critical IoT will definitely result in increased average revenue per user (ARPU) which would be a result of newly introduced services. On the other hand, the massive demand for connected things leads to overloading of certain cellular geographical areas which is a result of limited spectrum resources which have been licensed to cellular network operators. Research has shown that a deficit in broadband spectrum is likely to have approached 300 MHz in 2014, and the provision of additional spectrum to be considered for mobile broadband would also lead to an excessive increase of

100 billion dollars [127].

Cognitive Radio (CR) technology is a key network enabler for 5G mobile networks to utilise the limited and scarce spectrum resources in order to support the increasing and high demand for new service requirements of emerging and promising IoT applications. CR supports the capability of using or sharing the licensed spectrum in an opportunistic manner [128]. Dynamic spectrum access techniques allow the CR to operate in the best available channel. CR technology enables the identification of available free spectrum and also enables the detection of licensed users present in the system (spectrum sensing), the ability to select the best available channel (spectrum management), the ability to coordinate accessibility to the available channel with other users (spectrum sharing) and finally the ability to vacate the accessed channel on arrival of the licensed or primary user (spectrum mobility) [128].

CR technology can be used to augment next-generation cellular networks (such as LTE, WiMAX and future 5G mobile networks) to dynamically access newly introduced spectrum (such as TV White Space) which has been officially released by the Federal Communications Commission (FCC) for unlicensed operation in the TV Broadcast Bands [129]. This can be achieved by introducing a spectrum coordinator in the non-access stratum (NAS) which will enable cellular network technology to lease or access spectrum markets dynamically and to determine/identify secondary license-exempt spectrum opportunities which can be used in deploying IoT applications for a time in a given location. Another major consideration where CR technology can be an enabler to achieve the massive deployment of the IoT vision is the rural geographical area, which is generally known to have poor coverage. Since licensed spectrum is limited and scarce, cellular network operators in most cases prefer not to deploy their networks in rural areas owing to low population density distribution which is typically not cost effective owing to the limited number of network subscribers. With CR technology, white space spectrum, which has been proposed and made available for unlicensed use, can be explored for back-haul by cellular network providers in order to ensure that their cell towers are connected to their backbone networks, thereby providing and deploying connectivity solutions for IoT applications in unserved and underserved geographical areas which will support the vision of connecting more MTC devices for the IoT.

2.12 RESEARCH CHALLENGES AND FUTURE DIRECTION

The current demand for MTC connectivity to provide the various new services and applications for both the industrial and societal needs has introduced new challenges to meet the current requirements for the IoT vision. It is important to ensure that special consideration is given in order to address these challenges to support MTC devices, so that the security and QoS for both MTC devices and human-to-human (H2H) users that utilise the same network infrastructure is not compromised.

Network congestion and overload present a major challenge that needs to be addressed in the evolution of the IoT because the smart connected devices play a vital role in driving up the signaling load in the mobile network when compared to the traditional H2H traffic in cellular networks. Network congestion degrades the IoT performance and QoS. A challenging issue which is pertinent to MTC is the ability to accommodate the huge traffic that would be generated as a result of the massive number of MTC devices that would create congestion problems in the networks. Therefore, the IP will have to deal quite efficiently with the network congestion problem. Most networks provide a stable Internet connection, irrespective of the connection of the massive number of devices using TCP in the transport layer. However, the existing TCP implementations are not suitable for the IoT application scenarios, and fail to cope with the IoT traffic pattern since the traffic pattern of the IoT network is entirely different from that of the conventional networks.

CoAP has been developed by the IETF as a lightweight protocol for resource-constrained devices and lossy communication networks [91, 130, 131]. CoAP, being considered as an IP must be able to handle congestion control in order to maintain the network backbone. Enabled IoT networks would have to use different protocols to enable communication. There are different existing application protocols for the IoT. These protocols are designed to perform for different application scenarios. A lot of research is on-going by both the academics and industrials on congestion control mechanisms. Hence, without incurring high overheads, there is a need for efficient handling of network congestion and BER to compensate for the packet loss and delay in the IoT environment. In addition, the congestion control mechanism in IoT should be capable of assuring safe network operation with efficient network resource utilisation. Network congestion is unavoidable since with the massive number of connected things, the huge traffic that would be generated would result in network congestion and lead to high packet loss rate. Therefore, this demands an efficient QoS assurance system as a lightweight context-aware

congestion control (CACC) mechanism to manage and handle the network congestion problem and bit error rate in the IoT networks.

2.13 CHAPTER SUMMARY

The current expectation and future evolution of the IoT are promising to enable new services and quality of experience (QoE) across the user community. This is very challenging at the same time because of the resource-constrained nature of the network which has compelled the research community to ensure that the requirements for massive deployment of MTC applications are met for globally connected things.

This chapter has reviewed the unique features of the current state of IoT standard infrastructure.In section 2.2, the overview of WSNs unique features as building blocks for the IoT are presented and then followed by a discussion on functions of the WSNs transport protocols in section 2.3. Section 2.4 presents a discussion of the performance metrics for WSNs, including reliability, energy efficiency, quality of service, and fairness in resource allocation. In section 2.5, the overview of the IoT application requirements including those of the emerging IoT applications and the design requirements to enable massive deployment of the IoT use cases are discussed. A discussion of the existing IoT communication technology based on the non-cellular LPWA technologies, and the short-range networks is presented in section 2.6 and followed by a discussion of cellular LPWA-based solutions including eMTC, EC-GSM-IoT, and NB-IoT as connectivity landscape solutions for the future demand for Internet connectivity of Things in section 2.7. Section 2.8.1 presents a discussion of the Constrained Application Protocol as a unique approach for a Web application transfer protocol specifically designed to address the limitations with resource-constrained nodes and memory-constrained networks and this is followed by a discussion of congestion control in section 2.9. In section 2.10, the 5G new radio enhancements as a next-generation network for the IoT, including massive machine-type communications, enhanced mobile broadband, critical communications, network operation, and enhancement of vehicle-to-everything are discussed to meet the new service requirements. Section 2.11 presents a discussion of network enablers for the IoT, including SDWSN, NFV, and CR, to support dynamic data control, provide a centralized network system and to enable the adaptation of new service requirements to enable massive to critical IoT use cases with efficient coverage and high-capacity targets for lifetime resource-constrained devices. However, there still are open research challenges for effective control and management

of the IoT networks. For the future evolution of the IoT, it is therefore recommended to develop a context-aware congestion control (CACC) scheme for a lightweight CoAP/UDP-based IoT network as a multi-objective function that would support the exponential traffic growth pattern of the envisaged massive number of connected IoT networks as an open research direction in section 2.12.

CHAPTER 3 CONTEXT-AWARE CONGESTION CONTROL SCHEME

3.1 CHAPTER OVERVIEW

The Internet of Things (IoT) paradigm has drawn the attention of developers and the research community as a result of its applications that require the deployment of massive resource-constrained nodes. These network devices are subject to limited memory and hardware processing capacities. In addition, the communication technologies used by these constrained-devices are also faced with low data rates, as well as relatively high bit error rates (BER), resulting in congestion in constrained networks. Network congestion arises when the traffic load offered reaches the network capacity. In view of this, the Internet Engineering Task Force (IETF) CoRE Working Group (CWG) has proposed and designed the Constrained Application Protocol (CoAP) [91] as a lightweight RESTful protocol for IoT resource-constrained devices as a request-response interaction model. However, as stated in Chapter 2, there is a need for a context-aware congestion control (CACC) scheme as a multi-objective function for efficient resource utilisation. In line with the identified research gap, Chapter 3 investigates and fully presents the system model of the proposed CACC approach. As contribution to the body of knowledge, the system model and analysis presented in this Chapter have been published as a part of a journal article titled "CACC: Context-aware congestion control approach for lightweight CoAP/UDP-based Internet of Things traffic" [1].

The outline of the rest of this Chapter is as follows: Section 3.2 presents the overview of IoT and the need for congestion control (CC) in resource-constrained networks and this is followed by a discussion of related literature on CC in Section 3.3. The background relating to the benchmark protocols, including baseline CoAP CC specification and CoCoA+ for validation, is presented in

Section 3.4. The overall system model for the proposed CACC scheme is then described, detailing all interconnected tasks, including the three RTO estimation algorithm, dynamic SRTT and RTO overall estimation, context-aware RTO, RTT fluctuation aware RTO, and restricted RTO shrinkage in Section 3.5 while Section 3.6 presents an analysis of the proposed CACC model. Section 3.7 summarises the Chapter.

3.2 BACKGROUND

The emerging Internet of Things (IoT) is a revolutionary paradigm that combines several information and communication technologies (ICT) as a result of the advancements in the electronics and communications fields. This paradigm promises to connect billions of smart objects that add intelligence by collecting the information from the physical environment and exchanging the data globally through the Internet [6] to enhance the environment and to affect human lives positively. In this context, it is expected that the ICT, along with private organisations, will be a key enabler for the IoT to offer innovative solutions to address challenges in services and applications for both the population and industries at large [132, 133]. These network devices are characterized by limited memory and processing capacities, as well as low radio bandwidth and a relatively high bit error rate (BER), which are major consequences of congestion in constrained networks [134]. The need to catch up with the expectations of the IoT has resulted in the design of protocols and standards that would be utilized by constrained devices with limited hardware and computation capabilities. To this end, the Internet Engineering Task Force (IETF) has proposed different layers of the communication protocol stack for constrained devices in the IoT which operate in the lossy and low-powered networks (LLN). Amongst these, the IETF CoRE Working Group (CWG) standardized the Constrained Application Protocol (CoAP) [91], which enables the exchange of information within IoT networks.

The CoAP is designed to meet the requirements of LLN and to provide lightweight inter-operability among the global Internet services [4, 135]. By default, the CoAP has no end-to-end congestion control mechanism, and therefore it includes a separate CC scheme as part of the CoAP base specification. Still, the solution of a CC mechanism is not yet mature, especially when the traffic load on the IoT network approaches the network capacity. The mechanisms supported by the Transmission Control Protocol (TCP) in most conventional Internet applications support end-to-end CC. However, traditional Internet protocols are not suitable for resource-constrained environments due to the dynamic characteristics of

LLN [20, 136]. Moreover, the CoAP does not rely on TCP, as it operates over a datagram transport called User Datagram Protocol (UDP) to enable lightweight applications and to implement the CC scheme by itself. The standard CoAP has two abstract layers. The upper layer implements the request-response communication for RESTful model, whereas the lower layer implements the control mechanism over CoAP message transmission through the underlying UDP protocol [137], as depicted in Figure 3.1.

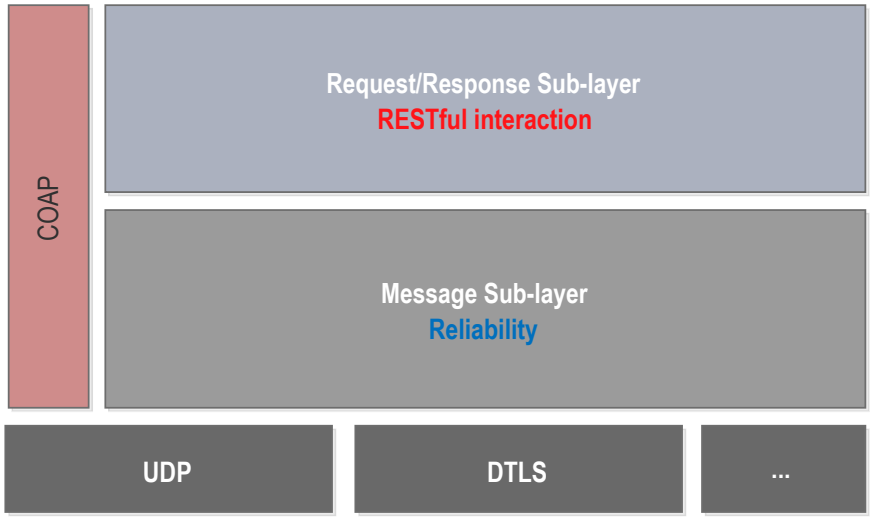

Figure 3.1. CoAP sublayers, taken from [1], with permission.

The traffic pattern of the IoT differs significantly from that of conventional networks. The CoAP employs Confirmable (CON), Non-confirmable (NON), Acknowledgement (ACK), and Reset (RST) messages for packet transmission. The CoAP implemented over UDP especially requires ACKs to ensure reliability when selecting the CON packet type. Even though, the IoT data size is small; the sensor devices communicate periodically to notify their measurements and ACKs. This periodic notification results in network congestion. From the core CoAP to the Advanced CC scheme for CoAP also called (CoCoA+), a simple CC mechanism is implemented by using retransmission timeout (RTO).

The standard CoAP CC under-performs when it is found to be too conservative in small networks or too aggressive in large networks [131, 138, 139], instead of adapting to the network status information. To improve on the base CoAP CC specification, new schemes have recently been proposed. Amongst

these is the CoCoA+ [134, 140] which is currently being standardized by the IETF CWG [105]. CoCoA+ makes the system sensitive to the network dynamics by incorporating a novel round-trip-time (RTT) estimation scheme which automatically sets the RTO value, along with a variable backoff factor (VBF) and aging mechanism for the transmission of CoAP messages. However, CoCoA+ is still deficient in choosing the exact RTO value during burst traffic because of the inability to determine the RTT of retransmitted request-response accurately, thereby resulting in unnecessary spurious retransmissions [141]. Without incurring high overhead, efficient handling of network congestion is essential to mitigate the packet loss and delay for efficient resource utilisation [11]. The IoT traffic patterns, together with their resource constrained-nature, present challenges to the efficiency of the CC mechanism for CoAP.

The major obstacle to identifying the actual RTT for packet retransmission is the unavailability of the context-aware information needed to identify the exact origin of the relevant packet. According to Karn‘s algorithm [106], this issue has been raised as a major concern. However, CoCoA+ still considers fixed RTO value for packet retransmission. Many CC techniques employ RTT as the leading metric to decide the RTO. However, the estimation is carried out without adapting its behaviour to the network status information which is available to CoAP. To address this problem in CC, we propose a context-aware congestion control (CACC) scheme for CoAP that utilizes the available retransmission count (RC) of the request-response communication model to estimate the actual RTT in order to determine the RTO value and effectively control congestion for constrained devices in the IoT networks. Also, the proposed CACC scheme applies the lower bound RTO restriction approach, which avoids unnecessary spurious retransmissions to adequately improve the efficiency of the scheme. To validate the performance of the proposed scheme against the baseline CoAP congestion control specification and CoCoA+, comprehensive model analysis and evaluation have been developed which use the Cooja simulation environment, a toolset of the Contiki OS. The results show that the proposed CACC scheme can reduce the overall number of retransmissions effectively with increased network throughput compared to that of baseline CoAP and CoCoA+. Therefore, the proposed work adapts the CC mechanism based on the available network status information.

3.3 RELATED WORK ON CONGESTION CONTROL

There are several works in Wireless ad hoc and Sensor Networks for CC. However, our work is based on the CC mechanisms in the UDP-application layer protocols for constrained networks.

The author in [20], emphasises the need to introduce a CC mechanism for CoAP UDP-based application layer protocols. [142], Eggert and Fairhurst, in their study, acknowledge categorically that the traditional IETF CC schemes are not suitable for CoAP as a low-data-volume application protocol. The authors recommend that an estimate of RTT should be maintained for the destination being communicated, or a conservative fixed value of 3 *s* should be assumed when there is no RTT estimate.

In [91], the Internet Requests for Comments, RFC 7252, which defines the CoAP, states that a base specification should be used for CC. This base specification of the CC does not take into consideration the actual RTT estimate values based on previous transactions. In [143], Dashkova et al. in their argument emphasize the need to introduce more advanced CC mechanisms to improve the network and energy resources of constrained nodes while maintaining maximum throughput.

In response to these observations, several protocols have been proposed recently to support the application layer in the IoT. An advanced CC protocol CoCoA for CoAP has been proposed [105, 137, 144] to deal effectively with the issues of CC in constrained IoT networks. To ascertain the reliability of end-to-end communication for constrained networks, the authors in [137] have analysed the performance evaluation for various CoAP CC schemes. Their results have shown that the recently proposed CoCoA+ CC scheme for CoAP [105] outperforms, and is better than or equal to the base CoAP CC mechanism for end-to-end communication.

In [140], for instance, CoCoA+ was evaluated in a typical IoT communication scenario in which General Packet Radio Service (GPRS) was used to connect IoT devices. In their analysis, the authors compared CoCoA+ to existing CC schemes available for TCP applications. Their results showed that CoCoA+ performed equally to or outperfored all other TCP-based algorithms. Similarly, in [139], the authors compared CoCoA+ against two TCP-based CC mechanisms on an emulated ZigBee-based network. The results showed that CoCoA+ and all other alternatives to CC were scalable and more efficient when compared to the default CoAP. In [145], the authors evaluated CC mechanisms implemented in the base CoAP and CoCoA+ for two different IoT traffic patterns. Their performance

analysis showed that in a network of bursty traffic, the performance of CoCoA+ could be significantly worse than the default CoAP owing to the inability to select a proper RTO value adequately. The authors in [146] proposed a new congestion control mechanism for CoAP based on the CAIA delay gradient (CDG) to predict and determine the network congestion, obtain the gradient of RTT over time and use a probabilistic backoff factor (PBF) to monitor and control the network congestion and to enable the adjustment of RTO based on the inferred condition. Results showed that, with the implementation of the delay gradients, it was quite effective to determine the accurate level of congestion and reduce the RTO with minimal delays and high packet-sending rates.

A Precise Congestion Control Algorithm (pCoCoA) [147] for CoAP has introduced the modified algorithm of CoCoA+ that overcomes the limitations in the existing algorithm. In pCoCoA, the use of a weak estimator is eliminated, and a transmission count is applied for matching the ACK messages with CON messages even during retransmission. This method also detects unnecessary retransmissions that occur in the network by comparing the transmission counter value of the retransmission with the transmission counter value of the ACK message. The pCoCoA has the advantage that it has reduced retransmissions and the ability to work in bursty traffic occurrence-based scenarios. However, considering the wide range of IoT scenarios, some of the fixed values used in this scheme might not be suitable. In the bid to improve on the CoCoA algorithm, the authors in [148], have proposed CoCoA 4-state-Strong as a modification to CoCoA that employed 4-state estimators in order to differentiate the wireless losses from the congestion losses. Their results have shown an improvement in the performance of maximizing throughput and the rate of packet loss at the same time. In [149] the authors propose a new rate-based congestion control for CoAP (CoAP-R) as an improvement to ascertain the performance of CoAP in bursty traffic environments. To control the network traffic, this mechanism monitors the sending rate of CoAP sources and use a rate-based scheme instead of the usual window-based mechanisms. This scheme is aimed at achieving maximum bandwidth and network resource allocation based on fairness. It has been observed that CoAP-R ensures a uniform distribution of network resources amongst all the senders with reduced delay compared to CoAP and CoCoA. For instance, in [150], the authors propose an enhancement of the CC mechanism for CoAP and CC/advanced which we call (E-CoCA), that uses the RC of the request packet to estimate the actual RTT by matching the request-response packet to determine the RTO value. The results have shown that the proposed mechanism improves the efficiency of CC when compared to the base specification of CoAP and CoCoA+.

A new mechanism called Fast-Slow RTO (FASOR) [151] based on retransmission timeout and a congestion control mechanism for CoAP is introduced to address the problem of packet loss which is either due to the wireless link environment or due to congestion by considering three unique features, including Self-adaptive retransmission timer backoff, Slow RTO computation, and Fast RTO computation mechanisms. One promising advantage of the scheme is the fact that, unlike traditional CoAP and CoCoA mechanisms, the FASOR scheme can handle a high congestion level, even in the buffer-bloat environment, but suffers from high latency cost owing to the slow RTOs involved in the scheme. In [152] a new mechanism based on dynamic model selection has been proposed by the author, which enables the extension of battery life. The mathematical model developed for CoAP makes use of MLE in order to predict both confirmable and non-confirmable application layer packet losses which are based on certain network layer parameters. In order to enable the switching of CoAP between transmission modes to limit the number of retransmissions and also reduce the level of data rates, as well as power consumption, the estimated confirmable and non-confirmable packet losses were dynamically compared with packet loss goals. Results show that the dynamic model can extend the battery life with minimal delay and transmission rates while improving the overall performance of the system.

A summary of the recently proposed CC mechanisms, with their unique characteristic features, is presented in Table 3.1

Table 3.1. Summary of congestion control algorithms, taken from [1], with permission.

Scheme	**RTO aging**	**Backoff method**	**RC**	**RTT estimators**	**Existing scheme**
CoAP	No	BBF	NO	No	None
CoCoA	Yes	VBF	NO	Yes-2	LinuxRTO
4-state strong	Yes	VBF	NO	Yes-4	CoCoA
CoCoA-S	Yes	VBF	No	Yes-1	CoCoA
E-CoCoA	Yes	VBF	YES	Yes-1	CoCoA
CoCoA++	No	PBF	Yes-TC	No-CDG	CoCoA

3.4 BACKGROUND ON BENCHMARK PROTOCOLS

This section presents an overview of the baseline protocols used for validation against the proposed CACC scheme.

3.4.1 The Base CoAP Congestion Control

CoAP is a Representational State Transfer style (RESTful) [107] protocol that uses GET, PUT, POST, and DELETE to maintain the servers' resources. As a lightweight RESTful application layer protocol for constrained devices, its interaction model (request-response) between clients and servers takes place asynchronously over an unreliable transport protocol called UDP. CoAP must ensure that the CC problem is taken care of by itself [140] to implement end-to-end reliability for CON messages. These messages require an ACK from the destination endpoint. Contrary to this, where the application does not require end-to-end reliability, NON-messages are used which do not require an ACK. CoAP makes use of a basic technique called RTO to identify packet losses. Basically, for a CON message to be sent from the sender to the destination endpoint, CoAP randomly selects an initial RTO value from a fixed interval of $[2,3]$ s for the initial message transmission [134]. The CoAP message is retransmitted up to four times before it is considered to have failed when the CoAP assumes that a loss has occurred at expiry of the timer, and the initiator of the message is yet to receive an ACK from the destination node.

In order to control the network congestion, the base CoAP CC specification uses a Binary Exponential Backoff (BEB) technique to double the RTO value for packet retransmission, which is insensitive to the network conditions. Base CoAP CC applies the NSTART parameter to determine the number of simultaneous exchanges that can be allowed towards a particular destination. With the base CoAP CC, NSTART is set to one, which can be used by most applications running CoAP. In summary, the base CoAP CC mechanism adapts the following:

i) Calculation of RTO for the initial transmission of a CON CoAP message.
ii) The application of classical BEB to the RTO before initiating retransmission of any CON CoAP message.
iii) The use of state information on destinations of CON CoAP messages.

3.4.2 CoCoA+: Advanced Congestion Control for CoAP

CoCoA+ is an advanced CC that enables the base CoAP CC to be dynamically sensitive to the network conditions. CoCoA+ utilises the RTT information which is obtainable from the ACK packets in order to adapt the RTO values for transmitted messages. The CoCoA+ technique makes use of two different RTO estimators: i) the strong RTO estimator (RTO_s) and ii) the weak RTO estimator (RTO_w). The strong RTO estimator is used to define RTT measurements for successful completion of transmissions that have taken place at the initial transmission, while a weak RTO estimator indicates RTT measurements for message exchanges that require at most two retransmissions as a result of packet loss at the initial transmission.

CoCoA+ mechanism takes into consideration three unique components, including adaptive RTO calculation, a variable backoff factor (VBF), and RTO aging mechanisms in order to determine dynamic RTO estimators for CoAP message transmissions. To achieve this, CoCoA+ utilises the RFC 6298 principles of the IETF for adaptive RTO calculation and measurement of RTT. In TCP implementations, the RFC 6298 is used for RTO computation using an exponentially weighted moving average of RTT and RTT variation estimates [140]. For efficient IoT communications, CoCoA+ mechanism employs this algorithm for its advanced congestion control. It is also worthwhile to mention that in TCP, the loss of packets is assumed to have resulted from network congestion. However, in IoT networks, a very high rate of packet loss is also expected to occur as a result of BER.

In RFC 6298, a strong RTO estimator is used as RTT measurements to indicate ACK received for packets before the sender node runs into retransmissions. To improve on this mechanism, CoCoA+ has introduced a weak RTO estimator that makes use of weak RTTs to represent the RTT measurements for packets that require at least two retransmissions. This enables the updating of either the weak or strong RTO (RTO_x) based on the weak or strong RTT measurements, by using the same principle of RFC 6298 as given in Equation 3.1 [140].

$$RTO_x = SRTT_x + K_x + RTTVAR_x \tag{3.1}$$

In Equation 3.1, x represents either a weak or strong, SRTT and RTTVAR represent the smoothed RTT and RTT variation, where K_x denotes K_{strong}, and K_{weak} with a value of 4 and 1 respectively. This gives a new value for the overall retransmission timeout RTO_{over} based on weighted average values as in Equation 3.2 [140].

$$RTO_{over} = \alpha \times RTO_x + (1 - \alpha) \times RTO_{over} \tag{3.2}$$

Where α represents 0.5 and 0.25 for both strong and weak RTO estimators respectively. To address the problem of small initial RTOs which could result in spurious retransmissions or large initial RTOs which could also lead to unnecessary increase in delay based on the use of BEB, CoCoA+ introduces a variable backoff factor (VBF) to adjust the backoff factor based on the initial RTO value for a transmission. To enable the VBF for an initial RTO value that is very small, either below 1 *s*, or for a large RTO value above 3 *s*, a backoff factor of either 3 or 1.5 is applied respectively for retransmissions. A VBF of 2, which is an equivalent of BEB, is used for a transaction that initiates with an RTO value between 1 and 3 *s*. The RTO aging mechanism is applied to update the estimated RTO values for a specific period due to the IoT network conditions which can cause the RTT to change quickly. In order to prevent the occurrance of bogus RTO values, CoCoA+ applies the aging mechanism to both small and large RTO estimations. This mechanism is used to enable an RTO that is small or large (either below 1 *s* or above 3 *s*, respectively), with no new measurement of RTT for 16 or 4 times the present RTO, respectively, then the RTO value can be adjusted based on the default initial value [140]. In summary, CoCoA+ adapts the following modifications as an improvement to the CC mechanism for CoAP. For a comprehensive illustration of this mechanism, the interested reader is referred to [134] for the rationale behind it.

i) A modification was introduced to the calculations of the weak estimator in order to reduce the impact of RTT_{weak} variations and their impact on RTO_{over}.
ii) The introduction of a Variable Backoff Factor (VBF) scheme as a replacement to BEB for retransmissions.
iii) The addition of the RTO aging mechanism that is used to update RTO estimators in situations where new RTT measurements are not taken for a long period and therefore the current RTO values become outdated.

3.5 CONTEXT-AWARE CONGESTION CONTROL

In this section, a detailed description of the proposed context-aware congestion control scheme is provided. The CACC model consists of a number of separate but interconnected tasks as shown in Figure 3.2

3.5.1 CACC: Proposed Work

The design of an efficient context-aware congestion control scheme for IoT traffic is paramount. The RTO is a critical factor in the CC mechanism of CoAP. After the RTO has elapsed, the sender considers it a packet loss and initiates the packet retransmission. Previous works that exploit the RTT have set RTO value automatically [141]. However, proper estimation of RTT and assignment of RTO value are difficult during congestion in a burst traffic environment. The main hurdle in the estimation of accurate RTT value is the unavailability of complete context information on the receiver side. It is essential to take two different policies of RTO recalculation into account to estimate the accurate RTT and RTO. Improper selection of the RTO value below the RTT tends to spurious retransmissions and packet losses due to congestion. On the contrary, too large an RTO value compared to RTT tends to unnecessary packet delay.

In order to cope with these issues, the proposed methodology contributes to context-aware RTO estimators along with three other RTO estimators. The network congestion is not the only reason behind the unexpected packet loss and delay, but also the network collision. The primary aim of the three RTO estimators is to differentiate the congestion and the collision dropping scenario to decide the appropriate RTO value. The proposed three RTO estimators decide how well a transmission has performed over time, instead of only relying on the RTO based on the last received acknowledgement packet. Three RTO estimators consider the success, delayed, and failed packets as Strong RTT (RTT_s), Weak RTT (RTT_w), and Failed RTT (RTT_f) respectively. The highly normalised value of the RTT_s and RTT_f over time reflects the scenario of collision dropping, whereas the high contribution of RTT_w and RTT_f reflects the congested environment. Hence, the three RTO estimators efficiently handle the network congestion problem and ensure the trade-off between packet loss and delay without incurring high overheads. The second contribution is a context-aware RTO estimator that takes the RTT variation correctly into account. The RTT variation below the lower bound shrinks the new RTO, whereas too

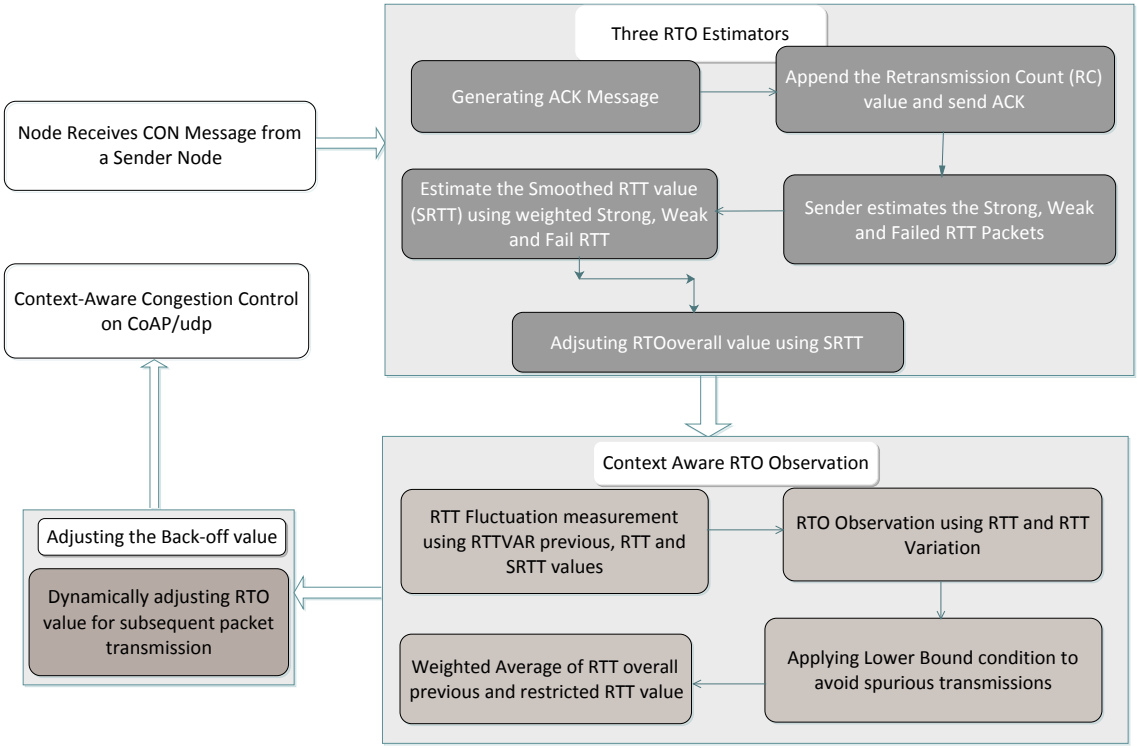

Figure 3.2. Block diagram of the proposed context-aware three RTO estimators, taken from [1], with permission.

high RTT variation tends to unnecessary delay. The context-aware RTO estimator takes into account the RC and lower bound RTT to avoid the excessively shrunk RTO value as well as unnecessary latency. Figure 3.2 depicts the block diagram of the proposed CACC Three RTO estimators over CoAP/UDP in the IoT environment.

3.5.2 Three RTO Estimation Algorithms

Most of the CC mechanisms implemented in CoAP assume that the packet losses are caused only by network congestion. In the IoT traffic, a high packet loss rate is also expected owing to network collision. The packet drop and delay occur when many nodes compete for the network resources, resulting in network congestion at both the node and communication links. The network congestion at the node level tends to packet delay, whereas the link level congestion tends to high data collision. Since the nodes overhear each other‘s radio transmissions in a densely populated area, it results in link level congestion. The link level congestion causes persistent collisions at the link layer and disturbs

the data flow. Congestion at node level produces packet losses only under very stressing network conditions, and therefore the CACC neglects packet drop caused by network congestion. However, congestion delay has a significant impact on the data flow. The CACC mechanism takes into account the collision dropping of link level and node level congestion delay in RTO measurement.

The CACC mechanism runs three RTO estimators which use immediate ACKs of the transmissions, a weak estimator that uses delayed ACKs of transmissions, and a failed estimator that uses no ACKs for the last four transmissions. The RTT_s represents the successful transmission of packets from the client nodes to the server node. Immediate ACKs is used to identify the RTT_s for which the packets have been successfully transmitted. The RTT_w represents a delayed transmission of packets from the client nodes to the server node. The delayed ACKs is used to identify the RTT_w for which the packets have been transmitted within the retransmission. The RTT_f represents the failed transmission of packets from the client nodes to the server node. No ACKs is used to identify the RTT_f for which the same packet has been retransmitted and failed to be delivered for up to four times. The combination of RTT_s and RTT_f represents the successful packet delivery as well as packet dropping. It signifies that there is a chance for link level packet collision, as some of the packets are delivered, and the remaining are dropped. Moreover, the congestion scenario turns the RTT_w value to high owing to the node level congestion delay. According to the principles of the proposed methodology, the RTT and adapted RTO measurements are implemented.

3.5.3 Dynamic SRTT and RTO Overall Estimation

The CC protocol exploits RTO to ensure reliability with minimum retransmissions. The default CC mechanism used in CoAP utilises the Smoothed RTT ($SRTT$) and RTT variation ($RTTVAR$). The CoAP decides on the RTT value for packet transmission according to the maximum latency and processing delay. After having received the first RTT sample, CACC employs the following formulae for RTO calculation, where K is equal to 4. The SRTT is a mean of previous RTTs. RTTVAR represents the variation in consecutive RTT values. In the CoAP protocol, a CON message mandates an ACK from the receiver, and if the ACK is not received, the packet is retransmitted four times by the sender node, before the transmission is considered to have failed.

$$\begin{aligned} SRTT &= RTT \\ RTTVAR &= \frac{RTT}{2} \\ RTO &= SRTT + K \times RTTVAR \end{aligned} \tag{3.3}$$

For subsequent transmissions, the last four retransmissions are considered to take a firm decision on RTO measurement. Equation 3.3 is improved by using the RTO_{over} and $RTTVAR$. The RTO_{over} is an average of the currently calculated $SRTT$ and the RTO_{over} value obtained in the previous step. To avoid a steep RTO increase/decrease after measuring a current RTO value, the RTO_{over} maintains the consistent increment/decrement in RTO value. The RTO_{over} is estimated by using equation 3.4.

$$RTO_{over} = 0.5 \times SRTT_x + 0.5 \times RTO_{over}(prev) \tag{3.4}$$

Where $SRTT_x$ is given in equation 3.5 as ;

$$SRTT_x = \begin{cases} \left[\left((1-\alpha)\frac{RTT_f}{FR}\right) + \left((\alpha)\frac{RTT_s}{SR}\right) \text{if} FR \&\& SR > 0 \&\& FR > SR\right] \\ \left[\left((1-\alpha)\frac{RTT_f}{FR}\right) + \left((\alpha)\frac{RTT_w}{WR}\right) \text{if} FR \&\& WR > 0 || FR < WR\right] \\ \left[\left((1-\alpha)\frac{RTT_s}{SR}\right) + \left((\alpha)\frac{RTT_w}{WR}\right) \text{if} SR > 0 \&\& SR > WR || SR > FR\right] \end{cases} \tag{3.5}$$

In subsequent transmissions, the $SRTT$ is measured by using RTT_s, RTT_w, and RTT_f over the last four (re)transmissions. The RTT_s, RTT_w, and RTT_f represent the RTT in a normal scenario, node level congestion delay, and link level collision dropping respectively. The measurement of $SRTT$ is modified to distinguish the scenarios of congestion and collision. Notably, RTT_s, « RTT_w « RTT_f, and RTT are the summation of RTT_s, RTT_w, and RTT_f over the last four transmissions. Where, FR, SR, and WR represent the frequency of RTT_f, RTT_s, and RTT_w over the last four transmissions respectively. The value of α is always greater than $(1-\alpha)$.

In the first scenario, the contribution of RTT_s, is greater than that of the RTT_f, since such case reflects the collision scenario. If the FR increases more than the SR, more packets are dropped owing to the

link level collision, and some of the packets are received properly on time. In that case, the congestion window is of minimum size since there is no use to increase the congestion window size in the collision environment. In another scenario, the RTT_w, factor contributes to higher value than the RTT_f,. A high value of WR represents the network delay owing to the node level congestion scenario. Therefore, the sender increases RTO at a maximum level. It reduces the impact of network congestion and eventually, the CACC in CoAP considerably improves the performance of IoT traffic. Moreover, the final case represents the normal scenario, where the SR is greater than others. Therefore, there is no need to increase the congestion window size. Where, x is either the combination of RTT_s and RTT_f, or RTT_w and RTT_f. By substituting equation 3.5 into 3.4, the value of RTO_{over} is estimated. It is essential to adapt the RTO to contextual information by using the estimation of three RTO. However, the $SRTT_x$ value is uncertain when there is an ambiguity in receiving an ACK, for which transmission the ACK packet has been received.

3.5.4 Context-Aware RTO by Applying Three RTO Estimator Adaptations

The standard CoAP protocol supports reliable communication using $SRTT_x$ and RTO_{over} based RTO measurement. The RTO measurement forces the receiver node to acknowledge the received data with an ACK packet before the timeout expires. Otherwise, the sender node retransmits the same packet up to four times. The nodes identify all CoAP acknowledgement messages using a Message ID (MID), and it assists the system to detect duplicate packets. In the congested scenario, the packet exchanges with the MID information that has run into retransmission and considers the transmission as the weak estimator. For the second, there is some ambiguity whether a response is a delayed ACK of transmission or immediate ACK of the retransmission. MID information alone is inadequate for the congestion scenarios when the sender plans to make multiple consecutive requests to the receiving node, and it requires modifications to the messages. The use of MID information alone tends to result in two problems when deciding the RTO. First, deciding on too large an RTO compared to RTT tends to unnecessary packet delay. Secondly, improper selection of RTO below the RTT tends to spurious retransmissions and packet losses owing to congestion. The parameter decided on is essential to enforce accurate estimation of RTT and retransmission timeout, RTO_x. Therefore, the proposed work takes into account the RC value, appended to the ACK packet. This provides feasibility to estimate the $SRTT_x$ value more accurately.

The proposed CACC utilises the field of RC to avoid both steep increment and excessive shrunk of the RTO. To append a new field such as RC, the option registry policy described in CoAP is utilized. With this field information, the nodes generate request, response, and ACK messages. The value of RC is assigned as zero when the retransmission does not occur. The RC value is incremented by one for every retransmitted packet. By utilizing the count information, the proposed CACC algorithm avoids unexpected delay and loss, due to proper $SRTT_x$ and RTO measurement. Figure 3.3 compares the transmissions of default CoAP and the proposed scheme. Figure 3.3 (a) shows the default CC mechanism of CoAP, and it shows that node A cannot achieve the correct RTT since there is no information in the ACK packet, for which the CON request was generated. Figure 3.3 (b) illustrates the implementation of CACC in CoAP, and it shows that node A can obtain the correct RTT because the ACK packet carries the RC information. Case 1 considers that the frequency of both the failed and weak RTT is 1. After sending the CON RC = 2, ACK for the first RC is received. Therefore, it is considered as weak RTT, and the first CON request is a failed RTT. However, the second case differs from the first one. First and second CON requests are failed RTT. For the third one, the ACK packet is received. According to the existing CC algorithms, the ACK for retransmission is considered to be the weak RTT.

In contrast, the proposed CACC considers it an immediate response to the CON request and strong RTT. By applying the comparison of transmission time of CON and reception time of ACK with the same RC value, CACC enables the sender node to measure the actual RTT of the packet. Moreover, the $RTTVAR_x$ is estimated with the knowledge of RTT fluctuation. The $RTTVAR_x$ based RTO_x measurement minimises the unnecessary extension, as well as shrinkage, of RTO in subsequent transmissions.

3.5.5 RTT Fluctuation Aware RTO

In CACC, the $SRTT_x$ acts as a low pass filter for subsequent RTT measurements. Therefore, RTT experiences considerable variation. The default CoAP employs RTTVAR as a factor to detect the changes in RTT and prevent it from skewing and steeping changes. To measure the RTO value, the RTT fluctuation over time also plays a central role. The difference between the transmission time of CON and the reception time of ACK with the same RC value is RTT. CACC enables the sender node to measure the actual RTO_x of the packet by using $SRTT_x$, RTT, and $RTTVAR_x$. The $RTTVAR_x$

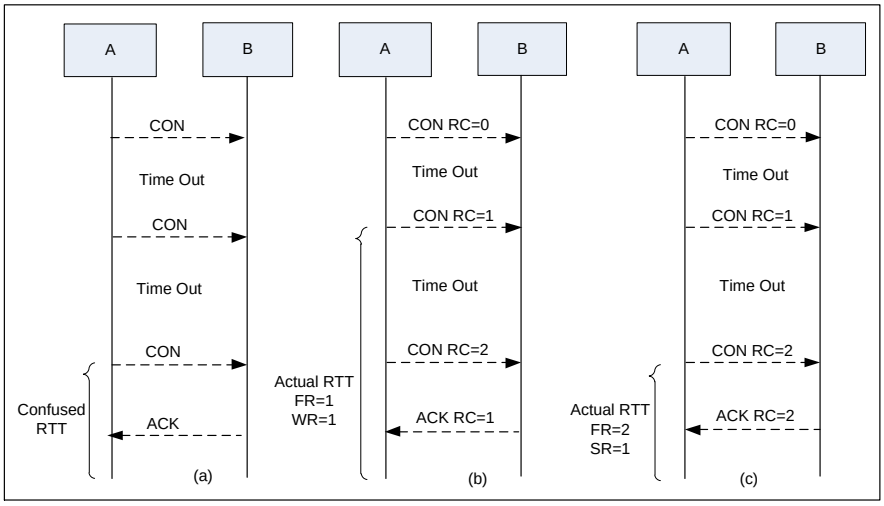

Figure 3.3. Flow of retransmission: A, default CoAP; B, Case 1: and C, Case 2: for three RTO estimator in CoAP, taken from [1], with permission.

keeps the history of RTT variation. Equation 3.6 returns the $RTTVAR_x$ value. The value of β is set as 0.125. In case 1 of Figure 3.3, both the RTT and $SRTT_x$ are nearly equal, and so the current value of $RTTVAR_x$ mostly depends on the past $RTTVAR_x$. In contrast, the second case has less RTT value than that of $SRTT_x$. Therefore, the value of $RTTVAR_x$ decreases.

$$RTTVAR_x = (1-\beta) \times RTTVAR_x(prev) + (\beta) \times (RTT - SRTT_x) \tag{3.6}$$

At the rapid increase of incoming packets, packets are delayed owing to congestion, and the RTT of each packet increases unexpectedly. This results in high $RTTVAR_x$ in a congestion scenario. In contrast, to reduce the impact of collision on congestion window size, a high contribution of strong RTT in $SRTT_x$ measurement degrades the value of $RTTVAR_x$. The RTT fluctuation-aware RTO estimation describes how the maximum value of RTO is limited.

The usage of context-aware measurement of $RTTVAR_x$ and $SRTT_x$ differentiates the congestion and collision scenarios successfully and effectively controls the tuning of RTO to a maximum value. Like the maximum RTO value, the lower bound of RTO also decides the efficiency of the CC mechanism.

Thus, proposed CACC successfully restricts the RTO lower bound value by using the RTO shrinkage scheme.

3.5.6 Restricted RTO Shrinkage

Small RTT value cannot cover the entire range of fluctuation in RTT. The negative variation in RTT shrinks the RTO more than the RTT value, and this scenario leads to spurious retransmission. This in turn leads to high collision and congestion. Therefore, the proposed methodology appends the condition of $max(-L, K \times RTTVAR_x)$ in restricted RTO (RTO_{rest}) measurement, as shown in equation 3.7. L is the lower bound of the RTT, and it is recommended to select the L value not greater than 10 s initially. In subsequent RTO measurement, the L value is increased by 10, if the negative variation in RTT increases. Otherwise, the L value is kept constant. The value of L is reset to 10 s when the positive variation in RTT is experienced. Therefore, The CACC algorithm decides on the minimum level of RTO shrinkage as well as spurious retransmissions of the CON message. Figure 3.4 shows the impact of lower bound RTO on the CC mechanism. Figure 3.4(a) depicts the problem of the CC mechanism with negative RTTVAR.

$$RTO_{rest}(x) = SRTT + \max(-L, K \times RTTVAR_x) \tag{3.7}$$

The value of K is always set as 4, and the RTO_x measurement employs the SRTT value. It significantly reduces the impact of RTTVAR on the weak RTO estimation. Owing to the minimal RTT value, the RTTVAR is reduced drastically in Figure 3.4(a). This causes a steep reduction in RTO and incurs spurious retransmissions. However, the proposed CACC approach reduces L value from RTT, and so the sender node receives the ACK message before its time expires. This process avoids the unnecessary packet retransmissions and delay.

3.5.7 CoAP Integration with the Context-Aware CC Scheme

To measure the RTT value for every retransmission accurately, the CACC perfectly balances RTO value between the $SRTT_x$ based RTO_{over} and $RTTVAR_x$ based RTO_{rest}. This scheme updates the RTO value

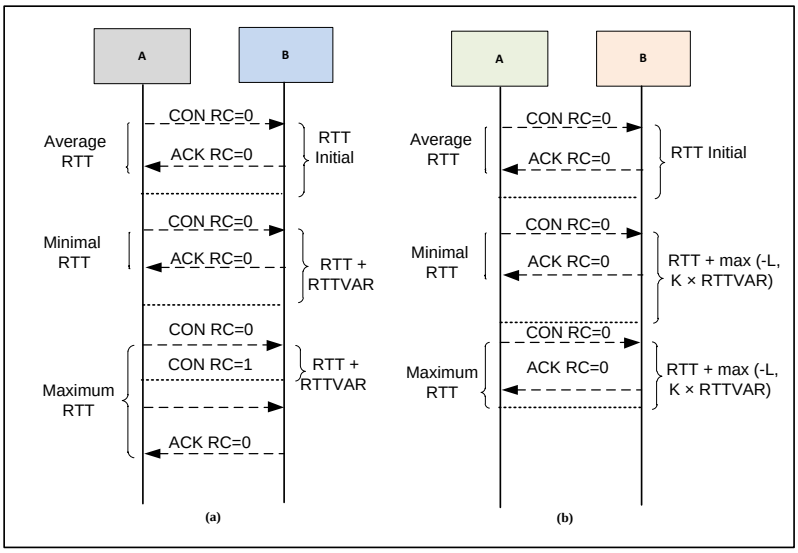

Figure 3.4. Impact of lower bound RTO on congestion control scheme, taken from [1], with permission.

to improve the packet transmission efficiency of CoAP. For subsequent transmissions, the following formula is applied to update the RTO value.

$$RTO_{new}(x) = 0.5 \times (RTO_{over} + 0.5 \times RTO_{rest}(x) \tag{3.8}$$

The aging concept is executed, when the estimated RTO values are not updated for an extended period. Minimum and maximum values of $RTO_{new}(x)$ should not be valid for a longer time. Considering the IoT environmental factors, the RTT changes fast. To avoid false RTO values due to such changes that are likely to occur, the proposed CACC scheme applies an aging mechanism to small and large RTO estimations based on CoCoA+ mechanism [140]. If an RTO estimation is small (below 1 s) or large (above 3 s), and also no new RTT measurement is performed for 16 times, the RTO value is assigned as the initial default value. Eventually, the context-aware CC in CoAP/UDP considerably improves the performance of IoT traffic by using three RTO estimators, lower-bound restriction on RTO value, RTT fluctuation-handling scheme, and aging concept.

3.6 ANALYSIS OF CONTEXT-AWARE CC ALGORITHM

The Transmission Rate (TR) defines the rate of transmitting the data packets from a client to the server. It is an important metric to assess the quality of the proposed CC mechanism in CoAP. In normal network conditions, the transmission rate depends on the RTO. However, in a congested network, the RTO plays a central role in maximizing the network performance. Let us consider that the total time taken by a client to transmit the data packets is denoted as T_{mp}. If the network congestion takes place during T_{mp}, then the relationship existing between the time interval for normal transmission T_{mn}, and the retransmission time due to loss of packet, T_{mrt}, decides on the protocol performance. Assuming that S_p, S_n, and S_{rt} represent the total number of packets sent during T_{mp}, T_{mn}, and T_{mrt}, respectively. Therefore;

$$T_{mp} = T_{mn} + T_{mrt} \tag{3.9}$$

$$S_p = S_n + S_{rt} \tag{3.10}$$

From equation 3.10, the summation of a total number of packets transmitted during normal network conditions and also transmitted in congested network conditions denotes the S_p. The data transmission time during normal network conditions is measured by using the average RTT, RTT_{avg}, and the number of packets, i. Therefore, the TR value is as follows:

$$TR = \frac{S_n T_{mrt}}{T_{mn} S_{rt}} \tag{3.11}$$

$$S_n = \frac{T_{mn}}{RTT_{avg}} \tag{3.12}$$

The network congestion causes continual retransmission timeouts. Considering that P_s denotes the probability of successful transmission between the client and server and k represents the number of retransmissions. Probability is analyzed as a value between 0 and 1 in all cases of transmissions . The probability of packet transmission failure (P_f) until the packets are successfully transmitted to the server is modelled as follows:

$$P_f = \frac{(1-P_s)^k}{P_s} \tag{3.13}$$

The expected value of successfully retransmitted packets S_{rt} is estimated by using equation 3.14.

$$E[S_{rt}] = \frac{(1-P_f)}{P_f} \tag{3.14}$$

The CACC sets the default RTO limit to 4. During retransmission, CoAP keeps doubling the initial RTO (RTO_{init}). However, the CACC increases the RTO when the possibility of network congestion is high. Otherwise, it retransmits the packet without increasing the RTO. Therefore, the waiting time for k packet retransmissions, RTO for CACC, is as follows:

$$RTO_{cacc} = RTO_{init} + \sum_{k=1}^{4} (RTO_{new})_k \tag{3.15}$$

When the IoT environment does not experience the network collision, increasing the value of RTO is meaningful, that means the only cause of packet dropping is network congestion. Otherwise, it wastes network resources. Therefore, the proposed CACC algorithm decides on the RTO value based on the network conditions by using strong, weak, and failed RTT measurements. Moreover, the usage of the RC factor assists the CACC algorithms to balance the RTO and protocol performance effectively in terms of delay and resource utilisation. The expected packet waiting time during congestion, $E[T_{rt}]$, is estimated by using equation 3.15 and 3.13.

$$E[T_{rt}] = \left(RTO_{init} + \sum_{k=1}^{4} (RTO_{new})_k \right) \times P_f \tag{3.16}$$

Therefore, the transmission rate of CACC is determined by using equation 3.16 and is 3.14, as follows:

$$TR = \frac{\left(P_f^2 [RTO_{init} + \sum_{k=1}^{4} (RTO_{new})_k] \right)}{RTT_{avg}(1 - P_f)} \tag{3.17}$$

This shows that the waiting time and packet retransmission time as a result of packet loss T_{mrt} rely on the network collision and RTO estimator in the proposed CACC scheme. Moreover, the total number of packets sent during T_{mrt} is maximally delivered successfully without reaching the maximum k value.

Furthermore, in default CoAP, the waiting time for k packet retransmissions, RTO is estimated by using equation 3.18.

$$RTO_{coap} = RTO_{init} + \sum_{k=1}^{4} 2^k (RTO_{init}) \tag{3.18}$$

In congested and collision IoT environments, the difference of RTO between the CoAP and CACC is great, since there is no need to increase the RTO in a collision environment as the CACC provides optimal RTO measurements in a congested IoT environment.

$$RTO_{diff} = RTO_{init} + \sum_{k=1}^{4} 2^k (RTO_{init}) - RTO_{init} + \sum_{k=1}^{4} (RTO_{new})_k \tag{3.19}$$

In the worst case, considering that the client node attempts four retransmissions to deliver data to the server, CoAP increases the RTO_{CoAP} nearly 2^{10} times RTO_{init}. However, in the same environment, CACC increases RTO more than four times RTO_{new}. Notably, the RTO_{new} varies slightly from the RTO_{init}. In such a case, equation (16) is rewritten as follows.

$$\left(RTO_{coap} \approx 2^{10} \times (RTO_{init})\right) >> (RTO_{cacc} \approx 2^{2} \times (RTO_{new}) \tag{3.20}$$

When the server delays sending the ACK packets, the CoAP fails to match the ACK packet with the corresponding data retransmission. However, with the knowledge of RC, CACC avoids both steep increment and excessive shrunk of the RTO. Even in the worst case, the CACC improves performance in both the CC and in packet delay.

3.7 CHAPTER SUMMARY

Congestion control remains a critical issue that demands attention for reliable communication in networks of resource-constrained nodes for efficient resource utilisation and for optimal network performance by the research community. In this Chapter, a context-aware congestion control scheme has been presented along with its system analysis for evaluation against the baseline CoAP congestion control specification and the Advanced congestion control mechanism (CoCoA+) as benchmark protocols for validation.

CHAPTER 4 PERFORMANCE EVALUATION AND DISCUSSION

4.1 CHAPTER OVERVIEW

This Chapter evaluates and discusses the network performance of implementing such a system model for congestion control (CC) in resource-constrained networks whose core components have been presented and discussed earlier. To validate the performance of this scheme, the proposed CACC is compared to that of a baseline CoAP CC specification and that of an advanced congestion control mechanism for CoAP (CoCoA+) as benchmark protocols. The aim of this evaluation is to present an efficient resource utilisation and derive optimal results of the proposed scheme for congestion control. The basis of this evaluation takes into consideration the performance metrics, including throughput, packet loss, delay and energy consumption, of the entire network to achieve results based on the proposed CACC to analyse the performance of the proposed system model. The results include evaluation involving three network configurations, grid, chain, and dumbbell topology. Within the scope of this research work, the validation of the proposed CACC scheme is solely based on simulation using Contiki OS and the Cooja simulator only against the benchmark protocols. The actual experimentation in this study is presented as part of future work owing to the unavailability of access to an IoT-test bed for congestion control at the time of this research work. As contribution to the body of knowledge, the work presented in this Chapter has been published as a journal article titled "CACC: Context-aware congestion control approach for lightweight CoAP/UDP-based Internet of Things traffic" [1].

The outline of the rest of this Chapter is as follows: The system implementation strategy, including the implementation tool, simulation setup, network topology, and CC performance metrics, is presented in

Section 4.2. In Section 4.3, the CACC results for validation against the benchmark protocol of several quality of service performance metrics from simulations are presented, while Section 4.4 discusses the results presented in detail, with subsections 6.4.2, 4.4.2, and 4.4.3, represent the discussion of results based on grid, chain, and dumbbell network topologies respectively. Section 4.5 summarises the Chapter.

4.2 IMPLEMENTATION TOOL

Cooja is a java-based Contiki operating system (OS) [153] simulation tool as an open source for both traditional wireless sensor networks and for the IoT, to connect tiny low-cost, low-power micro-controllers, and resource-constrained devices to the Internet and to implement most sensor motes. Figure 4.1 presents a comparison of the IETF communication protocol stack on the left against the Contiki implementation platform on the right, along with its specific layers. For the provision of IPv6 networking functionality, Contiki implements the uIPv6 stack. With the uIPv6, the Contiki-RPL enables the implementation of the routing protocol for low-power lossy IPv6 networks, as well as SICSlowpan, for the implementation of a 6LoWPAN adaptation layer [137]. The Erbium implementation of CoAP [135] also comes as part of the Contiki, in which CoAP is located at the top of the uIPv6 of the application layer [137]. To enable the routing of packets, formation of network, and for maintaining the network topology, RPL which is part of the uIPv6 stack, plays a major role for these functionalities. This platform enables the execution of Contiki OS-based code and a simulation output that can be collected, processed, and can obtain the performance metrics. This network simulator supports the simulation of sensor networks at three different levels, which are: the machine code instruction set, the application level and the operating system [154]. This platform enables Cooja motes to emulate an off-the-shelf wireless sensor node that takes into consideration the capability functions of real nodes in the simulation environment. The Cooja simulation tool, including the simulation script editor, the control panel and the serial socket (server) port connection plugin, are shown in Figure 4.2.

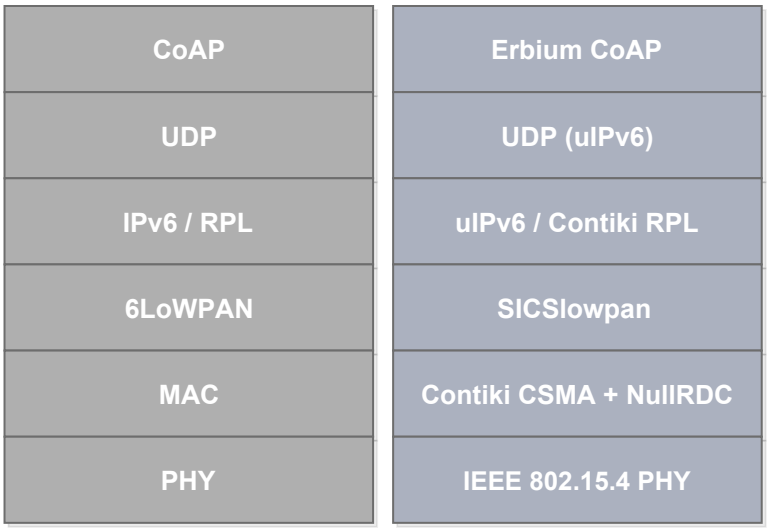

Figure 4.1. The default IETF communication protocol stack against its implementation in Contiki platform, taken from [1], with permission.

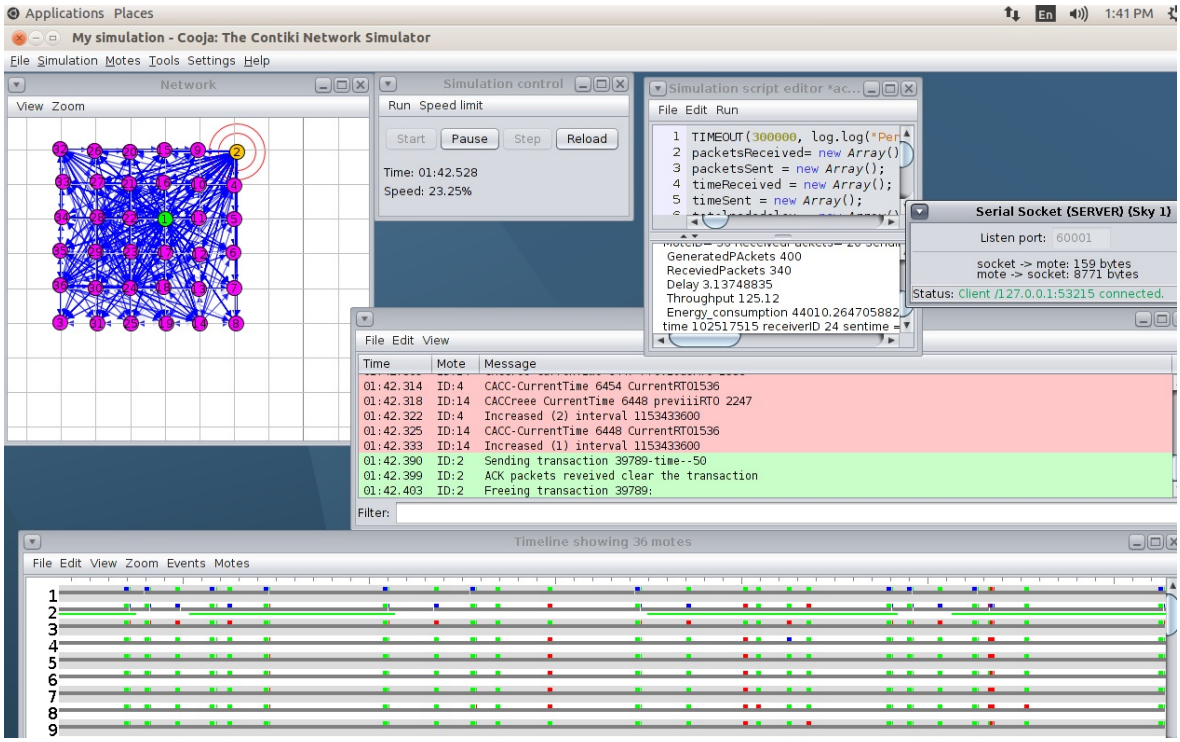

Figure 4.2. Contiki Cooja simulation environment.

4.2.1 Simulation Setup

With the Cooja simulator [154], two of the IEEE 802.15.4 capable motes with the same radio transceiver (CC2420) [155] have been carefully chosen for this simulation: the Tmote Sky from Moteiv [156], which is a hardware equivalent of the TelosB motes and the Z1 mote from Zolertia. The relevant specifications for the different motes used are defined in Table 4.1 which shows their unique capability features. The table shows the different RAM and ROM capabilities, and the MCU for the motes.

Table 4.1. Hardware Simulation Parameters for the Zolertia Z1 and Moteiv Tmote Sky Wireless Sensor Nodes, taken from [1], with permission.

Features	**Z1**	**Tmote Sky**
RAM	8 KB	10 KB
ROM	92 KB	48 KB
MCU	MSP430F1611	MSP430F2617
RADIO	CC2420	CC2420

Network topology: To evaluate the performance of CoAP CC mechanisms, we consider three different network topologies for the simulations. In each deployment, the number of nodes and their positioning differ, which will also be apparent in certain differences, such as in the number of direct neighbours of the nodes, the distance of routes between the source CoAP request and the destination, as well as the number of nodes to compete simultaneously for the radio channel. The network topology used to evaluate the performance analysis for the constant traffic scenario comprises: i) a grid of 36 nodes $(6x6)$, ii) a chain of 22 nodes and iii) a dumbbell topology with 20 nodes. Figure 4.3 represents a two-dimensional positioning for the various nodes in the three-network topology of the Cooja simulator Graphical User Interface (GUI). The communication range of the nodes is set to $50m$, with an interference range of $100m$. For radio transmissions, the simulation exploits a Unit Disk Graph Medium (UDGM) radio model with circular transmission and applies interference areas. When the UDGM radio model is used, a Static Link Delivery Ratio (LDR) setting is applied. Moreover, Radio Duty Cycling (RDC) of the MAC layer is turned off. RPL and UDP protocols are used in the network and transport layers respectively.

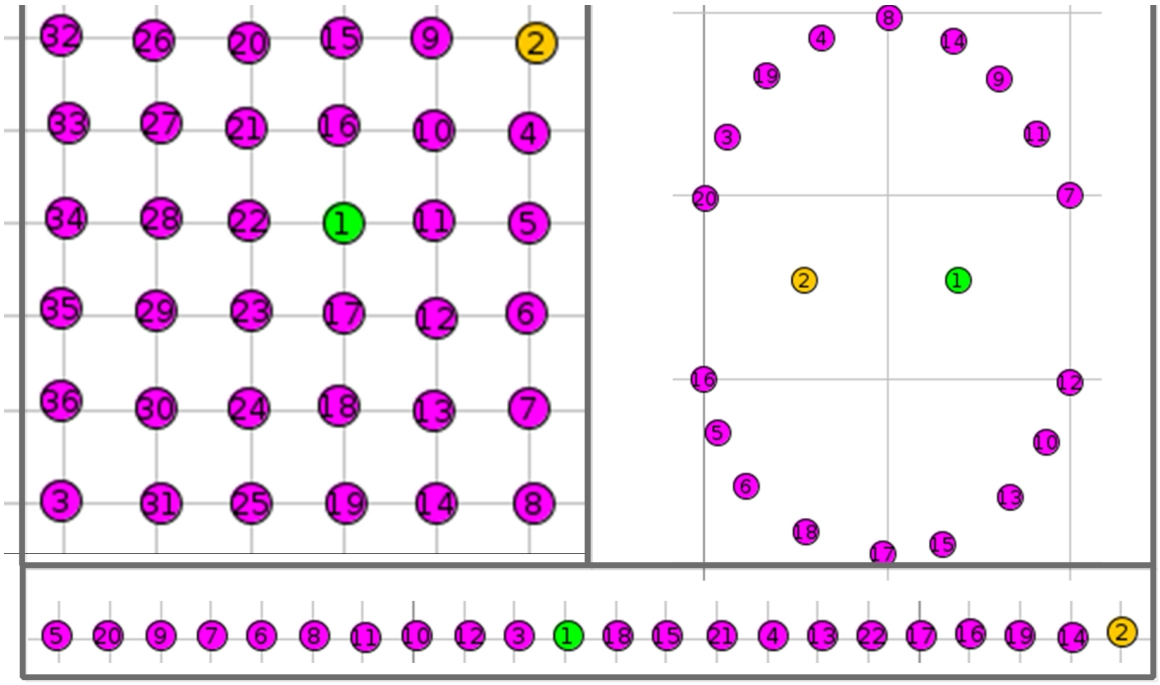

Figure 4.3. Starting from the upper left and going clockwise shows the three-network topology considered for the performance analysis (grid, dumbbell, and chain). 1, 2 represent RPL border routers and sink node respectively, and all other motes in the network topologies represent nodes for running the full Contiki OS stack and the Erbium implementation of CoAP taken from [1], with permission.

The CoAP messages are created by each CoAP node which periodically sends data at certain intervals towards a sink node in all the topology. In order to generate traffic of different levels, the periodic interval is varied from 1 *s* to 30 *s*. Periodic transmissions of messages from multiple nodes are assigned to a sink node to simulate the congested situation together with the collision network scenario. The duration of simulation for which the CoAP requests are generated is set to 300 *s*. After this time should have elapsed, all nodes stop spawning further CoAP requests, and subsequently, this leads to termination of the simulation phase. The following performance metrics have been evaluated for the various-network topology for the constant traffic scenario.

1. The first performance metric considered to evaluate the simulation results is system throughput. We define the throughput as the total number of delivered bits per second in a given time interval.
2. The second performance metric is the end-to-end delay of CoAP messages. The delay is the average time taken for a CoAP message to reach its final destination node from the moment it is sent at the application layer of the source node.

3. The third performance metric chosen for the analysis of the CC mechanism for CoAP is the message loss, which indicates the total number of undelivered data messages to the server.
4. And finally, the energy efficiency based on the amount of energy in joules consumed by all the nodes in the network.

4.3 SIMULATION RESULTS

To evaluate the performance of the congestion control scheme, a series of simulations s set up for the three different network topologies. This section presents the results of the evaluation for the congestion control mechanism in CoAP for a constant traffic scenario for the three different network topologies for varying data transmission intervals to analyse the performance of the congestion-control scheme for 300 *s* simulations running at 200 % speed, each with 36, 22, and 20 nodes for grid, chain, and dumbbell network topology respectively. Results are obtained while running baseline CoAP, CoCoA+, and also the proposed CACC. Then these results are evaluated by comparing them. The results obtained while considering the performance metrics for the various network topologies are presented in Figures 4.4, 4.5, 4.6, and 4.7 for grid network topology while Figures 4.8, 4.9, 4.10, and 4.11 represent chain network topology and Figures 4.12, 4.13, 4.14, and 4.15 for dumbbell network topology.

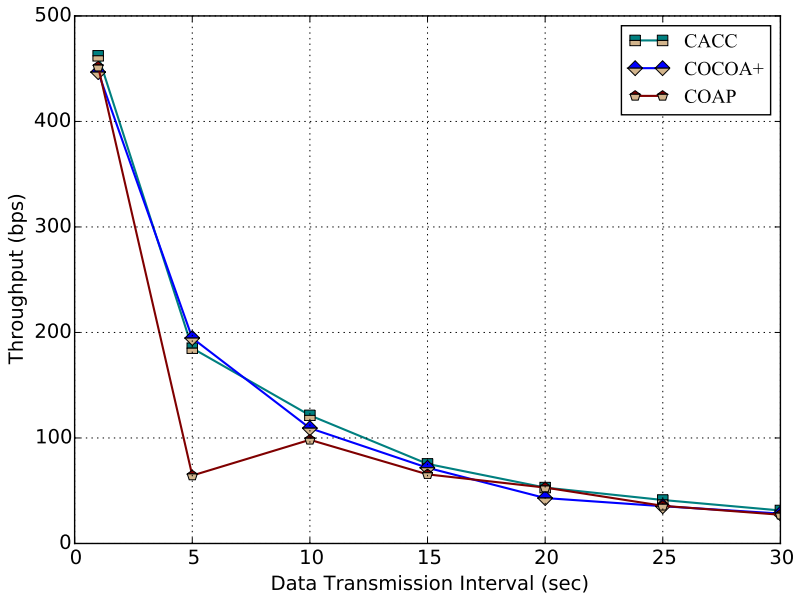

Figure 4.4. Throughput performance for data transmission interval in grid network topology, taken from [1], with permission.

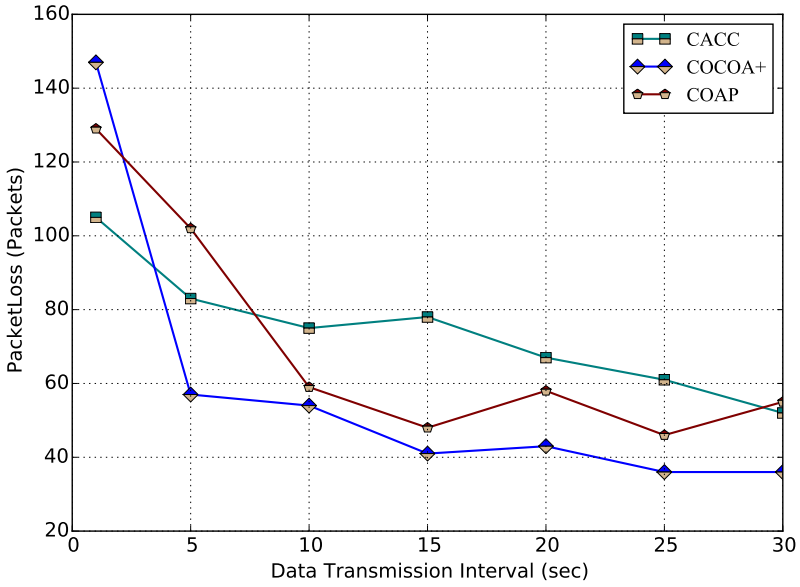

Figure 4.5. Packet loss performance for data transmission interval in grid network topology, taken from [1], with permission.

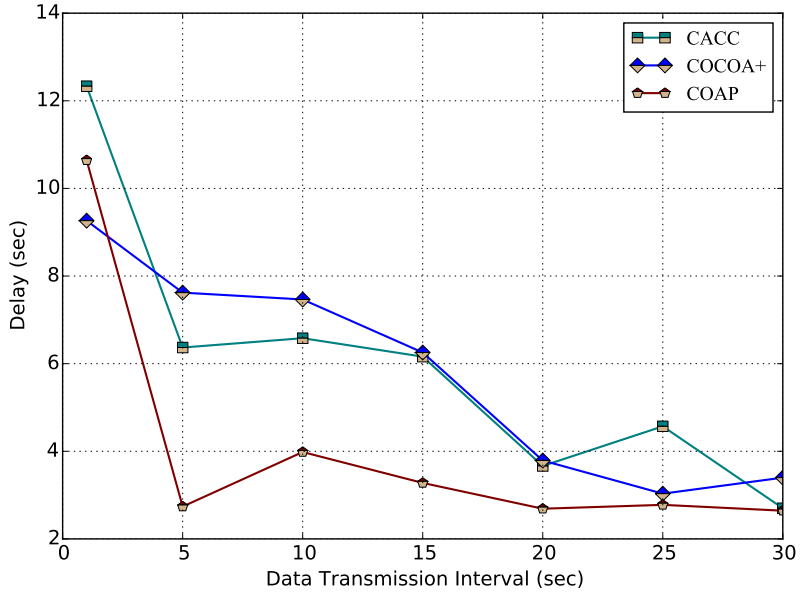

Figure 4.6. Delay performance for data transmission interval in grid network topology, taken from [1], with permission.

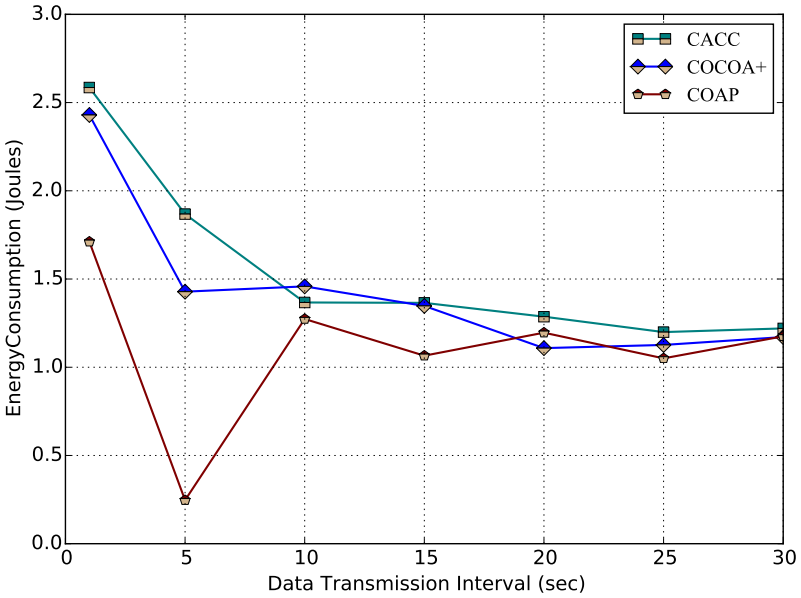

Figure 4.7. Energy consumption performance for data transmission interval in grid network topology, taken from [1], with permission.

Table 4.2. Summary of generated values for grid network topology, taken from [1], with permission.

DTx	**Default CoAP (%)**				**CoCoA+ (%)**				**CACC (%)**			
	TP	LPkts	D	EC	TP	LPkts	D	EC	TP	LPkts	D	EC
1	451.89	129	10.643	1.711	446.75	147	9.264	2.429	462.21	105	12.3326	2.585
5	64.4	102	2.738	0.246	194.67	57	7.621	1.428	185.10	83	6.368	1.868
10	98.2	138	3.98	1.27	109.2	108	7.46	1.45	121.4	75	6.58	1.367
15	65.5	105	3.27	1.065	71.76	88	6.25	1.34	75.4	78	6.15	1.364
20	52.9	67	2.68	1.19	43.05	94	3.78	1.10	52.9	94	3.78	1.28
25	35.69	76	2.77	1.05	35.3	77	3.033	1.126	41.26	61	4.56	1.19
30	27.23	63	2.64	1.17	28.3	60	3.39	1.17	31.2	52	2.70	1.22

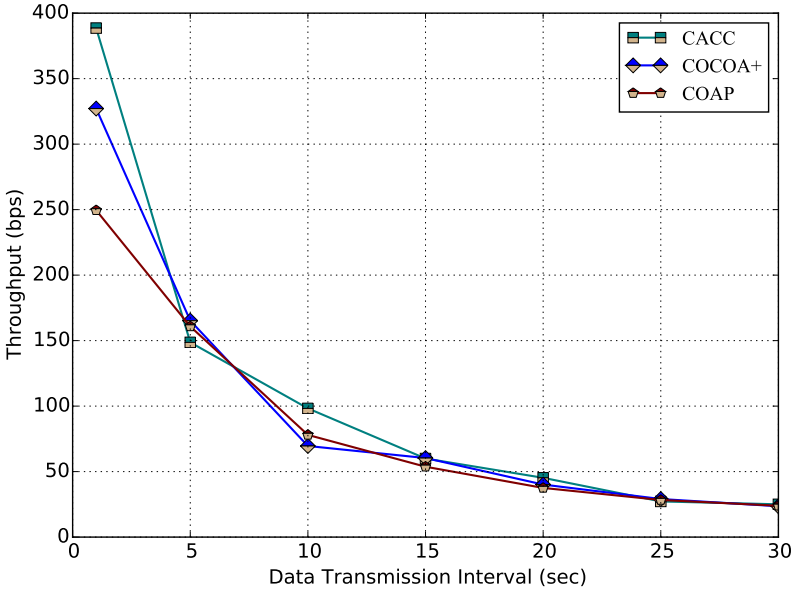

Figure 4.8. Throughput performance for data transmission interval in chain network topology, taken from [1], with permission.

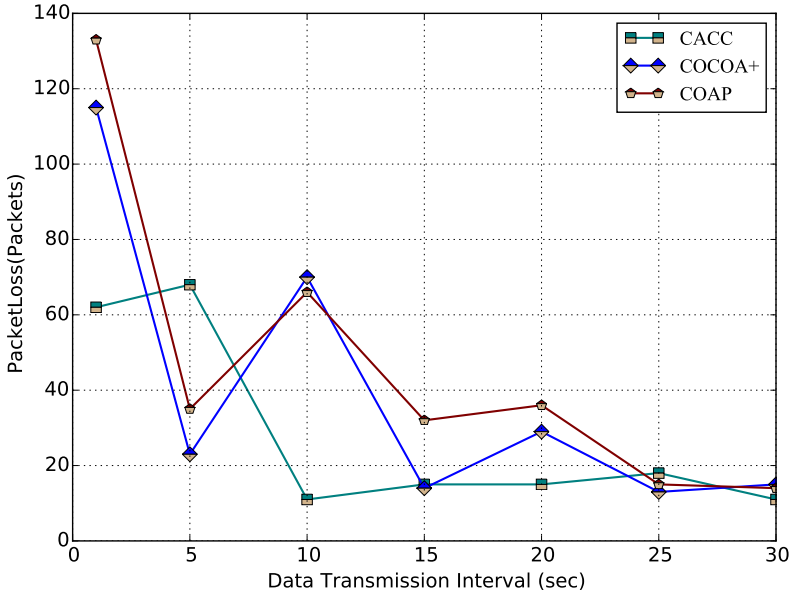

Figure 4.9. Packet loss performance for data transmission interval in chain network topology, taken from [1], with permission.

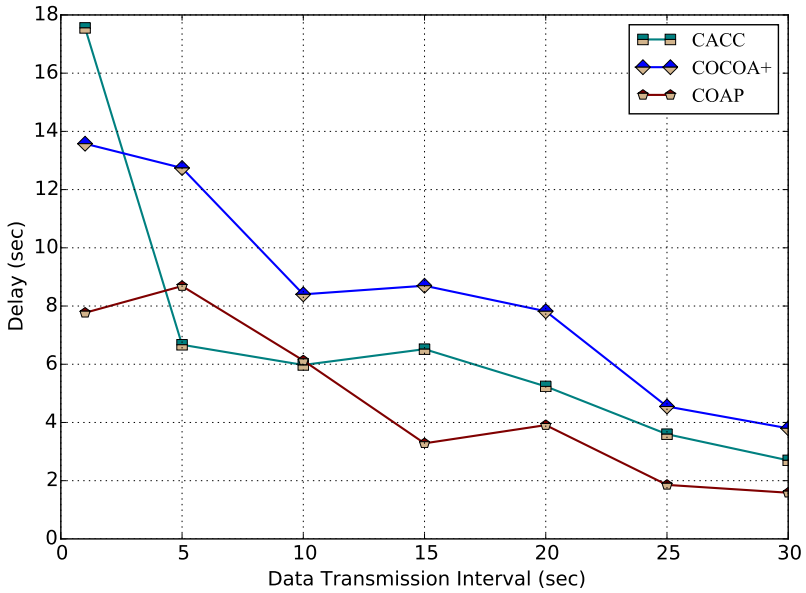

Figure 4.10. Delay performance for data transmission interval in chain network topology, taken from [1], with permission.

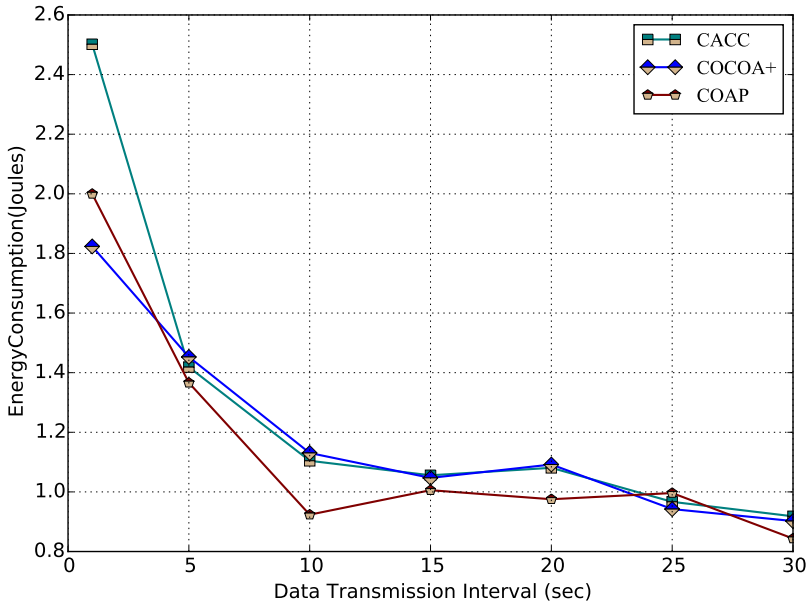

Figure 4.11. Energy consumption performance for data transmission interval in chain network topology, taken from [1], with permission.

Table 4.3. Summary of generated values for chain network topology, taken from [1], with permission.

DTx	Default CoAP (%)				CoCoA+ (%)				CACC (%)			
	TP	LPkts	D	EC	TP	LPkts	D	EC	TP	LPkts	D	EC
1	249.50	440	7.768	1.999	327.15	229	13.576	1.823	388.60	62	17.554	2.501
5	160.81	35	8.685	1.366	165.23	23	12.742	1.452	148.67	68	6.669	1.418
10	78.01	66	6.12	0.92	69.55	70	8.4	1.13	98.25	11	5.97	1.10
15	53.72	32	3.28	1.00	60.32	14	8.69	1.04	59.98	15	6.51	1.05
20	37.53	36	3.90	0.97	40.11	29	7.82	1.09	45.2	15	5.23	1.08
25	28.33	15	1.85	0.99	29.07	13	4.54	0.94	27.33	18	3.59	0.96
30	23.92	14	1.58	0.84	23.55	15	3.79	0.9	25.02	11	2.69	0.91

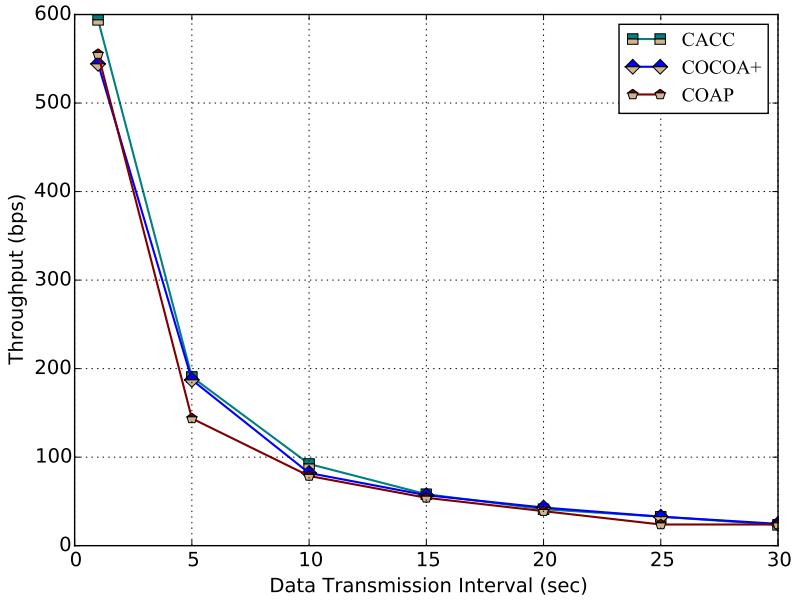

Figure 4.12. Throughput performance for data transmission interval in dumbbell network topology, taken from [1], with permission.

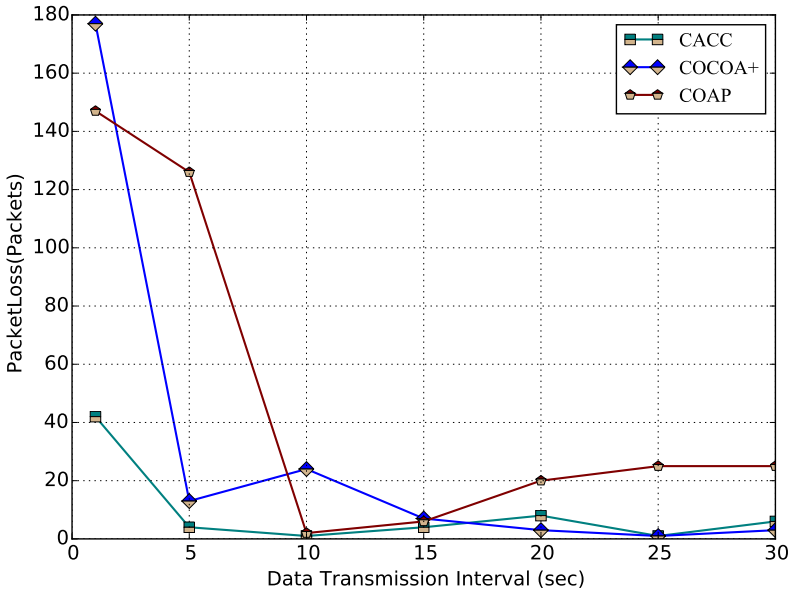

Figure 4.13. Packet loss performance for data transmission interval in dumbbell network topology, taken from [1], with permission.

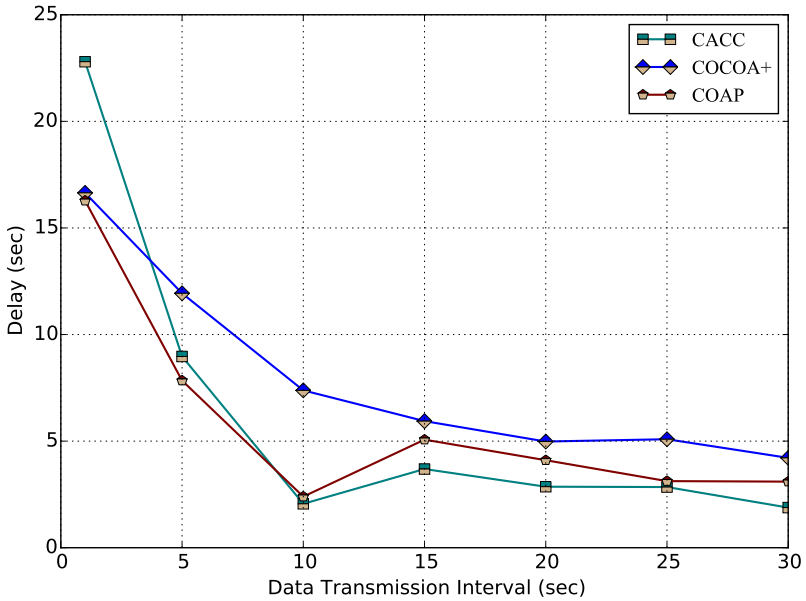

Figure 4.14. Delay performance for data transmission interval in dumbbell network topology, taken from [1], with permission.

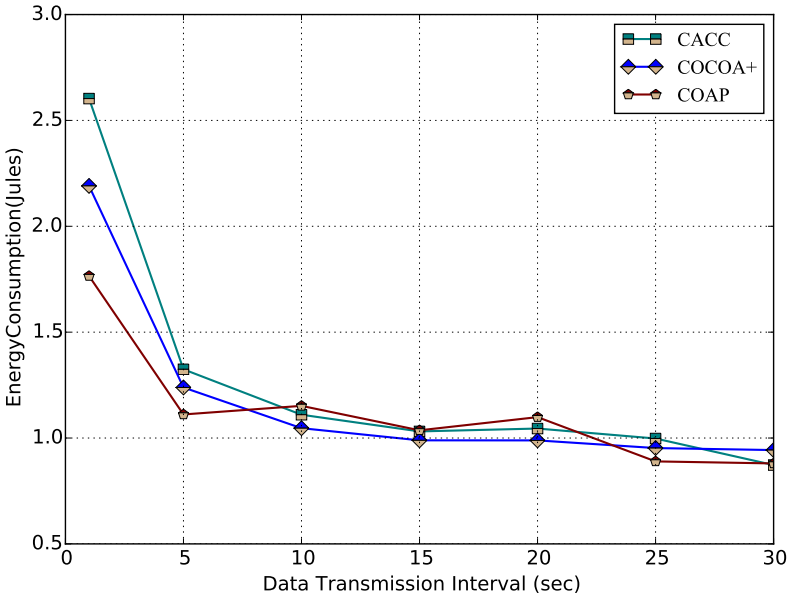

Figure 4.15. Energy consumption performance for data transmission interval in dumbbell network topology, taken from [1], with permission.

Table 4.4. Summary of generated values for dumbbell network topology, taken from [1], with permission.

DTx	**Default CoAP (%)**				**CoCoA+ (%)**				**CACC (%)**			
	TP	LPkts	D	EC	TP	LPkts	D	EC	TP	LPkts	D	EC
1	555.31	147	16.278	1.765	544.27	177	16.646	2.190	593.95	42	22.798	2.603
5	145.72	126	7.841	1.111	187.31	13	11.934	1.238	190.62	4	8.971	1.324
10	78.75	2	2.39	1.15	82.04	24	7.38	1.04	92.3	1	2.06	1.11
15	54.09	6	5.06	1.03	57.04	7	5.93	0.989	58.14	4	3.69	1.03
20	39	20	4.10	1.09	43.05	3	4.98	0.988	41.21	8	2.86	1.04
25	23.92	25	3.1	0.88	32.75	1	5.08	0.95	32.75	1	2.84	0.99
30	23.92	25	3.1	0.88	24.65	3	4.21	0.94	23.55	6	1.88	0.87

4.4 DISCUSSION

This section presents a discussion of the results obtained for evaluation of the different network scenarios during varying data transmission interval. Subsection 4.4.1 discusses the results obtained for grid network topology while Subsection 4.4.2 presents the discussion of the results obtained for chain

network topology and this is followed by the discussion of results for dumbbell network topology in Subsection 4.4.3.

4.4.1 Results based on periodic message transmission for grid network topology

In the grid topology, 36 clients are positioned to generate the periodic transmissions to compare the performance of the schemes. With reduced network traffic, the possibility of a network collision is less, and most of the packet loss is caused by network congestion only. The CACC outperforms CoCoA+ and base CoAP during various data transmission intervals as shown in Figure 4.4. The CACC delivers most of the data packets within a minimum number of retransmissions during high traffic on the network. As the data transmission interval increases from 1 *s* to 30 *s*, the network throughput starts to degrade due to less network traffic. The proposed CACC drops its throughput from 462.21 *bps* to 31.2 *bps*. The CoCoA+ performs better than or similar to the CoAP. However, to ensure low packet loss and less retransmission, the energy consumption and delay of CACC increases marginally with improved packet delivery compared to that of other baseline protocols.

Figure 4.5 shows a minimal packet loss for CACC compared to the existing works. For instance, the CACC, CoCoA+, and CoAP experience loss in 105, 147, and 129 packets respectively during high traffic on the network of 1 *s* data transmission interval. The proposed CACC avoids both steep increment and excessive shrunk of the RTO which avoids excessive retransmission in the network by utilising the field of RC in RTO measurement. Without incorporation of the RC factor, CoCoA+ and base CoAP protocols may fail to provide the exact status of the packet retransmission and this results in unnecessary packet loss.

Figure 4.6 shows delay performance for the propose and existing schemes. In a high-traffic environment, the number of successfully delivered packets increases the delay in the CACC mechanism. For instance, on a highly congested network with a great number of generated packets, the CACC, CoCoA+, and base CoAP successfully transmit 1256, 1214, and 1232 packets respectively. The CACC delivers in 12.332 *s*, CoCoA+ in 9.264 *s*, while the base CoAP delivers in 10.643 *s* for each successful transmitted packet. With the three estimators based on RTO measurement along with the overall estimator, the CACC updates RTO only in the congested network, but not in the collision environment. In a low-traffic environment, the CACC experiences more delay than CoCoA+. For instance, in a data transmission

interval of 25 *s*, the difference between CACC and CoCoA+ is 1.52 *s*. With improved packet delivery in terms of throughput, low packet loss and retransmission, the CACC compromises the delay, compared to other baseline protocols.

Figure 4.7 shows the level of energy consumption for the various schemes for the overall network. The proposed CACC achieves higher packet delivery with a compromise of energy consumption. It mainly targets the packet delivery factor, even when there is high congestion. For instance, with a data transmission interval of 1 *s*, the energy consumption of CACC is 2.585 Joules with successful transmission of 1256 packets, whereas the CoCoA+ utilises 2.429 Joules for the successful transmission of 1214 packets and the base CoAP uses 1.711 Joules for the successful transmission of 1232 packets. This shows clearly that the proposed CACC expends more energy in order to deliver the packets successfully with a minimal number of retransmissions.

Table 4.2 presents a summary of the generated values based on the various data transmission intervals for grid network topology.

4.4.2 Results based on periodic message transmission for chain network topology

The number of nodes of packet transmission increases the possibility of network congestion as well as collision, as analysed by varying the data transmission interval over the chain topology. The context-aware three RTO estimators improve the throughput in the CACC scheme over the existing baseline protocols as shown in Figure 4.8. When the packet interval is longer than 15 *s*, network congestion occurs with less probability only. In such a case, most of the communication is successful. Some of the data packets are retransmitted with CoCoA+ and base CoAP due to the unavailability of context-aware RTO measurement. In a high-traffic environment, CoCoA+ fails in differentiating the collision environment from the congested network, and this tends to fluctuate RTO significantly. This causes frequent retransmission and poor network throughput in CoCoA+. For a data transmission interval of 1 *s* at high levels of network traffic, CACC attains 388.60 *bps* throughput, whereas the CoCoA+ and base CoAP attain 327.15 *bps* and 249.50 *bps* respectively. As the data transmission interval increases, the throughput for various network topologies degrades significantly. The exponential increment of RTO in CoAP causes no retransmission until the RTO expires and leads to network under-utilisation and thereby reduces the network throughput.

Figure 4.9 represents the packet loss for the various mechanisms. The message interval is directly proportional to the network traffic and congestion. CoCoA+ increases the RTO value unnecessarily because of unawareness of network collision, and this may lead to RTO shrinkage and spurious retransmission to the server. The proposed CACC appends the L factor in RTO_{rest} measurement and prevents too much shrinkage in RTO, which improves the network throughput and reduces packet loss. CACC reduces the packet loss rate tremendously when compared to CoCoA+ and base CoAP. The proposed CACC mechanism detects the weak RTT in the request-response interaction model and estimates the smoothed RTT value accurately by incorporating the RC information to balance the fluctuation in IoT traffic effectively and to improve the packet delivery ratio considerably. For a data transmission interval of 1 s during high traffic on the network, base CoAP attains a total throughput of 249.50 bps, while CoCoA+ achieves a total of 327.15 bps and the proposed CACC attains 388.60 bps. As the data transmission interval increases from 1 s to 30 s, the system throughput degrades significantly.

Figure 4.10 shows delayed performance for the proposed and existing baseline protocols. The CACC outperforms CoCoA+, where the congested network environment starts to affect the protocol performance. As with the network traffic, improved performance in throughput is achieved by the CACC when compared to the CoCoA+ and base CoAP. The network delay in the proposed CACC increases with the number of successfully transmitted packets. CoCoA+ delivers the packets with multiple retransmissions and this results in much delay. The three RTO estimators make the CACC quite effective in improving the network performance. For instance, on a highly congested network of 1 s data transmission interval, the CACC and CoCoA+ successfully transmit 1056 and 889 packets in 17.554 s and 13.576 s respectively. However, the base CoAP successfully delivers only 678 packets using 7.768 s.

Figure 4.11 shows the results of the energy consumption level for CACC, CoCoA+, and base CoAP for different message intervals. As observed, since the proposed CACC delivers a considerable number of successfully transmitted packets due to the adjustment of RTO with context awareness, the energy consumption of CACC is considerable to ensure low packet loss and retransmission. However, the CoCoA+ attains less throughput, and at the same time consumes more energy when compared to the proposed CACC. For instance, in a 10 s data transmission interval, the CACC has a total energy consumption of 1.10 joules. However, CoCoA+ and base CoAP consume 1.13 joules and 0.92 joules respectively. Although, CoAP consumes less energy compared to other protocols, it delivers only a

few messages to the server.

Table 4.3 presents a summary of generated values based on the various data transmission intervals for chain network topology.

4.4.3 Results based on periodic message transmission for dumbbell network topology

Figure 4.12 shows the influence of the data transmission interval on the performance of CC mechanisms. Compared to other protocols, CACC demonstrates its performance by a considerable number of retransmissions, especially in a high traffic scenario. With high network traffic, CACC adapts to the network conditions and estimates the actual RTO value by incorporating the RC factor when determining the correct RTT value. While the data transmission interval increases, the throughput of the entire system decreases owing to the reduction in network traffic. For instance, the CACC attains 593.952 *bps* of throughput, whereas CoCoA+ and base CoAP attain 544.272 *bps* and 555.312 *bps* respectively for a data transmission interval of 1 *s*.

Figure 4.13 illustrates the performance of the proposed CACC over the existing baseline protocols when the data transmission interval is varied from 1 *s* to 30 *s*. CACC reduces the packet loss in most of the points by measuring RTO properly, which considerably reduces the spurious retransmissions, as well as an unnecessary delay, in communication. CACC successfully transmits most of the generated packets to the server and reduces the packet loss drastically when compared to baseline protocols. For a data transmission interval of 1 *s*, the proposed CACC scheme incurs loss of 42 packets, whereas CoCoA+ and base CoAP incur loss of 177 and 147 packets respectively. In a low-traffic environment, the message loss is reduced in CACC and CoCoA+, since there is less possibility of network congestion and collision. CoAP employs fixed RTO, which results in a significant number of packets lost as observed with the dumbbell network topology.

In Figure 4.14, CACC delivers a significant number of messages and at the same time maintains the delay. For a data transmission interval of 1 *s*, the proposed CACC scheme successfully transmits 1614 packets with a delay of 22.798 *s*, whereas CoCoA+ and base CoAP successfully transmit 1479 and 1509 packets with delays of 16.646 *s* and 16.278 *s* respectively. The accurate RTT measurement tends to measure the RTO accurately and avoids the skewing and abrupt changes in RTO. These changes in

CoCoA+ and base CoAP result in spurious retransmission and unnecessary delay in communication as observed with the dumbbell network topology.

Figure 4.15, shows that the energy consumption of CACC is marginal compared to that of base CoAP. Baseline CoAP and CoCoA+ protocols assume that network congestion causes all packet losses and delays. However, this is not correct in the case of a network collision. The energy consumption of CACC is reduced, compared to that of the CoAP. CACC consumes 2.603 joules for the successful transmission of 1614 packets with the CoAP consuming 1.765 joules for delivering 1509 packets while the CoCoA+ consumes 2.190 joules for successfully transmitting 1479 packets for a data transmission interval of 1 s. As the data transmission interval increases, the message delay decreases linearly and the amount of energy consumed by the entire network, based on each network topology, is also reduced marginally.

Table 4.4 presents a summary of generated values based on the various data transmission intervals for grid network topology.

4.5 CHAPTER SUMMARY

Congestion control remains a critical issue that demands attention for reliable communication in networks of resource-constrained nodes for efficient resource utilisation and optimal network performance from both the research community and industry at large. In this Chapter, we have carried out a performance evaluation of baseline CoAP CC and CoCoA+ against the new proposal, CACC. For this validation, the Cooja simulation environment was used to evaluate base CoAP within the Contiki communication protocol stack for constrained networks. For this analysis, three different network topologies were simulated with varying data transmission intervals.

The utilisation of strong, weak and failed RTT estimators enables the CACC successfully to identify the exact network status, RTT of request-response interactions to mitigate the issue of large fluctuated RTT estimates, and to provide adaptive congestion control. The incorporation of an RC factor in RTT estimation and lower-bound restriction prevent the issue of fluctuated IoT traffic. The resulting algorithm, named CACC, is demonstrated to be effective in reducing the overall number of retransmissions while guaranteeing higher throughput and minimized packet loss comparable to those obtained with baseline

CoAP and CoCoA+ in all the network scenarios. The proposed 3-state estimators help to differentiate the loss scenario due to congestion and collision and provide an efficient context-aware scheme in CoAP through underlying UDP. The results also show a significant performance improvement over both baseline CoAP and CoCoA+ for the various networks in different data transmission intervals. Possible future work is needed to decide on the optimal value of the number of retransmissions and max-age value based on the particle swarm optimization (PSO) algorithm to optimise the base CoAP default parameters for optimal solutions in the network congestion scenarios, and also the queuing analysis technique for congestion control in constrained-IoT networks.

CHAPTER 5 PARTICLE SWARM OPTIMISATION-BASED CONGESTION CONTROL

5.1 CHAPTER OVERVIEW

In the previous chapter, a context-aware congestion control approach is proposed and evaluated against baseline CoAP and CoCoA+. Results obtained show significant improvements in all network scenarios. However, the need for further research into congestion control mechanisms for resource-constrained IoT networks calls for exploration of potential techniques to optimise the CoAP default parameters for optimal solutions. This chapter substantiates the proposal for a Particle Swarm Optimisation (PSO)-based Adaptive Congestion Control Technique (PACT) that significantly enhances the performance of the base CoAP protocol by tuning the retransmission count to an optimal value along with a varying max-age value by enabling the Observe option in the CoAP protocol. As contribution to the body of knowledge, the work presented in this Chapter has been submitted and is currently under review as part of a journal article titled "PACT: An Optimisation-based adaptive congestion control technique for constrained application protocol" [157].

The rest of this chapter is organised as follows: In Section 5.2, a general overview is presented, followed by a discussion of related work on congestion control in Section 5.3. Section 5.4 delves into the general overview of the proposed (PACT) methodology, including the system and network model, preliminaries of base CoAP CC specification, an optimisation strategy for tuning the CoAP default parameters, a fitness measurement based on (PSO) algorithm and tuning inertia weight, and network congestion level measurement for various congestion scenarios.

5.2 BACKGROUND

The ever-growing Internet of Things (IoT) is paving the way for the global network infrastructure, and the number of devices interconnected to the IoT is proliferating, which in turn leads to several diverse smart applications and services [158]. In view of the tremendous potential of the IoT, it is predicted that more than 50 billion Internet-capable smart things will be connected over the Internet shortly and this will tend to revolutionize the global world [11]. The IoT connects numerous heterogeneous devices by applying a wide range of technologies that include communication, networking, and information processing and ensures inter-operability among global Internet services by integrating smart objects into the existing network and information systems [6, 138, 159]. In IoT network, the involvement of heterogeneous devices with a wide range of applications creates network congestion when the traffic load reaches its maximum capacity. Considering the massive number of devices to be connected, the design of an efficient and reliable IoT network is a requirement for processing the large amounts of data, irrespective of the limited computational and communication capabilities of IoT-connected devices. The traffic pattern in the IoT network is different from that of conventional networks, and congestion is a significant factor that degrades the performance of data communications [160]. Therefore it is necessary to introduce congestion control (CC) mechanisms into IoT to provide real-time quality of service (QoS).

CC mechanisms are employed to detect congestion in various network traffic patterns and to take preferable measures to avoid or remove congestion in the network [137]. In traditional networks, Transmission Control Protocol (TCP) provides end-to-end congestion control [161]. However, conventional Internet protocols are not suitable for resource-constrained environments. The varying communication patterns in different IoT applications make TCP inefficient and that leads to unnecessary delay, high overheads, and unfeasible connections. Hence, for lightweight applications, Constrained Application Protocol (CoAP) operating over user datagram protocol (UDP) is suitable as it handles congestion by itself. As a result, different protocols have been considered [162].

In the IoT, the application layer protocols are responsible for orchestrating the network and one of the standard protocols in the application layer is Constrained Application Protocol (CoAP) [163, 164]. CoAP [91] is a lightweight protocol stack which adapts the Representational State Transfer (REST) paradigm [165] for easy compatibility with the current Internet trend. CoAP has specialized features, including multicast support, very low overheads, and suitability for resource-constrained

lossy networks [91, 166, 167]. CoAP end-nodes exchange requests-responses by using messages exchanged asynchronously through User Datagram Protocol (UDP) [168]. Since UDP is unreliable for message transfer, CoAP must ensure and handle message reliability and congestion over the unreliable UDP transport protocol by itself. To this end, CoAP uses the confirmable message which requires an acknowledgment or a reset message to ascertain the message reliability from the destination end node. Typically, memory-constrained and CPU-constrained nodes are used in IoT communications with limited processing and memory capabilities, which makes CoAP suitable for these extreme constraints.

Accordingly, RFC 7641 [2], extends CoAP with Observe, a publish-subscribe mechanism for supporting unreliable CoAP communication with the introduction of an upper limit on the permitted rate of outgoing messages. By enabling an observe mechanism, servers can inform the interested parties about any changes in the state of resources. The standard CoAP CC specification doubles the retransmission timeout (RTO) at the expiration of the timeout. CoAP under-performs when it is found to be too conservative in small networks or too aggressive in large networks [131, 139], instead of adapting to the network status information. Another extension of CoAP CC specification is CoCoA [134, 140] which is currently being standardized by the IETF CWG [105], as an adaptive, advanced congestion control mechanism. CoCoA+ makes the system more sensitive to the network dynamics by incorporating a novel round-trip-time (RTT) estimation scheme which automatically sets the RTO value, along with a variable backoff factor (VBF) and aging mechanism for the transmission of CoAP messages [1]. It has been proven that CoCoA significantly improves the performance over CoAP in congested networks [139, 148]. However, CoCoA is still deficient in choosing the exact RTO value during burst traffic because of the inability to determine the RTT of retransmitted request-response accurately [150], thereby resulting in unnecessary spurious retransmissions [141].

This chapter presents a Particle Swarm Optimisation (PSO)-based Adaptive Congestion control Technique (PACT) that significantly improves the performance of the CoAP protocol by tuning the retransmission count to an optimal value, along with a varying max-age value, under various network conditions. The observe option is enabled in the CoAP protocol for max-age value. The optimal MAX_RETRANSMIT and max-age value are obtained by applying optimal parameter-driven simulations for different congestion levels,and employing the PSO technique, using the Cooja simulation environment, a toolset of the Contiki OS. The proposed PACT incorporates the random and optimal parameter-driven simulation based on different network congestion scenarios and applies the fitness

measure obtained from previous scenarios to update the velocity for which the fitness function is computed for the optimal number of retransmissions and max-age values. The proposed PACT-based algorithm leverages on the RTTVAR which is used to allocate the default values for CoAP timeout parameters in order to estimate the network congestion level, mitigate the network fluctuations, and to enhance the convergence for a stable network condition.

5.3 RELATED WORK ON CONGESTION CONTROL

Several works have been presented in Wireless Ad hoc and Sensor Networks for CC in constrained application protocol; notable contributions on CC for reliable exchanges of CoAP communications include [20, 134, 137, 139, 140, 144, 148]. The Internet Engineering Task Force (IETF) RFC 7641 [2] provides an extension of the CoAP base specification for CC in unreliable CoAP communications as CoAP with Observe, which uses a publish-subscribe mechanism, with the introduction of an upper limit for the permitted rate of outgoing messages. These schemes play an important role in providing CC for constrained networks.

The experimental evaluation of CC for the CoAP protocol is presented in [169] for unreliable CoAP communication without end-to-end reliability between connected devices over emulated GPRS/UMTS links and also in a real IEEE 802.15.4 multi-hop testbed of constrained devices. The experimental results show that CoCoA performs better by maintaining high performance in almost all the scenarios considered, due to its flexibility when compared to default CoAP protocol, whereas, owing to the requirement of an adjustment period, the rate control applied by CoCoA may not perform optimally, a result which is reflected in a slightly lower packet delivery ratio (PDR) due to the varying condition of the network [145, 148]. In [170], the authors propose a new rate-based CC for CoAP (CoAP-R) as an improvement to determine the performance of CoAP in bursty traffic environments. To control the network traffic, this mechanism monitors the sending rate of CoAP sources and uses a rate-based scheme instead of the usual window-based mechanisms. This scheme is aimed at achieving maximum bandwidth and network resource allocation based on fairness. It is observed that CoAP-R ensures uniform distribution of network resources amongst all the senders with reduced delay compared to CoAP and CoCoA. For instance, in [150], the authors proposed an enhancement of the CC mechanism for CoAP and CC/advanced which we called (E-CoCoA), that makes use of the RC of the request packet to estimate the actual RTT by matching the request-response packet to determine the RTO value.

The results show that the proposed mechanism improves the efficiency of CC when compared to the base specifications of CoAP and CoCoA+.

A new mechanism called Fast-Slow RTO (FASOR) [151] based on retransmission timeout and CC mechanism for CoAP is introduced to address the problem of packet loss, which is either due to the wireless link environment or to congestion, by considering three unique features, including self-adaptive retransmission timer backoff, Slow RTO computation, and Fast RTO computation mechanisms. One promising advantage of the scheme is the fact that, unlike traditional CoAP and CoCoA mechanisms, the FASOR scheme can handle a high congestion level, even in the buffer-bloat environment, but suffers from high-latency cost due to the slow RTOs involved in the scheme.

A Precise CC Algorithm (pCoCoA) [147] for CoAP introduces the modified algorithm of CoCoA+ that overcomes the limitations in the existing algorithm. In pCoCoA, the use of a weak estimator is eliminated, and a transmission count is applied to match the ACK messages with CON messages, even during retransmission. This method also detects unnecessary retransmissions in the network by comparing the transmission counter value of the retransmission with the transmission counter value of the ACK message. The advantage of pCoCoA is the fact that it has reduced retransmissions and the ability to work in bursty traffic occurrence-based scenarios. However, considering the wide range of IoT scenarios, some of the fixed values used in this scheme might not be suitable. In [1], the authors present a context-aware CC (CACC) approach for lightweight CoAP/UDP-Based IoT Traffic which improves the modification of CoCoA+ [134] by utilising mechanisms that include an RTO estimator, RC-based smoothed RTT observation, a lower-bound RTO restriction approach, the aging concept and a 3-RTO estimator to identify the exact network status and provide adaptive CC. The results show that the resulting scheme called CACC, is effective in reducing the overall number of retransmissions, while guaranteeing higher throughput and minimised packet loss compared to those obtained with the base CoAP and CoCoA+ in all the network scenarios. However, to the best of the knowledge on optimal parameter tuning, this is the only work on CC for CoAP communications based on the PSO-algorithm.

A summary of some of the proposed CC mechanisms, with their unique characteristic features, is presented in Table 5.1

Table 5.1. Summary of Congestion Control Algorithms with their unique features [1].

Scheme	RTO aging	Backoff method	RC	RTT estimators	Existing scheme
CoAP	No	BBF	No	No	None
CoCoA	Yes	VBF	No	Yes - 2	LinuxRTO
4-State Strong	Yes	VBF	No	Yes - 4	CoCoA
E-CoCoA	Yes	VBF	Yes	Yes - 1	CoCoA
CACC	Yes	VBF	Yes	Yes - 3	CoCoA

5.4 OVERVIEW OF PROPOSED PACT METHODOLOGY

The base CoAP CC specification is designed, based on the initial RTO and an exponential backoff for the RTO timer. Basically, for a reliable message exchange between the sender and destination endpoint, CoAP randomly chooses the initial RTO value from a fixed interval of $[2,3]$ s for the initial message transmission. However, the base CoAP is deficient in controlling the congestion properly in a buffer-bloat IoT environment, which leads to spurious retransmissions and unnecessary delays that result in wastage of network resource-capacity. Several approaches attempt to improve the CoAP protocol to solve this issue.

Figure 5.1 illustrates the idea of the proposed methodology. CoAP's main limitations rely in the variability of its performance, irrespective of its significant advantages, which is induced by the choice of its default parameters, including the number of retransmissions and max-age value. The inability to assign appropriate values to these parameters and their impact on the network's behaviour is still a difficult process for base CoAP CC specification. The number of retransmissions is tunable without a clear default value, while the max-age, is chosen as 5 and 20 s. Based on this, the proposed system identifies an appropriate MAX_RETRANSMIT and obtains a considerable configuration better than the default parameters for the various IoT-congested scenarios by employing the PSO-based technique. In the random parameter-driven simulation, the existing CoAP algorithm is performed using default parameters, and the process is then repeated for different congestion scenarios by varying the retransmission and *max-age* values using the optimisation algorithm. Furthermore, the system builds on low, medium, and high- traffic scenarios by varying the number of retransmissions and max-age value to analyse the impact of these parameters on base CoAP performance. The variation in

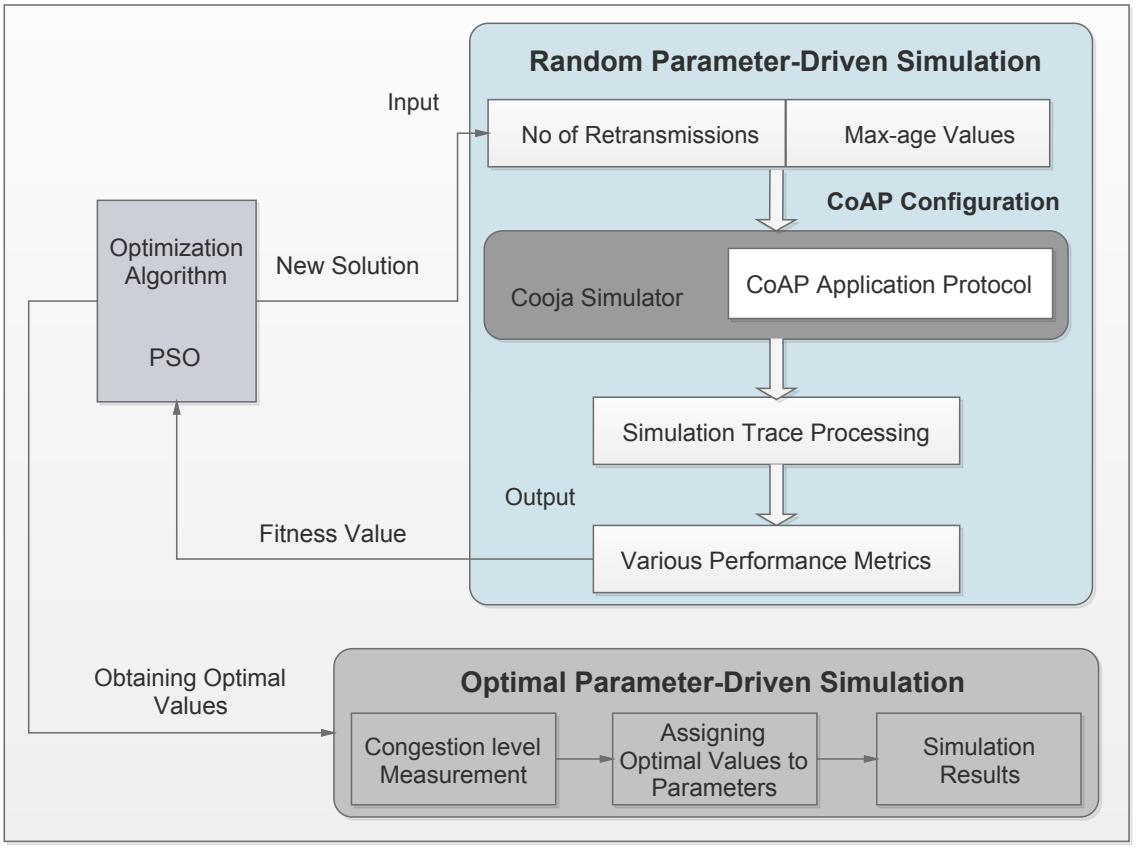

Figure 5.1. Proposed context-aware PSO methodology

these parameters is decided on based on a PSO fitness and velocity function. An optimal value of the MAX_RETRANSMIT is decided on to use the best fitness value. The velocity variation is adaptive to the network congestion level. Thus, the proposed PACT assigns an optimal value to the number of retransmissions with a fixed set of max-age values, according to the network congestion level, and improves performance of the base CoAP CC protocol in different IoT scenarios.

5.4.1 System and Network Model

CoAP is a stateless request-response protocol constructed in IEEE 802.15.4-based multi-hop networks of constrained devices. The standard CoAP model has two abstract layers. The upper layer implements the request-response communication for RESTful model, whereas the lower layer implements the control mechanism over CoAP message transmission through the underlying UDP protocol, which supports both reliable and unreliable modes in the network. In reliable transmission, the ACKs is

required to ensure reliability when selecting the CON packet type, while in unreliable mode, non-confirmable (NON) messages are used for unreliable communication where the sender does not expect an ACK as confirmation. In the Observe option, the NON-messages are mainly utilised during the transmission of notification messages. With the Observe, the server makes use of the max-age option to specify the exact expiry time for the CoAP response resource representation. The maximum message size is typically 128 bytes (block-wise) payload with a header size of 48 bytes. The primary CC specification of the CoAP lacks in supporting unreliable exchange of messages and is insensitive to network conditions. In order to overcome this limitation, the Observe extension is introduced to support CC in unreliable data communication.

Particle swarm optimization is an optimization base technique designed and developed by [171], and inspired based on the social behaviour of bird flocking and fish schooling. The PSO technique is a very simple concept, and easy to implements with the adjustment of few parameters including the number of particles (solutions), range of solutions, learning factors, inertia weight, V_{max} (which represents the maximum change for a particle during one iteration), global and local best positions. This technique is computationally inexpensive, and by the use of only primitive mathematical operators, the process can be achieved [172]. The basic equations which are generally utilised by the PSO technique are velocity and position update equations to optimize the problem as given by Equation 5.1, and 5.2 respectively [173]. These equations are modified in each iteration of PSO algorithm to achieve optimum solution.

$$V = V + C_1 \times RND \times (P_{best} - PRS) + C_2 \times RND \times (G_{best} - PRS) \tag{5.1}$$

$$PRS = PRS + V \tag{5.2}$$

Where V, RND, PRS, and P_{best} and G_{best} is the particle velocity, random number between (0,1), the current particle (solution), the local best position and global best position respectively, and C_1, C_2 are the leaning factors.

The PSO problem presents possible solutions (particles) that randomly send requests to a server node in the network. In this scenario, there is only one server in the network to respond to requests of the

connected CoAP nodes. The PSO technique learned from this, to apply and solve the optimisation problem. For this process, each single particle is a CoAP node in the network. This we can also call a solution. All solutions are characterised with fitness values which can be assessed by the fitness function to be optimised, and have velocities which direct the path of the solutions. The solutions request through the network to obtain the current optimum solutions. In the proposed methodology, CoAP with Observe protocol is used to consider both reliable and unreliable CoAP communication. The proposed PACT considers the process as a random and an optimal parameter-driven simulation. In the random parameter-driven simulation, the first iteration is based on the primary CC mechanism with a default transmission parameter. With PACT, the fitness values in PSO rely on the performance metrics obtained at the end of the simulation results from the perspective of validation. The velocity reflects the distance travelled by this particle at each iteration. Each particle (solution) has to maintain its position P_{best}, known as the local best position and the G_{best}, known as the global best position among all the particles. In the initial iteration of PSO, the current fitness value is considered as P_{best}, and one is assigned as G_{best}. In the progressing iterations, the P_{best} and G_{best} are updated based on the fitness value in consideration of optimal values of performance metrics. The velocity is added with MAX_RETRANSMIT for optimal tuning along with adjustment of the max-age value. During the simulation with adjusted parameters, the resultant value of the metrics is taken as a fitness value for the next iteration. The inertia weight, w, is adjusted, based on the network traffic, which controls the momentum of particle by weighing the contribution of the previous velocity in controlling the new velocity. In the proposed CC mechanism, round-trip-time (RTT) value is employed to determine the congestion level. For further iterations, the optimal number of retransmissions is obtained for different congestion scenarios. In the optimal parameter-driven simulation, the optimal tuned MAX_RETRANSMIT with the adjusted max-age are used, based on the network traffic, to control the congestion in the network.

5.4.2 Preliminaries of Base CoAP Specification

The CoAP is a web-specific protocol developed especially for resource-constrained sensor nodes and low-power networks (6lowPAN). The IoT network experiences high packet error rates, packet loss, and relatively small throughput. To mitigate these issues, CoAP implements the primary CC by using an exponential backoff mechanism. Initially, the client sends a CON message to a server node. The initial value of RTO in CoAP is assigned randomly within the interval between ACK_TIMEOUT and ACK_RANDOM_FACTOR, where ACK_TIMEOUT and ACK_RANDOM_FACTOR represent the

transmission parameters, and the default values of these factors are 2, and 1.5 *s* respectively. However, these factors are utilised only at the initial RTO measurement. Thus, the proposed methodology does not apply the PSO-based optimal value selection to those factors. After sending the CON message, the sender waits to receive an ACK before the RTO expires. If the sender node has not received the ACK before the RTO expires, CoAP doubles the RTO value. The sender node keeps on retransmitting the CON message until it reaches the maximum number of retransmissions MAX_RETRANSMIT and moreover the CoAP assigns MAX_RETRANSMIT value as 4. After four unsuccessful attempts of retransmission, CoAP enables the sender node to close the session. CoAP selects the initial RTO value randomly between ACK_TIMEOUT and (ACK_TIMEOUT x ACK_RANDOM_FACTOR).

Figure 5.2 illustrates the CC mechanism used in CoAP. Baseline CoAP CC specification enables the sender node to transmit the messages after having acknowledged the ACK message from the server. When congestion occurs and the ACK is not received, the RTO interval increases to double. The default CC scheme keeps doubling the RTO value each time until the client receives the ACK message. The client node increases the RTO value and retransmits the CON message up to four times and attempts to find the correct RTO value. Another important parameter in CoAP for consideration is freshness in terms of max-age value. When a response is considered to be fresh, it is used in further processes without requesting the server again and this improves the communication efficiency of CoAP. The mechanism for determining the freshness in CoAP is the exploitation of the expiry time by using the max-age value. The max-age value indicates that the response is invalid after its age has expired. By default, CoAP assigns the max-age value as 60 *s*. When a max-age value remains the same when its age is greater than 60 *s*, it is considered to be invalid. The shortcoming of the base specification scheme in CoAP is the fact that both the max-age value and MAX_RETRANSMIT are not set automatically. These parameters have to be assigned based on network congestion to improve the communication efficiency of CoAP. Table 5.2 presents a summary of default values chosen in the presence of congestion for CoAP delay and timeout parameters.

5.4.3 Optimisation Strategy for Tuning CoAP Parameters

The base CoAP protocol can perform better in different scenarios, but this is not an optimal solution for CC. The primary problem that recurrently emerges is due to the improper configuration of the default CoAP parameters. Therefore, the proposed PACT scheme utilises the optimisation algorithm

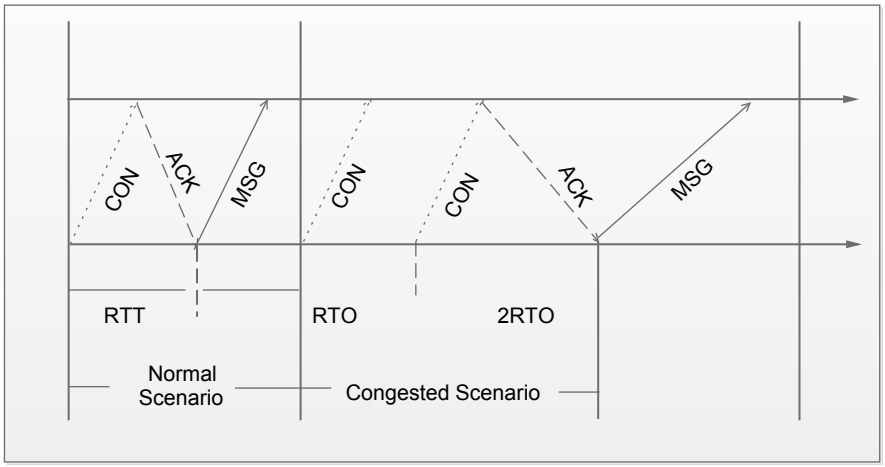

Figure 5.2. CoAP congestion control mechanism.

Table 5.2. Base values of CoAP delay and timeout parameters [2].

Parameter	Base value
ACK_TIMEOUT	2 s
ACK_RANDOM_FACTOR	1.5
MAX_RETRANSMIT	4
PROCESSING DELAY	2 s
DATA FRESHNESS	$\bar{T}_s = 60$ s
MAX_RTT	202 s
MAX_TRANSMIT_SPAN	45 s
DEFAULT_LEISURE	5 s

for deciding the optimal values for the parameters and providing an optimal configuration for CoAP in various IoT congestion scenarios. The proposed methodology consists of two parts: Random and Optimal Parameter-driven Simulations. The default CoAP specification exploits the random parameter-driven simulation, in which the base CoAP parameters are fixed with default values. The simulation of the IoT environment is carried out with the default values of these parameters. Based on the results obtained, the optimisation algorithm decides on the number of retransmissions and the max-age value to carry out the simulation. With the results of random parameter-driven simulations, the optimisation algorithm determines the optimal values for CoAP parameters. Assigning the same value to various congestion scenarios is not valid. Therefore, the proposed PACT considers the IoT congestion network

in three categories, namely, low, medium, and high-level congestion scenarios. For each category, the proposed PACT assigns the parameter values differently. The random parameter-driven simulation is a way of assigning the appropriate values to the parameters that regulate the performance of the base CoAP specification. This leads to optimal configurations for this protocol, tailored for various IoT scenarios.

5.4.3.1 Fitness Measurement using PSO Algorithm and Tuning Inertia Weight

The problem of data insufficiency at the server node during access by multiple clients simultaneously is a major cause of network congestion in resource-constrained networks. The proposed PACT creates several test scenarios to determine the network congestion levels. In the initial scenario, the default values are set for the CoAP parameters. The performance results realized from the first scenario are used to tune the default values of these parameters step by step. The proposed PACT continues this process for multiple times by using the PSO technique. The PSO technique applies the fitness measure of the results retrieved from the first scenario and updates the velocity to continue the process towards achieving an optimal number of retransmissions and max-age value. After the random parameter-driven simulation, it returns the simulation results of the packet delivery ratio PDR, the normalised overhead NO, and delay D of the communication scenario. This information is then used to compute the fitness function F as follows:

$$F = W_1 \times (PDR) + W_2 \times (-NO) + W_3 \times (-D) \times C \tag{5.3}$$

The objective of the PSO-based function is to maximize the packet delivery ratio and minimise both the packet loss and delay. For this reason, we formulate the system, such that the packet delivery ratio has a positive sign and the others have a negative sign. In Equation 5.3, factors W_1, W_2, and W_3 represent the influence of each metric on the resulted fitness value. These metrics are assigned with 0.5, 0.3, and 0.2, respectively. By applying such weighting factors, the packet delivery ratio takes priority over other metrics. The negative polarity in the weighing factors is considered for influencing the importance of each performance metric in both individual and combined perspectives. The delay is also multiplied by the constant value $(C = 0.01)$ to deal with a similar range of packet delivery ratios and normalised overheads. After the fitness for results of the random parameter-driven simulation has been measured,

the velocity value for each parameter is estimated by using the improved PSO based on the level of network congestion.

$$W_{\alpha} = \frac{\{(W_{max} - W_{min}) - (F_{max} - F_{min})\}}{(F_{avg} - F_{min})} \tag{5.4}$$

$$V_j^{SOL+1} = WV_j^{SOL} + c_1 r_1 (P_{Best} - X_j^{SOL}) + c_2 r_2 (G_{Best} - X_j^{SOL}) \tag{5.5}$$

$$IW = \begin{cases} \frac{[(W_{min} + (W_{\alpha})]}{2} \text{ if a,} \\ W_{min} + (W_{\alpha}) \text{ if b,} \\ W_{max} + (W_{\alpha}) \text{ if c.} \end{cases} \tag{5.6}$$

Where a, b, and c represent when congestion is low or normal, when congestion is medium, and when congestion is high respectively.

Here, the particles represent the possible values of several retransmissions and the max-age value. It is important to note that the PSO generates k solutions (SOL) and among them, the best solution is considered to be G_{Best}. The previous position of a particle acts as P_{Best} for the next solution. Moreover, X_j^{SOL} represents the position of a particle in a solution, and r_1 and r_2 are the random values, which are selected in the range of $[0,1]$. The factors c_1 and c_2 represent the acceleration constant. The fitness represents the quality of a solution for every particle movement. The inertia weight IW factor W, is estimated by using Equation 5.4, 5.5 and 5.6, where W_{max} and W_{min} denote the maximum and minimum values of W, and F_{min} and F_{avg} denote the minimum and average fitness of particles in previous solutions. Each particle moves from the current position to the next one at the estimated velocity used in Equation 5.5. The inertia weight controls the velocity of a particle in the next solution. Notably, the inertia weights for the first and second particles are selected randomly, since the terms $(W_{max} - W_{min})$, $(F_{max} - F_{min})$, and $(F_{avg} - F_{min})$ are zero for the second solution. From the third solution, the inertia weight starts to update according to the fitness of its solutions and improves the efficiency of the PSO-based algorithm.

A PSO algorithm employs the inertia weight to mitigate the effect of the previous velocity on the current one. Significant inertia weight leads the particle to a global search, while the smaller factor guides the particle to a local search by using the previously estimated best solution. In Equation 5.6, the inertia weight is updated based on the level of network congestion. For low network congestion, the proposed PACT forces the PSO to perform a local search only by increasing the inertia weight and velocity at a minimum level, compared to the medium and high congestion levels. The network with a high congestion level increases the inertia weight and the velocity of a particle significantly to perform a global search. Low inertia weight controls the velocity of a particle towards the solution. In the PSO-based algorithm, it is crucial to determine a proper termination criterion as this affects performance, solution quality and algorithmic efficiency. The termination of the algorithm is based on the closeness of the potential performance to the tuned retransmission count, provided that solution constraints are resolved. The number of retransmissions is controlled while tuning the parameter to avoid the possibility of getting a larger retransmission count, which in turn affects the performance efficiency of the system. Network congestion aware PSO-based factor tuning improves the efficiency of deciding on the base CoAP specification default parameters.

5.4.4 Network Congestion Level Measurement

The random parameter-driven simulation controls the network traffic to create various congestion scenarios differently. During the simulation of the real-time environment, it is hard to predict how many sensors will be connected with the server at a time. This creates significant challenges for the detection of network congestion levels. The key idea of the proposed PACT to estimate the level of network congestion is to utilise RTTVAR. The growth of RTT variance indicates the network state as a contention, and it leads to high packet loss and higher variability in communication delay. Therefore, the PACT leverages on the RTTVAR variable are proposed to define the characteristics of the network state, which are used in allocating the default values for the CoAP parameters, such as MAX_RETRANSMIT and max-age value.

$$RTO_{x(y)} = \begin{cases} \dfrac{[(SRTT_x - \gamma \times RTTVAR_x)]}{SRTT_x} \\ \dfrac{[(SRTT_x + \gamma \times RTTVAR_x)]}{SRTT_x} \end{cases} \tag{5.7}$$

Where: γ value is 1 if ACK is received in the first transmission, otherwise γ value varies from 1 to 3 if ACK is not received in the first transmission.

In Equation 5.7, the $RTO_{x(y)}$ represents the confidence level of the network congestion, as it is a measure based on RTT variability. By using Equation 5.6, we identify three characteristic regions as follows:

Case 1 (Low Congestion or Normal Scenario): $1 > RTO_{x(1)}$

In this state, the $RTO_{x(1)}$ denotes that there is no or less congestion in the network, as the negative RTTVAR leads to high $RTO_{x(1)}$. Consequently, the sending rate can be slightly increased.

Case 2 (Medium Congestion Scenario): $RTO_{x(1)} > 1 << RTO_{x(2)}$

This state represents the short-term RTT variability increment and this indicates the network congestion. Therefore, there is a need for transmitting messages in a controlled manner to avoid congestion.

Case 3 (High Congestion Scenario): $1 << RTO_{x(3)}$

In this state, the term $RTO_{x(3)}$ indicates the high congestion scenario, due to the huge variation in RTTVAR. Therefore, the CoAP has aggressively to reduce the sending rate and mitigate the impact of the congested network on message transmission.

5.4.4.1 Tuning the Optimal Number of retransmissions with Fixed Max-Age Values

The proposed PACT initially assigns a single sending rate to each message exchange and updates the sending rate on every ACK reception, based on the estimated level of network congestion. Specifically, the CoAP sender always maintains the maximum number of retransmissions and max-age value as 4 and 60 *s*, respectively. Since the proposed PACT is aimed at keeping the values of those parameters optimal, it employs the PSO-based optimal value search to achieve optimal maximum retransmission. The scheme uses max-age value 5 and 20 as these provide better performance with a tuned optimal

maximum retransmission value. This approach is known to mitigate the effect of network condition fluctuations and to facilitate the convergence to stable network behaviour.

Algorithm 1 explains how the inertia weight factor in PSO is updated based on network congestion. By applying the algorithm, PSO-based parameters are measured. The solution with the highest fitness is considered to be the optimal value in each network category. The proposed PACT estimates the congestion level of the network and sets the parameters of CoAP optimally. As MAX_RETRANSMIT is freely adjusted, including the possibility of obtaining more significant value than the default, it requires further adjustments to the time values such as ACK_RANDOM_FACTOR and ACK_TIMEOUT to avoid unnecessary waiting time. These parameters are adjusted based on the optimal MAX_RETRANSMIT value achieved during the optimal parameter-driven simulation. Therefore, the PSO-based CoAP parameter tuning assists in improving the efficiency and performance of the IoT network, compared to the CoAP with Observe in various congestion network scenarios.

5.5 CHAPTER SUMMARY

This chapter presents a Particle Swarm optimisation (PSO)-based Adaptive Congestion Control Technique (PACT) that significantly improves the performance of the base CoAP protocol by tuning the retransmission count to an optimal value along with a varying max-age value under various network conditions. The Observe option is enabled in the CoAP protocol for max-age value. The optimal MAX_RETRANSMIT and max-age value are obtained by applying optimal parameter-driven simulations for different congestion levels, employing the PSO technique, using the Cooja simulation environment, a toolset of the Contiki OS. The proposed PACT incorporates the random and optimal parameter-driven simulation based on different network congestion scenarios and applies the fitness measure obtained from previous scenarios to update the velocity for which the fitness function is computed for the optimal number of retransmissions and max-age value. The proposed PACT-based algorithm leverages on the RTTVAR which is used to allocate the default values for CoAP timeout parameters in order to estimate the network congestion level, mitigate the network fluctuations, and enhance the convergence for a stable network condition.

Algorithm 1 Network Congestion-Based Inertia Weight Update.

Input: Measurement of network congestion level

Output: Allocation of optimal values to number of retransmissions and max-age value

1: function for optimal allocation ($SRTT_x$, $RTTVAR_x$)
2: measurement of $RTO_{x(y)}$;
3: **If** $RTO_{x(1)} > 1$ **then**
4: Low Network Congestion;
5: **Else** if $RTO_{x(1)} > 1 << RTO_{x(2)}$ **then**
6: Medium Network Congestion;
7: **Else** if $1 << RTO_{x(3)}$ **then**
8: High Network Congestion;
9: **End if**
10: End function
11: **If** (network congestion = Low) **then**
12: Update the inertia weight using $(W_{min} + (W_\alpha))/2$;
13: Measurement of Velocity and Fitness;
14: **Else** if (network congestion = Medium) **then**
15: update the inertia weight using $W_{min} + (W_\alpha)$;
16: measurement of velocity and fitness;
17: **Else** if (network congestion = High) **then**
18: update the inertia weight using $W_{max} + (W_\alpha)$;
19: measurement of velocity and fitness;
20: **End if**

CHAPTER 6 PERFORMANCE EVALUATION AND DISCUSSION

6.1 CHAPTER OVERVIEW

This Chapter evaluates and discusses the network performance for implementing the proposed PACT-based congestion control (CC) in the Constrained Application Protocol whose core components have been presented and discussed in Chapter 5. The performance of the proposed PACT is validated against baseline CoAP with Observe to highlight its significance by running simulations using the Cooja simulator, a tool set of the Contiki-OS. The performance metrics considered for evaluation are packet loss, delay, and normalised overhead of the entire network and achieve results based on the proposed PACT to analyse the performance of the system model. CoCoA+ provides a flexible CC mechanism by combining an adaptive RTO calculation, with the use of weak RTTs, and an aging mechanism to optimise performance. The proposed PACT employs a flexible CC mechanism by tuning the parameters for corresponding network traffic. The results include evaluation involving two network configurations, random and grid-topology-based simulations. Within the scope of this research work, the validation of the proposed PACT model is based solely on simulation and the actual experimentation of this study is presented as part of future work owing to the unavailability of access to an IoT-test bed for congestion control at the time of this research work. However, both CoAP with Observe and CoCoA+ provide distinct CC mechanisms by considering a different set of parameters that have a significant impact on the network performance. The proposed PACT is compared to CoAP with Observe to illustrate the impact of tuned parameters in various network traffic scenarios. As contribution to the body of knowledge, the work presented in this Chapter has been submitted and is currently under review as part of a journal article titled "PACT: Optimisation-based adaptive congestion control technique for constrained application protocol" [157].

The outline for the rest of this Chapter is as follows: The system implementation strategy, including the implementation tool, simulation setup, network topology, and CC performance metrics are provided in Section 6.2. In Section 6.3, the PACT results for validation against the benchmark protocol of several quality of service performance metrics from simulations are presented while Section 6.4 presents a detailed discussion of the results achieved with Subsections 6.4.1, and 6.4.2 represents the discussion of results based on random, and grid network topologies respectively. Section 6.5 summarises the Chapter.

6.2 IMPLEMENTATION TOOL

Cooja is a Java-based Contiki operating system (OS) simulation tool as an open source for both traditional wireless sensor networks and for the IoT, to connect tiny low-cost, and low-power micro-controllers to the Internet. This platform enables the execution of Contiki OS-based code and a simulation output that can be collected and processed, and obtain the performance metrics. With this platform, a Cooja inbuilt java-script simulator editor called gedit can be used to set the simulation time and also to calculate the performance metrics. This network simulator supports the simulation of sensor networks at three different levels which are: the machine code instruction set, the application level and the operating system [154]. This platform enables Cooja motes to emulate an off-the-shelf wireless sensor node that takes into consideration the capability functions of real nodes in the simulation environment. The Cooja simulation tool, including the script editor, the control panel and the serial socket (server) port connection plugin, is shown in Figure 6.1.

6.2.1 Simulation Setup

The performance evaluation of the proposed PACT utilises the Cooja simulator of the Contiki-OS for constrained devices and the IoT to check the network performance of the proposed work. This platform permits the execution of Contiki OS-based code and enables Cooja motes to emulate an off-the-shelf wireless sensor node that takes into consideration the capability functions of real nodes in the simulation environment. The proposed work creates the sensor nodes and border router by using two of the IEEE 802.15.4 capable motes, Wismote and SkyMote, in a Cooja simulator, respectively. The simulation model of the proposed scheme consists of varying nodes in terms of 20, 30, 40, and 50 with the transmission range of each node as $50m$. The network is constructed with $100m \times 100m$ area.

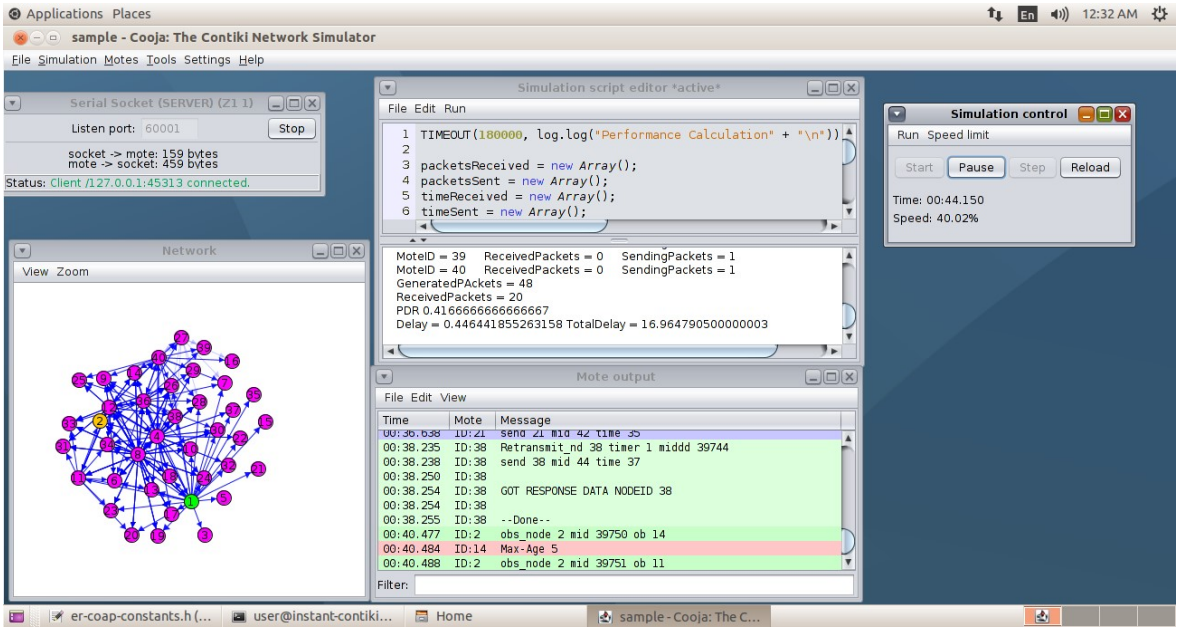

Figure 6.1. Contiki Cooja simulation environment showing the random network topology. 1 and 2 represent RPL border router and sink node respectively, while all other motes in the network represent nodes for running the full Contiki OS stack and the Erbium implementation of CoAP

For radio transmissions, the simulation exploits a Unit Disk Graph Medium (UDGM) radio model with circular transmission and applies interference areas. When applying the UDGM radio model, a static Link Delivery Ratio (LDR) setting is applied. The 802.15.4 MAC layer protocol and UDGM propagation model are used in the MAC and physical layer, respectively.

Moreover, RPL and UDP are used in the network and transport layer respectively. Different traffic loads are created by varying the packet intervals 4 *s*, 8 *s*, and 15 *s* for high, medium, and low traffic scenarios respectively. The total simulation time taken is 180 *s*. The proposed scheme shows the performance analysis for the variation of nodes in terms of 20 30, 40, and 50 for three traffic scenarios. The traffic scenarios are varied based on the number of packets generated per second, and for each scenario, their respective tuning factors such as the number of retransmissions and fixed max-age values of 5 and 20 are considered to achievev the performance of the proposed scheme. In the core CoAP with Observe extension, the default values of transmission parameters such as MAX_RETRANSMIT, ACK_TIMEOUT, and ACK_RANDOM_FACTOR are 4, 2, and 1.5 *s* respectively. The default max-age value is 60. The three traffic scenarios with the tuned parameters deliver improved maximum outputs in terms of packet delivery ratio, delay and normalised overhead. The simulation deploys

randomly distributed clients in two different test scenarios, such as random topology and grid topology. The hardware specifications with their unique capability features are presented in Table 6.1, and the details of the Cooja simulation parameter setup are presented in Table 6.2. The performance analysis of the proposed methodology is evaluated, based on the following metrics:

1. The first performance metric chosen for the analysis of the CC mechanism for CoAP is the packet loss, which indicates the total number of undelivered data packets to the server.
2. The second performance metric is the end-to-end delay of CoAP messages. The delay is the average time taken for a CoAP message to reach its final destination node from the moment it is sent at the application layer of the source node.
3. And finally, the normalised overhead is calculated based on the ratio of the number of observed request packets involved in the network by the total number of packets generated.

Table 6.1. Hardware Simulation Parameters for the WisMote and Moteiv Tmote Sky Wireless Sensor Nodes.

Features	**Wismote**	**Tmote Sky**
RAM	16 KB	10 KB
ROM	256 KB	48 KB
MCU	MSP430	MSP430F2617
RADIO	CC2520	CC2420

Table 6.2. Summary of simulation setup specifications.

Parameter	Value Specification
Congestion mechanisms	Base CoAP with Observe, PACT
Wireless channel model	Unit Disk Graph model, TR = 50m
Transport and network	UDP + uIPv6 + 6LoWPAN
Routing protocol	Routing Protocol for LLN (RPL)
Radio duty cycling (RDC)	Null-RDC
Media access control	(CSMA/CA)
max-age value	5, 20
Network area	$100m \times 100m$
Radio band	$2.4GHz$
Physical	$IEEE$ 802.15.4 PHY
Simulation time	180 s
Traffic loads	4, 8, and 15 s
Cooja simulation speed	200 %

6.3 SIMULATION RESULTS

In this section, we present the result validation for the proposed PACT against CoAP with Observe for different traffic scenarios as low (L), medium (M), and high (H) to evaluate the performance of the CC scheme. The set-up includes two 180 s simulations running at 200 % speed by varying the max-age values of 5 and 20, with tuned retransmissions, each with 20, 30, 40, and 50 nodes, to analyse the performance in a random and grid topology-based simulation, respectively. The simulation set generates results necessary to evaluate the performance in terms of packet loss (packets), delay in (s), and normalised overhead. The results obtained for the various performance metrics are presented in Figures 6.2, 6.3, 6.4, 6.5, 6.6, 6.7, 6.8, and 6.9 for random-based topology while Figures 6.10, 6.11, 6.12, 6.13, 6.14, 6.15, 6.16, and 6.17 are for grid-based topology.

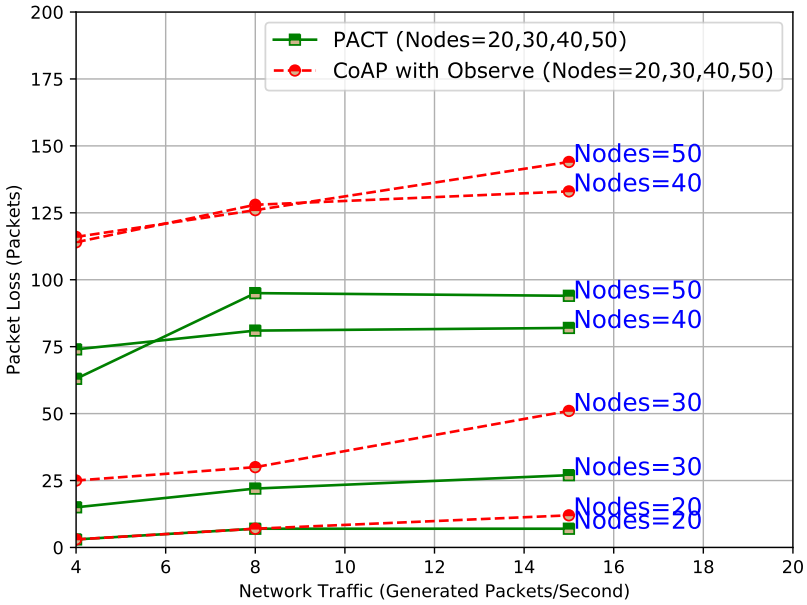

Figure 6.2. Packet Loss with Max-Age 5 for random topology.

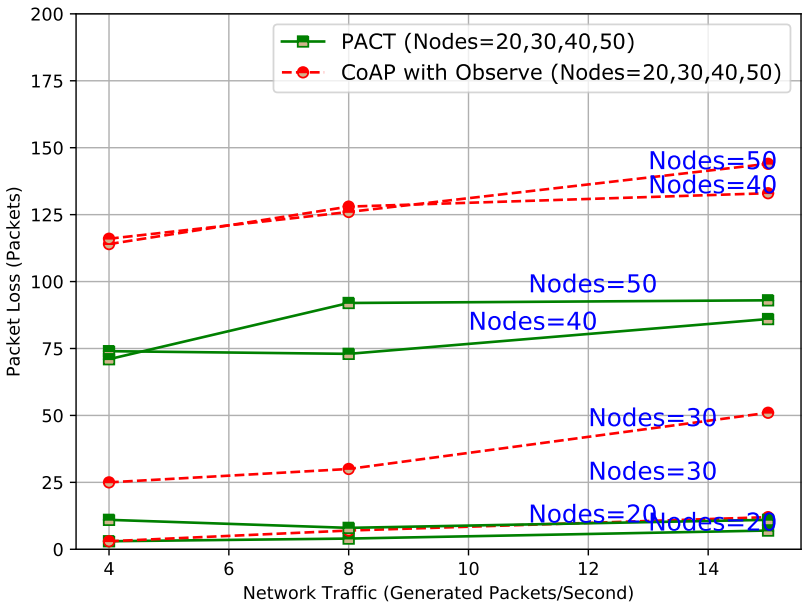

Figure 6.3. Packet Loss with Max-Age 20 for random topology.

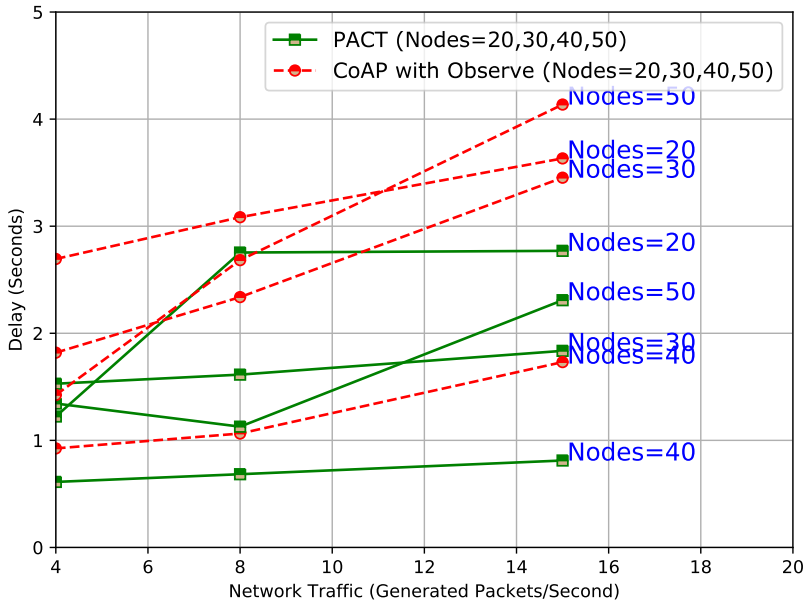

Figure 6.4. Delay with Max-Age 5 for random topology.

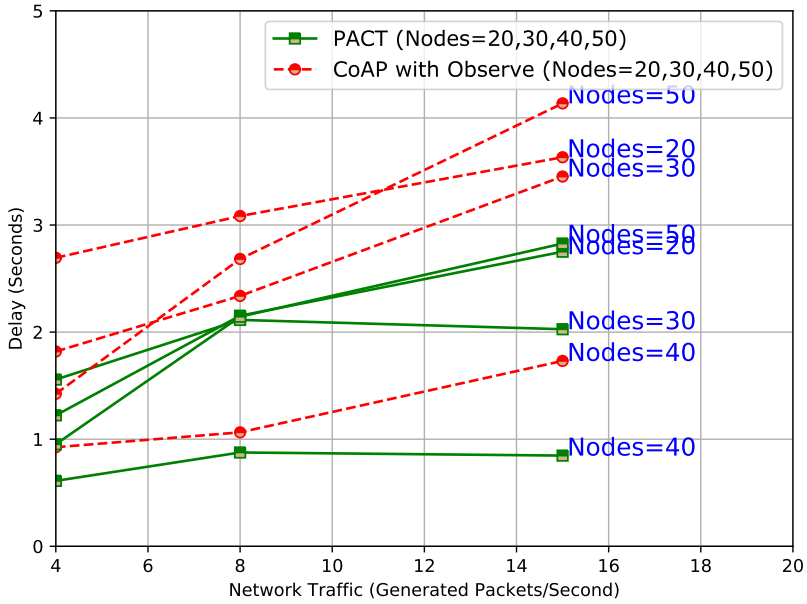

Figure 6.5. Delay with Max-Age 20 for random topology.

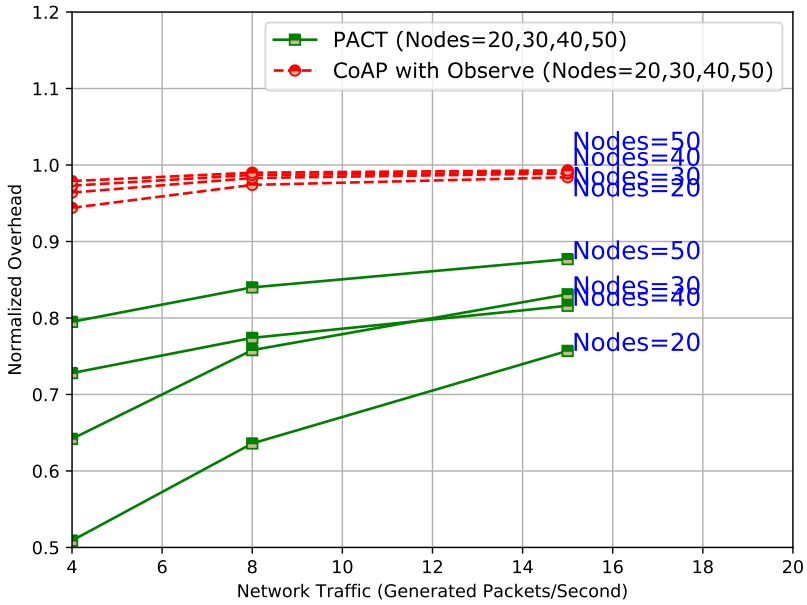

Figure 6.6. Normalised Overhead with Max-Age 5 for random topology.

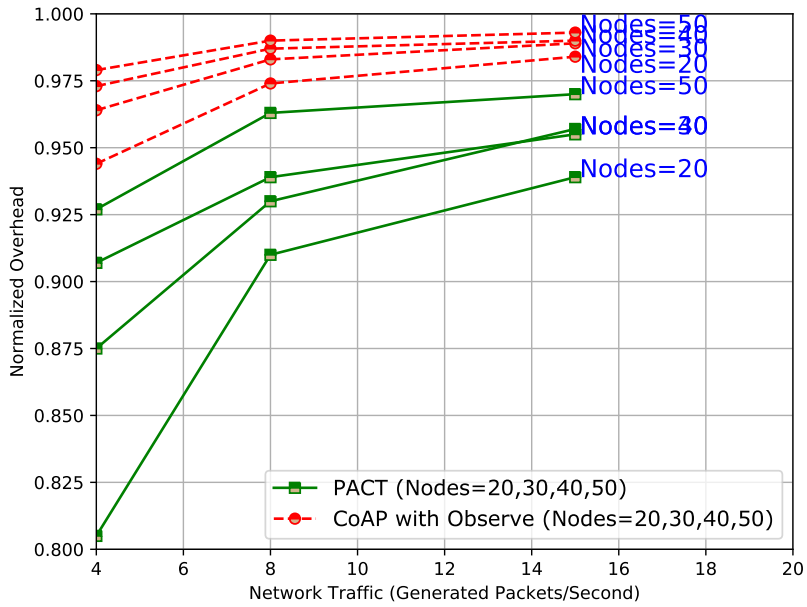

Figure 6.7. Normalised Overhead with Max-Age 20 for random topology.

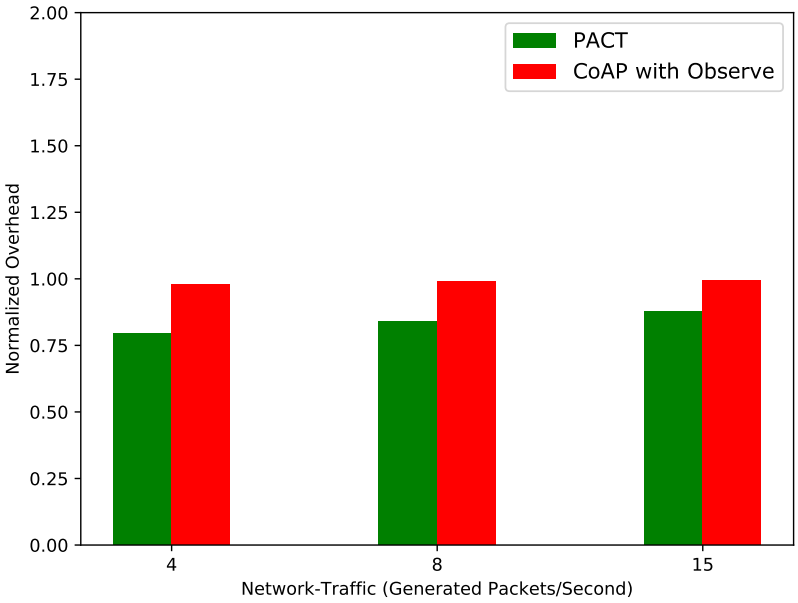

Figure 6.8. Normalised Overhead with Max-Age 5 for random topology.

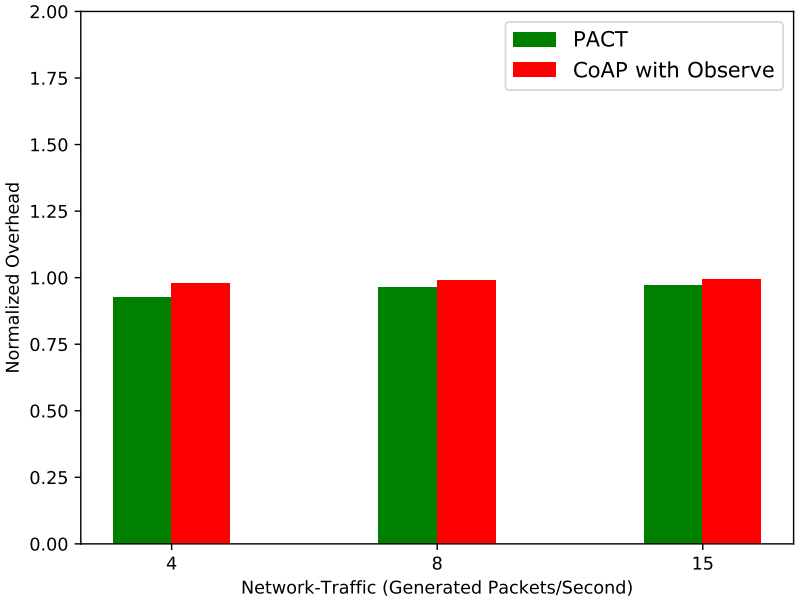

Figure 6.9. Normalised Overhead with Max-Age 20 for random topology.

Table 6.3. Normalised Overhead with Max-Age 5 for Random Topology.

	Base CoAP with Observe			PACT		
Nodes	L (%)	M (%)	H (%)	L (%)	M (%)	H (%)
20	0.944	0.974	0.984	0.509	0.636	0.757
30	0.964	0.983	0.989	0.642	0.758	0.831
40	0.973	0.987	0.990	0.728	0.774	0.816
50	0.979	0.990	0.993	0.795	0.840	0.877

Table 6.4. Normalised Overhead with Max-Age 20 for Random Topology.

	Base CoAP with Observe			PACT		
Nodes	L (%)	M (%)	H (%)	L (%)	M (%)	H (%)
20	0.944	0.974	0.984	0.805	0.910	0.939
30	0.964	0.983	0.989	0.875	0.930	0.957
40	0.973	0.987	0.990	0.907	0.939	0.955
50	0.979	0.990	0.993	0.927	0.963	0.970

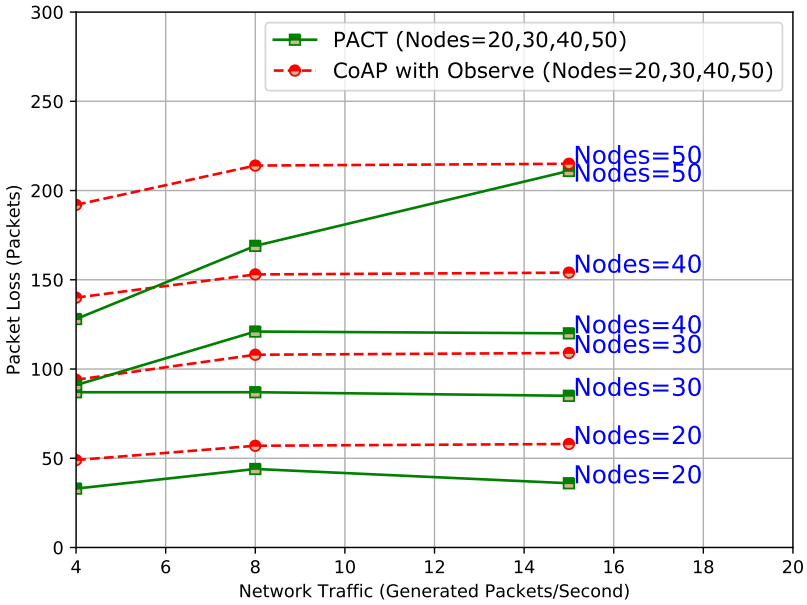

Figure 6.10. Packet Loss with Max-Age 5 for grid topology.

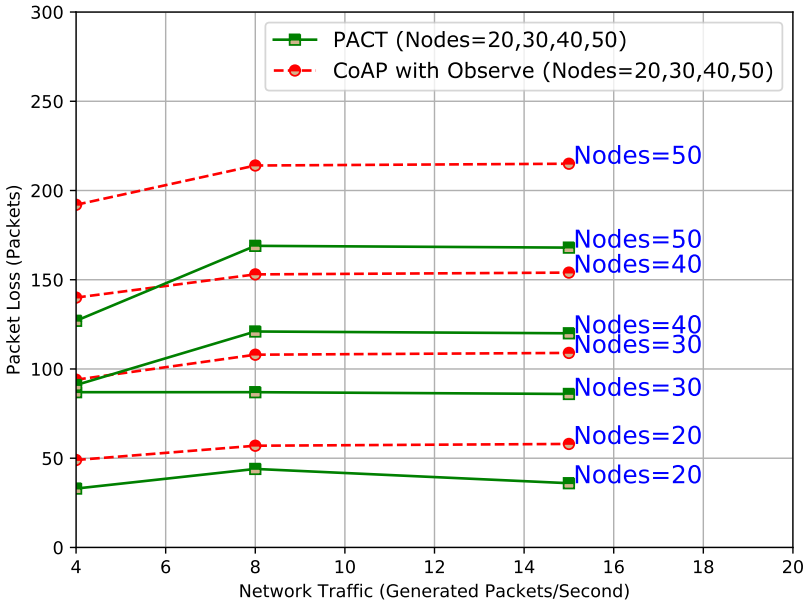

Figure 6.11. Packet Loss with Max-Age 20 for grid topology.

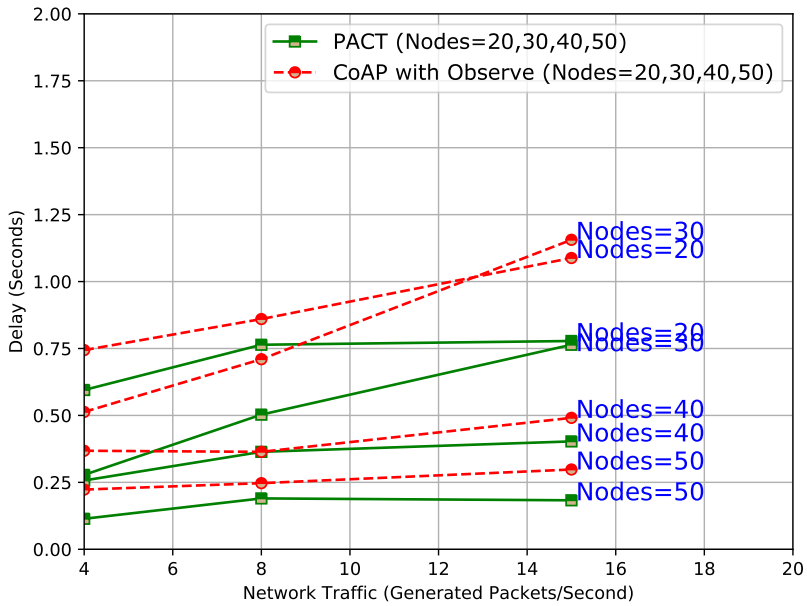

Figure 6.12. Delay with Max-Age 5 for grid topology.

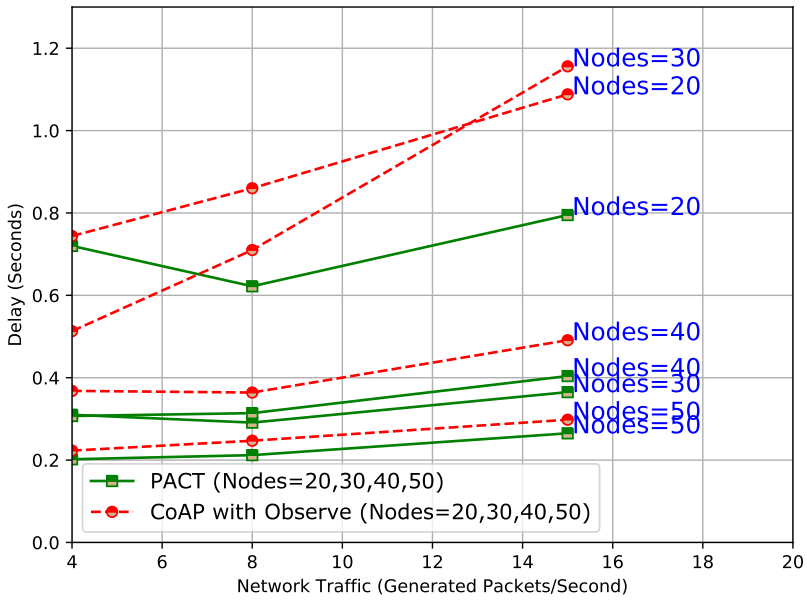

Figure 6.13. Delay with Max-Age 20 for grid topology.

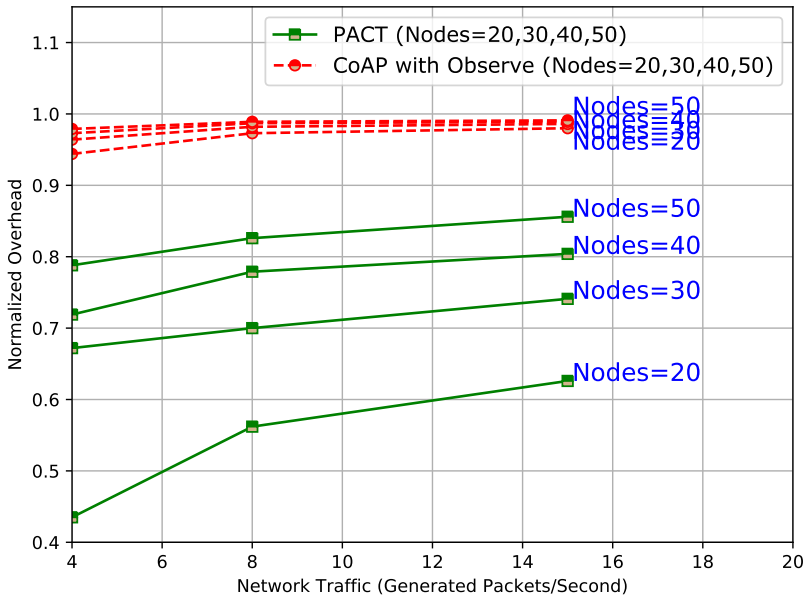

Figure 6.14. Normalised Overhead with Max-Age 5 for grid topology.

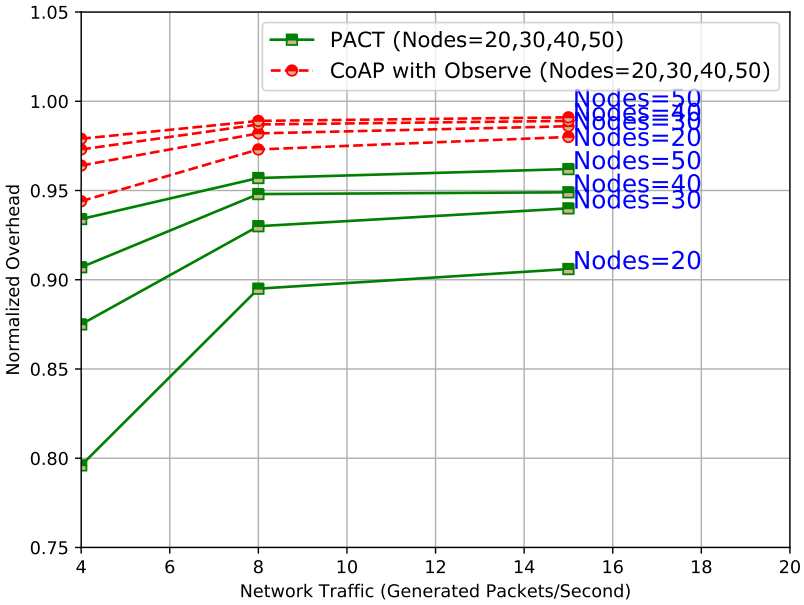

Figure 6.15. Normalised Overhead with Max-Age 20 for grid topology.

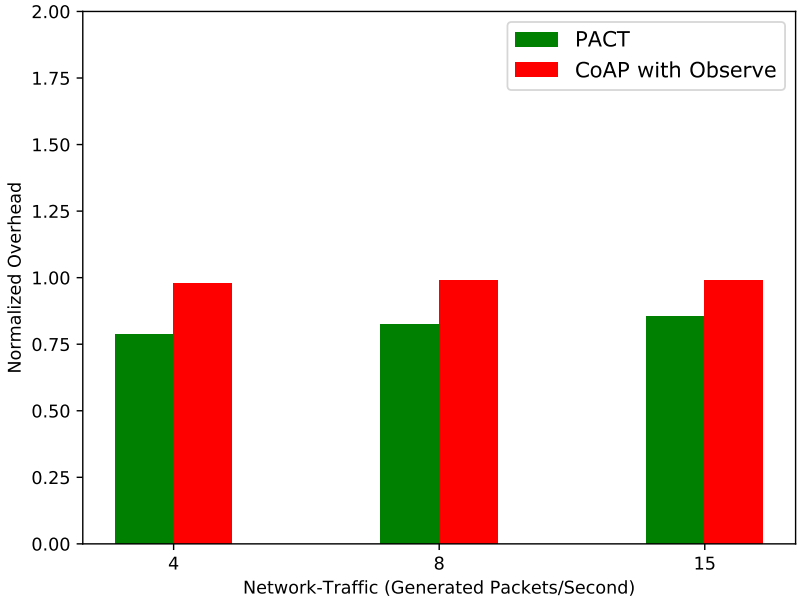

Figure 6.16. Normalised Overhead with Max-Age 5 for grid topology.

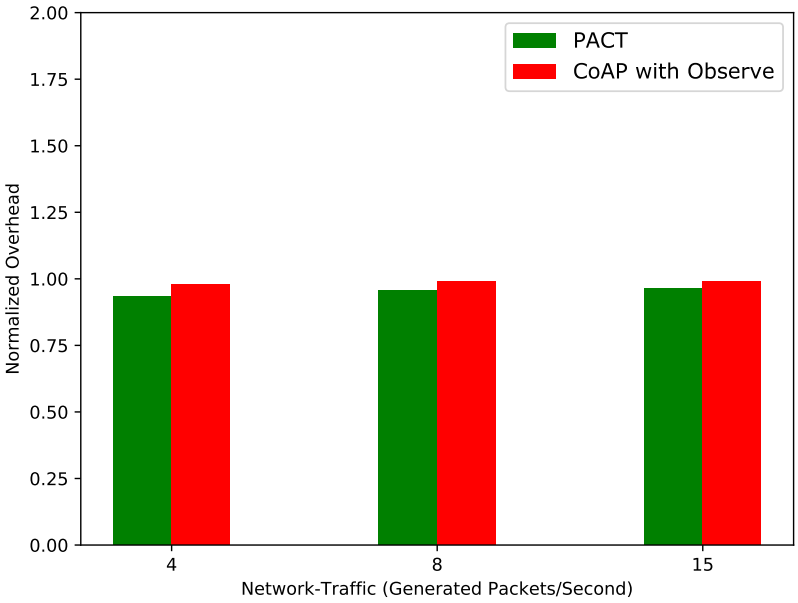

Figure 6.17. Normalised Overhead with Max-Age 20 for grid topology.

Table 6.5. Normalised Overhead with Max-Age 5 for grid topology.

	Base CoAP with Observe			**PACT**		
Nodes	L (%)	M (%)	H (%)	L (%)	M (%)	H (%)
20	0.944	0.973	0.980	0.435	0.562	0.626
30	0.964	0.983	0.989	0.672	0.700	0.741
40	0.973	0.987	0.989	0.719	0.779	0.804
50	0.979	0.989	0.991	0.788	0.826	0.856

Table 6.6. Normalised Overhead with Max-Age 20 for grid topology.

	Base CoAP with Observe			**PACT**		
Nodes	L (%)	M (%)	H (%)	L (%)	M (%)	H (%)
20	0.944	0.973	0.980	0.796	0.895	0.906
30	0.964	0.983	0.989	0.875	0.930	0.940
40	0.973	0.987	0.989	0.907	0.948	0.949
50	0.979	0.989	0.991	0.934	0.957	0.962

6.4 DISCUSSION

This section presents a discussion of the results obtained for performance evaluation of the proposed PACT against baseline CoAP with Observe in different network traffic scenarios. Subsection 6.4.1 discusses the results for random network topology and Subsection 6.4.2 presents the discussion for grid network topology.

6.4.1 Results based on network traffic with max-age value for random topology

The results obtained show that the implementation of PSO-based CC improves the performance of base CoAP with Observe parameters if it is developed in several key ways. The PSO-based algorithm considers the RTT and RTO values during packet retransmission in order to determine the congestion levels, which helps in the selection of an optimal retransmission count with efficient resource utilisation. Whereas baseline CoAP with Observe lacks in adjusting transmission parameters based on the network congestion levels, thereby resulting in high packet loss in all traffic scenarios, the PACT shows a notable improvement in packet loss, compared to baseline CoAP with Observe, as the connected nodes are increased from 20 to 50 by setting the optimal retransmission count with a max-age value of 5 and 20 as depicted in Figures 6.2 and 6.3 respectively, for different traffic scenarios. For a maximum number of nodes considered, the packet loss in the proposed PACT is less for 45 %, 24 %, and 34 % for the *L*, *M* and *H* congestion scenarios compared to baseline CoAP with Observe.

The proposed methodology shows a significant reduction in delay and normalised overhead compared to baseline CoAP with Observe in each case. For the system delay with max-age values of 5 and 20 as depicted in Figure 6.4 and 6.5 respectively, PACT maintains a minimised delay compared to baseline CoAP with Observe for node variation from 20 to 50. This shows a significant reduction in delay to 20 % and 32 % for medium and high-traffic scenarios respectively, even with the maximum number of nodes. In baseline CoAP with Observe, the MAX_RETRANSMIT is fixed at 4, irrespective of the network traffic scenarios, which leads to high packet loss and delay. However, in PACT, the PSO-based algorithm considers the results obtained for each varied number of retransmissions to choose the optimal MAX_RETRANSMIT, based on each network condition, which significantly improves the performance of the system. The number of nodes and the different traffic flows increase the possibility of network congestion, as well as a collision.

The normalised overhead results show the PSO-based algorithm controlling the overhead packets by reducing the load in the network, by using the optimal number of retransmissions and max-age values comparison to baseline CoAP with Observe. As depicted in Figures 6.8, and 6.9 with max-age values of 5 and 20 for different traffic scenarios, respectively, PACT shows a significant reduction in computation cost of normalised overhead compared to baseline CoAP with Observe for 50-connected nodes as a result of improved optimal transmission parameters. Tables 6.3, and 6.4 show a summary of the values generated for normalised overhead with max-age values of 5 and 20 respectively for varying traffic loads.

6.4.2 Results based on network traffic with max-age value for grid topology

In a resource-constrained network, packet loss remains one of the critical performance metrics, most especially for applications that require immediate reactions, short notification periods, and data integrity. The performance analysis reveals that the packet loss of the proposed PACT is significantly reduced compared to that of base CoAP with Observe as shown in Figures 6.10, and 6.11 with a max-age value of 5 and 20, respectively. The results show a notable improvement in packet loss even as the nodes are increased from 20 to 50. The proposed PACT experiences packet loss in low, medium and high traffic scenarios as 33, 44, and 36 packets respectively, whereas, in the baseline CoAP with Observe there are 49, 57, and 58 packets respectively, in a similar network scenario. It is obvious that the significant decrease in the packet loss realized in the PSO-based algorithm is due to the avoidance of both steep increment and excessive shrunk of the RTO, by utilizing the field of retransmission count.

The proposed PACT shows a significant improvement in reduction for both the delay and normalised overhead even as the nodes are increased from 20 to 50 in various network traffic scenarios. In Figures 6.12, and 6.13 the PACT outperforms baseline CoAP with Observe, especially at small packet intervals, where the congested network environment starts to build up and affect the protocol performance. For instance, for 20 connected nodes, the delay experienced between the PSO-based algorithm and the baseline CoAP with Observe in the low-traffic scenario is 0.14 s, while the delay difference in the medium traffic scenario is 0.09 s. The proposed *PACT* experiences 0.31 s less delay compared to the base *CoAP* with Observe protocol in high network traffic. The fixed parameter values in CoAP with Observe and insensitivity to network conditions lead to longer delays and packet loss. However, the consideration of the congestion level and application of optimal transmission parameters provide

efficient and better packet delivery with reduced delay in the proposed PSO-based CC.

The performance evaluation for normalised overhead between the proposed PACT and baseline CoAP with Observe for varying traffic scenarios is shown in Figures 6.14, and 6.15 with max-age values of 5 and 20, respectively. The PSO-based CC shows a reduced computational overhead for both max-age 5 and 20 compared to baseline CoAP with Observe. For instance, at 50 connected nodes with max-age 5, the proposed PACT shows a lesser overhead of 0.191 in a low-traffic scenario compared to that of the baseline CoAP with Observe. As shown in Figures 6.16, and 6.17 with max-age values of 5 and 20 for different traffic conditions respectively, PACT shows reduced computation cost for normalised overhead compared to those obtained with baseline CoAP with Observe for a 50-nodes connected network. Tables 6.5, and 6.6 present a summary of the values with max-age values of 5 and 20, respectively. The use of optimal transmission parameters by the PSO-based algorithm helps to improve the default CoAP parameters, improves the delivery of packet loss with reduced delay, and minimises the computational overhead, thereby improving the efficiency of the overall system performance.

6.5 CHAPTER SUMMARY

CC remains a critical issue that demands continuous research for reliable communication in resource-constrained networks for efficient resource utilisation and for optimal network performance. In this chapter a solution has been proposed for congestion control that significantly improves the reliability of base CoAP protocol at various congestion levels with a varying number of connected nodes. The proposed PACT methodology utilises PSO-based optimisation for transmission parameters such as MAX_RETRANSMIT with varying max-age values by applying random and optimal parameter-driven simulation. The PSO-based algorithm determines the optimal number of retransmissions for different traffic scenarios, and these parameters, in turn, yield better performance, even when the network condition fluctuates.

The performance validation of the proposed scheme shows significant improvement in packet loss, delay, and computational overhead compared to the results obtained with baseline CoAP with Observe. Possible future work would be further enhancement of the performance of congestion control mechanisms by obtaining optimal parameters to develop a system framework for Internet of Things service

provisioning as a three-state system base on queuing analysis and an intelligent machine-learning-based congestion control approach to Internet of Things traffic that will improve the performance of certain Internet communication components for an efficient user-friendly service experience.

CHAPTER 7 CONCLUSION AND FUTURE WORK

7.1 CONCLUSION

The emerging Internet of Things (IoT) has attracted immense attention and recognition in the research community in recent times by its amazing promises of several applications across the industry and society at large. Wireless sensor network (WSNs) is a vital networking component of the IoT since most of the smart sensor nodes will constitute the building blocks for the IoT. With recent advancements in research on Micro-eletcronics, Microelectromechanical Systems (MEMS) have brought about the deployment of these sensor nodes to become inexpensive, very intelligent, and amenable to easy networking through wireless and wired communications for the emerging and promising IoT paradigm. The massive number of smart devices envisaged to be connected across the IoT will definitely play a vital role in driving up the signaling load in the mobile network and will lead to network congestion. Network congestion degrades the performance of the IoT system and in the worst case, blacks out the entire network. The main reason for the IoT congestion is bursty traffic, which represents the scenario of high-bandwidth data transmission over a short period, resulting from the huge number of connected devices. The IoT traffic patterns, together with the constrained wireless links, pose severe challenges for the existing transmission control protocol (TCP) congestion control mechanisms. The conventional techniques assume that a packet loss occurs because of network congestion. However, owing to the resource-constrained nature of IoT networks, a high packet drop is also possible because of the bit error-rate (BER). Having first identified a knowledgeable research gap, without incurring high overheads, efficient handling of network congestion and BER is essential to trade-off the packet loss and unnecessary delay in the IoT. Therefore, the congestion control mechanism in IoT must be capable of assuring safe network operation with efficient network resource utilisation. Therefore, the

challenge to understudy the research work presented in this thesis has emerged.

The thesis write-up has been presented in a very logical manner, based on the principles of technical reporting required in the field of engineering, most especially the electronic engineering domain. This concluding part of the thesis focuses more on the most relevant ideas, contribution to scientific knowledge and important findings, which have been reported in the previous chapters of the thesis write-up. Then, future recommendations are made or proffered for possible considerations.

Chapter one of this work presents a succinct introduction to the entire research study. It provides the necessary guideline and capitalises on these through which the entire research work has eventually unfolded. In this chapter, the problem statement for the research work is established, in order to identify and proffer viable possible solutions to such investigated problems. Furthermore, the chapter clearly presents the research objectives to be achieved for the research work. Finally, the chapter concludes by providing a list of the contributions to scientific knowledge, and a list of possible publications in the course of this research work.

In Chapter two, a comprehensive literature survey of 5G networks for the Internet of Things: Communication technologies and challenges has been conducted. The study has delved into the body of knowledge on the Internet of Things as a promising paradigm for emerging and critical applications. A massive number of devices will be connected through wireless or wired communication. Therefore, this review presents the various IoT connectivity landscape solutions with diverse service requirements to achieve the modern IoT vision. In addition, the third-generation partnership projects of cellular-based low-power wide-area solutions to support and enable new service requirements for Massive to Critical IoT use cases are presented in detail, including extended coverage of global systems for mobile communications for the Internet of Things (EC-GSM-IoT), enhanced machine-type communications (eMTC) to enhance effective communication of existing cellular technologies such as the global system for mobile communications (GSM), the long-term evolution (LTE) networks, and narrowband-Internet of Things. It further delves into 5G new radio enhancements for new service requirements envisaged to support the exponential traffic growth to enable the IoT. Therefore, in this research study, the challenges and open research directions pertinent to the deployment of massive to critical IoT applications are also presented to come up with an efficient context-aware congestion control mechanism for resource-constrained IoT networks.

Chapter three further investigates the problem of congestion control and, in line with the aforementioned need for efficient resource utilisation, this research work proposes a context-aware congestion-control approach for lightweight CoAP/UDP-based Internet of Things traffic for resource-constrained networks. After having carefully identified the various components of congestion control in resource-constrained IoT networks by extending the piggyback response feature of the Constrained Application Protocol (CoAP), the chapter develops a system model that takes into consideration a number of separate but interconnected tasks as identified components, including the three-RTO estimation algorithm, dynamic SRTT and RTO overall estimation, the context-aware RTO, RTT fluctuation aware RTO, and restricted RTO shrinkage for analysis of the context-aware congestion control scheme.

When the proposed concepts have been implemented in a Cooja simulation environment, a toolset of the Contiki OS has been used to evaluate the overall network performance in terms of throughput, packet loss, delay and energy consumption against baseline CoAP congestion control specification and CoCoA+ protocol. Results from the implementation have been used to compare the benchmark protocols used for performance validation. A discussion of these is presented in Chapter four.

Chapter five calls for further research into congestion-control techniques that optimize the base CoAP parameters for optimal solutions. Therefore, this aspect of the research study investigates the various components to proffer optimal solutions and develop a system and network model. The model incorporates both random and optimal parameter-driven simulations to optimize the base CoAP parameters in order to adapt to the network traffic conditions. The random parameter-driven simulation builds on its first iteration on the base congestion control specification with a default transmission parameter.The fitness values in the PSO relies on the performance metrics obtained at the end of running the simulation. For each particle as a solution, the velocity represents the distance travelled by each particle as P_{best}, known as the local best position and the G_{best}, also called the global best position among all the particles.The proposed algorithm employs the round-trip-time (RTT) value to determine the network congestion level. In the optimal parameter-driven simulation, the optimal tuned MAX_RETRANSMIT, with the adjusted max-age values, is used, based on the traffic, to control and mitigate congestion in the network. The proposed PACT is validated against baseline CoAP with Observe using the Cooja simulation environment, a toolset of the Contiki OS to ascertain and evaluate the overall network performance of the system in terms of packet loss, delay and normalised overhead. The results obtained and the discussion show a significant improvement compared with baseline CoAP with Observe as presented in Chapter six.

In summary, the validation of the proposed mechanisms show a significant improvement in comparison with existing congestion control schemes. In a context-aware congestion control algorithm with a 3-RTO estimators, the proposed scheme shows to be effective in reducing the overall number of retransmissions while maintaining higher throughput and minimize packet loss comparable to those obtained with base CoAP and CoCoA+ in all the networks scenarios. The particle swarm optimisation-based technique for congestion control in CoAP shows significant performance improvement in terms of packet loss, delay, and normalized overhead when compare to CoAP with Observe under different network topologies. It is hope that future considerations for congestion control in resource-constrained environments will continue to prevail for the ever-growing Internet of Things networks.

The concluding part of this thesis presents a summary of contributions, and further recommendations for future works in the research area for congestion control in a resource-constrained Internet of Things network.

7.1.1 Summary of Contributions

Throughout this research study, several contributions have been made to scientific research knowledge and these can be summarised as follows:

- A comprehensive literature review surveying the emerging low-power wide-area connectivity solutions, including EC-GSM-IoT, eMTC, and NB-IoT, and other existing technologies along with their design requirements for the Internet of Things with the main focus on 5G mobile networks envisaged to support the exponential traffic growth and new service requirements, including mMTC, eMBB, critical communications, and network operations towards enabling efficient Internet of Things use cases. The outcome of this contribution has resulted in the publication of a survey paper detailing contributions across the various landscape LPWA connectivity solutions as novel communication technologies to meet the Internet of Things diverse application requirements. Based on the findings, the study has also identified crucial research gaps in the area of congestion control for the Internet of Things as a multi-objective function for efficient resource utilisation. The current statistics of research, google scholar and other platforms show that the paper in question is increasingly being downloaded, read, cited, and recommended to researchers in the field of the Internet of Things.

- A design and implementation of a context-aware congestion control scheme in Constrained Application Protocol for resource-constrained Internet of Things networks. As part of this approach, the model utilises strong, weak, and failed round-trip-time (RTT), retransmission-count-based smoothed RTT observation, lower bound retransmission timeout (RTO) restriction approach, and aging mechanism to identify the exact network status and provide adaptive congestion control based on the available information about the network status. This work has been demonstrated by using Contiki OS and the Cooja simulator. Performance evaluation has shown that the simulation results are effective in reducing the overall number of retransmissions while guaranteeing higher throughput and minimising packet loss compared with baseline CoAP congestion control specification and CoCoA+ as benchmark protocols for validation. The outcome of these evaluations has been published as a journal article in the Transactions on Emerging Telecommunications Technologies.
- A design and implementation for a particle swarm optimisation (PSO)-based algorithm for congestion control in Constrained Application Protocol for optimal solutions that significantly improves the performance of the baseline CoAP protocol by tuning the retransmission count to an optimal value along with a varying max-age value under various network conditions. The Observe option is enabled in the CoAP protocol for max-age value. By applying the optimal parameter-driven simulations at different congestion levels, the optimal MAX_RETRANSMIT and max-age value are obtained, by employing the PSO technique using Cooja simulation environment, a toolset of the Contiki-OS. The algorithm has been implemented and evaluated against baseline CoAP with Observe. Performance evaluation results have shown improvements in packet loss, delay, and normalised overhead. The contributions made by this study have been compiled and are currently under review for publication.

7.2 RECOMMENDATIONS FOR FUTURE WORK

Although, in the literature review and current work presented, a lot has been investigated and several contributions have already been made to the body of scientific knowledge about congestion control for the Internet of Things, it is very important to state that there is still some more that is open for study in this research area. The following are therefore possible recommendations for future consideration by the research community in this field of study. The recommendations are presented as possible future solutions to address the ever-growing problem of congestion control in resource-constrained Internet of Things networks.

7.2.1 Recommendation based on the use of queuing analysis for mitigating congestion control

As the entire network load of the system exceeds the Internet of Things service capacity, the overall quality of service performance of the Internet of Things service continues to degrade significantly. Considering the resource-constrained nature of the Internet of Things, a queuing management mechanisms for congestion control becomes very critical for Internet of Things service provisioning. Recently, research works in [174, 175] introduce modeling of IoT-based networks. The authors in [174] present an approach to modeling the reliability and cost of service composition on the Internet of Things based on the Markov Decision Process. Herrero, in [175] presents a two-state Markov model that can be used to identify the network packet loss based on application, and also to estimate the effect of lossy wireless channels in CoAP-based applications. A large amount of data has to be collected, transmitted, and processed on the Internet of Things for service provisioning. As a result of the massive number of smart devices to be connected to the Internet, the resource-constrained nature of the Internet of Things and also the finite capacities of system servers, congestion control plays an important role in the Internet of Things to meet requirements of the service performance. Modeling and analysing congestion control for efficient service access to the Internet of Things are very challenging and should be met in order to accommodate the ever-increasingly diverse range of traffic that is expected to confront the end-users. These issues call for a system framework for Internet of Things service provisioning as a three-state Markov system for congestion control in CoAP for resource-constrained IoT-based networks as shown in Fig. 7.1. Then a system model for congestion control for Internet of Things traffic needs to be developed, based on the system framework.

7.2.2 Recommendation based on the use of intelligent reinforcement learning for mitigating congestion control

Having envisaged the massive number of smart devices that will be connected via the Internet, the volume of data to be disseminated will increase significantly.Therefore, the need for an intelligent processing and analysis of this voluminous data is a major key to developing these smart Internet of Things applications across the industry and society at large. Future study is needed to investigate the intelligent reinforcement of learning for congestion control in resource-constrained Internet of Things networks. This calls for a proposal for a novel framework for an intelligent machine-learning-based

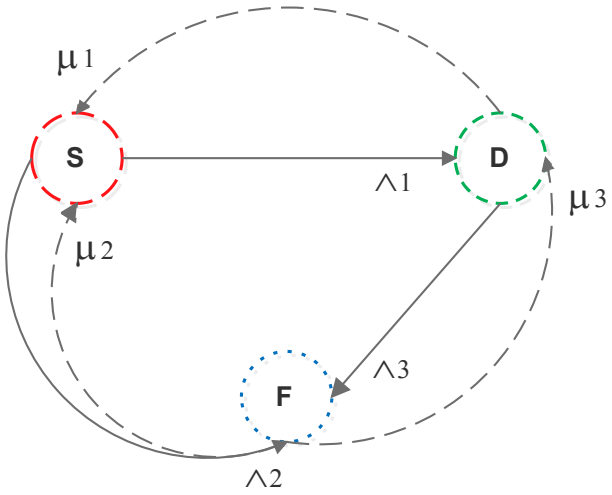

Figure 7.1. A three-state Markov model for congestion control in CoAP for IoT-based networks. Where S, D, and F represent the probability of successful, delayed, and failed transmissions respectively.

congestion control approach to Internet of Things traffic that will improve the performance of certain Internet communication components for an efficient user-friendly service experience.

REFERENCES

[1] G. A. Akpakwu, G. P. Hancke, and A. M. Abu-Mahfouz, "CACC: Context-aware congestion control approach for lightweight CoAP/UDP-based Internet of Things traffic," *Transactions on Emerging Telecommunications Technologies*, vol. 31, no. 2, pp. 1–19, 2020.

[2] K. Hartke, "Observing resources in the constrained application protocol (CoAP)," *Internet Engineering Task Force (IETF), RFC 7641*, Sept 2015.

[3] 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," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 3, pp. 510–527, 2016.

[4] E. Borgia, "The Internet of Things vision: Key features, applications and open issues," *Computer Communications*, vol. 54, no. 12, pp. 1–31, 2014.

[5] R. Want, B. N. Schilit, and S. Jenson, "Enabling the Internet of Things," *Computer*, vol. 48, no. 1, pp. 28–35, 2015.

[6] L. Atzori, A. Iera, and G. Morabito, "The Internet of Things: A survey," *Computer Networks*, vol. 54, no. 15, pp. 2787–2805, 2010.

[7] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of Things: A survey on enabling technologies, protocols, and applications," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 4, pp. 2347–2376, 2015.

[8] O. Vermesan, P. Friess, P. Guillemin, S. Gusmeroli, H. Sundmaeker, A. Bassi, I. S. Jubert, M. Mazura, M. Harrison, M. Eisenhauer *et al.*, "Internet of Things strategic research roadmap," *Internet of Things-Global Technological and Societal Trends*, vol. 1, no. 6, pp. 9–52, 2011.

[9] 3GPP, "TS 22.368 Service requirements for machine-type communications," *v. 15.5.0*, 2012.

[10] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of Things (IoT): A vision, architectural elements, and future directions," *Future Generation Computer Systems*, vol. 29, no. 7, pp. 1645–1660, 2013.

[11] G. A. Akpakwu, B. J. Silva, G. P. Hancke, and A. M. Abu-Mahfouz, "A survey on 5G networks for the Internet of Things: Communication technologies and challenges," *IEEE Access*, vol. 6, no. 2, pp. 3619–3647, 2017.

[12] Ericsson, "Cellular networks for massive IoT-enabling low power wide area applications," *Ericsson White Paper*, vol. 1, no. 1, pp. 1–13, 2016.

[13] Ericsson, "Cellular networks for massive IoT," *Ericsson White Paper*, vol. 1, no. 1, pp. 1–16, 2020.

[14] N. Oyj, "LTE evolution for IoT connectivity," *Nokia Corporation White Paper*, 2016.

[15] E. Berthelsen and J. Morrish, "Forecasting the Internet of Things revenue opportunity," *Machina Research, Tech. Rep.*, 2015.

[16] H. Boyes, B. Hallaq, J. Cunningham, and T. Watson, "The industrial Internet of Things (IIoT): An analysis framework," *Computers in industry*, vol. 101, pp. 1–12, 2018.

[17] H. I. Kobo, A. M. Abu-Mahfouz, and G. P. Hancke, "A survey on software-defined wireless sensor networks: Challenges and design requirements," *IEEE Access*, vol. 5, no. 2, pp. 1872–1899, 2017.

[18] Y. G. Iyer, S. Gandham, and S. Venkatesan, "STCP: A generic transport layer protocol for wireless sensor networks," in *Proceedings of the 14th International Conference on Computer Communications and Networks (ICCCN)*, Oct 2005, pp. 449–454.

[19] F. Javed, M. K. Afzal, M. Sharif, and B.-S. Kim, "Internet of Things (IoT) operating systems support, networking technologies, applications, and challenges: A comparative review," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 3, pp. 2062–2100, 2018.

[20] L. Eggert, "Congestion control for the constrained application protocol (coap)," 2011, draft-eggert-core-congestion-control-01 (work in progress) [Online]. Available: https://tools.ietf.org/id/draft-eggert-core-congestion-control-01.txt (accessed Jan. 4, 2019).

[21] O. B. Akan and I. F. Akyildiz, "Event-to-sink reliable transport in wireless sensor networks," *IEEE/ACM transactions on networking*, vol. 13, no. 5, pp. 1003–1016, 2005.

[22] A. Dhraief, A. Belghith, H. Mathkour, and K. Drira, "An m2m gateway-centric architecture for autonomic healing and optimising of machine-to-machine overlay networks," *International Journal of Ad Hoc and Ubiquitous Computing*, vol. 26, no. 1, pp. 12–28, 2017.

[23] R. K. R. Kummitha and N. Crutzen, "How do we understand smart cities? an evolutionary perspective," *Cities*, vol. 67, no. 5.

[24] Z. M. Fadlullah, M. M. Fouda, N. Kato, A. Takeuchi, N. Iwasaki, and Y. Nozaki, "Toward intelligent machine-to-machine communications in smart grid," *IEEE Communications Magazine*, vol. 49, no. 4, pp. 60–65, 2011.

[25] M. Faheem, S. B. H. Shah, R. A. Butt, B. Raza, M. Anwar, M. W. Ashraf, M. A. Ngadi, and V. C. Gungor, "Smart grid communication and information technologies in the perspective of industry 4.0: Opportunities and challenges," *Computer Science Review*, vol. 30, no. 8, pp. 1–30, 2018.

[26] N. Komninos, E. Philippou, and A. Pitsillides, "Survey in smart grid and smart home security: Issues, challenges and countermeasures," *IEEE Communications Surveys & Tutorials*, vol. 16,

no. 4, pp. 1933–1954, 2014.

[27] W. Ali, G. Dustgeer, M. Awais, and M. A. Shah, "Iot based smart home: Security challenges, security requirements and solutions," in *Proceedings of the 23rd International Conference on Automation and Computing (ICAC)*. IEEE, 2017, pp. 1–6.

[28] A. M. Abu-Mahfouz, T. O. Olwal, A. M. Kurien, J. L. Munda, and K. Djouani, "Toward developing a distributed autonomous energy management system (DAEMS)," in *Proceedings of the IEEE AFRICON*, Sep 2015, pp. 1–6.

[29] P. Dongbaare, S. D. Chowdhury, T. Olwal, and A. Abu-Mahfouz, "Smart energy management system based on an automated distributed load limiting mechanism and multi-power switching technique," in *Proceedings of the 51st Int. Univ. Power Eng. Conf*, Sep 2016, pp. 1–x.

[30] B. Silva, R. M. Fisher, A. Kumar, and G. P. Hancke, "Experimental link quality characterization of wireless sensor networks for underground monitoring," *IEEE Transactions on Industrial Informatics*, vol. 11, no. 5, pp. 1099–1110, 2015.

[31] B. Cheng, L. Cui, W. Jia, W. Zhao, and P. H. Gerhard, "Multiple region of interest coverage in camera sensor networks for tele-intensive care units," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 6, pp. 2331–2341, 2016.

[32] K. S. E. Phala, A. Kumar, and G. P. Hancke, "Air quality monitoring system based on ISO/IEC/IEEE 21451 standards," *IEEE Sensors Journal*, vol. 16, no. 12, pp. 5037–5045, 2016.

[33] A. M. Abu-Mahfouz, Y. Hamam, P. R. Page, K. B. Adedeji, A. O. Anele, and E. Todini, "Real-time dynamic hydraulic model of water distribution networks," *Water*, vol. 11, no. 3, p. 470, 2019.

[34] J. Jermyn, R. P. Jover, I. Murynets, M. Istomin, and S. Stolfo, "Scalability of machine to machine systems and the Internet of Things on LTE mobile networks," in *Proceedings of the 16th IEEE International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, Jun 2015, pp. 1–9.

[35] F. Moreu, R. E. Kim, and B. Spencer Jr, "Railroad bridge monitoring using wireless smart sensors," *Structural Control and Health Monitoring*, vol. 24, no. 2, p. e1863, 2017.

[36] D. J. Cook, A. S. Crandall, B. L. Thomas, and N. C. Krishnan, "CASAS: A smart home in a box," *Computer*, vol. 46, no. 7, pp. 62–69, 2012.

[37] S. Mahmud, S. Ahmed, and K. Shikder, "A smart home automation and metering system using Internet of Things (IoT)," in *PInternational Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)*. IEEE, 2019, pp. 451–454.

[38] M. B. Yassein, W. Mardini, and A. Khalil, "Smart homes automation using Z-wave protocol," in *Proceedings of the International Conference on Engineering & MIS (ICEMIS)*, Sep 2016, pp. 1–6.

[39] S. Li, S. Zhao, P. Yang, P. Andriotis, L. Xu, and Q. Sun, "Distributed consensus algorithm for events detection in cyber-physical systems," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 2299–2308, 2019.

[40] L. Yongfu, S. Dihua, L. Weining, and Z. Xuebo, "A service-oriented architecture for the transportation cyber-physical systems," in *Proceedings of the 31st Chinese Control Conference*, Jul 2012, pp. 7674–7678.

[41] C. Urmson, "The self-driving car logs more miles on new wheels," 2012, [Online]. Available: https://googleblog.blogspot.com/2012/08/the-self-driving-car-logs-more-miles-on.html (accessed Sept. 3, 2017).

[42] B. N. Silva, M. Khan, and K. Han, "Towards sustainable smart cities: A review of trends, architectures, components, and open challenges in smart cities," *Sustainable Cities and Society*, vol. 38, no. 2, pp. 697–713, 2018.

[43] A. M. S. Osman, "A novel big data analytics framework for smart cities," *Future Generation Computer Systems*, vol. 91, pp. 620–633, 2019.

[44] G. P. Hancke, B. D. C. Silva, and G. P. Hancke Jr, "The role of advanced sensing in smart cities," *Sensors*, vol. 13, no. 1, pp. 393–425, 2013.

[45] D. Amaxilatis, I. Chatzigiannakis, C. Tselios, N. Tsironis, N. Niakas, and S. Papadogeorgos, "A smart water metering deployment based on the fog computing paradigm," *Applied Sciences*, vol. 10, no. 6, p. 1965, 2020.

[46] C. F. Pasluosta, H. Gassner, J. Winkler, J. Klucken, and B. M. Eskofier, "An emerging era in the management of Parkinson's disease: Wearable technologies and the Internet of Things," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 6, pp. 1873–1881, 2015.

[47] S. B. Baker, W. Xiang, and I. Atkinson, "Internet of Things for smart healthcare: Technologies, challenges, and opportunities," *IEEE Access*, vol. 5, no. 11.

[48] Radical-7, "Breakthrough measurements [Internet]," 2013, radical monitor, Masimo Corp., [Online]. Available: http://vasinc.net/pdf/LAB7293a_Sell_Sheet_Radical-7.pdf (accessed Sept. 3, 2017).

[49] C. Nay, "Sensors remind doctors to wash up," 2013, IBM Res., [Online]. Available: http://ibm.com/blogs/research/2013/11/sensors-remind-doctors-to-wash-up/ (accessed Sept. 3, 2017).

[50] K. Michaelsen, J. L. Sanders, S. M. Zimmer, and G. M. Bump, "Overcoming patient barriers to discussing physician hand hygiene: Do patients prefer electronic reminders to other methods?" *Infection Control & Hospital Epidemiology*, vol. 34, no. 9, pp. 929–934, 2013.

[51] E. Sisinni, A. Saifullah, S. Han, U. Jennehag, and M. Gidlund, "Industrial Internet of Things: Challenges, opportunities, and directions," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 11, pp. 4724–4734, 2018.

[52] V. Ç. Güngör and G. P. Hancke, *Industrial wireless sensor networks: Applications, protocols, and standards*. CRC Press, 2013.

[53] T. Chang, P. Tuset-Peiro, X. Vilajosana, and T. Watteyne, "OpenWSN & OpenMote: Demo'ing a complete ecosystem for the industrial Internet of Things," in *Proceedings of the 13th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, Jun 2016, pp. 1–3.

[54] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of Things for smart cities," *IEEE Internet of Things Journal*, vol. 1, no. 1, pp. 22–32, 2014.

[55] Y. Mehmood, F. Ahmad, I. Yaqoob, A. Adnane, M. Imran, and S. Guizani, "Internet-of-things-based smart cities: Recent advances and challenges," *IEEE Communications Magazine*, vol. 55, no. 9, pp. 16–24, 2017.

[56] 5G Americas, "LTE and 5G technologies enabling the Internet of Things," Tech. Rep., Dec 2016.

[57] H. Karl and A. Willig, *Protocols and architectures for wireless sensor networks*. John Wiley & Sons, 2007.

[58] W. Ejaz, M. Naeem, A. Shahid, A. Anpalagan, and M. Jo, "Efficient energy management for the Internet of Things in smart cities," *IEEE Communications Magazine*, vol. 55, no. 1, pp. 84–91, 2017.

[59] J. Zhou, Z. Cao, X. Dong, and A. V. Vasilakos, "Security and privacy for cloud-based IoT: Challenges," *IEEE Communications Magazine*, vol. 55, no. 1, pp. 26–33, 2017.

[60] L. Vangelista, A. Zanella, and M. Zorzi, "Long-range IoT technologies: The dawn of LoRa," in *Proceedings of the Future Access Enablers of Ubiquitous and Intelligent Infrastructures*, Sep 2015, pp. 51–58.

[61] N. Sornin, M. Luis, T. Eirich, T. Kramp, and O. Hersent, "LoRaWAN specification," *LoRa Alliance*, 2015.

[62] J. Xu, J. Yao, L. Wang, Z. Ming, K. Wu, and L. Chen, "Narrowband Internet of Things: Evolutions, technologies, and open issues," *IEEE Internet of Things Journal*, vol. 5, no. 3, pp. 1449–1462, 2017.

[63] A. M. Yousuf, E. M. Rochester, B. Ousat, and M. Ghaderi, "Throughput, coverage and scalability of LoRa LPWAN for Internet of Things," in *Proceedings of the IEEE/ACM 26th International Symposium on Quality of Service (IWQoS)*. IEEE, 2018, pp. 1–10.

[64] J. C. Zuniga and B. Ponsard, "Sigfox system description," *LPWAN@ IETF97*, vol. 25, Nov 2016.

[65] M. Centenaro, L. Vangelista, A. Zanella, and M. Zorzi, "Long-range communications in unlicensed bands: The rising stars in the IoT and smart city scenarios," *IEEE Wireless Communications*, vol. 23, no. 5, pp. 60–67, 2016.

[66] J. Burns, S. Kirtay, and P. Marks, "Future use of licence exempt radio spectrum," *Plum Consulting, London, UK, Tech. Rep*, 2015.

[67] J.-P. Bardyn, T. Melly, O. Seller, and N. Sornin, "IoT: The era of LPWAN is starting now," in *Proceedings of the 42nd European Solid-State Circuits Conference (ESSCIRC)*, Sep 2016, pp. 25–30.

[68] Ingenu Tech, "RPMA technology for the Internet of Things," 2017, [Online]. Available: https://theinternetofthings.report/Resources/Whitepapers/4cbc5e5e-6ef8-4455-b8cd-f6e3888624cb_RPMA%20Technology.pdf (accessed Sept. 3, 2017).

[69] T. J. Myers, D. T. Werner, K. C. Sinsuan, J. R. Wilson, S. L. Reuland, P. M. Singler, and M. J. Huovila, "Light monitoring system using a random phase multiple access system," Jul. 2 2013, US Patent 8,477,830.

[70] R. G. Cid-Fuentes, M. Y. Naderi, R. Doost-Mohammady, K. R. Chowdhury, A. Cabellos-Aparicio, and E. Alarcón, "Leveraging deliberately generated interferences for multi-sensor

wireless RF power transmission," in *Proceedings of the IEEE Global Communications Conference (GLOBECOM)*, Dec 2015, pp. 1–6.

[71] W. Ayoub, A. E. Samhat, F. Nouvel, M. Mroue, and J.-C. Prévotet, "Internet of mobile things: Overview of LoRaWAN, DASH7, and NB-IoT in LPWANs standards and supported mobility," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 2, pp. 1561–1581, 2018.

[72] O. Cetinkaya and O. B. Akan, "A DASH7-based power metering system," in *Proceedings of the 12th Annual IEEE Consumer Communications and Networking Conference (CCNC)*, Jan 2015, pp. 406–411.

[73] W. Webb, "Weightless: The technology to finally realise the M2M vision," *International Journal of Interdisciplinary Telecommunications and Networking (IJITN)*, vol. 4, no. 2, pp. 30–37, 2012.

[74] K.-H. Chang, "Bluetooth: a viable solution for IoT? [industry perspectives]," *IEEE Wireless Communications*, vol. 21, no. 6, pp. 6–7, 2014.

[75] M. Marawaha, P. Jha, R. Razdan, J. Dukes, S. Sheehan, and E. O. Nuallain, "Performance evaluation of Bluetooth low energy communication," *Journal of Information Sciences and Computing Technologies*, vol. 7, no. 2, pp. 718–725, 2018.

[76] M. Siekkinen, M. Hiienkari, J. K. Nurminen, and J. Nieminen, "How low energy is Bluetooth low energy? Comparative measurements with Zigbee/802.15.4," in *Proceedings of the IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, Apr 2012, pp. 232–237.

[77] *IEEE Standard for part 15.4: Wireless medium access control (MAC) and physical layer (PHY) specifications for low-rate wireless personal area networks (LR-WPANs)*, IEEE Standard 802.4-2003, 2003.

[78] ZigBee Alliance, "Zigbee specifications," 2012, [Online]. Available: http://www.ZigBee.org (accessed Sept. 8, 2017).

[79] *IEEE Standard for information technology–Telecommunications and information exchange between systems–Local and metropolitan area networks–Specific requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 6: Wireless Access in Vehicular Environments*, IEEE Standard 802.11ah-2010 (Amendment to IEEE Standard 802.11-2010), 2010.

[80] S. Andreev, O. Galinina, A. Pyattaev, M. Gerasimenko, T. Tirronen, J. Torsner, J. Sachs, M. Dohler, and Y. Koucheryavy, "Understanding the IoT connectivity landscape: A contemporary M2M radio technology roadmap," *IEEE Communications Magazine*, vol. 53, no. 9, pp. 32–40, 2015.

[81] U. Raza, P. Kulkarni, and M. Sooriyabandara, "Low power wide area networks: An overview," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 2, pp. 855–873, 2017.

[82] T. Adame, A. Bel, B. Bellalta, J. Barcelo, and M. Oliver, "IEEE 802.11AH: The WiFi approach for M2M communications," *IEEE Wireless Communications*, vol. 21, no. 6, pp. 144–152, 2014.

[83] A. Attour, T. Burger-Helmchen, I. Capdevila, and M. I. Zarlenga, "Smart city or smart citizens? The Barcelona case," *Journal of Strategy and Management*, vol. 1, no. 8, pp. 1–17, 2015.

[84] S. Dimatteo, P. Hui, B. Han, and V. O. Li, "Cellular traffic offloading through WiFi networks," in *Proceedings of the Eighth IEEE International Conference on Mobile Ad-Hoc and Sensor Systems*, Oct 2011, pp. 192–201.

[85] 3GPP, "TR 36.888 Study on provision of low-cost machine-type communications (MTC) user equipments (UEs) based on LTE," *v. 12. 0.0*, June 2013.

[86] 3GPP, "TR 45.820 Cellular system support for ultra low complexity and low throughput Internet of Things (CIoT)," *v. 13. 1.0*, Nov 2015.

[87] Ericsson and Nokia Networks, "Further LTE physical layer enhancements for MTC," 2016, rP-141660, 3GPP TSG RAN Meeting # 65, Sept. [Online]. Available:

https://www.3gpp.org/ftp/tsg_ran/tsg_ran/TSGR_65/Docs/RP-141660.zip. (accessed Sept. 3, 2017).

[88] P. Stuckmann, *The GSM evolution: Mobile packet data services*. John Wiley & Sons, 2002.

[89] E. Dahlman, S. Parkvall, and J. Skold, *4G: LTE/LTE-Advanced for mobile broadband*. Oxford, Academic Press, 2013.

[90] N. Mangalvedhe, R. Ratasuk, and A. Ghosh, "NB-IoT deployment study for low power wide area cellular IoT," in *Proceedings of the IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Sep 2016, pp. 1–6.

[91] Z. Shelby, K. Hartke, and C. Bormann, "The constrained application protocol (CoAP)," *Internet Engineering Task Force (IETF), RFC 7252*, Jun 2014.

[92] C. Bormann, M. Ersue, and A. Keranen, "Terminology for constrained-node networks," *Internet Engineering Task Force (IETF), RFC 7228*, May 2014.

[93] R. T. Fielding, "Architectural styles and the design of network-based software architectures," Ph.D. dissertation, University of California, Irvine, 2000.

[94] H. Balakrishnan, V. N. Padmanabhan, S. Seshan, and R. H. Katz, "A comparison of mechanisms for improving TCP performance over wireless links," *IEEE/ACM Transactions on Networking*, vol. 5, no. 6, pp. 756–769, 1997.

[95] G. Holland and N. Vaidya, "Analysis of TCP performance over mobile ad hoc networks," *Wireless Networks*, vol. 8, no. 2-3, pp. 275–288, 2002.

[96] J. W. Hui and D. E. Culler, "IP is dead, long live IP for wireless sensor networks," in *Proceedings of the 6th ACM Conference on Embedded Network Sensor Systems*, Nov 2008, pp. 15–28.

[97] M. Kovatsch, "Scalable web technology for the Internet of Things," Ph.D. dissertation, ETH Zurich, 2015.

[98] R. Fielding and J. Reschke, "Hypertext transfer protocol (HTTP/1.1): Message syntax and routing," *Internet Engineering Task Force (IETF), RFC 7230*, Jun 2014.

[99] J. Hui and R. Kelsey, "Multicast protocol for low power and lossy networks (MPL)," 2014, draft-Ietf-Roll-Trickle-Mcast-09, Internet Engineering Task Force (IETF), [Online]. Available: https://tools.ietf.org/pdf/draft-ietf-roll-trickle-mcast-09.pdf (accessed June 8, 2020).

[100] B. Fenner, M. Handley, H. Holbrook, I. Kouvelas, R. Parekh, Z. Zhang, and L. Zheng, "Protocol independent multicast-sparse mode (PIM-SM): Protocol specification (Revised)." *Internet Engineering Task Force (IETF), RFC 7761*, Mar 2016.

[101] A. Rahman and E. Dijk, "Group communication for the constrained application protocol (CoAP)," *Internet Engineering Task Force (IETF), RFC 7390*, Oct 2014.

[102] T. Dierks and E. Rescorla, "The transport layer security (TLS) protocol version 1.2. (Proposed Standard)," *Network Working Group (NWG), RFC 5246*, Aug 2008.

[103] E. Rescorla and N. Modadugu, "Datagram transport layer security version 1.2 (Proposed Standard)," *Internet Engineering Task Force (IETF), RFC 6347*, Jan 2012.

[104] P. Wouters, H. Tschofenig, J. Gilmore, S. Weiler, and T. Kivinen, "Using raw public keys in transport layer security (TLS) and datagram transport layer security (DTLS)," *Internet Engineering Task Force (IETF), RFC 7250*, Jun 2014.

[105] C. Bormann, A. Betzler, C. Gomez, and I. Demirkol, "CoAP simple congestion control/advanced," 2014, draft-bormann-core-cocoa-02, Internet Engineering Task Force (IETF) CoRE Working Group, [Online]. Available: https://tools.ietf.org/id/draft-bormann-core-cocoa-02.txt (accessed June 8, 2018).

[106] P. Karn and C. Partridge, "Improving round-trip time estimates in reliable transport protocols," *ACM SIGCOMM Computer Communication Review*, vol. 17, no. 5, pp. 2–7, 1987.

[107] V. Paxson, M. Allman, J. Chu, and M. Sargent, "Computing TCP's retransmission timer," RFC 6298., Tech. Rep., 2011.

[108] A. Betzler, C. Gomez, I. Demirkol, and M. Kovatsch, "Congestion control for CoAP cloud services," in *Proceedings of the IEEE Emerging Technology and Factory Automation (ETFA)*, Sep 2014, pp. 1–6.

[109] 3GPP, "TR 22.891 Feasibility study on new service and market technology enablers," *v. 14.2.0*, 2016.

[110] 3GPP, "TR 22.861 Feasibility study on new service and market technology enablers for massive Internet of Things," *v. 14.1.0*, 2016.

[111] 3GPP, "TR 22.863 Feasibility study on new service and market technology enablers for enhanced mobile broadband," *Stage 1, v. 14.1.0*, 2016.

[112] 3GPP, "TR 22.862 Feasibility study on new service and market technology enablers for critical communications," *Stage 1, v. 14.1.0*, 2016.

[113] 3GPP, "TR 22.864 Feasibility study on new service and market technology enablers for network operation," *Stage 1, v. 15.0.0*, 2016.

[114] NGMN Alliance, "NGMN 5G White Paper," *Next Generation Mobile Networks*, vol. 1, no. 2, pp. 1–125, 2015.

[115] M. Ndiaye, G. P. Hancke, and A. M. Abu-Mahfouz, "Software defined networking for improved wireless sensor network management: A survey," *Sensors*, vol. 17, no. 5, p. 1031, 2017.

[116] K. M. Modieginyane, B. B. Letswamotse, R. Malekian, and A. M. Abu-Mahfouz, "Software defined wireless sensor networks application opportunities for efficient network management: A survey," *Computers & Electrical Engineering*, vol. 66, no. 2, pp. 274–287, 2018.

[117] I. F. Akyildiz, A. Lee, P. Wang, M. Luo, and W. Chou, "A roadmap for traffic engineering in SDN-OpenFlow networks," *Computer Networks*, vol. 71, no. 10, pp. 1–30, 2014.

[118] W. Xia, Y. Wen, C. H. Foh, D. Niyato, and H. Xie, "A survey on software-defined networking," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 1, pp. 27–51, 2014.

[119] I. F. Akyildiz, P. Wang, and S.-C. Lin, "SoftAir: A software defined networking architecture for 5G wireless systems," *Computer Networks*, vol. 85, no. 7, pp. 1–18, 2015.

[120] F. Granelli, A. A. Gebremariam, M. Usman, F. Cugini, V. Stamati, M. Alitska, and P. Chatzimisios, "Software defined and virtualized wireless access in future wireless networks: Scenarios and standards," *IEEE Communications Magazine*, vol. 53, no. 6, pp. 26–34, 2015.

[121] H.-H. Cho, C.-F. Lai, T. K. Shih, and H.-C. Chao, "Integration of SDR and SDN for 5G," *IEEE Access*, vol. 2, no. 9, pp. 1196–1204, 2014.

[122] A. Hakiri and P. Berthou, "Leveraging SDN for the 5G networks: Trends, prospects and challenges," *arXiv preprint arXiv:1506.02876*, vol. 1, no. 6, pp. 1–23, 2015.

[123] W. H. Chin, Z. Fan, and R. Haines, "Emerging technologies and research challenges for 5G wireless networks," *IEEE Wireless Communications*, vol. 21, no. 2, pp. 106–112, 2014.

[124] M. Jacobsson and C. Orfanidis, "Using software-defined networking principles for wireless sensor networks," in *Proceedings of the SNCNW*, May 2015, pp. 1–5.

[125] I. F. Akyildiz, S.-C. Lin, and P. Wang, "Wireless software-defined networks (W-SDNs) and network function virtualization (NFV) for 5G cellular systems: An overview and qualitative evaluation," *Computer Networks*, vol. 93, no. 12, pp. 66–79, 2015.

[126] Ericsson, "Cloud RAN," *Ericsson White Paper*, pp. 1–11, 2015.

[127] *Mobile broadband: The benefits of additional spectrum*, Federal Communications Commission (OBI tech. paper), 2010.

[128] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "Next generation/dynamic spectrum access/cognitive radio wireless networks: A survey," *Computer Networks*, vol. 50, no. 13, pp. 2127–2159, 2006.

[129] *Unlicensed operations in the TV broadcast bands, second memorandum opinion and order*, Federal Communications Commission (ET Docket 10-174), 2010.

[130] D. Thangavel, X. Ma, A. Valera, H.-X. Tan, and C. K.-Y. Tan, "Performance evaluation of MQTT and CoAP via a common middleware," in *Proceedings of the IEEE ninth International Conference on Intelligent Sensors, Sensor Setworks and Information Processing (ISSNIP)*, Apr 2014, pp. 1–6.

[131] T. Pötsch, K. Kuladinithi, M. Becker, P. Trenkamp, and C. Goerg, "Performance evaluation of CoAP using RPL and LPL in TinyOS," in *Proceedings of the 5th International Conference on New Technologies, Mobility and Security (NTMS)*, May 2012, pp. 1–5.

[132] T. Heo, K. Kim, H. Kim, C. Lee, J. H. Ryu, Y. T. Leem, J. A. Jun, C. Pyo, S.-M. Yoo, and J. Ko, "Escaping from ancient Rome! applications and challenges for designing smart cities," *Transactions on Emerging Telecommunications Technologies*, vol. 25, no. 1, pp. 109–119, 2014.

[133] R. Petrolo, V. Loscri, and N. Mitton, "Towards a smart city based on cloud of things, a survey on the smart city vision and paradigms," *Transactions on Emerging Telecommunications Technologies*, vol. 28, no. 1, pp. 1–12, 2017.

[134] A. Betzler, C. Gomez, I. Demirkol, and J. Paradells, "CoCoA+: An advanced congestion control mechanism for CoAP," *Ad Hoc Networks*, vol. 33, pp. 126–139, 2015.

[135] M. Kovatsch, S. Duquennoy, and A. Dunkels, "A low-power CoAP for Contiki," in *Proceedings of the IEEE Eighth International Conference on Mobile Ad-Hoc and Sensor Systems*, Oct 2011, pp. 855–860.

[136] W. Shang, Y. Yu, R. Droms, and L. Zhang, "Challenges in IoT networking via TCP/IP architecture," *Technical Report NDN-0038. NDN Project*, 2016.

[137] A. Betzler, C. Gomez, I. Demirkol, and J. Paradells, "Congestion control in reliable CoAP communication," in *Proceedings of the 16th ACM International Conference on Modeling, Analysis & Simulation of Wireless and Mobile Systems*, Nov 2013, pp. 365–372.

[138] W. Colitti, K. Steenhaut, N. De Caro, B. Buta, and V. Dobrota, "Evaluation of constrained application protocol for wireless sensor networks," in *Proceedings of the 18th IEEE Workshop on Local & Metropolitan Area Networks (LANMAN)*, Oct 2011, pp. 1–6.

[139] I. Järvinen, L. Daniel, and M. Kojo, "Experimental evaluation of alternative congestion control algorithms for constrained application protocol (CoAP)," in *Proceedings of the 2nd IEEE World Forum on Internet of Things (WF-IoT)*, Dec 2015, pp. 453–458.

[140] A. Betzler, C. Gomez, I. Demirkol, and J. Paradells, "CoAP congestion control for the Internet of Things," *IEEE Communications Magazine*, vol. 54, no. 7, pp. 154–160, 2016.

[141] S. Bolettieri, C. Vallati, G. Tanganelli, and E. Mingozzi, "Highlighting some shortcomings of the CoCoA+ congestion control algorithm," in *Proceedings of the International Conference on Ad-Hoc Networks and Wireless*, Sep 2017, pp. 213–220.

[142] L. Eggert and G. Fairhurst, "Unicast UDP usage guidelines for application designers," *Internet Society*, 2008.

[143] E. Dashkova and A. Gurtov, "Survey on congestion control mechanisms for wireless sensor networks," in *Internet of Things, Smart Spaces, and Next Generation Networking*. Springer, 2012, pp. 75–85.

[144] A. Betzler, C. Gomez, I. Demirkol, and M. Kovatsch, "Congestion control for CoAP cloud services," in *Proceedings of the IEEE Emerging Technology and Factory Automation (ETFA)*, Sep 2014, pp. 1–6.

[145] E. Ancillotti and R. Bruno, "Comparison of CoAP and CoCoA+ congestion control mechanisms for different IoT application scenarios," in *Proceedings of the IEEE Symposium on Computers and Communications (ISCC)*, Jul 2017, pp. 1186–1192.

[146] V. Rathod, N. Jeppu, S. Sastry, S. Singala, and M. P. Tahiliani, "CoCoA++: Delay gradient based congestion control for Internet of Things," *Future Generation Computer Systems*, vol. 100, no. 11, pp. 1053–1072, 2019.

[147] S. Bolettieri, G. Tanganelli, C. Vallati, and E. Mingozzi, "pCoCoA: A precise congestion control algorithm for CoAP," *Ad Hoc Networks*, vol. 80, no. 11, pp. 116–129, 2018.

[148] R. Bhalerao, S. S. Subramanian, and J. Pasquale, "An analysis and improvement of congestion control in the CoAP Internet-of-Things protocol," in *Proceedings of the 13th IEEE Annual Consumer Communications & Networking Conference (CCNC)*, Jan 2016, pp. 889–894.

[149] E. Ancillotti, R. Bruno, C. Vallati, and E. Mingozzi, "Design and evaluation of a rate-based congestion control mechanism in CoAP for IoT applications," in *Proceedings of the 19th IEEE International Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM)*, Jun 2018, pp. 14–15.

[150] J. J. Lee, K. T. Kim, and H. Y. Youn, "Enhancement of congestion control of constrained application protocol/congestion control/advanced for Internet of Things environment," *International Journal of Distributed Sensor Networks*, vol. 12, no. 11, pp. 1–13, 2016.

[151] I. Jarvinen, I. Raitahila, Z. Cao, and M. Kojo, "Fasor retransmission timeout and congestion control mechanism for CoAP," in *Proceedings of the IEEE Global Communications Conference (GLOBECOM)*, Dec 2018, pp. 1–7.

[152] R. Herrero, "Dynamic CoAP mode control in real time wireless IoT networks," *IEEE Internet of Things Journal*, vol. 6, no. 1, pp. 801–807, 2018.

[153] A. Dunkels, B. Gronvall, and T. Voigt, "Contiki - a lightweight and flexible operating system for tiny networked sensors," in *Proccedings of the Annual IEEE International Conference on Local Computer Networks*, Nov 2004, pp. 455–462.

[154] F. Osterlind, A. Dunkels, J. Eriksson, N. Finne, and T. Voigt, "Cross-level sensor network simulation with Cooja," in *Proceedings of the 31st IEEE Conference on Local Computer*

Networks, Nov 2006, pp. 641–648.

[155] Texas Instruments, "Chipcon products: CC2420 Datasheet: 2.4 GHz IEEE 802.15.4/ZigBee-ready RF Transceiver," Mar 2013, [Online]. Available: https://www.ti.com/lit/ds/symlink/cc2420.pdf (accessed Jan. 4, 2019).

[156] Moteiv Coorperation, "Tmote Sky: Ultra low power IEEE 802.15.4 compliant wireless sensor module," 2006, [Online]. Available: http://www. sentilla. com/pdf/eol/tmote-skydatasheet. pdf. (accessed Nov. 25, 2017).

[157] G. A. Akpakwu, G. P. Hancke, and A. M. Abu-Mahfouz, "PACT: An Optimisation-based adaptive congestion control technique for constrained application protocol," *International Journal of Network Management*, 2020, under review.

[158] A. Paul and R. Jeyaraj, "Internet of Things: A primer," *Human Behavior and Emerging Technologies*, vol. 1, no. 1, pp. 37–47, 2019.

[159] E. Siow, T. Tiropanis, and W. Hall, "Analytics for the Internet of Things: A survey," *ACM Computing Surveys (CSUR)*, vol. 51, no. 4, p. 74, 2018.

[160] F. Kiani, "A survey on management frameworks and open challenges in IoT," *Wireless Communications and Mobile Computing*, vol. 2018, no. 8, pp. 1–33, 2018.

[161] A. Sehgal, V. Perelman, S. Kuryla, and J. Schonwalder, "Management of resource constrained devices in the Internet of Things," *IEEE Communications Magazine*, vol. 50, no. 12, pp. 144–149, 2012.

[162] S. Verma, Y. Kawamoto, Z. M. Fadlullah, H. Nishiyama, and N. Kato, "A survey on network methodologies for real-time analytics of massive IoT data and open research issues," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 3, pp. 1457–1477, 2017.

[163] A. A. O. Bahashwan and S. Manickam, "A brief review of messaging protocol standards for Internet of Things (IoT)," *Journal of Cyber Security and Mobility*, vol. 8, no. 1, pp. 1–14, 2019.

[164] V. Karagiannis, P. Chatzimisios, F. Vazquez-Gallego, and J. Alonso-Zarate, "A survey on application layer protocols for the Internet of Things," *Transactions on IoT and Cloud Computing*, vol. 3, no. 1, pp. 11–17, 2015.

[165] C. Bormann, A. P. Castellani, and Z. Shelby, "CoAP: An application protocol for billions of tiny Internet nodes," *IEEE Internet Computing*, vol. 16, no. 2, pp. 62–67, 2012.

[166] N. Kushalnagar, G. Montenegro, C. Schumacher *et al.*, "IPv6 over low-power wireless personal area networks (6LoWPANs): Overview, assumptions, problem statement, and goals," *Internet Engineering Task Force (IETF), RFC 4919 (Informational)*, 2007.

[167] F. Ouakasse and S. Rakrak, "A comparative study of MQTT and CoAP application layer protocols via performances evaluation," *Journal of Engineering and Applied Sciences*, vol. 13, no. 15, pp. 6053–6061, 2018.

[168] J. Misic, M. Z. Ali, and V. B. Misic, "Protocol architectures for IoT domains," *IEEE Network*, vol. 32, no. 4, pp. 81–87, 2018.

[169] A. Betzler, J. Isern, C. Gomez, I. Demirkol, and J. Paradells, "Experimental evaluation of congestion control for CoAP communications without end-to-end reliability," *Ad Hoc Networks*, vol. 52, no. 12, pp. 183–194, 2016.

[170] E. Ancillotti, S. Bolettieri, and R. Bruno, "RTT-based congestion control for the Internet of Things," in *Proceedings of the International Conference on Wired/Wireless Internet Communication*, Jun 2018, pp. 3–15.

[171] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of the 95th IEEE International Conference on Neural Networks (ICNN)*, Nov 1995, pp. 1942–1948.

[172] R. Eberhart and J. Kennedy, "A new optimizer using particle swarm theory," in *Proceedings of the Sixth IEEE International Symposium on Micro Machine and Human Science (MHS's)*, Oct 1995, pp. 39–43.

[173] X. Hu, "PSO Tutorial," 2006, [Online].Available: http://www.swarmintelligence.org/tutorials.php (accessed June 13, 2019).

[174] L. Li, Z. Jin, G. Li, L. Zheng, and Q. Wei, "Modeling and analyzing the reliability and cost of service composition in the IoT: A probabilistic approach," in *Proceedings of the 19th IEEE International Conference on Web Services*, Jun 2012, pp. 584–591.

[175] R. Herrero, "Analytical model of iot coap traffic," *Digital Communications and Networks*, vol. 5, no. 2, pp. 63–68, 2019.