Reliable Architecture for Future Smart Communities.

Malarski, Krzysztof Mateusz

Publication date:
2021

Document Version
Publisher's PDF, also known as Version of record

Citation (APA):
Reliable Architecture for Future Smart Communities
PhD thesis

Author: Krzysztof Mateusz Malarski

Kongens Lyngby, 2021
Supervisor: Assoc. Prof. Sarah Ruepp
Co-supervisors: Assoc. Prof. Martin Nordal Petersen
& Assoc. Prof. Michael Stübert Berger

DTU Fotonik
Department of Photonics Engineering
Network Technologies and Service Platforms Group
Technical University of Denmark
Ørsted's Plads
Building 343
2800 Kongens Lyngby, Denmark
Summary

Smart communities of the future require a reliable communications infrastructure, which is able to provide seamless connectivity and at the same time is resistant to failures and secured against hacker attacks. A new subset of Internet of Things (IoT) technologies, called Low-Power Wide Area Network (LP-WAN), has been developed to address such needs. However, the current design of the most prominent LP-WAN technologies is not fully ready to live up to the smart community expectations. This thesis describes the efforts that evaluate the coverage, failure-safety and security of the LP-WAN protocols: Narrowband Internet of Things (NB-IoT), LTE for Machine-Type Communications (LTE-M), Long Range Wide Area Network (LoRaWAN) and Sigfox, and explores the possibilities to enhance their capabilities in those aspects. Particularly, in this work multiple coverage field trials and measurement campaigns were conducted in an outdoor and an underground deep-indoor environment. As a result, a new empirical dataset, as well as deep-indoor channel predictor models, were derived. Moreover, a lightweight Radio Access Network (RAN) enhancement providing evolved node-B (eNB) failure tolerance in Cellular Internet of Things (CIoT) networks was proposed. Furthermore, security breaches of NB-IoT, LoRaWAN and Sigfox were identified. A new research direction of Sustainable Security for Internet of Things (SSIoT) was introduced and Group Object Security for Constrained RESTful Environments (OSCORE) security protocol for constrained devices was implemented and experimentally evaluated.

The outcome of the research work can be summarised as follows: 1) satisfactory LP-WAN coverage could be observed in all the measured scenarios. Environmental feature engineering based on empirical data can be helpful to identify new factors influencing the radio signal attenuation and, consequently, derive new and more accurate path-loss predictors. 2) A Device-to-Device (D2D)-based scheme for CIoT, which introduces eNB failure-tolerance is considered a promising way of increasing the robustness of the IoT system at a minimal resource cost; 3) NB-IoT, LoRaWAN and Sigfox protocols have shown their security weaknesses, which ought to be continuously analysed and addressed in the standardisation process. Future IoT security works should follow the path of Sustainable Security for Internet of Things (SSIoT), and an example of a robust, but lightweight and sustainable security solution for IoT application-layer is Group OSCORE protocol, which was experimentally verified.
This dissertation is submitted in partial fulfilment of the requirements for a degree of Doctor of Philosophy (Ph.D) at the Technical University of Denmark (DTU), February 2021.

Kongens Lyngby, 28th February 2021

Krzysztof Mateusz Malarski
First and foremost, I find no appropriate words to express my endless gratitude to God for all talents and opportunities he never stops to grant me. I dedicate this work to His greater glory.

I would like to thank my supervisors, Associate Professors: Sarah Ruepp, Martin Nordal Petersen and Michael Stübert Berger for their guidance, feedback and fruitful discussions. I appreciate not only their research supervision, but also a lot of interesting insights into the process of acquiring research grants, disseminating IoT research to the industry, and a lot more.

Many thanks to Associate Professors: Maria Kihl, Danilo de Souza Santos and Professors: Erik Larsson and Angelo Perkusich for welcoming me in their Universities, where I had a pleasure to have external research stays. I highly appreciate fruitful collaboration done with other PhD students: Jakob Thrane, Michele De Donno, Martin Gunnarsson and Dalton Ceza Gomes Valadares. It was a great experience and I enjoyed learning a lot in a short time!

I would like to express my gratitude towards all Network Technologies and Service Platforms Group colleagues, especially Professor Lars Dittmann, Associate Professors Henrik Lehmann Christiansen, Henrik Wessing, Ying Yan and Jose Soler for help and friendly, family-like atmosphere. I would like to thank those students, with whom I shared several years of my daily life, especially to my ”PHD-roommates”: Eder Ollora Zaballa, Kalpit Dilip Ballal and Anas Mohammad Al Shalyan. I thank Jakob Thrane and Martin Gunnarsson for proof-reading of this thesis.

I am wholeheartedly grateful to my dearest wife, my brother, my parents and the whole family for continuous support and motivation.
List of publications

During the course of this PhD study the following articles were published:

Peer-reviewed conference papers


Journal articles


Dataset contributions


Other articles published during the PhD study


Greek

3GPP Third Generation Partnership Project.
AAD Additional Authenticated Data.
ABP Activation By Personalisation.
ADR Adaptive Data Rate.
AEAD Authenticated Encryption with Associated Data.
AES Advanced Encryption Standard.
API Application Programming Interface.
APN Access Point Name.
ARD Automatic Relevance Determination.
AS Application Server.
AuC Authentication Centre.
BLE Bluetooth Low Energy.
BLER BLock Error Rate.
BPSK Binary Phase Shift Keying.
BS Base Station.
CBOR Concise Binary Object Representation.
CE Coverage Enhancement.
CIoT Cellular Internet of Things.
CoAP Constrained Application Protocol.
COSE CBOR Object Signing and Encryption.
COTS  Commercial Off-The-Shelf.

CP  Control Plane.

CPU  Central Processing Unit.

CSMA  Carrier Sense Multiple Access.

CV  Cross-Validation.

D-BPSK  Differential Binary Phase Shift Keying.

D2D  Device-to-Device.

DDoS  Distributed Denial of Service.

DL  Downlink.

DoS  Denial of Service.

DRX  Discontinuous Reception.

DTLS  Datagram Transport Layer Security.

E-CID  Enhanced Cell-ID.

E2E  End-to-End.

eDRX  extended Discontinuous Reception.

eMTC  enhanced Machine-Type Communications.

eNB  evolved node-B.

EPC  Evolved Packet Core.

EPS  Evolved Packet System.

FDD  Frequency Division Duplex.

GFSK  Gaussian Frequency Shift Keying.

Gid  Group ID.

GL  Group Leader.

GM  Group Manager.

GNSS  Global Navigation Satellite System.

GPR  Gaussian Process Regression.
GPRS  General Packet Radio Service.

GPS  Global Positioning System.

GSM  Global System for Mobile communications.

HKDF  HMAC-based Key Derivation Function.

HSM  Hardware Security Module.

HTTP  HyperText Transfer Protocol.

I2I  Indoor-to-Indoor.

ICMP  Internet Control Message Protocol.

IMEI  International Mobile Equipment Identifier.

IoT  Internet of Things.

IP  Internet Protocol.

ISM  Industrial, Scientific and Medical.

IV  Initialisation Vector.

JTAG  Joint Test Action Group.

LAN  Local Access Network.

LIDAR  Light Detection And Ranging.

LoRa  Long Range.

LoRaWAN  Long Range Wide Area Network.

LP-WAN  Low-Power Wide Area Network.

LTE  Long-Term Evolution.

LTE-M  LTE for Machine-Type Communications.

MAC  Message Authentication Code.

MAE  Mean Absolute Error.

MCL  Maximum Coupling Loss.

MIB  Master Information Block.
MIC  Message Integrity Code.
MitM  Man-in-the-middle.
ML    Machine Learning.
MME   Mobility Management Entity.
MNO   Mobile Network Operator.
MSE   Mean Square Error.
MTC   Machine-Type Communications.
MTTF  Mean Time To Failure.
MTTR  Mean Time To Repair.
NAS   Non Access Stratum.
NAT   Network Address Translation.
NB-IoT Narrowband Internet of Things.
NLOS  Non-Line-Of-Sight.
NPBCH Narrowband Physical Broadcast Channel.
NPDCCH Narrowband Physical Downlink Control Channel.
NPDSCH Narrowband Physical Downlink Shared Channel.
NPSS  Narrowband Primary Synchronisation Signal.
NRS   Narrowband Reference Signal.
NS    Network Server.
NSSS  Narrowband Secondary Synchronisation Signal.
O2DI  Outdoor-to-Deep-Indoor.
O2I   Outdoor-to-Indoor.
O2O   Outdoor-to-Outdoor.
OFDMA Orthogonal Frequency Division Multiple Access.
OLS   Ordinary Least Squares.
OS    Operating System.
OSCORE  Object Security for Constrained RESTful Environments.

OTAA  Over The Air Activation.

OTDOA  Observed Time Difference of Arrival.

PC  Personal Computer.

PDCP  Packet Data Convergence Protocol.

PDP  Packet Data Protocol.

PDU  Protocol Data Unit.

PGW  Packet Data Gateway.

PoC  Proof-of-Concept.

PRB  Physical Resource Block.

ProSe  Proximity Services.

PSM  Power Saving Mode.

QAM  Quadrature Amplitude Modulation.

QoS  Quality of Service.

QPSK  Quadrature Phase Shift Keying.

RA  Random Access.

RACH  Random Access Channel.

RAM  Random Access Memory.

RAN  Radio Access Network.

RAT  Radio Access Technology.

RFE  Recursive Feature Elimination.

RMSE  Residual Mean Square Error.

ROM  Read-Only Memory.

RRC  Radio Resource Control.

RSRP  Received Signal Reference Power.

RSSI  Received Signal Strength Indicator.
RTT Round Trip Time.
RUE Relay UE.
SDR Software-Defined Radio.
S-GW Serving Gateway.
SCEF Service Capability Exposure Function.
SC-FDMA Single Carrier Frequency Division Multiple Access.
SF Spreading Factor.
SIM Subscriber Identity Module.
SL Sidelink.
SN Sequence Number.
SNR Signal-to-Noise Ratio.
SPF Semtech Packet Forwarder.
SSE Sum of Squares Error.
SSIoT Sustainable Security for Internet of Things.
TCP Transmission Control Protocol.
TDD Time Division Duplex.
TTN The Things Network.
UART Universal Asynchronous Receiver/Transmitter.
UDP User Datagram Protocol.
UE User Equipment.
UL Uplink.
UP User Plane.
URI Uniform Resource Identifier.
USB Universal Serial Bus.
USIM Universal Subscriber Identity Module.
V2X Vehicle-to-Anything.
VUE Victim UE.
I Ubiquitous connectivity

1 Excellent coverage: the biggest LP-WAN promise 7
   1.1 IoT communication standards ............................ 7
   1.1.1 LoRaWAN ................................................. 8
       1.1.1.1 LoRaWAN version 1.1 .......................... 9
   1.1.2 Sigfox .................................................. 10
   1.1.3 NB-IoT .................................................. 12
       1.1.3.1 NB-IoT evolution ................................. 14
   1.1.4 LTE-M (eMTC) .......................................... 15
       1.1.4.1 Similarities to NB-IoT .......................... 15
       1.1.4.2 Differences to NB-IoT ......................... 16
## Contents

1.1.5 Comparison of communication range ........................................... 16  
1.2 LP-WAN coverage issues ................................................................. 17  
1.2.1 Contributions ............................................................................ 18

2 Outdoor LP-WAN coverage: a maritime perspective 19  
2.1 Harbour scenario ........................................................................... 20  
2.2 First maritime trial: NB-IoT ......................................................... 20  
2.2.1 Results ...................................................................................... 21  
2.2.2 NB-IoT vs. LoRaWAN ............................................................... 23  
2.3 Second maritime trial: NB-IoT and LTE-M ..................................... 23  
2.3.1 Results ...................................................................................... 23  
2.4 Discussion ...................................................................................... 25  
2.5 Summary ......................................................................................... 27

3 On the extreme LP-WAN deep-indoor coverage 29  
3.1 Underground sensor deployment: a killer scenario for wireless communications ........................................... 30  
3.1.1 Related work ............................................................................. 31  
3.2 Investigating deep-indoor coverage ................................................. 32  
3.2.1 NB-IoT and LoRaWAN trial ....................................................... 32  
3.2.1.1 Measurements .................................................................... 32  
3.2.2 NB-IoT underground trial: test of the coverage and validation of the existing path loss models ........................................... 34  
3.2.2.1 Experiment description ......................................................... 35  
3.2.2.2 Results ................................................................................. 37  
3.3 Towards new models for underground IoT coverage planning .......... 38  
3.3.1 Methodology ............................................................................. 39  
3.3.1.1 Why statistical modelling? ..................................................... 39  
3.3.1.2 From raw samples to new models ........................................ 40  
3.3.1.3 Statistical toolkit .................................................................. 41  
3.3.2 Underground measurement campaign ....................................... 44  
3.3.3 Feature engineering ................................................................... 45  
3.3.3.1 Distance-related features ...................................................... 47  
3.3.3.2 Tunnel-related features ........................................................ 47  
3.3.4 Feature analysis ......................................................................... 48  
3.3.4.1 Linear analysis ..................................................................... 48  
3.3.4.2 Non-linear analysis ............................................................... 50  
3.3.5 New path-loss models ................................................................. 52  
3.3.6 Discussion .................................................................................. 55  
3.3.6.1 The value of the collected data ............................................ 55  
3.3.6.2 Versatility of the study ........................................................ 56  
3.3.6.3 The meaning of feature engineering .................................... 57  
3.3.6.4 The relevance of the study for particular LP-WANs ......... 57  
3.4 Summary ....................................................................................... 59
# Contents

<table>
<thead>
<tr>
<th>Part I Summary</th>
<th>61</th>
</tr>
</thead>
<tbody>
<tr>
<td>II Failure-tolerance</td>
<td>63</td>
</tr>
<tr>
<td>4 When the system is put to the test</td>
<td>65</td>
</tr>
<tr>
<td>4.1 RAN failure scenario</td>
<td>66</td>
</tr>
<tr>
<td>4.2 Possible solutions</td>
<td>67</td>
</tr>
<tr>
<td>4.3 Contributions</td>
<td>68</td>
</tr>
<tr>
<td>5 Towards D2D-enabled IoT resilience</td>
<td>71</td>
</tr>
<tr>
<td>5.1 D2D with NB-IoT: challenges and opportunities</td>
<td>71</td>
</tr>
<tr>
<td>5.1.1 Definition and overview</td>
<td>71</td>
</tr>
<tr>
<td>5.1.2 Device discovery</td>
<td>72</td>
</tr>
<tr>
<td>5.1.3 Applying D2D in NB-IoT system</td>
<td>73</td>
</tr>
<tr>
<td>5.1.3.1 Challenges</td>
<td>73</td>
</tr>
<tr>
<td>5.1.4 3GPP efforts</td>
<td>77</td>
</tr>
<tr>
<td>5.1.4.1 3GPP standardisation priorities</td>
<td>78</td>
</tr>
<tr>
<td>5.2 A D2D-based failure-tolerant scheme for cellular IoT</td>
<td>79</td>
</tr>
<tr>
<td>5.2.1 Related work</td>
<td>79</td>
</tr>
<tr>
<td>5.2.2 Design description</td>
<td>80</td>
</tr>
<tr>
<td>5.2.2.1 Assumptions and requirements</td>
<td>81</td>
</tr>
<tr>
<td>5.2.3 System behaviour</td>
<td>82</td>
</tr>
<tr>
<td>5.2.4 Signalling patterns</td>
<td>83</td>
</tr>
<tr>
<td>5.2.5 Implementation considerations</td>
<td>84</td>
</tr>
<tr>
<td>5.2.5.1 Implementation on real-life equipment</td>
<td>87</td>
</tr>
<tr>
<td>5.2.6 Discussion</td>
<td>87</td>
</tr>
<tr>
<td>5.2.6.1 A non-CIoT perspective</td>
<td>88</td>
</tr>
<tr>
<td>Part II Summary</td>
<td>89</td>
</tr>
<tr>
<td>III Security</td>
<td>91</td>
</tr>
<tr>
<td>6 Smart community in danger</td>
<td>93</td>
</tr>
<tr>
<td>6.1 Security goals</td>
<td>95</td>
</tr>
<tr>
<td>6.2 Contributions</td>
<td>95</td>
</tr>
<tr>
<td>7 LP-WAN security analysis</td>
<td>97</td>
</tr>
<tr>
<td>7.1 Related work</td>
<td>97</td>
</tr>
<tr>
<td>7.2 General LP-WAN security issues</td>
<td>98</td>
</tr>
<tr>
<td>7.3 Technology-specific security analysis</td>
<td>99</td>
</tr>
<tr>
<td>7.3.1 LoRaWAN</td>
<td>100</td>
</tr>
<tr>
<td>7.3.1.1 LoRaWAN packet forging</td>
<td>102</td>
</tr>
<tr>
<td>7.3.2 Sigfox</td>
<td>104</td>
</tr>
</tbody>
</table>
7.3.2.1 Sigfox replay attack + device DoS 108
7.3.3 NB-IoT and LTE-M 109
7.3.3.1 Attack on NB-IoT using malicious UE 112
7.4 Discussion 113
7.4.1 The significance of the experiments 115
7.5 Summary 115

8 Securing IoT communication 117
8.1 Sustainable Security for IoT (SSIoT) 117
8.1.1 SSIoT building blocks 118
8.1.1.1 Green and Sustainable Computing 118
8.1.1.2 Green and Sustainable Internet of Things 119
8.1.1.3 Green and Sustainable Security 119
8.1.2 SSIoT concept 119
8.2 Lightweight application-layer security: protecting CoAP communications 121
8.2.1 CoAP overview 122
8.2.1.1 CoAP features 122
8.2.1.2 CoAP group communications 123
8.2.1.3 CoAP security with DTLS 124
8.2.1.4 Group CoAP security issues 125
8.2.2 Lightweight DTLS alternative: OSCORE 126
8.2.3 Novel OSCORE flavour: Group OSCORE 128
8.2.3.1 Security context amendments 129
8.2.3.2 The Group Manager role 129
8.2.3.3 Cryptography protocol differences 130
8.2.4 Performance evaluation experiment 130
8.2.4.1 Memory occupancy analysis 131
8.2.4.2 Round-trip-time experiment 133
8.2.5 Discussion 135
8.3 Summary 136

Part III Summary 137

Conclusion 139

Appendices 143

Appendix A Submitted papers 145

Bibliography 151
# List of Figures

0.1 IoT-powered smart community and the selected use-cases: 1) Sensor networks (smart farming); 2) Natural disaster (base station failure, traffic rerouted through the neighbouring sensors); 3) Autonomous driving (Vehicle-to-Anything (V2X) communication); 4) Smart home/Smart office.  

0.2 Main pillars of a reliable IoT architecture for a smart community.  

1.1 LoRaWAN architecture (inspired by [14]).  
1.2 LoRaWAN 1.1 architecture (based on Figure 2 in [15]).  
1.3 Sigfox message transmission (inspired by Figure 4-9 from [16]).  
1.4 Sigfox star architecture (based on [17]).  
1.5 The modes of NB-IoT deployment (based on [18]).  
1.6 NB-IoT architecture (on the basis of [19]).  
1.7 Communication range comparison of IoT technologies  

2.1 The locations of the measurement points for the NB-IoT trial [2].  
2.2 LoRaWAN RSSI and SNR samples captured in the harbour [1].  
2.3 NB-IoT RSSI and SNR samples captured in the harbour [2].  
2.4 RSRP measurements of LTE-M and NB-IoT.  
2.5 RSRP measurements of NB-IoT taken in 2018 vs 2020.  

3.1 Xirio Coverage simulation results for NB-IoT [3].  
3.2 Xirio Coverage simulation results for LoRaWAN [3].  
3.3 Measurement locations [3]. LoRaWAN gateway and NB-IoT eNB denoted as green X and orange pentagon, respectively.  
3.4 LoRaWAN and NB-IoT measurements vs. simulation estimations (based on [3]).  
3.5 NB-IoT resource grid (guard-band and stand-alone deployments) [5].  
3.6 Measurement trial locations [5].  
3.7 Experimental setup [5].  
3.8 RSSI heatmap in an above-ground and underground location [5].  
3.9 Measured NB-IoT RSSI, together with 3GPP 38.901 estimations [5].
3.10 Deep-indoor IoT coverage planning [8]. Traditionally, the results of the modelling may motivate infrastructure modifications, such relay placement and re-positioning of the sensors. However, in deep-indoor situations, such as basements coverage modelling is difficult and impractical. Currently available statistical models exhibit high accuracy at indoor deployment located at level 0 and above but are not properly calibrated at level -1 and -2. 39
3.11 Deriving a new deep-indoor path-loss component from the measured data [9]. 41
3.12 $PL_{in}$ calculation [9]. 42
3.13 The cross-section of the above-ground campus area [9]. 44
3.14 Layout of DTU Lyngby Campus tunnel system [9]. eNB is represented by orange symbol. Received Signal Reference Power (RSRP) heatmap shows along which corridors the samples were taken. 46
3.15 RSRP in dBm versus 3D distance, compared with 3GPP prediction [9]. 46
3.16 Overview of distance-related features [8]. The parameters lay along the same, straight line between the transmitter and the receiver. 47
3.17 Tunnel features in an example 3D tunnel fragment [9]. In this particular example, $n_c$ is equal to 1, as is the number of corridor entrances within the distance of $d_{thr}$ from the measurement point. 48
3.18 Evaluation of the selected linear regression models [9]. 50
3.19 Mean Absolute Error comparison of predictors utilising different distance features [8]. 53
3.20 Mean Absolute Error comparison of 3GPP 38.901 model and the new proposals: linear regression an GPR [9]. Median values inside the boxes. 53
3.21 $PL_{in}$ with respect to 2D indoor distance, predicted on the test dataset (25% samples) [9]. 54

4.1 The considered cellular IoT RAN failure scenario. Note that once the $eNB1$ fails, the $UEs$ in cell1 cannot re-connect to the network, as the closest base station, $eNB2$, is out of range. 66
4.2 The scope of Part II contributions in terms of a generic IoT architecture. The numbers represent the corresponding sections of this document. 68

5.1 The categories of D2D communications. 72
5.2 Overview of D2D NB-IoT system aspects [7]. 76
5.3 Third Generation Partnership Project (3GPP) Proximity Services (ProSe) deployment scenarios [20]. Note that in "Out of coverage" and "Partial coverage" scenarios, direct D2D communications and direct discovery may only occur for public-safety use-cases. 77
5.4 3GPP D2D (ProSe) Architecture for no mobility scenario [20]. 78
5.5 Considered RAN failure-recovery situation. The recovery D2D link is established between the VUEs and the RUE, which may belong to another IoT deployment. 81
5.6 Pre-disaster configuration signalling. 83
5.7 Communication signalling at the time of a disaster. .......................... 84
5.8 A modified RRC state machine of a GL VUE. The diagram is based on Figure 30 in [21]. .................................................. 85
5.9 A modified RRC state machine of a RUE. The diagram is based on Figure 30 in [21]. .................................................. 86

6.1 The fundamentals of CIA principle in data security [22]. ................. 95
6.2 The scope of Part III contributions in terms of a generic IoT architecture. The numbers represent the corresponding sections of this document. ..... 96

7.1 Simplified architecture of a IoT device (sensor, gateway) [4]. .......... 98
7.2 LP-WAN general security threats. .............................................. 100
7.3 Packet format and security mechanism in LoRaWAN (based on [23, 24]). 101
7.4 Key generation in LoRaWAN [23]. ............................................. 101
7.5 MIC bruteforcing workflow [4]. ............................................... 105
7.6 The results of the packet forging attack [4]. ................................ 106
7.7 Security in Sigfox network (inspired by [25]). Black - obligatory security features, Gray - optional features. ........................................ 107
7.8 Sigfox packet format in uplink, based on Figure 2 from [25]. ............ 107
7.9 The results of Sigfox replay attack [4]. ....................................... 109
7.10 Control Plane protocols and security (simplified) [26]. E - encryption, I - integrity protection. .............................................. 110
7.11 User Plane protocols and security (simplified) [26]. E - encryption, I - integrity protection. .............................................. 110
7.12 Control plane NAS packet format [27]. ..................................... 110
7.13 PDCP packet formats [28]. ................................................... 111
7.14 A simplified scenario of NB-IoT attack with a malicious device [4]. .... 112
7.15 The results of the NB-IoT attack [4]. ....................................... 113

8.1 The illustration of SSIoT concept (based on [6]). .......................... 120
8.2 SSIoT pillars (based on [6]). .................................................. 121
8.3 CoAP message format [29]. .................................................. 123
8.4 DTLS Architecture (inspired by [30]). ...................................... 125
8.5 The format of the OSCORE option [31]. The total length can be maximum 255 bytes. .................................................. 127
8.6 OSCORE Client and Server security contexts [31, 32] ..................... 127
8.7 Experimental setup diagram. .................................................. 131
8.8 Memory occupancy measured on Zoul Firefly Rev. A. Marked areas denote stack usage (in RAM usage) or memory occupied by the cryptography (in ROM usage). .................................................. 132
8.9 Memory occupancy measured on TI Simplelink CC1352R1 launchpad. Marked areas denote stack usage (in RAM usage) or memory occupied by the cryptography (in ROM usage). .................................................. 132
8.10 Round-trip-time measurements taken for various payload sizes. ........ 134
<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Overview of NB-IoT enhancements [33–36]</td>
<td>14</td>
</tr>
<tr>
<td>1.2</td>
<td>LTE-M and NB-IoT feature comparison</td>
<td>16</td>
</tr>
<tr>
<td>2.1</td>
<td>Description of the measurement points [2]</td>
<td>21</td>
</tr>
<tr>
<td>3.1</td>
<td>Base station simulation parameters [3]</td>
<td>32</td>
</tr>
<tr>
<td>3.2</td>
<td>Number of samples per location [5]</td>
<td>37</td>
</tr>
<tr>
<td>3.3</td>
<td>Experiment parameters [9]</td>
<td>45</td>
</tr>
<tr>
<td>3.4</td>
<td>Automatic feature selection [9]</td>
<td>49</td>
</tr>
<tr>
<td>3.5</td>
<td>ARD results for different corridor scenarios [9]</td>
<td>51</td>
</tr>
<tr>
<td>3.6</td>
<td>ARD results for 75% of the shuffled data [9]</td>
<td>52</td>
</tr>
<tr>
<td>7.1</td>
<td>Summary of LPWAN attacks [4]</td>
<td>114</td>
</tr>
<tr>
<td>8.1</td>
<td>Memory occupancy of OSCORE/Group OSCORE cryptography (bytes)</td>
<td>131</td>
</tr>
</tbody>
</table>
Introduction

Our world relies on connectivity, mobility and data. Modern economies demand that more and more processes in factories, farms, institutions and residential areas are connected and digitised, which paves the way for automation and optimisation of everyday life. The society of the future consists of Smart Communities, defined by Coe Paquet and Roy [37] as:

"a geographical area ranging in size from a neighbourhood to a multicountry region within which citizens, organisations and governing institutions deploy New Information and Communications Technologies to transform their region in significant and fundamental ways."

Such powerful impact of smart communities requires a communications infrastructure that connects people, machines and virtually anything that may be controlled, monitored or tracked. Thus, Internet of Things (IoT), a natural enabler for smart communities, has attracted significant attention of academia and industry. As shown in Figure 0.1, thanks to a reliable IoT architecture multiple critical applications of Smart Communities can be realised. However, assuring reliability of IoT deployments remains a challenge for scientists, network engineers and network operators.

The main focus of this thesis is to explore how coverage, failure-safety and security of low-power IoT architectures can be improved, so that they may become suitable for Smart Communities of the future. The research efforts described in this dissertation are oriented around Low-Power Wide Area Network (LP-WAN) technologies, a new family of IoT created with the thought of providing extreme coverage, ultra-low power consumption and extended battery life. Since the scope of LP-WAN is limited to low data-rate applications with relaxed latency requirements, the research presented in this document addresses only those smart community scenarios where low-power sensor networks are involved: smart city, intelligent agriculture, asset tracking, critical monitoring and remote metering.

The author’s view on the essence of a reliable IoT communications architecture is depicted in Figure 0.2. First and foremost, robust connectivity is an essential condition of any wireless network, as no information can be exchanged if the devices cannot reach out to one another. However, even when ubiquitous coverage is achieved, the

---

1Up to 10 seconds end-to-end delay is acceptable for LP-WANs [38], whereas critical 5G applications with strict requirements will operate with less than 1 milisecond latency [39].
system may (and in real life will) experience unpredictable events, such as natural disasters or hardware failure, resulting in partial or total communications disruption. If no resilience or failure-safety mechanisms are implemented, the system cannot be considered ready to serve truly critical applications. Finally, even seamless connectivity and service continuity guarantees are not enough to enable use-cases with sensitive data, as the creativity of hackers and malicious parties might make itself visible by means of data modification or stealth, infrastructure damage and, as a result, loss of revenue and reputation. Therefore, the work presented in this thesis addresses all three pillars of reliability and seeks to pave the way for further advances in the field of trustworthy and efficient smart communities of the future.
The core content of this dissertation is divided into 3 parts. Each part addresses one of the aspects illustrated in Figure 0.2. Every part consists of several chapters, the first of which typically introduces the main theoretical concepts and state-of-the-art research efforts; the remaining chapters of the part discuss the author’s contributions, juxtaposed with some relevant findings from the scientific community. Finally, the part is summarised in a final section, where the author additionally puts the outcome of the research in a perspective and highlights the most interesting future research directions.

In Part I, the connectivity challenge of LP-WAN is discussed. Chapter 1 provides an overview of the coverage promises of various low-power IoT standards, and sheds the light on the most critical issues that still remain unresolved by the researchers. Then, the aspect of network coverage is divided into two realms: outdoor and deep-indoor, and discussed in depth in Chapter 2 and Chapter 3, respectively. These chapters introduce the following contributions: outdoor Narrowband Internet of Things (NB-IoT) and LTE for Machine-Type Communications (LTE-M) field trials (Chapter 2), underground NB-IoT measurement campaigns and a Machine Learning (ML)-aided study on Third Generation Partnership Project (3GPP) deep-indoor path-loss models’ calibration and enhancement (Chapter 3). In Chapter 2, it is proven that both Long Range Wide Area Network (LoRaWAN), NB-IoT and LTE-M technologies are able to provide seamless connectivity in a challenging outdoor area of a harbour, and they do not require a densely deployed grid of base stations or gateways to achieve this goal. In Chapter 3, it is shown that sub-GHz IoT standards, such as NB-IoT can also be successful in underground areas. Moreover, even though the existing channel models for deep-indoor signal propagation are not precise enough for reliable coverage modelling, it is possible to improve the accuracy of the statistical models by calibrating them with empirical underground signal data and utilising advanced environmental features.

In Part II, the reader is presented a study on failure-proof operation of low-power IoT infrastructures. The examination concentrates on the problem of a evolved node-B (eNB) failure in a remote deployment of IoT devices, which as a result lose the connectivity with the core network. The description of the considered emergency scenario is presented in Chapter 4. Chapter 5 contains an authors proposal of Cellular Internet of Things (CIoT) enhancement, which ensures service continuity under the aforementioned failure situation. The communications can be sustained, as the affected devices utilise a relay User Equipment (UE), which has agreed beforehand to forward the uplink data. The Device-to-Device (D2D)-based emergency link is established only at the time of the failure and is terminated as soon as the broken eNB returns to operation. In Chapter 5, the requirements, assumptions, behaviour, signalling and implementation considerations are provided. It is found that in comparison to other ideas found in the literature, the proposed solution is more lightweight and energy-efficient.

In Part III, the focus is kept on how to ensure secure IoT communication. Chapter 6 reveals the most severe dangers that the low-power sensor network may face, as well as the consequences of not being prepared to mitigate the risks of malicious actions.
A presentation of LP-WAN security vulnerabilities, exhibited by Proof-of-Concept (PoC) attacks, are included in Chapter 7. In Chapter 8, the author introduces the newly defined research concept of Sustainable Security for Internet of Things (SSIoT) and explains the features of the research works and technology proposals that falls into SSIoT. An example of such solution is Group Object Security for Constrained RESTful Environments (OSCORE), a lightweight security protocol for Constrained Application Protocol (CoAP) group communications, and Chapter 8 includes the first publicly available experimental performance evaluation of Group OSCORE. The protocol is implemented by the author in Contiki-NG operating system and publicly available. The design is tested on 2 commercial off-the-shelf hardware platforms: Zolertia Firefly Rev. A [40] and TI Simplelink CC1352R1 [41]. The collected data regarding the Round Trip Time (RTT), memory occupancy and energy consumption show that Group OSCORE constitutes a reasonable overhead for constrained devices and is an attractive application-layer security solution for low-power IoT.

The final Chapter named Conclusion, summarises the whole document, highlights the conclusions on the research contributions and outlines possible future research directions.
Part I

Ubiquitous connectivity
In network planning, one of the most crucial performance indicators is \textit{coverage}. Intuitively, one may understand coverage as the geographical span within which the receiver can successfully demodulate radio messages sent by the transmitter. In its cellular IoT performance evaluation methodology, 3GPP specifies that the device is under coverage as long as it can correctly demodulate all the control and data signals with the BLock Error Rate (BLER) of 10\% and the user data rate of at least \textbf{160 bps} \cite{42}. From the perspective of a smart connected community, knowing the coverage situation in the area determines which services can be realised, i.e. whether the required bit-rate can be achieved, as well as the energy costs of the operation, as poor signal conditions typically forces the transceivers to: switch to robust modulation schemes providing smaller throughput, increase the transmission power and/or repeat the transmission.

This chapter describes the most popular IoT technologies and compares them in terms of coverage performance. Short-range IoT standards are only briefly introduced, while the LP-WANs, which this thesis focuses on, are explained in details. Finally, the known LP-WAN coverage issues are introduced.

1.1 IoT communication standards

Before the time of LP-WANs, the IoT technology landscape included 2 families of standards:

1. Short-range unlicensed-spectrum. Designed primarily for smart home and smart factory applications or personal area networks, these technologies are energy-efficient, operate in the open frequency channels around 2.4 GHz and offer
Excellent coverage: the biggest LP-WAN promise

Communication ranges no longer than 200-250 m. The most popular standards: Zigbee, WiFi and Bluetooth Low Energy (BLE).

2. Long-range licensed-spectrum. Global System for Mobile communications (GSM), also known as 2G, has become obsolete for mobile broadband users, due to, among others, insufficient data rate supported. However, even being older than the short-range standards, GSM provides long-range communication (several kilometres range observed in [43], theoretically even 35 kilometres [44]), robust data exchange\(^1\) and the GSM radio modules are still commonly available at low cost. That is why the IoT industry has initially decided to take advantage of the remaining GSM infrastructure and apply 2G technology for low-data long-range use-cases. The drawback of 2G is suboptimal power consumption, which in the context of IoT becomes a severe trouble.

In this situation, Low-Power Wide Area Network standards have successfully addressed the market need of a solution that would combine the energy-efficiency of short-range IoT with the long-range coverage and robustness of the cellular networks, but with higher deployment flexibility. This chapter discusses radio capabilities of those LP-WANs that are the most relevant for the dissertation work: LoRaWAN, Sigfox, LTE for Machine-Type Communications and Narrowband Internet of Things.

1.1.1 LoRaWAN

LoRaWAN is an open-source network protocol formulated and maintained by LoRa Alliance [45]. The system bases on proprietary LoRa radio, that has been used for military purposes and can be characterised by very high tolerance to signal interference. As an End-to-End (E2E) solution, LoRaWAN enables the Mobile Network Operators (MNOs) to deploy the core and access network infrastructure and provide it as a service, but one may also construct the whole network on his/her own; the only caveat is the duty-cycle restriction that must be obeyed\(^2\).

LoRaWAN radio, LoRa, uses frequency chirp spread spectrum modulation, owned by Semtech. The radio module chips the data signal into a chirp signal that continuously varies in frequency at a certain Spreading Factor (SF). The resultant modulation consumes little power, enables long ranges and exhibits resistance to multipath propagation, fading and Doppler shift [46].

LoRaWAN specification [23] defines 3 types of end-devices, discriminated in terms of transmission and energy consumption limitations:

1. Class A devices support both Uplink and Downlink communication between the gateway and the LoRaWAN device. The device can send the uplink message

---

\(^1\)note that GSM900 operates in sub-GHz spectrum, where the signal penetrates obstacles better than in the 2.4 GHz spectrum.

\(^2\)In Europe, 868 MHz band is subjected to 1% duty-cycle rule, i.e. a device can only transmit during 1% of the time, and for the rest 99% it can only receive the signal.
at any time, and it then opens two receiver windows after 1 second and 2 seconds, respectively. If the downlink is not complete by this time, then it has to be queued in the Network Server (NS) and only sent after next uplink transmission, as the end-device enters a sleep mode. Class A transmission is a default mode of communication for most energy-constrained devices, which do not require frequent downlink transmissions.

2. Class B is a further extension of Class A. This transmission method is used by devices which require to receive downlink transmissions frequently. Class B devices listen for the beacon frame from the gateway at fixed interval and register themselves a slot for receiving downlink message.

3. Class C devices have no restrictions on the power consumption. They are listening for any downlink transmissions on the default downlink channel. Class C devices uses similar uplink transmission method as that of Class A, but here the device never stops receiving the messages, unless it is transmitting.

The access network consists of end-devices and gateways that acts as simple message forwarders. Therefore, there is no attachment of the sensor to the gateway, i.e. the constrained endpoint does not need to exchange special signalling messages with the gateway and become registered in order to successfully communicate; instead, all the gateways that have received the packet from a sensor send it "blindly" to the NS. Therefore, the network needs to de-duplicate the incoming packets at the server side. Apart from message de-duplication, the NS controls and manages the network nodes and decides upon: radio configuration, data format and sequence numbers. Moreover, the NS provides the user applications with the IoT data by interacting with Application Server (AS). The topology of LoRaWAN (star of stars) can be observed in Figure 1.1.

1.1.1 LoRaWAN version 1.1.

Version 1.1 of the standard has brought design changes making the new version of LoRaWAN not fully compatible with the older releases [47]. Two main areas of enhancement are:

1. Inter-network roaming, so that full control over the end-device can be passed from one infrastructure to another. This function required that the new standard defines 3 main NS functionalities: home NS, forwarding NS and serving NS; moreover, a new entity called Join Server was introduced (see Figure 1.2).

2. Miscellaneous security enhancements.

At the time of writing this dissertation, no equipment supporting version 1.1 of LoRaWAN could be found and accessed. Therefore, the experiments and considerations presented in this document focus mainly on the older, well-established LoRaWAN version.
1.1.2 Sigfox

Called by its founders a 0G technology, Sigfox was introduced in 2010 as a "bare-minimum" standard allowing for ultra-low power applications involving little data sent infrequently. At the time of writing this thesis, there has been 15.2 million connected devices in 71 countries [48].

Even though Sigfox operates in unlicensed spectrum (868 MHz in Europe) and its radio specification has been released to the public, it is considered a proprietary technology, as the infrastructure (gateways and the backend network) are owned by the Sigfox company and one cannot make use of any Sigfox device unless registered and subscribed to the backend. The subscription options range from 1 to 144 uplink messages per day and up to 4 downlink messages per day; the upper message limits are dictated by the 1% duty cycle in the 868 MHz band.

Figure 1.3 shows the mode of transmission in Sigfox radio. Every message can be of up to 12 bytes of payload and is sent one or three times over different frequency channels\(^3\). Sigfox does not apply TCP/IP protocol stack, a dominant solution for today’s Internet. Instead, a lightweight proprietary protocol is used. The reason for such a decision is, that IP is not optimised for small amounts of data and minimal power consumption; as a result, sending just 12 bytes of payload requires 40-bytes header. On the other hand, Sigfox protocol introduces much smaller overhead (14 bytes), and does not require any signalling to establish and maintain the communication channels, thus reducing power consumption and simplifying the processing of the packets.

\(^3\)This is the reason for 144 message per day at maximum. Every message is sent over 2 seconds; 2 seconds * 3 * 144 = 864 seconds = 1% of 1 hour.
Sigfox utilises Differential Binary Phase Shift Keying (D-BPSK) modulation in uplink, resulting in a bitrate of 100 bps and Gaussian Frequency Shift Keying (GFSK) in downlink, providing 600 bps. Low bit-rate and Ultra Narrow Band transmission (100Hz, the whole Sigfox operational bandwidth is only 192 KHz), makes the technology long range (Maximum Coupling Loss (MCL) up to 163.3 dB [49]) and robust against interference [16]. In Sigfox, it is typical to connect "uplink-only" devices, but even if they support downlink, in general Sigfox is very uplink oriented, as only up to 4 messages per day can be received by the end-device (in the highest subscription plan).

Figure 1.4 shows the network architecture of Sigfox. As in LoRaWAN, the gateways (base stations) demodulate packets from all sensors in the range and blindly forward the data to the backend, which in turn pushes the collected data towards the customer Application Servers.
Excellent coverage: the biggest LP-WAN promise

Figure 1.3. Sigfox message transmission (inspired by Figure 4-9 from [16]).

1.1.3 NB-IoT

Narrowband Internet of Things is a cellular Machine-Type Communications (MTC) technology created with the thought of those MNOs that could not (or wished not to) re-use GSM carriers for IoT and were interested in including the new use-case within Long-Term Evolution (LTE) carriers. The result of the standardisation consensus is a LTE-based (though not fully LTE-compatible) system with 3 different deployment possibilities. As the bandwidth of NB-IoT fits in a single Physical Resource Block (PRB), it can operate either together with the classical LTE, within LTE guardband or standalone, as presented in Figure 1.5.

The requirements for NB-IoT include: 52547 devices handled by a single cell, device cost less than $5, 10 years battery lifetime and MCL of 164dB. Such high coupling losses correspond to a 20 dB increase with respect to LTE and can be
translated to seven-fold coverage improvement. It could be achieved by increasing Power Spectral Density (preserving the TX power on the LTE-level, but with only 180 kHz bandwidth) and introducing message repetition scheme with up to 2048 repetitions in the downlink and 128 repetitions in the uplink [50]. NB-IoT uses Orthogonal Frequency Division Multiple Access (OFDMA) with 15 kHz subcarrier spacing in the downlink and Single Carrier Frequency Division Multiple Access (SC-FDMA) in the uplink with 15 kHz and 3.75 kHz spacing possibilities. Contrary to LTE, only Binary Phase Shift Keying (BPSK) and Quadrature Phase Shift Keying (QPSK) modulation formats are supported in NB-IoT, which limits the data rate on one hand, but ensures communication robustness at the same time [51]. Note that even though NB-IoT can be deployed in-band with LTE, its physical layer is not compatible with classical LTE.

Up to 10 years of battery lifetime promise has been made possible with the aid of the following concepts:

- Power Saving Mode (PSM), being a mode of operation when most of the hardware components are off and the device remains unreachable from the network (though registered).
- extended Discontinuous Reception (eDRX). The use of Discontinuous Reception (DRX) generally allows the UE not to listen for the Downlink (DL) messages continuously, but only periodically. eDRX further limits the occurrences of such UE wake-ups, thus optimising the power consumption.

Although LTE was designed with the thought of Internet Protocol (IP) data only, NB-IoT system enables also non-IP packets and defines a dedicated entity, Service Capability Exposure Function (SCEF), for handling them. Moreover, both IP and non-IP data can be delivered via user or control plane [19], as shown in Figure 1.6. The technology uses 2 mechanisms responsible for information transfer:

1. Control Plane optimisation, the data can be piggybacked in Radio Resource Control (RRC) signalling messages, i.e. within Non Access Stratum (NAS). At the Evolved Packet System (EPS), non-IP payload is forwarded to the user servers via SCEF, whilst the IP packets are sent through Serving Gateway (S-GW) and Packet Data Gateway (PGW).
14

1 Excellent coverage: the biggest LP-WAN promise

Figure 1.6. NB-IoT architecture (on the basis of [19]).

2. User Plane optimisation, where (non-)IP user data traverse the path: UE-eNB-S-GW-PGW-application server [51].

1.1.3.1 NB-IoT evolution

Since the time of its first release (Release 13), NB-IoT has been evolving across the new releases. The nature of the changes can be seen in Table 1.1. In Release 14, multicast communication scheme, called Single-Cell Point-to-Multipoint (SC-PTM),

<table>
<thead>
<tr>
<th>Feature</th>
<th>R13</th>
<th>R14</th>
<th>R15</th>
<th>R16</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bit-rate (UL/DL)</td>
<td>67 kbps/32 kbps</td>
<td>159 kbps/127 kbps</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Handover</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>UE power classes</td>
<td>23 dBm</td>
<td>23 dBm, 14 dBm</td>
<td>23 dBm, 14 dBm</td>
<td>23 dBm, 14 dBm</td>
</tr>
<tr>
<td>Localisation</td>
<td>No (CID)</td>
<td>E-CID, OTDOA</td>
<td>E-CID, OTDOA</td>
<td>E-CID, OTDOA</td>
</tr>
<tr>
<td>Control Plane Optimisation</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>User Plane Optimisation</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Non-IP data delivery</td>
<td>Yes</td>
<td>(unreliable only)</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>QoS Guarantees</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Wake-up signalling</td>
<td>No</td>
<td>No</td>
<td>Yes (single)</td>
<td>Yes (single and group)</td>
</tr>
<tr>
<td>Multicast</td>
<td>No</td>
<td>Yes (SC-PTM)</td>
<td>Yes (SC-PTM)</td>
<td>Yes (SC-PTM)</td>
</tr>
<tr>
<td>Duplex mode</td>
<td>Half-duplex</td>
<td>Half-duplex</td>
<td>Half-duplex</td>
<td>Half-duplex</td>
</tr>
<tr>
<td>Small cell support</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Table 1.1. Overview of NB-IoT enhancements [33–36].
was introduced. Additionally, the UE handover became possible and Enhanced Cell-ID (E-CID) and Observed Time Difference of Arrival (OTDOA) localisation was enabled. For scenarios with good radio conditions, a more energy-saving UE class was defined (14 dBm of TX power) [34].

Release 15 brought a number of signalling optimisations serving further minimisation of UE energy consumption. For example, wake-up signals allows the device to only monitor for paging occasions when there is an actual message to receive. Small cell support enables a heterogeneous network with more dense deployment and optimised coverage enhancement mode. Another important improvement is the ability for the network to differentiate more of the UEs parameters, which enables more intelligent scheduling, based e.g. on the battery life of the device. Moreover, signalling optimisation in control and user planes allows for so called, early data transmission, i.e. piggybacking the application data in the RRC connection establishment signalling.

In Release 16, an important step forward was to introduce a basic support of Quality of Service (QoS), making NB-IoT more suitable for critical services. Another enhancements targeted energy-efficiency of the sensors. UE-specific DRX allows for more tailored power consumption minimisation, and the possibility of preconfigured uplink resources in idle mode introduces a new UL scenario, where the device sends a message without any Random Access (RA) signalling.

1.1.4 LTE-M (eMTC)

Apart from NB-IoT, another 3GPP standard for cellular IoT has been proposed in release 13 under the name LTE for Machine-Type Communications (LTE-M). The technology was intended to improve the coverage of LTE, at the same reducing the energy consumption and radio module complexity and not exceeding 1/3 of the cost of a basic LTE terminal [38]. Also known as enhanced Machine-Type Communications (eMTC), LTE-M has been established at the same time as NB-IoT and both standards share certain design principles and technology choices; therefore, this section describes the features of LTE-M with respect to NB-IoT, also summarised in Table 1.2.

1.1.4.1 Similarities to NB-IoT

LTE-M achieves the improvement of coverage by repeating messages; two Coverage Enhancement (CE) modes are defined: mode A enables up to 32 repetitions, and mode B supports even 2048 repetitions. The operation of mode B allows for substantial coverage extension of 20 dB, with respect to the classical LTE [52]. Moreover, as is the case of NB-IoT, LTE-M utilises PSM and eDRX mechanisms to maximise the time during which the UE stays in deep-sleep mode. The multiple access approach in eMTC follows the LTE: OFDMA in the downlink and SC-FDMA in the uplink. Finally, the optimisation of control and user planes in the network architecture, allowing for both
IP and non-IP packets to traverse the EPS are also applicable to LTE-M (see Figure 1.6).

1.1.4.2 Differences to NB-IoT

While NB-IoT requires 1 PRB (180 kHz) and enables multiple deployment options (see Figure 1.5), and its physical layer is not fully compatible with LTE, LTE-M can only be deployed in-band with LTE, as it is fully LTE-compliant and occupies the bandwidth of 6 PRBs (1.4 MHz), also referred to as a *narrowband*. Consequently, LTE-M inherits more of the LTE functionality than NB-IoT, supporting for example voice service, OTDOA positioning and mobility in the RRC-connected state. The available data rate is also higher in LTE-M, reaching 1 Mbps in the ideal case [38], due to wider bandwidth and the use of 16 Quadrature Amplitude Modulation (QAM). However, LTE-M (cat. M1) devices are more complex than cat NB UEs due to higher resource requirements of the LTE-M system.

1.1.5 Comparison of communication range

The communication range of the IoT standards considered in this chapter is compared in Figure 1.7. It is clear to see that the technologies can be divided into short-range

---

Table 1.2. LTE-M and NB-IoT feature comparison.

<table>
<thead>
<tr>
<th>Feature</th>
<th>NB-IoT</th>
<th>LTE-M</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bandwidth</td>
<td>180 kHz</td>
<td>1.4 MHz</td>
</tr>
<tr>
<td>Deployment</td>
<td>In-band, Guard-band, Standalone</td>
<td>In-band</td>
</tr>
<tr>
<td>Uplink peak data rate (peak/instantaneous peak)</td>
<td>62.5 kbps / 250 kbps</td>
<td>375 kbps / 1000 kbps</td>
</tr>
<tr>
<td>Modulation</td>
<td>BPSK, QPSK</td>
<td>QPSK, 16QAM</td>
</tr>
<tr>
<td>PHY layer compatibility with LTE</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Maximum Coupling Loss</td>
<td>164 dB</td>
<td>164 dB</td>
</tr>
<tr>
<td>Duplex mode</td>
<td>Half-duplex</td>
<td>Half-duplex, Full-duplex</td>
</tr>
<tr>
<td>Voice service</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>OTDOA positioning</td>
<td>Yes (Rel 14+)</td>
<td>Yes</td>
</tr>
</tbody>
</table>

---

4In fact, the 3GPP optimisations are defined for CIoT systems. The term CIoT includes both LTE-M and NB-IoT.
and LP-WAN groups. One can notice that it is not correct to think about the standards as direct competitors; in fact, they have been designed to serve various purposes and even within LP-WAN, where the coverage is more comparable, the technologies vary in the supported bitrate.

## 1.2 LP-WAN coverage issues

As mentioned in the previous section, the main purpose of LP-WAN technologies was to enable lightweight, long-range applications in a more efficient way than the obsolete GSM solution. As a matter of fact, the initial simulations of the LP-WANs has proven their outdoor, indoor and even deep-indoor connectivity and the coverage being significantly better than in the case of GSM [53,54]. However, in order to acquire certainty about ubiquitous coverage across various scenarios, one needs to perform thorough simulations and, even more important, real-life measurements considering especially the challenging outdoor, indoor and deep-indoor situations. The drive test of NB-IoT and LoRa reveals that both technologies can exchange data outdoors up to 5 km away from the base station, and the signals were also received in the underground area [55]. Albeit, it is difficult to evaluate the results, because the paper does not provide any details about the measurement setup. In [56], NB-IoT and LoRa coverage was simulated using real-life MNO data in the Parma region, Italy. The results indicates NB-IoT as the more robust solution, but it is visible that regardless the technology the coverage becomes significantly poorer in the deep-indoor scenario. In [57], the performance of LoRa was tested in various environments; while in the open area strong signal was still received at distances bigger than 1 km, in dense

![Figure 1.7. Communication range comparison of IoT technologies](image-url)
Excellent coverage: the biggest LP-WAN promise

forest and urban setting the connectivity was completely lost at 700 m and 350 m, respectively.

The aforementioned references confirm that conducting measurement campaigns in problematic surroundings helps capture performance limitations of the LP-WANs and realise that the standards considered in this work are not ready to offer truly robust coverage across the variety of realistic outdoor, indoor and deep-indoor scenarios. Particularly, the deep-indoor use-case constitutes the biggest challenge for all the technologies. It has to be mentioned that the reason for a communications solution to experience poorer coverage performance can be different to environmental profile or distance from the transmitter; it has been shown that the coverage of the unlicensed spectrum LP-WANs (LoRaWAN and Sigfox) can be threatened by the fact that the frequency band is heavily occupied by other devices [58], whilst NB-IoT performance depends on system configuration [59].

1.2.1 Contributions

Inspired by the state-of-the-art efforts evaluating the LP-WAN standards, the following problems were addressed:

1. Coverage degradation in a challenging scenario with Non-Line-Of-Sight (NLOS) conditions, e.g. a dense forest ([57]),

2. Poorer performance and coverage holes in deep-indoor situations ([54,56,60]).

In Chapter 2, a description of the outdoor measurement studies, evaluating LoRaWAN and NB-IoT coverage in a challenging harbour use-case is provided. Apart from filling the research gap relevant to problem 1, the work additionally strives to verify whether these technologies could be successfully applied for maritime applications, such as harbour occupancy system. Problem 2 is studied thoroughly in Chapter 3, where NB-IoT signal propagation in underground tunnel system is analysed. The research efforts in Chapter 3 describe the measurement campaigns and explain the process of new deep-indoor path-loss modelling through environmental parameter engineering.
CHAPTER 2

Outdoor LP-WAN coverage: a maritime perspective

New low-power IoT technologies have given smart communities lightweight and robust communication solutions that can be an enabling factor for numerous scenarios where residential, commercial and industrial objects are interconnected (e.g. intelligent traffic lights, waste management, asset tracking, remote patient monitoring). However, in real-life settings, both rural and urban areas encompass various objects (buildings, pedestrians, animals, vegetation, among others) that act as obstacles in wireless communications, thus constituting Non-Line-Of-Sight (NLOS) radio paths. In NLOS conditions, radio waves traverse the air along complicated, non-straight lines and are subject to complex physical phenomena as multipath propagation, diffraction, fading and reflection. The multitude of physical effects accompanying wireless transmission make it difficult to predict with absolute certainty how exactly the signal would behave in an arbitrary environment.

This thesis considers one of the NLOS outdoor scenarios, namely harbour environment. The choice of the maritime area can be explained by the fact that the economy of Denmark, a land with long coastline and big number of harbours, relies to a great extend on maritime industry and the country’s largest export share comes from shipping [61]. Moreover, the profile of the coastal environment and the presence of both masts, boats, harbour buildings and open water affect the propagation properties of the radio signal in the way that can be difficult to predict by means of computer simulations. The studies mentioned in this chapter evaluate the LP-WAN coverage performance in the harbour scenario and pave the way for more educated choices on the IoT radio technology in challenging outdoor settings. Part of the work presented here was included in the following conference publications:


## 2.1 Harbour scenario

For every smart community with access to the sea, rivers or lakes, coastal neighbourhood is a place of operation for multiple services related to tourism and free-time activities on one side, and maritime traffic and fleet control on the other side. Even in a non-industrial harbour for private yachts and boats one may identify necessary maintenance operations that can be time-consuming for the harbourmaster, if done manually. For example, controlling the occupancy of the harbour requires a lot of communication with the sailors (to get notified about entering or leaving the port, or if anybody has not come back as planned) and occasional physical inspection of the harbour slots (to verify the actual status of the harbour). An automated LP-WAN based solution would be a natural step towards the optimisation of maritime processes. However, from the point of view of wireless communication, harbour scenarios are considered challenging, as one needs to take into account the presence of various types of obstacles (boats, yacht masts, marina structures) and the open water, which affects radio wave propagation and requires tailored channel models [62]. At the time of the first trial (June 2018), to the best of the author’s knowledge, no publicly available experimental trials of NB-IoT in the maritime environment were present.

## 2.2 First maritime trial: NB-IoT

A NB-IoT measurement trial was performed to empirically verify whether NB-IoT technology can provide robust connectivity in a maritime environment, and thus be considered suitable for harbour applications. The trial was inspired by the equivalent outdoor study of LoRaWAN technology [1], where Received Signal Strength Indicator (RSSI) and Signal-to-Noise Ratio (SNR) samples were taken at 9 distinct locations in a Danish harbour Svanemøllehavn [63]. Those can be seen in Figure 2.1, but contrary to the NB-IoT eNB, located more than 2 kilometres towards the West, the LoRaWAN gateway was deployed 1 metre away from point number 9.

*Figure 2.1. The locations of the measurement points for the NB-IoT trial [2].*
2.2 First maritime trial: NB-IoT

Figure 2.2 shows that the LoRaWAN signal could be successfully received at all the measurement points, though at the farthest one the communications was successful only for Spreading Factor (SF) 7. It can be seen that the end-device maintained connectivity with the gateway for distances up to 919 metres, and this implies that harbours of area similar to Svanemøllehaven may be LoRaWAN covered using a single gateway. The experiment described in this section was conducted to verify the same for NB-IoT.

For that reason, the measurement area and the positions of the data acquisition were kept identical as in [1]. During the trial, a u-blox Sodaq Sara N211 board was connected to the NB-IoT network and it was capturing RSSI and SNR values of the DL signal from the eNB. The UE was subscribed to TDC network, serving NB-IoT in guard-band of Band 20 (820.5 MHz). The eNB was situated approximately 2.4 km from the western harbour edge. At each of the measurement points a total of 200 samples were captured: 100 samples 180 cm above the ground level and 100 samples at the height of 10 cm. The details of the measurement points are shown in Figure 2.1 and in Table 2.1.

2.2.1 Results

Figure 2.3 presents the acquired signal samples. One can notice that at all chosen positions the strength of the signal allows for successful communications, as the RSSI values are higher than -120 dBm, while the sensitivity of the receiving device is -135 dBm\(^1\). Furthermore, in most of the cases, the signal is stronger at the 180 cm level than at the 10 cm level. A possible reason might be the proximity of water, being responsible for signal absorption and reflection [64]. At the same time, the locations close to open water (3, 7 and 9) experience less distinct RSSI drop across different heights, than the places surrounded by other objects; in fact, the SNR of the signal


<table>
<thead>
<tr>
<th>Point #</th>
<th>LOS distance to BS (m)</th>
<th>Surroundings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2474</td>
<td>boats</td>
</tr>
<tr>
<td>2</td>
<td>2481</td>
<td>boats</td>
</tr>
<tr>
<td>3</td>
<td>2484</td>
<td>water</td>
</tr>
<tr>
<td>4</td>
<td>2673</td>
<td>boats</td>
</tr>
<tr>
<td>5</td>
<td>2836</td>
<td>parked cars</td>
</tr>
<tr>
<td>6</td>
<td>3164</td>
<td>buildings</td>
</tr>
<tr>
<td>7</td>
<td>3195</td>
<td>water</td>
</tr>
<tr>
<td>8</td>
<td>3359</td>
<td>buildings</td>
</tr>
<tr>
<td>9</td>
<td>3439</td>
<td>water</td>
</tr>
</tbody>
</table>

Table 2.1. Description of the measurement points [2].
close to the water surface is slightly higher than at higher positions, which was also observed in another measurement study in a tropical aquaculture environment [65]. It can be seen that neither RSSI nor SNR decreases proportionally to the distance to the Base Station (BS). The worst signal quality (SNR) was captured at the points in the neighbourhood of marina premises and office buildings; this could have been caused by multipath fading effect. Although the weather conditions, such as temperature and humidity, have proven to influence the strength of the signal outdoors [66], in the presented case, they were not measured and instead were assumed stable throughout the study.
2.2.2 NB-IoT vs. LoRaWAN

Comparing the collected results with the findings in [1] it may seem that NB-IoT offered better connectivity in Svanemøllehavn area, because the RSSI and SNR values in LoRaWAN case were lower, and at the same time the LoRaWAN gateway was placed within the harbour area, whereas the eNB of NB-IoT was located 2.4 km farther away. However, such conclusion would not be correct, since the eNB was transmitting from more than 20 metres height, whilst the LoRaWAN gateway could only be placed 2 metres above the ground. Moreover, the details of the eNB radio transmitter parameters were not given, which makes direct comparison difficult. Nevertheless, it can be concluded that in the real-life conditions, both the technologies can serve the maritime use-cases providing robust coverage.

2.3 Second maritime trial: NB-IoT and LTE-M

In September 2020, a new measurement trial was done in the same harbour area. The same measurement locations and the methodology was preserved, however this time both NB-IoT and LTE-M coverage was tested by means of u-blox SARA R410M. The firmware of the device did not support SNR and RSSI measurements, therefore RSRP samples were taken. During the measurements, the device was tuned to measure the downlink signal of Band 20 (795 MHz for In-Band deployed NB-IoT, 796 MHz for LTE-M) using one technology at a time. In this trial, the Subscriber Identity Module (SIM) card used was subscribed to Telia operator, as the available cards of TDC did not operate with LTE-M, nor NB-IoT.

2.3.1 Results

Signal power measurements of both cellular IoT technologies are presented in Figure 2.4. Some of the sample groups (all NB-IoT readings at point 4 and NB-IoT readings at point 7 taken at 10 cm level) freeze at a constant value of -44 dBm (horizontal black line in the figure), which according to the module's manual signifies: "signal power of 44dBm or stronger". Such high values were considered erroneous and those samples were disregarded from further analysis. It can be seen that the device was under coverage in all the chosen locations. LTE-M signal was always stronger at the height of 180 cm than 10 cm, but in the case of NB-IoT the opposite trend was discovered for points: 2, 6 and 9. Close to the ground level, NB-IoT signal was stronger than LTE-M signal, but at some of the points (5, 8 and 9) the RSRP of the LTE-M was higher than the NB-IoT one. Noteworthy, LTE-M measurements are significantly more stable (closer to the median value) than NB-IoT samples at some of the points (3, 6, 9).
It has to be mentioned that in 2163 out of 3600 samples taken the end-device received signals from multiple base stations; only the strongest DL signal was considered in such cases.

The NB-IoT samples described in Section 2.2 were used in Figure 2.5 next to the RSRP readings from the second trial. One can observe significant difference in the DL signal power between the new and the old study. The cause of much stronger signal in 2020 is the new infrastructure deployment of the MNO - based on the publicly available data [67], the closest Telia NB-IoT eNB was found only 600 metres away from the harbour centre. The proximity of the base station can also partly explain that the behaviour of the signal is different; for example, in the new trial one can notice very high RSRP at point 4 and, contrary to the old trial, at points 6 and 9 the received signal was stronger at 10 cm than at 180 cm. A more detailed signal behaviour comparison of both NB-IoT trials is troublesome due to multiple differences in the experiment conditions: 1) different MNO, operating frequency, and NB-IoT deployment mode 2) different NB-IoT module used (the antenna was the same), 3) different deployment situation (shorter distance to the BS).
2.4 Discussion

The collected signal samples allow for a performance showcase of 3 LP-WAN technologies: LoRaWAN, NB-IoT, and LTE-M. The former two standards were evaluated in early 2018, whereas NB-IoT was tested again, together with LTE-M, in the fall of the year 2020. Therefore, the collected signal strength samples come from multiple technologies, operators and environmental situation, and might be used as an important part of an empirical dataset for further analyses of outdoor sub-GHz signal behaviour in NLOS radio conditions. Nevertheless, the satisfactory performance of all the considered IoT standards has been proven in a real-life setting.

The area chosen for the tests is a small yacht harbour with low-height premises, situated in dense urban surroundings. The results obtained in this study might not be extrapolated to large-scale industrial harbours and marinas in the rural environment, as significantly different distribution of harbour machinery and huge vessels, or the non-urban neighbourhood would influence the LP-WAN signal propagation in a different way. Nevertheless, it is important to investigate the performance of IoT coverage in maritime environments separately from the urban and rural ones, as harbours, embracing both boats, buildings, people, as well as open water surroundings, constitute an unique NLOS scenario, which cannot be directly compared to building-dominated
cities and rural areas, covered with forests or farms.

Even though the presented measurement trials provide valuable insights into LP-WAN operation in a challenging outdoor scenario, they can by no means be seen as an exhaustive comparative coverage study. First of all, the chosen locations for taking the radio samples represent various types of possible obstacles, however more of them could have been chosen for a more fine-grained measurement campaign. Secondly, performing the trials in multiple harbours would hinder the peculiarities of the chosen area, that introduce bias to the acquired data. Moreover, the presented LP-WAN outdoor performance comparison is incomplete, as Sigfox technology is not included. One can also realise that only downlink signalling is analysed, while it would be of high interest to investigate uplink communication as well.

The limitation of the amount of samples taken and only one harbour chosen can be explained by the constraint of time dedicated to this study. However, as far as the involved LP-WAN technologies are concerned, the following issues prevented a more complete study:

- **No access to the proprietary network infrastructure.** Both cellular IoT standards (NB-IoT and LTE-M) and Sigfox operate on MNO’s premises that the subscriber can communicate with, but has no control over them. It means that no configuration or direct access to the eNB/gateway was granted, which made it difficult to design a reliable uplink connectivity study. In the experiment, the RSSI, SNR and RSRP values were collected at the end-device via AT commands; this was not possible at the gateway side. In fact, even the details of the current eNB/Sigfox gateway type and transmitter settings were not known, thus limiting how detailed the comparison study could be.

- **No information about the gateway locations.** In the case of Sigfox, the connectivity was provided by the local MNO, which did not provide information about the gateway deployment in the studied harbour area. Therefore, it was difficult to perform a measurement study equivalent to the one presented in this chapter, due to lack of reference positions of the gateways and inability to estimate the straight-line distances between the sensor and the gateway.

- **Technologies not supported at the time of the study.** The cellular IoT samples were taken using the SIM cards of two operators, however, in an inconsistent way. The inability to compare the measurements from both the operators and taken at the same time was caused by the fact that: 1) At the time the first trial was performed, LTE-M technology was not yet enabled by any Danish network operator; 2) at the time of the second trial, the SIM card of the operator used in the first trial was not able to attach to the network via NB-IoT anymore. Huge infrastructural and service changes related to the nation-wide 5G rollout [68] could have resulted in a deactivation or change of the subscription type of our SIM cards. At the same time, the MNO failed to provide new cards within the timespan of this project.
2.5 Summary

In this chapter, outdoor coverage of LP-WANs was evaluated in one of the environments characterised by NLOS paths and challenging radio conditions: urban harbour area. A measurement trial of downlink signal strength and quality of NB-IoT and LTE-M was performed and then compared to an equivalent experiment with LoRaWAN. The results clearly indicate that applying either of the aforementioned LP-WANs for maritime IoT use-cases is a good idea, as robust coverage can be assured with minimal infrastructure cost (1 base station/gateway covers the whole area of Svanemøllehavn). Unfortunately, including Sigfox in the analysis was not possible due to insufficient reference information regarding the local infrastructure.

Even though the availability of particular LP-WAN technologies in the given location may change substantially over time, outdoor trials can help to evaluate the IoT communication in a real-life situations, where the peculiarities of the surrounding objects and terrain and the presence of other wireless devices may constitute such a testing environment that simulations cannot easily produce. Additionally, one may discover the most current connectivity options available in the given area (e.g. during the first trial, only NB-IoT was reachable, while the second study could also include LTE-M).

It is beneficial to repeat the outdoor coverage experiment whenever a new version of the technology becomes available (e.g. LoRaWAN 1.1), the Mobile Network Operator shares its deployment and radio information (e.g. about the radio parameters Sigfox gateways and their locations) or the local deployment situation changes (e.g. Telia operator begins to serve NB-IoT next to TDC). Moreover, the signal strength samples may be used to formulate a new, or improve the existing channel model, thus contributing to more accurate outdoor coverage modelling.
In the previous chapter, it was shown that LP-WAN communication can provide robust coverage in outdoor applications, also when the closest base station is several kilometres away from the sensors. In fact, long-distance outdoor deployment is not the only challenge the LP-WANs were created to address. Another scenario, where coverage holes have traditionally been taken into account is the deep-indoor scenario. The radio conditions in such an environment are so poor, that the applications such as gas or water metering could not be possible prior to the advent of IoT technologies supporting extreme coverage. In this work, deep-indoor situation is identified when the transmitter and the receiver are separated from each other by multiple walls and/or the ground level (e.g. transmission from an above-ground transmitter to an underground receiver).

The research presented in this thesis concentrates on the underground behaviour of NB-IoT signal. However, the findings only concern radio propagation and attenuation and therefore they can also be relevant for other LP-WANs operating in sub-GHz frequency band. An initial simple simulation is followed by tunnel measurement campaigns and finally, the collected data are statistically analysed and with the aid of ML tools, new propositions for deep-indoor path-loss modelling component are derived. The efforts described in this chapter were also published as the following items:


### 3.1 Underground sensor deployment: a killer scenario for wireless communications

Wireless transmission in underground areas (tunnels, parking lots, mines, storage zones, etc.) often needs to face huge attenuation, if not blockage, of the signal. First of all, Non-Line-Of-Sight communications is a norm, unless signal repeaters are additionally deployed underground\(^1\). The signal reflections in a long tunnel, for example, can easily interfere with one another and cancel out or amplify, thus influencing the overall noise level. Consequently, the shape of the tunnel, its cross-section, and the presence of big objects, acting as obstacles, have substantial impact on signal attenuation, as identified in [69]. In their survey, the authors also mention the factor of electromagnetic properties of the materials, from which the tunnel is built; however, the influence of such factor is found to be less significant. Apart from radio wave behaviour inside the underground area, one ought to consider the losses the signal is subjected to during ground penetration; these can be affected by certain soil characteristics, such as moisture level [70]. Furthermore, the effect of the above-ground profile of the terrain needs to be considered as well.

All in all, it is not easy (sometimes even infeasible) to successfully provision robust coverage in deep-indoor areas. However, overcoming this challenge has been under interest of academia, due to numerous use-cases under the umbrella of smart communities, relying on remote access to sensors located underground, for instance: urban drainage system [71], monitoring of cable system [72] or mine safety warning system [73]. One may also find a possible interest of the industry in the family of remote metering applications, where updating the gas or water consumption status in the cordless way allows for huge time and expense minimisation, as workers do not need to read the data in hard-to-reach locations.

\(^1\)In such a case, the problem of reaching out the above-ground base-station remains, the only difference being the entity “in trouble”: the repeater instead of the sensor.
In order to plan and realise robust and economically feasible IoT applications in a deep-indoor scenario, one must have the following at hand:

- Optimised hardware and software design, guaranteeing minimal power consumption.
- A tailored communications technology, so that connectivity underground can be expected.
- Thorough knowledge about signal behaviour in the target area, allowing for correct sensor placement and infrastructure planning.

The goal of the work described in this chapter is to acquire understanding IoT signal propagation in the underground environment, which enables more optimal deployment and communication technology selection.

### 3.1.1 Related work

Lauridsen et al. [53] estimated the coverage of General Packet Radio Service (GPRS), Long Range (LoRa), NB-IoT and Sigfox in a simulation study. The comparison of the LP-WANs depends on the choice of constant losses, representing deep-indoor scenario. In the strictest case (deep-indoor penalty of 30 dB), none of the technologies can cover the tested area; the outage areas for NB-IoT, Sigfox and LoRa are 8%, 13% and 20%, respectively. On the basis of a Ray-Tracing simulation and a single-point measurement, the authors of [74] conclude that NB-IoT and LoRaWAN can be applied in underground scenarios with up to 1 metre of depth. The simulation study, presented in [70] shows that as long as LoRaWAN devices are buried no deeper than 70 cm under the ground level, they can be reachable from more than 1 km. Malik et al. [75] performed a NB-IoT measurement campaign and experienced no connectivity underground, whenever the receiver was more than 400 m away from the base station. The authors of [76] performed Ray-Tracing simulations and claimed that NB-IoT and LTE-M cannot provide robust deep-indoor coverage, unless additional enhancements, e.g. multi-hop relaying, are introduced.

A new LP-WAN channel model derived based on the empirical measurements collected in 3 different locations, is described in [77]. The authors construct a generic statistical formula accounting for short- and long-term path-loss variations. On the other hand, all the measurements were done with up to 500 metres distance between the transmitter and the receiver, and the amount of deep-indoor (underground) samples used to formulate the model was not given.

One might notice that most of the current research efforts on LP-WAN deep-indoor connectivity are based on simulations, rather than measurement campaigns. The simulation studies provide useful insights into the performance of the IoT standards, but do not grasp the whole complexity of real-life signal behaviour (deep-indoor scenario is simplified to a constant loss penalty). On the other hand, the measurement campaigns done so far encompass limited area and consist of small amounts of samples.
Therefore, large-scale empirical dataset of LP-WAN signal measured underground is an essential asset in the process of deep-indoor coverage modelling.

3.2 Investigating deep-indoor coverage

3.2.1 NB-IoT and LoRaWAN trial

<table>
<thead>
<tr>
<th>Parameter</th>
<th>NB-IoT</th>
<th>LoRaWAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location (Lat/Lon)</td>
<td>55.785192/12.52381976</td>
<td>55.784122/12.523057</td>
</tr>
<tr>
<td>Frequency</td>
<td>847 MHz</td>
<td>867 MHz</td>
</tr>
<tr>
<td>Gain</td>
<td>16 dB</td>
<td>13 dB</td>
</tr>
<tr>
<td>Height</td>
<td>40 m</td>
<td>20 m</td>
</tr>
<tr>
<td>TX Power</td>
<td>49 dBm</td>
<td>28 dBm</td>
</tr>
<tr>
<td>Additional losses</td>
<td>2.768 dB</td>
<td>2 dB</td>
</tr>
<tr>
<td>RX sensitivity</td>
<td>-135 dBm</td>
<td>-137 dBm</td>
</tr>
</tbody>
</table>

Table 3.1. Base station simulation parameters [3].

In the initial work, published in [3], a simulation and measurement trial was performed to investigate NB-IoT and LoRaWAN indoor coverage capabilities in DTU Lyngby Campus area. In the simulation part, a free version of an online network coverage simulation tool, Xirio Online, was utilised [78]. Table 3.1 presents the input parameters; in both cases the path-loss model used for computations was Okumura-Hata model, tuned to "urban - small city" scenario with the transmitter situated on a rooftop of the building. The purpose of the simulation was to observe theoretical coverage boundaries in the Campus area. These can be observed in Figures 3.1 and 3.2. Note that white colour corresponds to RSSI values of -40 dBm or more, light green refers to RSSI values between -60 dBm and -40 dBm, dark green denotes RSSI between -90 dBm and -60 dBm and blue signifies values between the receiver sensitivity threshold and -90 dBm.

The results show that considerable area can be successfully covered by a single base station or gateway. The range denoted by dark green colour corresponds to at least -90 dBm of RSSI, thus even if the sensor was located in a basement, the connectivity would be preserved. In the case of NB-IoT, such region has a radius of 5 km, while in the case of LoRaWAN it is 1.5 km.

3.2.1.1 Measurements

For the sake of comparison and verification of the simulation results with empirical data, a measurement trial was conducted, taking 100 RSSI samples at the selected indoor locations within the DTU Campus (see Figure 3.3). The samples were captured at different floors and in some places, denoted in blue, it was also possible to measure
3.2 Investigating deep-indoor coverage

Figure 3.1. Xirio Coverage simulation results for NB-IoT [3].

Figure 3.2. Xirio Coverage simulation results for LoRaWAN [3].

the signal in the basement. The equipment used in the trial was: u-Blox Sodaq Sara-N211 NB-IoT device, Seeeduino LoRaWAN board and Kerlink Wirnet Station LoRaWAN gateway (NB-IoT eNB was provided by the local MNO).

Figure 3.4 shows the measured RSSI with respect to the distance to the base station, together with reference green lines symbolising Xirio simulator signal strength estimations (-40 dBm in Figure 3.4a and -90 dBm in Figure 3.4b. Clearly, the results of the simulation are not realistic, thus not useful for deep-indoor coverage planning; the measured RSSI data are 30-40 dB lower than in Xirio tool. One of the reasons is, that the simulator does not take into account the building penetration loss component; similarly, the chosen path-loss model is well-suited only for Outdoor-to-Outdoor scenarios and thus underestimates the losses in the considered case.

The measurements illustrated in Figure 3.4 do not include the basements of the buildings. In fact, the
connectivity was lost every time the device placed underground was farther than 300 m away from the transmitter. Nevertheless, NB-IoT signal strength is noticeably higher than the one of LoRaWAN. However, the NB-IoT eNB transmitted at higher power with the antenna of higher gain.

All in all, NB-IoT might appear more suitable for underground scenarios than LoRaWAN, but the aforementioned study did not bring any strong evidence of that, as the basement measurements could not be collected due to connectivity problems. However, it was shown that simulation tools, available at the time of the study, need to be used with care, as they tend to represent the LP-WAN coverage too optimistically.

### 3.2.2 NB-IoT underground trial: test of the coverage and validation of the existing path loss models

Unfortunately, the first study did not result in a "true" deep-indoor dataset, therefore a new measurement campaign was designed. Apart from collecting real-life signal strength data, the goal of the study was to evaluate some of the existing statistical path-loss models in the underground use-case. A closer look was taken at 3GPP 38.901 models, defined for wide range of frequencies (from 500 MHz to 100 GHz) and encompassing both Outdoor-to-Outdoor (O2O), Outdoor-to-Indoor (O2I) and Indoor-to-Indoor (I2I) communications scenarios [79]. Specifically, the investigated overall O2I path loss formula looks as follows:

\[
PL = PL_b + PL_{tw} + PL_{in} + N(0, \sigma_P^2), \tag{3.1}
\]

where \(PL_b\) is the basic outdoor path loss, \(PL_{tw}\) is the building penetration loss through the external wall, dependent on the material the building is constructed of,
3.2 Investigating deep-indoor coverage

\( \sigma_p^2 \) is the standard deviation for the penetration loss and \( PL_{in} \) is the loss that is dependent on the depth into the building. In this work, the focus is on \( PL_{in} \), thus (deep) indoor path-loss component in the Outdoor-to-Deep-Indoor (O2DI) scenario. The 3GPP standard defines \( PL_{in} \) as:

\[
PL_{in} = 0.5d_{2D-in},
\]

where \( d_{2D-in} \) is the indoor distance in 2D, to the outer wall. Since the model in its current state is very simple (with only 1 variable) and has not been calibrated with underground empirical data, a natural doubt arises whether the proposal of 3GPP 38.901 is accurate for path-loss prediction in deep-indoor scenarios.

To the best of the author’s knowledge, nobody had addressed such problem in the publicly available experimental work before. Thus, in order to fill that knowledge gap, a series of RSSI measurements was performed along a fragment of DTU Lyngby Campus underground tunnel system interconnecting most of the university buildings.

3.2.2.1 Experiment description

The quantity considered in the study was strength of downlink reference signal of the NB-IoT Release 13 system, deployed in guard-band of band 20 (820.5 MHz). In general, the downlink transmission in NB-IoT employs the following channels [51]:

1. Narrowband Physical Broadcast Channel (NPBCH) that conveys essential operation information in a form of Master Information Block (MIB),

2. Narrowband Physical Downlink Control Channel (NPDCCH) that contains scheduling information for uplink and downlink,

3. Narrowband Physical Downlink Shared Channel (NPDSCH), carrying random access response, paging indication and user data.

Besides, 3 downlink signals are defined:

1. Narrowband Primary Synchronisation Signal (NPSS), applied for synchronisation between the UE and the eNB,

2. Narrowband Secondary Synchronisation Signal (NSSS), by means of which the end-device learns cell identity and the frame structure,

3. Narrowband Reference Signal (NRS), used as a phase reference for demodulation of the downlink channels

In the experiment, the RSSI of NRS was measured. The mapping of NRS, NPDCCH and NPDSCH on a PRB resource grid in guardband NB-IoT deployment are illustrated in Figure 3.5.

The samples were acquired in two buildings of DTU Lyngby Campus (340 and 343), ground and -1 floors, and several tunnel corridors on levels -1 and -2; all the
locations, together with the eNB position, can be seen in Figure 3.6. The measurement procedure can be described as follows: a trolley with Rohde & Schwartz TSMW radio scanner, Arduino Mega 2560, 2 LiLight Detection And Ranging (LiDAR) sensors and a portable computer was moved along the area. At the time the trolley was in motion, the radio scanner was continuously capturing the RSSI of the downlink reference signal from the closest above-ground NB-IoT base station. TSMW was additionally recording the waveform of the signal. Since the area of study was out of Global Positioning System (GPS) coverage, a local LiDAR-based positioning solution was used. 2 lasers, pointing at the walls were measuring the distances to the walls and forwarding them to the Arduino board, responsible for translating the raw laser readings into XY co-ordinates and sending the timestamped locations to the computer. As both the positioning and radio data were collected by the laptop with timestamps, it was possible to correlate the samples with the corresponding positions in the room or tunnel. The setup is depicted in Figure 3.7.

During the post-processing in MATLAB, the samples from the TSMW and the XY positions were merged into a single dataset that also contained the measurements’ GPS positions. The conversion was possible, as the GPS co-ordinates of both the eNB and room/tunnel entry were known. Moreover, one was able to determine the distance between the measurement point and the outer-most wall that was closest to the eNB in the straight line; such metric is referred to as indoor distance.

Figure 3.5. NB-IoT resource grid (guard-band and stand-alone deployments) [5].

Figure 3.6. Measurement trial locations [5].

Figure 3.7. Experimental setup [5].
Table 3.2 summarises the number of samples collected during the study.

**Table 3.2.** Number of samples per location [5]

<table>
<thead>
<tr>
<th>Location</th>
<th>Number of samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Building 340</td>
<td>2172</td>
</tr>
<tr>
<td>Building 343</td>
<td>12256</td>
</tr>
<tr>
<td>Tunnel entrance</td>
<td>713</td>
</tr>
<tr>
<td>Tunnel level -1</td>
<td>803</td>
</tr>
<tr>
<td>Tunnel level -2</td>
<td>6246</td>
</tr>
<tr>
<td><strong>Total # of samples</strong></td>
<td><strong>22190</strong></td>
</tr>
</tbody>
</table>

### 3.2.2.2 Results

The outcome of collecting the signal strength data is illustrated in Figure 3.8. The shape of the heatmap reflects the movement patterns of the trolley, scanning the $25m^2$ room (Figure 3.8a) and a fragment of the underground tunnel system (Figure 3.8b), where the setup traversed the corridor only once. In the ground floor scenario, even though not 100% of the area is covered by the trial, one can observe that the RSSI is generally lower close to the walls and higher at some spots in the middle of the room. Moreover, the signal was substantially weaker in the tunnel than in the room (most of the values are around 10dBm lower), but the standard deviation underground is smaller than on the ground floor.

The data from ground floor of building 343 and the -2 level part of the tunnel are presented in Figure 3.9. The plots compare the observed RSSI to indoor distance relationship with the corresponding 3GPP 38.901 predictions. Not surprisingly, the

![Figure 3.8](image-url)

(a) Staircase, ground floor.  
(b) Tunnel, level -1.

**Figure 3.8.** RSSI heatmap in an above-ground and underground location [5].
existing model proofs its path-loss estimation accuracy for an ordinary indoor setup, where the measured room is separated from the outside world by a single wall (the outer building wall). On the other hand, 3GPP 38.901 fails to predict the signal behaviour in the underground scenario, as instead of linear drop with the increasing indoor distance, an indistinct value fluctuation between -82 dBm and -84 dBm is visible. This may imply that in the O2DI communications, the total link budget is more complicated than the formulae the 3GPP 38.901 model suggests (Equations 3.1 and 3.2). Specifically, the deep-indoor path-loss component is not suitable for successful underground prediction and needs to be refined.

3.3 Towards new models for underground IoT coverage planning

Accurate deep-indoor coverage modelling is a challenging task to accomplish. Based on the findings presented in the previous section, one may notice that the statistical model relying on simplistic linear relationship and a single parameter of indoor distance cannot reflect the complexity of IoT signal behaviour in underground areas. Unfortunately, it is difficult to identify all the plethora of features that influence signal propagation in underground areas. The main reason is the lack of awareness of the peculiarities of a considered environment: the parameters of the ground and the tunnel structure, precise positions of the devices placed underground, the exact distribution of objects inside the considered tunnel, and many others. The problem of coverage planning is depicted in Figure 3.10. All in all, collecting the valuable empirical data from the deep-indoor scenarios would help improve the existing path-loss

Figure 3.9. Measured NB-IoT RSSI, together with 3GPP 38.901 estimations [5].
models and increase the feasibility of the underground coverage modelling.

This section describes multi-stage research efforts, published in [8] and [9]. The main purpose of this exploration study was to investigate how the features or parameters that could be derived from the communications scenario can be used to deepen the understanding of deep-indoor IoT signal behaviour. Analogically, it was expected to find out that some of the features hold no significance and should not be applied in the modelling of the path-loss channel.

A large-scale NB-IoT measurement campaign was performed in DTU Lyngby Campus tunnel system. The measurement points were precisely located with the aid of the high-resolution LIDAR plan of the tunnels. Based on the acquired data, several features were engineered and analysed statistically in terms of usability for signal power prediction. In the conference work [8], the parameters related to transmitter-receiver distance were explained and analysed in the linear domain. In the journal article [9], more features were added and analysed both from linear and non-linear perspectives. Moreover, on the enhancement of 3GPP 38.901, new models for deep-indoor path-loss prediction were derived using the most relevant features. The performance of the proposed models was compared to the indoor component of 3GPP 38.901 (see Equation 3.2), used throughout the whole study as a baseline.

3.3.1 Methodology

3.3.1.1 Why statistical modelling?

Coverage modelling is an important process that is indispensable at real-life IoT deployments, as it allows for the estimations of radio connectivity situation in the considered area without the need of costly manual testing (see Figure 3.10). It is de-
sirable to use versatile modelling tools that are capable of predicting the coverage in various scenarios (ideally, in an arbitrary situation) utilising input data that is feasible to acquire. In general, one may apply modellers based on Ray Tracing approach or statistical models. As far as deep-indoor use-cases are concerned, the former solution may not be convenient, as Ray Tracing requires detailed geographical data describing the scenario, which is not always available; furthermore, the models themselves are complicated, resulting in high software sophistication (thus, high price) and computational burden. On the other hand, statistical models, though much simpler, are able to provide comparable predictive performance to the Ray Tracing methods [80]. Since in statistical modelling real-world phenomena are represented as mathematical formulae, the models are easy to implement. Moreover, if the parameters used in the model are not characteristic to a particular environment, then the said model can be considered universal, i.e. applicable to a variety of scenarios. To optimise a statistical model, one needs to address the following challenges:

- Select optimal (i.e. the most explanatory) parameters and the amount thereof, in order to tune the model applicability and complexity;
- Calibrate model coefficients with experimental data, to reduce the bias of the model; the more samples and the more heterogeneous the calibration data, the more robust the model becomes.

3.3.1.2 From raw samples to new models

The procedure that has led to a new deep-indoor model proposals can be seen in Figure 3.11. Having access to the underground tunnel areas, the radio signal measurements were taken. Thanks to the availability of the LIDAR data, the feature candidates were derived and then their relevance in terms of predicting the strength of the signal was studied by means of the statistical tools described in Section 3.3.1.3. The proposed path-loss indoor \( (PL_{\text{in}}) \) terms are the models including the most relevant feature candidates and calibrated with the collected samples.

The new estimators of \( PL_{\text{in}} \) were found according to the following certain assumptions towards the link budget in the considered communication scenario (O2DI). While in general, the link budget consists of multiple components:

- transmit power \( P_{TX} \),
- antenna gain of the transmitter \( (G_{TX}) \) and the receiver \( (G_{RX}) \),
- noise figures of the communicating entities: \( L_{TX} \) and \( L_{RX} \),
- other, miscellaneous losses \( L_{misc} \),

in the study, it was not possible to know the values of the transmitter (eNB). For that reason, the link budget calculations included educated guesses of \( G_{TX} \) and \( L_{TX} \). \( L_{misc} \) term, corresponding to any other factors that were unknown (e.g. antenna...
3.3 Towards new models for underground IoT coverage planning

![Diagram](image)

**Figure 3.11.** Deriving a new deep-indoor path-loss component from the measured data [9].

radiation pattern, antenna polarisation) was defined as a constant to be adjusted. Specifically, the value of $L_{misc}$ was found through minimisation of the Mean Square Error (MSE) of the measurements towards a link budget, where the considered constant accounts for both the unknown gains and losses, and any inaccuracies in the assumptions. Calibrating $L_{misc}$ in such a way could be done, as TX/RX gains and losses in general were assumed invariant throughout the measurement campaign. Figure 3.12 presents how the raw signal power samples were used to produce $PL_{in}$ data.

3.3.1.3 Statistical toolkit

The parameters derived in this study were analysed by means of feature filtering, Recursive Feature Elimination (RFE), Lasso Regularisation, Ordinary Least Squares (OLS) regression and Gaussian Process Regression (GPR). OLS and GPR were also used to propose novel deep-indoor path-loss models.

**Feature filtering** is a simple method of feature selection, based on removal of a subset of parameters that do not fulfil the filtering condition. In this work, the filtering was applied after finding the Pearson correlation coefficients between the inputs and the output. The features which Pearson correlation value below 0.3 were discarded.

**Recursive Feature Elimination** is a method that selects the most important input parameters using model accuracy metric. Starting with all the features given as input, RFE builds new models from the features, compares the accuracy, removes the least relevant features and continues building new models using the remaining parameters. The procedure terminates when the desired number of features is achieved [81].
In this work, the optimal number of features \( N \) was found by testing the RFE on all possible amounts of features; then RFE was employed again to select \( N \) most important features.

**Least Absolute Shrinkage and Selection Operator (LASSO)** is a method that selects model features by minimising Sum of Squares Error (SSE), however with an additional upper bound on the sum of the absolute values of the features’ coefficients. LASSO uses a regularisation (or shrinkage) coefficient to penalise irrelevant features by diminishing their coefficients. In the case it becomes a zero, a feature is removed from model training process \([82]\). In this work, LASSO regularisation was used for feature selection; all non-zeroed remaining parameters were considered relevant.

**Linear Regression** is a method of statistical inference applicable to situations when the underlying phenomenon resemble linear relationship \([83]\). The technique is simple to use and interpret, since the output \( y \) is modelled as linear combination of the input terms \( x \), calibrated by the value of the assigned weights \( \omega \). Considering \( M \) features, the mathematical representation of linear regression is as follows:

\[
y = f(x, \omega) = \omega_0 + \omega_1 x_1 + \cdots + \omega_M x_M \tag{3.3}
\]

Optimisation of a linear regression model can be done by tuning the weights of
the inputs, so as to minimise the deviations between the predicted and the observed
outputs. The optimisation approach applied in this work, called Ordinary Least
Squares, aims at minimising Sum of Squares Error, defined as:

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2,$$  \hspace{1cm} (3.4)

where \(n\) is the number of output samples, \(y_i\) is the actual observation and \(\hat{y}_i\) is
the predicted output value \([84]\).

Linear regression was used to investigate statistical relevance of the engineered
features and then to propose a new linear term for \(PL_{in}\) prediction. For the former
purpose, linear models consisting of different combinations of the features as \(x\) and
the measured RSRP as \(y\) were trained; the measures of goodness of fit coefficient \(R^2\) and Residual Mean Square Error (RMSE) were examined to find the most relevant
parameters. For the latter goal, a new linear regression model was trained, using the
most relevant features.

**Gaussian Process** is a probabilistic method suitable for regression and classification. GPR can be useful for statistical inference of linear and non-linear relationship
of any number of dimensions, provided that both the input and the output follow
gaussian distribution \([85]\). The immediate outcome of GPR is a distribution of function (a process), which can be sampled to give possible input-output relationship (functions). Since the process is a probabilistic distribution, the significance of its
samples-functions can be evaluated by studying the given probabilities. Regression
with Gaussian Processes is based on the fact that the following operations do not
alter the characteristics of a gaussian distribution:

- Marginalisation, when some of the dimensions are discarded,
- Conditioning under the training data \([85]\).

Let us consider a dataset of \(M\) samples. GPR first creates M-dimensional distribu-
tion, represented by mean \(\bar{f}_*\) and covariance matrix \(cov(f_*)\). The latter one is derived
by means of a kernel function, which also provides similarity measure between any
two input points. In this work, the applied kernel function, called squared-exponential
kernel, is represented as:

$$cov(f(x), f(x_*)) = K(X, X_*) = \exp \left\{ \left( -\frac{1}{2} |x - x_*|^2 \right) \right\}$$ \hspace{1cm} (3.5)

GPR creates a predictive function distribution given the training and test data:

$$f_* | X, y, X_* \sim N(f_*, cov(f_*))$$ \hspace{1cm} (3.6)

where the mean is defined as:
\[ \hat{f}_* = K(X_*, X)[K(X, X) + \sigma_n^2 I]^{-1} y \] (3.7)

Finally, sampling from the distribution described in Equation 3.6 different predicted output functions can be obtained. The mean of the samples is the prediction of the output variable given the test inputs. Noteworthy, the hyperparameters of the chosen kernel function (squared-exponential) correspond to the distance in the input space at which the function values become uncorrelated. This enables Automatic Relevance Determination (ARD), a straightforward method of feature relevance analysis, as it is enough to look at the hyperparameter values: the smaller the length-scale of an input feature, the more relevant the feature is [85].

In this work, GPR with ARD was employed to enable feature analysis in non-linear domain. Furthermore, a new \( PL_{in} \) prediction GPR model was created with the same features as the linear regression model. More details on the new models and the comparison with 3GPP 38.901 standard is presented in Section 3.3.5.

### 3.3.2 Underground measurement campaign

The measured sub-GHz signal power and other UE radio statistics come from 1048 underground positions of DTU underground tunnel system. The measured area (1.6 km in total) was divided into a set of equidistant points, such that the signal data were taken only at these points, in a stationary way\(^2\). The distance between consecutive locations was 1 or 2 metres. The above-ground profile of the Campus is illustrated in Figure 3.13 and the layout of the measurement area is shown in Figure 3.14. The tunnels are situated under the area partly covered by parking with high vegetation and partly by university buildings, which are 2-4 floors high and are built from brick or glass/steel structure. The measurements were collected only in the main corridors connecting groups of buildings, but never in the corridor with either end at the entrance of a building. It means that every measurement trial started and ended at a certain distance away from the entrance to the tunnel system; thus, the influence of the tunnel entrances was not examined.

The equipment used in the campaign was Rohde&Schwartz TSMW network tester, u-blox Sodaq SARA N211 NB-IoT device, a laptop and a gel rechargeable battery. The hardware was mounted on a trolley, similarly as in Figure 3.7, but instead of LIDAR lasers and Arduino board, the NB-IoT device was used. The antennae of the NB-IoT sensor and the TSMW were fixed vertically on the trolley. At each measurement point, while the trolley was standing still, the network tester measured 1 million of IQ

---

\(^2\)Contrary to the previous measurement experiment, where the samples were taken continuously over the measured area
radio samples, using low-pass filtering around the operating frequency. Afterwards, the UE took 10 radio statistic measurements, the average of which were later used to account for shadowing and large-scale fading impairments. The radio parameters of the experiment are summarised in Table 3.3.

The measurement campaign described in this section can be considered unique, since even though the underground area was out of Global Navigation Satellite System (GNSS) coverage, all the measurement points could be precisely localised. The positioning solution relied on high-resolution LIDAR plan of the entire tunnel system. The LIDAR dataset consisting of \((x,y,z)\) coordinates enabled precise (1 cm of uncertainty) localisation based on the following procedure:

1. The exact location of the first and the last point of each corridor interval was documented by photos.
2. The start and end positions of the measurement series were identified in the LIDAR point cloud and the GPS coordinates were computed.
3. Knowing the amount of points, the exact start/end position and the interval between consecutive points, the remaining points were localised by interpolation.

In Figure 3.14, one can see the RSRP values at the exact locations in the tunnels; the plot of signal power against the 3D eNB-to-UE distance can be observed in Figure 3.15.

3.3.3 Feature engineering

In a O2DI scenario, the transmitted signal has to traverse air, outdoor obstacles, the ground, tunnel walls and underground obstacles before it reaches the receiver. The

<table>
<thead>
<tr>
<th>Table 3.3. Experiment parameters [9]</th>
</tr>
</thead>
<tbody>
<tr>
<td># of measurement points</td>
</tr>
<tr>
<td>TSMW/UE measurements per point</td>
</tr>
<tr>
<td>Operating frequency</td>
</tr>
<tr>
<td>Bandwidth</td>
</tr>
<tr>
<td>Noise figure (TX/RX)</td>
</tr>
<tr>
<td>TX power</td>
</tr>
<tr>
<td>Receiver antenna type</td>
</tr>
<tr>
<td>Receiver antenna position</td>
</tr>
<tr>
<td>TX/RX antenna gain</td>
</tr>
</tbody>
</table>
Figure 3.14. Layout of DTU Lyngby Campus tunnel system [9]. eNB is represented by orange symbol. RSRP heatmap shows along which corridors the samples were taken.

Figure 3.15. RSRP in dBm versus 3D distance, compared with 3GPP prediction [9].
relationship between the signal power and the TX/RX distance is complex and far from linear, as visible in Figure 3.15. Undoubtedly, the properties of the media the radio wave penetrates have significant influence on the signal behaviour. However, in real life it can be impossible to know such parameters, especially those related to the construction details of underground structures. Instead, different parameters can be engineered for path-loss modelling, but this requires at least accurate sample positioning. In this work, feature engineering was possible thanks to the access to the LIDAR dataset of the considered underground tunnel system.

3.3.3.1 Distance-related features

The exact GPS position of the base station was known, therefore, 3D distance between the eNB and the localised measurement points \((d_{3D})\) could be calculated\(^3\). With the aid of the tunnel dataset and 3D trigonometry azimuth angle \(\theta\) and elevation angle \(\phi\) could be found, which enabled the derivation of indoor distance \(d_{in,2D}\) (the distance to the outer-most tunnel wall towards the transmitter) and penetration distance \(d_{pen,3D}\), being the distance between the tunnel corner and the ground surface. The aforementioned metrics are illustrated in Figure 3.16.

**Figure 3.16.** Overview of distance-related features [8]. The parameters lay along the same, straight line between the transmitter and the receiver.

3.3.3.2 Tunnel-related features

Apart from the features related to the straight-line distance between the UE and the eNB, a number of tunnel-related features was formulated. The name of the second category means that the parameters refer to the tunnel characteristics only, and no above-ground data were involved in their definition. Thanks to the knowledge of the corridors’ shape, distances to the tunnel walls \(d_{wall,x}, d_{wall,y}\) and tunnel ceiling \(d_{wall,z}\) could be derived. The coordinates of the farthest tunnel corner in a given corridor were identified; based on those, the distance and angle to such a corner were obtained

\(^3\)For this purpose, a translation from the co-ordinate system of the LIDAR data to the GPS co-ordinate system was performed.
On the extreme LP-WAN deep-indoor coverage (denoted $d_{\text{corner}}$ and $a_{\text{corner}}$, respectively). Considering the tunnel along which the measurement points were located as the main tunnel, the positions of the entrances to other tunnels, crossing the main tunnel, were identified. Therefore, it became possible to derive the metric of average distance to the nearest corridor $d_{\text{cor,avg}}$, formulated as follows:

$$d_{\text{cor,avg}} = \frac{d_{\text{cor,closest}} + d_{\text{cor,farthest}}}{2}$$ (3.8)

where $d_{\text{cor,closest}}$ and $d_{\text{cor,farthest}}$ are the distances between the measurement point and the closest/farthest corridor entrance in the given main tunnel, respectively. Another formulated feature, $n_c$, represents the number of close corridors, i.e. the amount of corridor entrances crossing the main tunnel that were within the distance of $d_{\text{thr}}$ from the measurement point. The value assigned to $d_{\text{thr}}$ for $n_c$ engineering was 20 metres.

The meaning of the tunnel-related parameters can be better understood by observing Figure 3.17.

### 3.3.4 Feature analysis

In this section, it is shown how the environmental features described in Section 3.3.3 were evaluated in terms of relevance for deep-indoor path-loss prediction. The statistical methods introduced in Section 3.3.1.3 were applied.

#### 3.3.4.1 Linear analysis

Table 3.4 presents the outcome of automatic feature selection with the aid of: Feature filtering based on Pearson correlation, RFE and LASSO regularisation. The differences in both the amount of selected parameters and which ones were chosen clearly reflect the character of each method. In the case of simple filtering, where only the features correlated with the output (RSRP) by more than 0.3 were left, only $d_{3D}$ and $d_{\text{cor,avg}}$ are significant. As a result of the RFE technique, where only the least important metrics are removed at each steps of the complex selection process, and $R^2$ coefficient

![Figure 3.17. Tunnel features in an example 3D tunnel fragment [9].](image) In this particular example, $n_c$ is equal to 1, as is the number of corridor entrances within the distance of $d_{\text{thr}}$ from the measurement point.
is considered, all parameters, except $n_c$, remain relevant. Finally, the LASSO approach, being somewhat in between the filtering and the RFE, selects iteratively 5 best features: $d_{pen,3D}$, $d_{cor,avg}$, $d_{corner}$, $d_{wall,x}$ and $d_{3D}$. It is important to note that irrespective of the approach, the 3D distance and the average distance to the corridor are always relevant.

In addition to automatic methods, a manual **backward feature selection** process was performed. The performance of each intermediate model, represented by its RMSE and $R^2$ metrics was compared. The procedure continued as long as the new models containing less input features, were not experiencing a significant drop of the performance. Beginning with all the parameter candidates, it was possible to remove 9 features with only $d_{3D}$, $d_{in,2D}$, $d_{pen,3D}$, $d_{cor,avg}$, and $d_{wall,x}$ remaining at the cost of 1.3 dB increase of the RMSE. On the other hand, constructing a model encompassing only $d_{cor,avg}$, $n_c$, $d_{corner}$ and $d_{wall,y}$ yielded virtually identical performance (RMSE increased by 0.004 dB) to the model including all tunnel-related parameters.

In the next stage of the work, the considered features were subjected to **linear regression analysis**, where a selection of single- and multi-variable models were compared against one another using RMSE criterion. The outcomes of such comparison can be seen in Figure 3.18. Each single-variable model was created using one of the feature candidates\(^4\). $d_{3D}$ and $d_{cor,avg}$ models exhibit the lowest RMSE and are the most important for explaining the measured signal power, what agrees with the results from Table 3.4. Noteworthy, the azimuth and elevation angles ($\theta$ and $\phi$) explain the output RSRP significantly; this does not mean that these angles can be good candidates for $PL_{in}$ prediction, but instead these features point to geo-statistical parameters, which were not discovered and turned into feature candidates, but which do influence the underground behaviour of sub-GHz signal. The results indicate that $d_{pen,3D}$ and $d_{in,2D}$ are not explanatory towards the received signal power in the linear realm; this is interesting knowing that 3GPP 38.901 standard employs $d_{in,2D}$ as the sole parameter in its indoor path-loss component. Out of the tunnel-related features, $d_{cor,avg}$ explains the RSRP significantly better than the distances to the tunnel walls

\(^4\) $a_{cor}$ was not included in the comparison, as it was used to derive $d_{wall,x}$ and $d_{wall,y}$.

**Table 3.4. Automatic feature selection** [9]

<table>
<thead>
<tr>
<th>Method</th>
<th># of features selected</th>
<th>Features selected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Filtering</td>
<td>2</td>
<td>$d_{3D}$, $d_{cor,avg}$</td>
</tr>
<tr>
<td>RFE</td>
<td>13</td>
<td>$d_{3D}$, $d_{in,2D}$, $d_{pen,3D}$, $d_{cor,avg}$, $d_{corner}$, $a_{corner}$, $\cos(\phi)$, $\sin(\phi)$, $\cos(\theta)$, $\sin(\theta)$, $d_{wall,x}$, $d_{wall,y}$, $d_{wall,z}$</td>
</tr>
<tr>
<td>Lasso</td>
<td>5</td>
<td>$d_{pen,3D}$, $d_{cor,avg}$, $d_{corner}$, $d_{wall,x}$, $d_{3D}$</td>
</tr>
</tbody>
</table>
and edges.

Figure 3.18b compares the following multi-variable models:

- *all_feats*, where all the candidates are included,
- *dists_only*, encompassing all distance-related features: \(d_{3D}, d_{in,2D}\) and \(d_{pen,3D}\),
- *tun_angles*, using both the tunnel-related parameters and the elevation/azimuth angles (i.e. *all_feats* - *dists_only*),
- *tun_only* model containing only tunnel-related features,
- *bs_all* involving the features obtained by the manual backward feature selection, where the starting point was all features,
- *bs_tun*, implementing the result of the manual backward feature selection, starting from tunnel-related metrics.

In the Figure, it can be seen that the best model includes all the features, and its RMSE achieves the value of 15.1 dB. Thus, one can recognise all the engineered parameters as meaningful. After a more detailed comparison of feature groups it can also be observed that the distance-related category is more explanatory to the RSRP than the tunnel-related group. This is no longer true if \(\theta\) and \(\phi\) angles are considered together with the tunnel features. Comparing *tun_only* and *bs_tun* models leads to a remark that 3 of the tunnel-related parameters are marginally significant in signal power prediction from the linear perspective.

### 3.3.4.2 Non-linear analysis

The evaluation of the proposed metrics in the non-linear realm was conducted by means of Automatic Relevance Determination, being one of the properties of Gaussian Process Regression. The features included in the investigation were: \(d_{in,2D}\),

\[
\begin{align*}
\text{RMSE [dB]} & \quad \text{d}_{3D} \\
\text{cor, avg} & \quad \text{sin}(\phi) \\
\text{cos}(\theta) & \quad \text{cos}(\phi) \\
\text{d}_{wall, y} & \quad \text{d}_{corner} \\
\text{d}_{wall, x} & \quad \text{n}_{c} \\
\text{d}_{in, 2D} & \quad \text{d}_{pen, 3D} \\
\text{d}_{wall, z} & \quad \text{sin}(\theta)
\end{align*}
\]

![Image](a) Single-variable models.  ![Image](b) Multi-variable models.

**Figure 3.18.** Evaluation of the selected linear regression models [9].
3.3 Towards new models for underground IoT coverage planning

$d_{\text{pen,3D}}$, $d_{\text{cor,avg}}$, $n_c$, $d_{\text{corner}}$, $d_{\text{wall,x}}$, $d_{\text{wall,y}}$ and $d_{\text{wall,z}}$. Since it was generally intuitive, expected and shown in the linear analysis that 3D distance would have a major impact on the signal behaviour, $d_{\text{3D}}$ was not considered for ARD. The angles of elevation and azimuth were also discarded, as they were not supposed to be final feature candidates; it was enough to observe in Section 3.3.4 that they explain a noticeable portion of RSRP variance and that the phenomena behind them (consequently, new parameters) need to be addressed.

In order to be compatible by ARD, the features were first standardised and centred around zero mean. The tests were performed using squared-exponential kernel function with no assumptions regarding the prior distribution.

1. Test - investigating specific corridor effect

The measurement area, a part of the underground tunnel system, consisted of 9 distinct intervals. Each independent trial was conducted along 1 interval (corridor). Therefore, the Gaussian Process was modelled 9 times, every time excluding one of the intervals. A Cross-Validation (CV) was then applied and the amount of data splits were adjusted to each of the case, so that that a single split contained no more than 10 samples. For example, in one of the scenarios, the remaining amount of samples after discarding one corridor was 895; the dataset was then divided into 179 splits, yielding 5 samples per split. At each CV split, the hyperparameters of the model were tuned with the aid of Adam optimiser, and until the loss function (i.e. marginal likelihood) converged.

The results of the first ARD analysis is presented in Table 3.5. The values of the length-scales corresponding to the considered features are put together with the corresponding relevance rank, which is reciprocal to the hyperparameter value. The most important parameter is $d_{\text{m,2D}}$, that got a rank lower than 1 in only 1 out of 9 scenarios. On the other hand, the least relevant feature, $d_{\text{wall,z}}$ was ranked last in 8 cases. Interestingly, $d_{\text{m,2D}}$, $d_{\text{wall,x}}$, $d_{\text{pen,3D}}$ and $d_{\text{cor,avg}}$ were the most significant features chosen by ARD. If one excludes $d_{\text{3D}}$, then it can be seen that the results from the non-linear test agree with the linear backward selection model $bs\_all$, described in Section 3.3.4. This indicates that these 4 parameters are relevant to RSRP modelling, both in linear and non-linear domains. Besides, both distance- and tunnel-related features are indispensable to explain the behaviour of sub-GHz radio signal

### Table 3.5. ARD results for different corridor scenarios [9]

<table>
<thead>
<tr>
<th>Scenario</th>
<th>$d_{\text{m,2D}}$</th>
<th>$d_{\text{pen,3D}}$</th>
<th>$d_{\text{cor,avg}}$</th>
<th>$n_c$</th>
<th>$d_{\text{corner}}$</th>
<th>$d_{\text{wall,x}}$</th>
<th>$d_{\text{wall,y}}$</th>
<th>$d_{\text{wall,z}}$</th>
<th>RMSE[dB]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.233</td>
<td>1.338</td>
<td>1.092</td>
<td>4</td>
<td>3.315</td>
<td>8.038</td>
<td>0.41</td>
<td>6</td>
<td>9.572</td>
</tr>
<tr>
<td>2</td>
<td>0.263</td>
<td>1.358</td>
<td>1.132</td>
<td>4</td>
<td>3.355</td>
<td>8.412</td>
<td>0.443</td>
<td>2</td>
<td>9.448</td>
</tr>
<tr>
<td>3</td>
<td>0.274</td>
<td>1.054</td>
<td>7.159</td>
<td>5</td>
<td>7.318</td>
<td>6.540</td>
<td>0.28</td>
<td>2</td>
<td>9.967</td>
</tr>
<tr>
<td>4</td>
<td>0.255</td>
<td>1.048</td>
<td>1.773</td>
<td>4</td>
<td>4.081</td>
<td>7.263</td>
<td>0.459</td>
<td>2</td>
<td>6.241</td>
</tr>
<tr>
<td>5</td>
<td>0.262</td>
<td>1.050</td>
<td>2.269</td>
<td>4</td>
<td>4.093</td>
<td>7.509</td>
<td>0.549</td>
<td>3</td>
<td>6.913</td>
</tr>
<tr>
<td>6</td>
<td>0.262</td>
<td>1.047</td>
<td>1.825</td>
<td>4</td>
<td>4.092</td>
<td>7.744</td>
<td>0.625</td>
<td>2</td>
<td>8.066</td>
</tr>
<tr>
<td>7</td>
<td>0.222</td>
<td>2.718</td>
<td>5.037</td>
<td>3</td>
<td>5.693</td>
<td>9.485</td>
<td>0.553</td>
<td>2</td>
<td>8.602</td>
</tr>
<tr>
<td>8</td>
<td>0.62</td>
<td>1.605</td>
<td>2.482</td>
<td>4</td>
<td>3.223</td>
<td>7.357</td>
<td>0.249</td>
<td>1</td>
<td>7.244</td>
</tr>
<tr>
<td>9</td>
<td>0.202</td>
<td>1.669</td>
<td>2.555</td>
<td>4</td>
<td>2.725</td>
<td>9.018</td>
<td>0.42</td>
<td>2</td>
<td>4.872</td>
</tr>
<tr>
<td>AVG RANK</td>
<td>1.222</td>
<td>3.556</td>
<td>3.888</td>
<td>5.111</td>
<td>6.777</td>
<td>2</td>
<td>7.655</td>
<td>7.888</td>
<td></td>
</tr>
</tbody>
</table>

In order to be compatible by ARD, the features were first standardised and centred around zero mean. The tests were performed using squared-exponential kernel function with no assumptions regarding the prior distribution.

1. Test - investigating specific corridor effect

The measurement area, a part of the underground tunnel system, consisted of 9 distinct intervals. Each independent trial was conducted along 1 interval (corridor). Therefore, the Gaussian Process was modelled 9 times, every time excluding one of the intervals. A Cross-Validation (CV) was then applied and the amount of data splits were adjusted to each of the case, so that that a single split contained no more than 10 samples. For example, in one of the scenarios, the remaining amount of samples after discarding one corridor was 895; the dataset was then divided into 179 splits, yielding 5 samples per split. At each CV split, the hyperparameters of the model were tuned with the aid of Adam optimiser, and until the loss function (i.e. marginal likelihood) converged.

The results of the first ARD analysis is presented in Table 3.5. The values of the length-scales corresponding to the considered features are put together with the corresponding relevance rank, which is reciprocal to the hyperparameter value. The most important parameter is $d_{\text{m,2D}}$, that got a rank lower than 1 in only 1 out of 9 scenarios. On the other hand, the least relevant feature, $d_{\text{wall,z}}$ was ranked last in 8 cases. Interestingly, $d_{\text{m,2D}}$, $d_{\text{wall,x}}$, $d_{\text{pen,3D}}$ and $d_{\text{cor,avg}}$ were the most significant features chosen by ARD. If one excludes $d_{\text{3D}}$, then it can be seen that the results from the non-linear test agree with the linear backward selection model $bs\_all$, described in Section 3.3.4. This indicates that these 4 parameters are relevant to RSRP modelling, both in linear and non-linear domains. Besides, both distance- and tunnel-related features are indispensable to explain the behaviour of sub-GHz radio signal

### Table 3.5. ARD results for different corridor scenarios [9]

<table>
<thead>
<tr>
<th>Scenario</th>
<th>$d_{\text{m,2D}}$</th>
<th>$d_{\text{pen,3D}}$</th>
<th>$d_{\text{cor,avg}}$</th>
<th>$n_c$</th>
<th>$d_{\text{corner}}$</th>
<th>$d_{\text{wall,x}}$</th>
<th>$d_{\text{wall,y}}$</th>
<th>$d_{\text{wall,z}}$</th>
<th>RMSE[dB]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.233</td>
<td>1.338</td>
<td>1.092</td>
<td>4</td>
<td>3.315</td>
<td>8.038</td>
<td>0.41</td>
<td>6</td>
<td>9.572</td>
</tr>
<tr>
<td>2</td>
<td>0.263</td>
<td>1.358</td>
<td>1.132</td>
<td>4</td>
<td>3.355</td>
<td>8.412</td>
<td>0.443</td>
<td>2</td>
<td>9.448</td>
</tr>
<tr>
<td>3</td>
<td>0.274</td>
<td>1.054</td>
<td>7.159</td>
<td>5</td>
<td>7.318</td>
<td>6.540</td>
<td>0.28</td>
<td>2</td>
<td>9.967</td>
</tr>
<tr>
<td>4</td>
<td>0.255</td>
<td>1.048</td>
<td>1.773</td>
<td>4</td>
<td>4.081</td>
<td>7.263</td>
<td>0.459</td>
<td>2</td>
<td>6.241</td>
</tr>
<tr>
<td>5</td>
<td>0.262</td>
<td>1.050</td>
<td>2.269</td>
<td>4</td>
<td>4.093</td>
<td>7.509</td>
<td>0.549</td>
<td>3</td>
<td>6.913</td>
</tr>
<tr>
<td>6</td>
<td>0.262</td>
<td>1.047</td>
<td>1.825</td>
<td>4</td>
<td>4.092</td>
<td>7.744</td>
<td>0.625</td>
<td>2</td>
<td>8.066</td>
</tr>
<tr>
<td>7</td>
<td>0.222</td>
<td>2.718</td>
<td>5.037</td>
<td>3</td>
<td>5.693</td>
<td>9.485</td>
<td>0.553</td>
<td>2</td>
<td>8.602</td>
</tr>
<tr>
<td>8</td>
<td>0.62</td>
<td>1.605</td>
<td>2.482</td>
<td>4</td>
<td>3.223</td>
<td>7.357</td>
<td>0.249</td>
<td>1</td>
<td>7.244</td>
</tr>
<tr>
<td>9</td>
<td>0.202</td>
<td>1.669</td>
<td>2.555</td>
<td>4</td>
<td>2.725</td>
<td>9.018</td>
<td>0.42</td>
<td>2</td>
<td>4.872</td>
</tr>
<tr>
<td>AVG RANK</td>
<td>1.222</td>
<td>3.556</td>
<td>3.888</td>
<td>5.111</td>
<td>6.777</td>
<td>2</td>
<td>7.655</td>
<td>7.888</td>
<td></td>
</tr>
</tbody>
</table>
in underground environments.

2. Test - training on shuffled data

In the second experiment, the whole measurement dataset was split into a 75% training set and a 25% test set. The GPR model was trained on the 75% part and the corresponding ARD results are shown in Table 3.6. It is visible that the relevance ranking obtained in this experiment differs from any of the scenarios from Table 3.5. For example, $d_{\text{wall},y}$ has been promoted to the first rank, while $d_{\text{pen},3D}$ appears much less relevant than in the previous analysis.

3.3.5 New path-loss models

Since the starting point of this work was to evaluate the suitability of 3GPP 38.901 standard for O2DI scenarios, then the natural first step on new model proposal was to investigate the indoor prediction component (Equation 3.2) with alternative parameters in place of indoor distance. The distance-related features: $d_{\text{in},2D}$, $d_{\text{in},3D}$, $d_{\text{pen},2D}$, $d_{\text{pen},3D}$ and the combination of $d_{\text{pen},3D}$ and $d_{\text{in},3D}$ were applied to the indoor 3GPP model and the prediction comparison of thus created models is presented in Figure 3.19. The boxplot additionally includes the O2I model without the $PL_{\text{in}}$ component at all, denoted as none. Any of the proposed variables increases the Mean Absolute Error (MAE) of the prediction by $\approx 2$ dB to $\approx 12$ dB.

The aforementioned findings motivated the decision of formulating new $PL_{\text{in}}$ models, by means of linear regression and Gaussian Processes. First, the measured RSRP values were transformed into $PL_{\text{in}}$ values, according to the procedure from Figure 3.12. As explained in Section 3.3.1.2, the assumptions followed in this work considered antenna gains, noise figures, Outdoor-to-Indoor path-loss and miscellaneous losses constant (the value of $L_{\text{misc}}$ was found to be 5.6 dB, thanks to the link-budget calibration). Furthermore, the invariability of the above-ground conditions throughout the measurement campaign was also assumed. As a consequence, only $PL_{\text{in}}$ term remained variable and became the output of the new models. Having the same output (prediction quantity) enabled the proposed predictors to be compared with the 3GPP model directly.

Figure 3.20 illustrates the MAE of the new models:

- The linear regression model, which can be described by the following equation:

$$PL_{\text{in}}[dB] = 11.0773 - 0.1362 \times d_{\text{in},2D} - 6.9658 \times d_{\text{wall},x} + 4.055 \times d_{\text{wall},y} \quad (3.9)$$

<table>
<thead>
<tr>
<th>$d_{\text{wall},x}$</th>
<th>$d_{\text{wall},y}$</th>
<th>$d_{\text{wall},z}$</th>
<th>$d_{\text{cor},\text{avg}}$</th>
<th>$d_{\text{corner}}$</th>
<th>$d_{\text{wall},x}$</th>
<th>$d_{\text{wall},y}$</th>
<th>$d_{\text{wall},z}$</th>
<th>$RMSE[\text{dB}]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.347</td>
<td>12.468</td>
<td>2.168</td>
<td>4.900</td>
<td>8.012</td>
<td>0.369</td>
<td>0.356</td>
<td>25.852</td>
<td>7.877</td>
</tr>
</tbody>
</table>
3.3 Towards new models for underground IoT coverage planning

Figure 3.19. Mean Absolute Error comparison of predictors utilising different distance features [8].

Figure 3.20. Mean Absolute Error comparison of 3GPP 38.901 model and the new proposals: linear regression and GPR [9]. Median values inside the boxes.
• The GPR model, utilising the same features as the linear model.

The linear regression proposal provides the MAE improvement of 1.8 dB with respect to 3GPP 38.901 and error of the non-linear estimator is lower by 2.3 dB compared to the linear model. Thus, the GPR model corresponds to 4.1 dB MAE improvement with respect to the 3GPP baseline. Taking into account solely the data from the measurement campaign described in this chapter, it can be observed that:

• The underlying relationship between the signal power and the input parameters may be better explained by non-linear modelling;

• Even in the linear realm, features other than 3D distance, such as the distances to the tunnel walls, can be relevant to understanding the signal behaviour.

\[ P_{L_{in}} \] predictions wrt. \( d_{in,2D} \)

![Image](image.png)

**Figure 3.21.** \( P_{L_{in}} \) with respect to 2D indoor distance, predicted on the test dataset (25% samples) [9].

In order to visualise the performance of the compared models, their predictions on the test data versus \( d_{in,2D} \) are shown in Figure 3.21. Since the 3GPP 38.901 document only considers 2D indoor distance up to 25 metres, the plot only shows thus limited values. The baseline predictions of the observations are clearly poor, whereas the proposed models offer noticeably higher accuracy. Nevertheless, some
clusters of the measurements (such as those for $d_{in,2D}$ between 5 and 10 metres) could not be predicted by any model. Interestingly, the GPR model seems to fit the output better than the linear models, for $d_{in,2D}$ more than 15 metres. It has to be noticed that the spread of the observed $PL_{in}$ is large, which constitutes additional challenge for the compared models. The reason for such value spread can lay in radio propagation phenomena related to particular corridors (e.g. wave guiding), that were not captured in the presented modelling approach.

3.3.6 Discussion

Feature analysis revealed that the metrics introduced in this work were meaningful in terms of signal behaviour explanation. Some of them ($d_{in,2D}$, $d_{wall,x}$ and $d_{wall,y}$) were selected as inputs to the novel deep-indoor path-loss predictors, since their outstanding performance could be verified by means of multiple statistical methods both in linear and non-linear realms. Interestingly, not all the results of the feature analysis using different methods were in agreement with one another. Most notably, the linear analysis based on OLS regression did not render the distance-related features: $d_{in,2D}$, $d_{in,3D}$, $d_{pen,2D}$ and $d_{pen,3D}$ explanatory for deep-indoor signal prediction. Moreover, as shown in Figure 3.19, applying any of these as the parameter of the current form of indoor component of 3GPP 38.901 path-loss model resulted in higher prediction error. On the other hand, the non-linear investigation of feature relevance with the aid of ARD revealed that one of the distance-related parameters, $d_{in,2D}$, in fact is a significant factor in deep-indoor path-loss prediction and later on, this feature became an ingredient of the proposed models. This shows that:

1. Even very thorough linear feature relevance examination cannot guarantee reliable results, unless supplemented by non-linear analysis;

2. In the considered case, replacing the parameter in the 3GPP 38.901 formula was not sufficient to effectively improve the model performance. Instead, a newly formulated the deep-indoor path-loss prediction component could provide better accuracy, both being a linear and a non-linear predictor.

Some of the engineered features, such as $\theta$ and $\phi$ angles, were helpful to derive more advanced parameters (e.g. penetration distances), and at the same time to understand that the process of feature engineering needs to be continued, as significant share of the output variance remains unexplained.

3.3.6.1 The value of the collected data

It has to be underlined that regardless of the results of feature analysis and new models formulation, the collected datasets of underground radio measurements constitute a valuable contribution to the future efforts of deep-indoor channel models calibration. To the best of the author’s knowledge, at the time of performing the experiments
described in this chapter, the related work items did not present publicly available empirical data collected underground that are similar to the dataset published in [10] in terms of size (amount of samples) and where all the samples were localised with such precision. The measured area is only a part of the tunnel system, however, based on the captured signal samples, one can learn that there exists a plethora of physical phenomena and location-specific factors affecting the attenuation of the signal in the tunnels. Consequently, the way towards complete understanding of how the radio wave behaves in such scenarios is very long and difficult. Nevertheless, the examination presented in this thesis can be seen as a step forward, as the evaluation of a subset of possible environmental parameters, with respect to signal attenuation, could be achieved. Furthermore, the dataset includes several other radio statistics (such as physical layer throughput) that were not investigated in this work, but could constitute a basis for research efforts towards more optimal underground IoT deployments.

3.3.6.2 Versatility of the study

The statistical models constructed from the derived features predicted $PL_{tn}$ more accurate than the 3GPP 38.901 formula. Albeit, the validation of the findings is somewhat problematic. The biggest obstacle is the fact that the statistical inference and model training is based on an empirical dataset collected in one and the same underground environment. It has to be mentioned that the samples were taken across several corridors of diverse length, orientation and the profile of the above-ground terrain; in each of the tunnel intervals, different signal behaviour was observed. Nonetheless, certain geographical settings, such as: the distribution of above-ground objects (buildings, vegetation), transmitter deployment, tunnel structure and the placement of underground obstacles remained constant throughout the study. Undoubtedly, the validity of the feature analysis and the prediction accuracy of the proposed models would be greatly improved if more radio measurements were acquired from multiple deep-indoor scenarios (e.g. another underground tunnel or parking area). Unfortunately, enriching the empirical dataset is difficult for the following reasons:

1. **Limited access.** Deep-indoor environments often refers to places, where only qualified personnel can be allowed to enter, for example: mines, train/metro tunnels, underground storage. Conducting measurements in such areas may be totally forbidden, or restricted in terms of time and types of activities in the area (e.g. one could enter and measure the signal during metro renovation only).

2. **Additional authorisation/supervision needed.** In addition to the above, even if the access is granted under some circumstances, an authorised worker is required to assist the measurement team, or e.g. let them in. This was the case in the DTU Campus and, due to some periods of unavailability of the authorised person, the amount of data collected was limited.
3. Lack of precise indoor positioning. Even if the measurements can be freely taken, they would not be useful to the extent of this study unless accurate locations of the samples can be found. In this work, the tunnel system had been scanned by LIDAR devices and the point cloud dataset representing the area could be used. However, one cannot assume that for other scenarios, such detailed LIDAR plan would be available, and additionally no GNSS coverage in the underground area must be expected.

3.3.6.3 The meaning of feature engineering

As presented in Tables 3.5 and 3.6, the relevance of the investigated features varied when different corridors were excluded or when all data were shuffled. Such discrepancies were difficult to interpret and studying them ended with no conclusion. This leads to a remark that the engineered parameters explains only a part of the signal behaviour complexity in a deep-indoor scenario, thus the prediction accuracy of the proposed models is limited. Consequently, the approach of deriving a novel, more optimised path-loss model should be continued by engineering more environmental features explaining the remaining variance of the output. Particularly promising could be the family of parameters related to the spatial profile of the underground area, e.g.: the cross-section of the open corridor space or the deployment of objects in the tunnel. Besides, climate factors, such as temperature, are interesting features candidates, as it has been experimentally observed that they can impact the signal strength [86].

All the linear feature selection methods indicated 3D distance as the most important one. At the same time non-linear ARD technique chose 2D indoor distance and distances to the tunnel walls. This indicates that to predict the path-loss in deep-indoor environments, one should determine both the distance to the transmitter and the deployment of the receiver with respect to the tunnel walls.

3.3.6.4 The relevance of the study for particular LP-WANs

The work described in this chapter aimed at improving the accuracy of signal behaviour modelling in underground scenarios, in order to improve the effectiveness of deep-indoor coverage planning. In this section, the nuances of LP-WAN technologies are summarised to observe the effect the underground coverage situation has on the operation of particular LP-WANs.

Considering operating frequency, the signal measurements, captured at 820.5 MHz (NB-IoT Band 20 Guard-Band), can be meaningful for the low-power IoT standards considered in this thesis: NB-IoT, LTE-M (796 MHz in Band 20), LoRaWAN and Sigfox (868 MHz in Europe). However, depending on their design principles, the technologies react differently to a change of radio conditions.
NB-IoT and LTE-M

These two cellular IoT standards enable excellent link budget by: high power spectral efficiency and uplink/downlink message repetition mechanism [38]. In theory, both technologies support up to 164 dB of coupling loss, although in a simulation experiment [87] it was observed that LTE-M performs more robust connectivity and better energy-efficiency in good radio conditions and high data exchange case, while NB-IoT wins in the case of very poor signal strength.

The repetition scheme utilised by both standards enables up to 128 repetitions in Uplink (UL) and 2048 in DL. The network defines up to 3 CE levels, and depending on the perceived signal strength and quality a certain CE level is chosen and it determines the actual number of UL/DL repetitions. In underground environments, the RSRP values are typically very low, which forces the system to enable "extreme coverage" CE, which leads to the increase of repetitions. From the sensor’s perspective, sending each packet many times inevitably drains the battery and, if such situation persists, the resultant battery lifetime may be even too short to claim feasibility of an underground service.

LoRaWAN

The robust coverage in LoRaWAN is provided by means of frequency chirp modulation (LoRa). Additionally, Adaptive Data Rate mechanism adjusts the spreading factor, thus the current bit-rate according to the actual coverage situation. In the case of the best radio conditions, corresponding to SF7, the overall bit-rate reaches 27 kbps, whereas in the poorest signal conditions the SF12 only allows for 0.3 kbps transmission rate. Since the drop in the capacity corresponds to longer time to transmit the message and higher energy consumption, LoRaWAN networks not employing the Adaptive Data Rate (ADR) mechanism may benefit from precise coverage planning, which helps deploying LoRaWAN devices in the spots where the highest spreading factor can be avoided.

Apart from energy-efficiency, there are certain features of LoRaWAN that may not function properly under very bad coverage spots. Underground sensors are Class A devices, where downlink transmission can only be realised after successful arrival of the uplink packet. This means that if the application requires any data to be delivered to the end-device (e.g. a configuration packet), this may not be successful if the UL message had not been correctly demodulated at the gateway. Moreover, if the service design includes reliability based on retransmissions (e.g. sensor expecting ACK from the network), data loss might still be unavoidable, as the overall number of messages is limited by the 1% duty cycle restriction that the overall time-on-air

5A deep-indoor application may turn out to be infeasible if the device requires often battery replacement, but is deployed in a hard-to-reach location; in any case, changing the battery of thousands of sensors is a costly and time-consuming process.

6Value for Europe, 868 MHz ISM band
3.4 Summary

In this chapter, the complicated nature of sub-GHz radio signal propagation and attenuation in deep-indoor scenarios was examined. Since the results of an initial simulation trial did not correspond to the real observations, multiple NB-IoT signal power measurement studies in deep-indoor environment were conducted. The captured signal behaviour was compared to one of the publicly available O2I path-loss models. **3GPP 38.901 was evaluated in the underground scenario** and it was observed that insufficient calibration with empirical data imposes limitations to the model accuracy. Since the investigated model failed to predict the attenuation in the tunnel area, the next research efforts focused on providing significant amount of deep-indoor data, formulating new environmental parameters and evaluating their relevance for underground path-loss prediction. A large-scale measurement campaign collected new RSRP samples in the tunnel system, and a number of features, related to the tunnel geometry and the straight-line distance between the UE and eNB was formulated. Based on linear statistical methods and non-linear Gaussian Process Regression it was found that both distance- and tunnel-related metrics are explanatory to the deep-indoor path-loss and by means of 2D indoor distance and the distances to the tunnel wall one may construct linear and non-linear models outperforming 3GPP 38.901 in path-loss prediction.

Although the applicability of the proposed deep-indoor path-loss models is limited, as they have not been validated with empirical underground data from multiple environments, the DTU tunnel measurement campaign has brought valuable radio

Sigfox

Sigfox technology build up its link budget on high gateway sensitivity and robust D-BPSK modulation scheme; excellent coverage is assured, because the signal occupies extremely narrow bandwidth [49] and, by default, is repeated 3 times using different timing and frequency channel, as can be seen Figure 1.3\(^7\). No additional mechanism is defined to be activated, when the radio conditions detected are particularly poor. Therefore, unlike in the other aforementioned LP-WANs, the reliability and energy-efficiency of Sigfox system would not necessarily be improved in the case the coverage prediction in the underground scenario is more precise. However, the technology can still benefit from improved coverage planning, as avoiding coverage holes and ensuring successful Sigfox message exchange becomes possible.

\(^7\)The 3 repetitions can be optionally removed, saving power of the device, but sacrificing the coverage [16].

cannot violate [88].
statistics data that can (and should) be further studied. The dataset collected as a result of the campaign has been made available via IEEE DataPort [10]. Nevertheless, the captured signal samples revealed that sub-GHz communications, which LP-WAN standards based upon, can be successfully established in an Outdoor-to-Deep-Indoor scenario over the distance of at least 1 kilometre, with the aid of a simple constrained UE equipped with a standard IoT antenna.

Footnote: Note that only RSRP statistic was utilised in this study; other quantities include: instantaneous throughput at the MAC and RLC layers, SNR and cell information.
Part I Summary

Networked applications under the umbrella of smart communities tolerate no compromise in terms of coverage, both in outdoor, or indoor situations. That fundamental demand for ubiquitous connectivity was promised to be met by several LP-WAN technologies: LoRaWAN, NB-IoT, LTE-M and Sigfox. The work presented in this part of the dissertation aimed at verifying the connectivity performance of the aforementioned standards against real-life, challenging environments outdoor (maritime premises) and deep-indoor (underground).

Based on multiple field trials it can be concluded that LoRaWAN, NB-IoT and LTE-M can offer successful connectivity in a harbour area with minimal access network density. Repeating the experiments can help update the knowledge regarding current coverage situation and supported technologies in the considered area. Furthermore, the extensive measurement campaigns in a large underground tunnel system proved that sub-GHz signal (relevant to all the LP-WANs considered in this thesis), originated from a neighbouring above-ground base station could be successfully received and demodulated in the tunnels, where the receiver was an off-the-shelf NB-IoT sensor board. One could also learn that the existing 3GPP path-loss model, designed for ample range of frequencies, including sub-GHz ones, is not suitable for underground attenuation prediction, since its indoor path-loss component is too simple to reflect how the signal really behaves.

Taking advantage of the newly collected empirical data, several parameters related to straight-line transmitter-receiver distances and tunnel geometry were engineered and evaluated by means of multiple linear and non-linear statistical methods. This work cannot be considered exhaustive, as only a subset of possible features was derived; however, the results of the analysis presented in this thesis could already provide useful remarks regarding the relevance of the formulated parameters. It was even possible to propose a linear and a non-linear deep-indoor path-loss predictor utilising the most significant features, both of them showing higher prediction accuracy than the 3GPP model. This shows that the approach of feature engineering has a great potential in achieving more precise channel modelling, which in turn is essential for successful IoT deployment planning, that smart communities benefit from.

Hard to overestimate is the new data asset acquired throughout the study: an unique, precisely localised, large-scale set of samples including several radio statistics, out of which RSRP was utilised in the statistical analysis described in this thesis. The publicly available dataset addresses the lack of empirical deep-indoor signal mea-
measurements that can be used to calibrate statistical channel models and enables a wide range of future studies (for example, analysing the features against UE throughput). The dataset and feature engineering could be possible thanks to the access of high-resolution LIDAR plan of the underground area, which can be further explored, so that more advanced parameters can be derived.
Part II

Failure-tolerance
CHAPTER 4

When the system is put to the test

Every communications system, including an IoT network powering smart communities, will face a risk of failure throughout its entire lifetime. From the network engineering perspective, a *failure* refers to any situation in which either a piece of hardware, software or a communication link in the network ceases to operate properly [89]. The reasons for the faulty system behaviour may be: erroneous design and/or implementation, lack of testing and verification, poor components integration, malicious activities and natural factors, to name a few. The impact of a failure on the IoT network depends on the element affected:

- A fault of an *end-node* typically results in the corruption of data acquisition from that endpoint and loss of connection to the sensors or devices that the corrupted node interfaces with.

- A failure of a *base station* or a *gateway* may eliminate from the communications all the end-devices attached to it or in the range. CIoT technologies are more harmed by this kind of problem, as the communicating UEs need to be attached to an eNB.

- A failure of the *backend (core) network*. Faulty core network (backend/Network Server/Evolved Packet Core) virtually renders the whole IoT infrastructure out of order, as the data cannot be delivered to the Application Servers and the core intelligence of the deployment is unavailable.

A network may be considered *failure-tolerant*, if it is able to maintain its operation, at least to a limited extend, even upon and immediately after the event of failure. Achieving failure-tolerance is a challenging process, requiring resilient infrastructure design and implementation. Applying software and hardware redundancy is a traditional approach of computer networks; however, the low-cost, low-power and highly distributed nature of IoT limits the feasibility of such an idea [90]. Moreover, the constrained devices are battery powered, rely on wireless communication and collectively represent heterogeneous ecosystem, the state of which (e.g. currently running services, connected members) changes frequently, which further complicates IoT fault-tolerance provisioning [91].
The complexity and multi-layer structure of IoT networks necessitates that resilience solutions protects the system at all tiers and layers. The survey presented in [91] analyses 60 articles on fault-tolerant IoT, focusing on various network architecture styles, communication patterns, performance goals, etc. From the end-device perspective, an important step towards reliability is to ensure that the firmware is running on corruption-free memory. A hybrid approach, described in [92], consists of a robust software deployment mechanism, omitting faulty memory sectors, and software-defined recovery scheme, able to handle random bit flips using error-localising coding. Hu et al. [93] addresses the fault-tolerance of IoT from the application level. The proposed language-agnostic framework facilitates exception handling related to failure-recovery scenarios and can be incorporated in the application code. A lightweight and accurate failure detection method, described in [94], examines the analogue signal footprint of the IoT devices in the first moments after switching off. Since such footprints can be used to unambiguously define a sensor, the authors implemented a method of distinguishing correctly operating sensors from faulty ones, showing 99% detection accuracy in an agriculture deployment.

4.1 RAN failure scenario

This part of the thesis focuses on the problem of connectivity loss between the end-device and the base station of the Cellular Internet of Things (e.g. NB-IoT or LTE-M), i.e. within the Radio Access Network (RAN). In particular, the considered situation
includes a critical IoT application (i.e. involving sensitive environmental data, such as fault alarms, to be delivered to the monitoring centre within a specific time interval), which is realised by resource-constrained UEs. The deployment of the devices with respect to the closest base station corresponds to a "difficult radio conditions" scenario; one may identify such a case as either a distant outdoor deployment or a deep-indoor deployment. Considering the case where the UE is attached (connected) to only 1 eNB, the base station is the single point of failure, and its unavailability makes the IoT endpoints unreachable from the core network until the problem is resolved (see Figure 4.1). The time to repair the eNB may typically take 24 hours, but in the hard-to-reach deployments or in the case of massive damage, the process may last many days. Such a long outage time is unacceptable for critical monitoring applications, when not delivering alarm notifications would lead to dangerous events or failures in the monitored environment. All in all, a failure-tolerant solution is necessary for the CIoT technologies to be applicable in fairly remote and at the same time reliable and critical IoT applications.

4.2 Possible solutions

To this end, one may identify two main approaches that preserve connectivity between the harmed UEs and the network:

1. **Multi-RAT fall-back.** This idea combines multiple radio technologies on a single end-device, which on one hand improves the resilience and deployment flexibility, as well as the spectrum of potential use-cases, but on the other hand introduces higher power consumption and complexity level of the device. The authors of show on the example of NB-IoT + LoRaWAN prototype device that multi-Radio Access Technology (RAT) approach can be a promising enabler of reliable Smart City applications, however, many challenges related to technology integration, resource allocation and economical feasibility (the device becomes bigger, more expensive and requires bigger battery) have to be addressed. Li, Ota and Dong proposes a resilient architecture supporting LoRaWAN fall-back interface, when the other technologies (Bluetooth, WiFi) fail. However, the approach is only tested on relatively powerful Raspberry Pi devices and its complexity may be prohibitive for constrained LP-WAN equipment.

2. **Principal communication technology enhancement.** Contrary to supplying the end-device with an additional radio module, this approach relies on modifying the primary communication standard in such a way that the system

---

1 This situation holds either when only 1 eNB is in the range of the UE, or when there are multiple eNBs in the range of the UE, but only one of them supports the considered CIoT technology.

2 Educated guess.
is able to sustain the connectivity after the failure. For example Ref. [99] introduces a disaster-recovery D2D-based signalling behaviour for cellular networks, which provides significant energy savings and enables reachability of the UEs affected by the infrastructure break-down. In the context of CIoT, Lianghai et al. [100] propose a way to utilise sidelink communications and a novel context-aware algorithm in a D2D-enabled data exchange for massive MTC. The authors validate the enhancement via simulations, which show positive impact on service reliability and the battery lifetime of the devices. Focusing on a single radio standard helps to avoid some of the multi-RAT challenges: integration, interworking and energy-efficiency. Moreover, if D2D is used, radio redundancy can still be achieved to some extent (different frequency channels used).

4.3 Contributions

The work presented in this part of the dissertation addresses the scenario presented in Figure 4.1. Approach 2 was chosen, which means that RAT-redundancy was not considered. Instead, we proposed an enhancement to RAN of cellular IoT, which can utilise Device-to-Device communications as a fall-back uplink transmission path in the case where the only available base station experiences a failure. If the proposal is applied to the system, critical sensor data can be delivered to the network via D2D link, thanks to a relay device that forwards the data from the victim UEs towards the application server. On the other hand, if the currently available NB-IoT/LTE-M standards are applied ”as-is”, the communications between the victim deployment

![Diagram](Figure 4.2)

**Figure 4.2.** The scope of Part II contributions in terms of a generic IoT architecture. The numbers represent the corresponding sections of this document.
and the core network cannot be established, until the faulty eNB is repaired. Chapter 5 begins with a definition and description of D2D and its features, as well as implementation opportunities and challenges in the context of NB-IoT use-cases. In the remaining part of Chapter 5, the proposed reliability enhancement is described in details: the assumptions, the requirements and the behaviour of the network are explained. Practical considerations regarding the implementation of the proposed scheme in a network simulator are also included in Chapter 5, as is the discussion about the potential and applicability to NB-IoT and LTE-M standards. Finally, it is examined how the proposed concept could be beneficial for non-CIoT systems, such as Sigfox and LoRaWAN. The relevance of the Part II contributions in the context of an IoT architecture is shown in Figure 4.2.
This chapter describes the approach of improving Cellular Internet of Things resilience by applying D2D communications. First, the D2D concept is introduced in Section 5.1. The idea is presented in the context of cellular networks, and the considerations regarding possible benefits and implementation challenges are oriented around NB-IoT system and use-cases. Then, a novel NB-IoT enhancement based on D2D is described in Section 5.2. The work described in this chapter was also published in the following peer-reviewed papers:


- **Malarski, K. M., Ballal, K. D. & Ruepp, S.** (2021). D2D-enabled Failure-tolerance in Cellular IoT. Submitted to 12th International Conference on the Network of the Future (NoF2021). (Submitted, see Appendix A)

### 5.1 D2D with NB-IoT: challenges and opportunities

#### 5.1.1 Definition and overview

Device-to-Device (D2D) communications, also listed as one of the key IoT enablers, is a wireless networking paradigm in which the end-devices can connect with one-another directly, i.e. without (or with limited) assistance of the network infrastructure. A natural use-case for such an approach arises when an UE is close to another UE, so it becomes possible to establish a direct link between them. Depending on the desired performance goal, D2D communications can be applied alongside the regular link and increase the throughput, offload the access network or improve energy-efficiency [101,102], but D2D can also serve as a backup path, increasing the availability of the system [99].

As shown in Figure 5.1, D2D can be divided into 2 categories [103,104]:


- **Malarski, K. M., Ballal, K. D. & Ruepp, S.** (2021). D2D-enabled Failure-tolerance in Cellular IoT. Submitted to 12th International Conference on the Network of the Future (NoF2021). (Submitted, see Appendix A)
1. **Inband communications**, which works in the licensed spectrum and needs to share the frequency resources with non-D2D devices, providing better spectrum utilisation. Under the umbrella of inband D2D, one may further distinguish:

   a) **Underlay D2D communications**, where the resources are dynamically allocated both to the D2D and to the cellular users. The art of fair, fast and robust allocation is a challenge in this scheme.

   b) **Overlay D2D communications**, in which certain pool of resources is dedicated to D2D only. The main drawback of this approach is sub-optimal utilisation of the radio blocks, when D2D traffic volume is low.

2. **Outband communications**, operating in the unlicensed spectrum, so that the infrastructure never needs to share the resources of the cellular users with the D2D UEs. On the other hand, interference from other devices (especially those violating the duty-free cycle rule) may degrade the D2D performance.

### 5.1.2 Device discovery

During a so called, **discovery** procedure, the D2D candidates can learn about one another’s presence and communication capabilities. This knowledge allows for D2D link establishment, based on the particular method and requirements of the data exchange. At the discovery stage, the endpoints may propagate discovery messages containing the preferences of themselves (announcing devices), or monitor the environment looking for the announcements from the neighbours (monitoring devices).

If the network infrastructure assists in the discovery process, it is called **network-centric** discovery. Since the control and the supervision of the D2D service resides in the hands of the MNO, the operator can precisely monitor wireless traffic patterns and easily implement e.g. charging functionalities. On the other hand, such a mode of operation exhibits 2 main disadvantages: it necessitates that the UEs are attached to the network and is not fully scalable, as the network resources can be unavailable to massive D2D discoveries. An alternative is **device-centric** discovery, where the D2D peers are the only entities involved in the process [105]. This kind of discovery is, for natural reasons, the only possibility whenever the UEs are out of coverage.
5.1.3 Applying D2D in NB-IoT system

As noticed by Militano et al. [106], D2D communications can be a valuable enhancement to NB-IoT systems. Among the effects listed by the authors, the following are particularly relevant for this thesis:

- *Improved system availability*, since the additional D2D path enables end-to-end connectivity even when a part of the network infrastructure (e.g. a base station) has failed.

- *Optimised power usage*, which might be particularly observable in poor cellular coverage scenarios.

- *Increased scalability*, due to less cellular connections needed.

- *Multi-RAT operation*, which in turn can lead to further optimisations of energy-efficiency, coverage, reliability and radio resource utilisation.

In the design of NB-IoT, the extended coverage is achieved by repeating the messages. However, increasing the number of repetitions results in increased delay and power consumption [107]. The deployment in which the UE is forced to transmit the maximum number of uplink repetitions (128) makes it difficult to assure multiple years on battery and the low cost of operation. In this case, the possibility of relaying the UL data via another neighbouring UE and using the D2D link removes the burden of numerous repetitions and lowers the energy usage of the end-device.

5.1.3.1 Challenges

Even though the promises of D2D are appealing, the resultant enhanced system becomes more complex and must be designed and configured even more carefully. This section discusses the most urgent issues arising from NB-IoT-D2D fusion.

Network topology and D2D path

The topology of the access network, as well as the path along which the remote UEs connect to the base stations, have significant influence on system performance. A 2-hop path (remote UE-relay UE-eNB) can be considered the simplest case. Nau- man et al. [108] proposes that the selection of the relay UE is permanent and only performed once, while Li et al. [109] recommends that the remote UEs look for the relay candidates every time the channel conditions change. Ref. [110] describes a multi-hop approach based on mesh topology, where the discovery process is realised with the aid of direct mode beacon exchange. In terms of network topology, the following must be taken into account at the design process:

- **Star networks**, where the (single) relay UE aggregates the uplink packets from the remote UEs can be beneficial due to its low complexity and ease
of management. The D2D path helps achieving lower E2E delays, under the condition that the radio conditions between the end-device and the relay allow for "repetition-free" communications. The cons of this approach are: the relay UE being the bottleneck and the single point of failure of the topology, and the fact that the energy usage of the remote UEs depends heavily on the relay selection.

- **Mesh networks** with flexible, multi-hop communications can improve the availability and energy-efficiency\(^1\). The drawback is the complexity of the solution, which becomes difficult to manage and efficiently implement on heavily-constrained devices.

**D2D radio technology**

The standards of cellular IoT approves the application of foreign (i.e. non-cellular) RATs as the D2D radios in NB-IoT system [111]. As a combination of a 3GPP and a non-3GPP radio typically results in frequency diversity, higher reliability and decreased interference can be obtained. Moreover, if both the relay and the remote UEs are in cellular coverage, both the regular and the D2D paths can contribute to higher data rates (if used simultaneously) or a 1+1/1:1\(^2\) communications resiliency scheme. At the same time, a device employing multiple radio modules becomes bigger, consumes more power and inherits the security issues of all of the embraced technologies, as well as the new vulnerabilities, stemming from their combination.

**Power consumption**

As mentioned before, the main reason behind the energy-saving effect of D2D communications in a NB-IoT system is the opportunity of uplink transmissions with less or even zero repetitions. However, from the perspective of other nodes in the network (e.g. the relay UE), lowering the power consumption can be troublesome. The end-devices adopt power saving schemes (PSM and eDRX), which turn the node into a deep-sleep mode for longest possible time. During the PSM state, the radio module and most of the circuitry remain off, enabling low current consumption, but prohibiting the device from receiving and transmitting data. In the situations where relay UEs implement the PSM and eDRX, the D2D communications suffers huge delays when the relay is out of reach due to staying in a deep-sleep. Therefore, minimising the on-time of the relay devices (especially the constrained ones) is one of the main optimisation objectives towards energy-efficient D2D [104].

As far as device discovery is concerned, it constitutes a noticeable energy overhead to NB-IoT devices and additionally requires that the UEs remain awaken for

---

\(^1\)From the remote UEs perspective.

\(^2\)1+1 scenario - the data are duplicated over the regular and the backup paths. 1:1 scenario - the data traverse the ordinary path, but can be redirected to the backup path, when necessary.
the duration of the discovery. Otherwise, trying to preserve the deep-sleep opportunities as much as possible may substantially prolong the exchange of discovery messages. Therefore, the energy-latency trade-off in the context of D2D discovery needs to be addressed [112]. Besides, in massive D2D deployments the additional device discovery signalling may result in signalling storms and contribute to the problem of interference, collisions and increased energy consumption. To this end, quick discovery schemes should be developed to be able to arrange D2D communications more efficiently and in a scalable way [105].

Security

The considered D2D+NB-IoT system, as any wireless network, should be protected against malicious activities targeting Confidentiality, Integrity and Availability (CIA). In particular, the data might be captured or modified by means of node impersonation, but also due to service discontinuity under e.g. user mobility [113]. A problem of trust arises in the case of opportunistic methods, involving devices characterised by different type, ownership and behaviour history. According to the proposal of Militano et al. [106,114], the data can be securely uploaded in a cooperative way. Multiple end-devices can form a coalition, which is constructed on the basis of user trust; the trust metric is mutually tracked and evaluated. The scheme proves its effectiveness in eliminating malicious UEs from the legitimate communications; however, the situations in which the eNB cannot be trusted were not considered. Even though this is not expected to occur often, a successful compromise of a base station would enable the hacker to infiltrate the trust database about the whole deployment and arbitrarily assign the trust values of the UEs.

Participants of D2D

In IoT applications where no mobility is needed, stationary UEs are typically owned by a single body (e.g. a farmer, a factory owner, a municipality). Predictable positioning of the actual deployment and possibly also the neighbouring ones facilitates network configuration and, for instance, D2D relay selection. The situation becomes more challenging when the end-devices are in motion (e.g. asset tracking) and might stay in proximity of foreign nodes, which may not support, or simply lack interest in, D2D communications. This is understandable, as for the third-party UE, acting as a relay corresponds to increased resource drainage. In this light, it is essential to derive a method of encouragement for the external devices to join D2D. An incentive scheme, based on monetary per-packet rewards to the third-party relay UEs, is proposed in [115]. The system also keeps track on the relays’ activity in order to detect cheating in a form of forging the transmission receipt messages.

Apart from attracting the devices from foreign deployments to aid the D2D communications, a crucial issue is to select the proper type of equipment that would suit relaying. The trade-off to face includes a more sophisticated units, e.g. LTE smartphones [110], which are more powerful and are suitable for complicated algorithms,
but on the other hand do not support extreme coverage. On the other side, low-power and long-range UEs, such as NB-IoT boards, have only little computational resources to spare and need to adhere to strict power-saving cycles in order to ensure long battery lifetime.

**Radio resource allocation**

In the case of in-band and guard-band deployments, NB-IoT terminals share the resource blocks with all other subscribers of the LTE network. The optimal resource allocation approach is conditioned on which mode of communication is selected for D2D: underlay or overlay. In the latter case, a certain pool of resources is statically assigned to the proximity services. Noteworthy, if no D2D transmission takes place, those radio resources are effectively wasted, harming the efficiency and scalability of the infrastructure. This may be avoided in the underlay D2D scenario, where the cellular and D2D users are sharing the available PRBs dynamically. As far as the implementation matters are concerned, one should expect that static allocation is chosen in real-life deployments, as the complexity of the optimisation algorithms for underlay resource allocation is high. For instance, in a multi-hop D2D scheme with pre-configured paths optimising the problem of efficient resource allocation towards maximising the E2E throughput is already a NP-hard problem [115].

**Other issues**

Every promising D2D scheme must be scalable, which means that the network performs on similar level even when either the overall number of UEs, the amount of the remote UEs, the number of relay UEs, or all of them, increase. Chen et

---

*Figure 5.2. Overview of D2D NB-IoT system aspects [7].*
al. showed that the rise of the number of relays can annihilate the benefit which D2D provides [116]. Moreover, if device discovery is based on beacon signalling, the amount of wireless traffic required by the discovery process may lead to increased rate of interference. In this context, an interesting idea of group-based device discovery is presented in [117]. Each device in a cluster can advertise the presence of all the group members, which enables more optimised deep-sleep cycles and lower interference.

5.1.4 3GPP efforts

D2D is not coupled to any particular technology, however, its application in a particular system must be defined, as it may introduce significant changes to the network. Already in Release 12, 3GPP has been focusing on direct (i.e. Evolved Packet Core (EPC)-free) UE interactions in the emergency cases, when the eNBs were not available due to, for example, a natural disaster. A scheme for device discovery and D2D communications (network-assisted and fully direct) were defined for public safety, as well as commercial, scenarios. All in all, the original feature, called Proximity Services (ProSe), enables 3 kinds of deployment, depicted in Figure 5.3.

Figure 5.4 presents a simplified LTE network architecture with D2D functionalities included. The new network element, ProSe Function is included, together with ProSe App and ProSe App Server. ProSe Function contains the core logic behind network-assisted device discovery and D2D communications. Both the end-devices and the core network implement special software, utilising the proximity service (ProSe Application and ProSe Application Server). Not surprisingly, the new features require new interfaces to connect the new processes and network entities: PC1, PC2, PC3, PC4 and PC5. The latter one is responsible for peer-to-peer data exchange between UEs.

![Figure 5.3. 3GPP ProSe deployment scenarios](image-url)

Figure 5.3. 3GPP ProSe deployment scenarios [20]. Note that in "Out of coverage" and "Partial coverage" scenarios, direct D2D communications and direct discovery may only occur for public-safety use-cases.
via Sidelink, which, taking radio resources from the UL, also follows the configuration of the uplink as much as possible (e.g. using Frequency Division Duplex (FDD) or Time Division Duplex (TDD) and SC-FDMA technique).

The introduction of Sidelink (SL) came with new physical channels for data traffic, control signalling and synchronisation [118]. Regarding the radio resource allocation, 2 modes are defined:

- **Mode 1**, where the eNB dynamically assigns which PRBs can be used for Sidelink (SL) communications,
- **Mode 2**, in which the resources are pre-assigned, so that the UEs need not to rely on the eNB control. This mode is the only possible way of operation in the "Out of coverage" case [118].

### 3GPP standardisation priorities

The first 3GPP release introducing ProSe (Release 12) considered public-safety D2D communications in 3 scenarios, depicted in Figure 5.3; discovery was only allowed for in-coverage case [20]. Release 13 added the possibility of device discovery in the remaining scenarios, and introduced Layer-3 relaying. The IP-layer relay UEs in Release 13 do not distinguish the traffic from the remote UEs from own packets, which are forwarded to the EPC on the same basis. This leads to security problems, as the relay can access the remote UEs’ payload, and to QoS provisioning issues. Those were addressed in Release 14, where Layer-2 relaying was introduced [119]. In the newer scheme, the remote device is recognisable by the core network, which allows for QoS mechanisms, and the relay UE cannot access the remote packets, protected by a

![Figure 5.4. 3GPP D2D (ProSe) Architecture for no mobility scenario [20].](image-url)
5.2 A D2D-based failure-tolerant scheme for cellular IoT

This section describes a solution applicable to cellular IoT, which improves the reliability of the Radio Access Network, and is tailored to the scenario illustrated in Figure 4.1. As written in Section 4.3, the proposed UE enhancement utilises a single RAT (i.e. NB-IoT/LTE-M) that benefits from D2D communications when the connectivity between the end-devices and the eNB is corrupted due to the base station failure or compromise. During the emergency situation, the affected (out of RAN coverage) devices send the application packets to the relay UEs via a D2D link. The relays, being attached to the network, forward the received data to the core network, and the desired destinations. Thus, the service can remain uninterrupted even if the access network fails.

5.2.1 Related work

Based on the analysis of the relevant 3GPP documentation, it was concluded that applying D2D (or, in the 3GPP terminology, ProSe) for the aforementioned use-case has never been standardised. The standards describing ProSe define the system for direct discovery and communications for LTE-grade terminals, originally for Public-Safety cases only. The complexity of ProSe can be overwhelming for LP-WAN devices, such as NB-IoT sensors, especially in those applications where the only situation, in which the D2D service is required, is the eNB failure scenario. Unfortunately, the enhancements of ProSe, tailored for resource-constrained devices within CIoT

---

**Footnote:** Even the TR 36.746 document has not been updated in Release 16; the newest version dates back to April 2018.
category have only been described at the general (recommendation) level, and, to the best of the author’s knowledge, no follow-up standardisation work has been conducted to shape the recommendations into a de-facto standard.

Hunukumbure et al. [99] propose D2D service applied to a cellular infrastructure as a means to increase the level of system reliability, improve energy-efficiency and diminish network congestion. Instead of a traditional, contention-based signalling via Random Access Channel (RACH), needed for the UEs to obtain a grant to transmit UL data, the base station broadcasts the messages announcing D2D mode of transmission between the UEs, realised by means of a Carrier Sense Multiple Access (CSMA) with collision avoidance. However, the study focuses on public-safety cellular network with no regard to IoT.

Lianghai et al. introduce a D2D scheme for CIoT that partly corresponds to the ProSe architecture [100]. In their idea, the direct communications are network-controlled, and the role of the UE (remote UE, relay UE) can be assigned based on the received context-information (battery-level, location, etc.). Once the devices are clustered and assigned roles, the UEs can attach to the network, (re)configure the transmission mode and send uplink messages, utilising the new signalling behaviour, defined by the authors. The results of a system-level simulation imply that the proposed system may enhance the availability and improve the battery life of the sensors. Unfortunately, the simulation study was performed on pure LTE radio, thus without taking into account the nature of NB-IoT and LTE-M, such as PSM/eDRX modes, message repetitions, etc. Moreover, the work described in [100] strives to apply D2D in cellular IoT for regular communications, while the goal in this dissertation is to use the proximity service for failure-recovery only.

To the best of the knowledge of this thesis’ author, no D2D-based solution for CIoT, tailored to aid RAN failure-recovery and active only at the time of emergency was proposed and made available to the public.

5.2.2 Design description

The scenario of operation is shown in Figure 5.5. A deployment of end-devices experiencing connectivity failure with the only neighbouring eNB (also called Victim UEs (VUEs)) takes advantage of the fact that a UE attached to another base station and not belonging to the VUE deployment can be reachable by the VUEs by means of a direct communication link. In this way, the foreign device acts as a relay (Relay UE (RUE)) and forwards the uplink data from the VUEs toward the network. Such a configuration, in which the affected devices recognise and utilise a relay in their neighbourhood, has been set prior to the emergency situation. As soon as the broken or compromised eNB is back in operation, the D2D communications with the RUE is terminated, and the VUEs resume sending their data to the recovered base station. Additionally, in order to avoid message flooding and relay overloading in large-scale deployments, only the selected VUEs, called Group Leaders (GLs), com-
5.2 A D2D-based failure-tolerant scheme for cellular IoT

![Diagram of RAN failure-recovery situation](image)

**Figure 5.5.** Considered RAN failure-recovery situation. The recovery D2D link is established between the VUEs and the RUE, which may belong to another IoT deployment.

Communicate with the relay, sending aggregated uplink data from all the VUEs belonging to a given group.

### 5.2.2.1 Assumptions and requirements

Contrary to D2D-aided IoT communications solutions, such as e.g. in [100], the scheme presented in this section is use-case specific and only refers to emergency D2D link being enabled in order to sustain the service under the failure of the eNB, and disabled as soon as the faulty or compromised base station is again in operation. Therefore, the following assumptions and requirements have been formulated:

- The considered network deployment serves a **critical application**, i.e. the reliable operation of the service must be ensured to avoid infrastructure damage, health risks and other dangers. This is important in the view of 3GPP, originally allowing out-of-coverage D2D communications only in public-safety scenario. It is namely assumed that the importance of the IoT application can justify the possibility of out-of-coverage D2D, i.e. without supervision of the core network.

- The considered IoT sensors are **stationary**, therefore mobility issues are out of scope of this work.

- Prior to the failure, it was possible to find **at least 1 relay candidate within the range of the VUEs**, that has become a RUE, able to forward the critical packets at the time of failure.
The VUEs are located in such a way that any 2 UEs within the deployment are within the good coverage (CE 0) range to each other. It means they can reach each other without the need of increased number of repetitions.

The deployment of VUEs is divided into groups, each one having 1 Group Leader (GL) (note that a GL is also a VUE) that locally aggregates the messages from all the group members and then send the data to the relay over the D2D link.

The Mean Time To Repair (MTTR) for the eNB is 24 hours, and the failure is very unlikely (Mean Time To Failure (MTTF) = 1 year).

The application does not require lots of data and frequent transmissions (not more than 1 message per 10 minutes = 28.8kB per day per device)

The requirement for this failure-recovery solution is to sustain service availability until the repair of the eNB failure. Therefore, the complexity must be lowest possible, and the resource penalty should be minimised, both for the Relay UE (I priority) and the VUEs (II priority). This goal corresponds to the assumption that all the devices, including the relay are Commercial Off-The-Shelf NB-IoT or LTE-M boards.

5.2.3 System behaviour

This section describes the action sequence of the system under the failure scenario.

1. The victim devices running critical service are D2D capable, and there exist at least 1 D2D and relay capable device from a foreign deployment that is attached to a different eNB. Thanks to the configuration signalling (see Section 5.2.4), which has occurred beforehand, the victim devices store the information about the RUE: its IP address and radio parameters for D2D. The RUE is also aware of the VUEs, by storing the Group IDs of the Group Leaders.

2. At the time of sending a message, the victim devices realise the failure. This can be experienced by e.g. lack of NRS signalling for a given CellID during the UE’s active time.

3. The victim devices send a “relay activation request” message to the relay UE. Only the GLs send the messages to the relays, while the other VUEs send the data to the GLs acting as aggregation points.

4. After a timeout, the GLs send the aggregated messages to the relay, in an unacknowledged way.

5. The RUE passes the communication towards the core network.
6. As soon as any VUE realises the eNB restoration (e.g., restored signalling), it notifies the corresponding GL, which in turns propagates the event to the remaining GLs. One of the GLs sends a “relay deactivation request” to the relay UE; all GLs forward all pending data to the eNB. Any victim device that has not transmitted its critical message, but has realised the restoration of the eNB, will not transmit it to the GL, but to the eNB.

7. The RUE releases all radio and Central Processing Unit (CPU) resources dedicated to relaying immediately after receiving relay deactivation request and sending the acknowledgement.

5.2.4 Signalling patterns

The new failure-recovery approach requires the introduction of tailored messages, mentioned in Section 5.2.3. Figure 5.6 shows how a new GL can register to the network and be assigned a relay. The EPC is responsible for keeping track of all the devices in the deployment and maintains the appropriate hierarchy between them, knowing the relations between VUEs and GLs, and between GLs and RUEs. As shown in Figure 5.6, the optimal relay can be chosen based on its remaining battery power and signal quality. The network can select the relay from the pool of those devices that have advertised the support of the functionality at their attachment time. On the other hand, the assignment of GLs roles happens when the deployment commences its operation, since such roles may necessitate e.g. installing higher-capacity battery on the selected UEs. Therefore, reassignment of GLs is not considered. A GL sends a relay attachment request message, containing the ID of the group the device represents. The result of the relay assignment is a relay attachment response downlink message,

---

**Figure 5.6.** Pre-disaster configuration signalling.
carrying the radio configuration needed for the GL to contact the RUE at the moment of failure without any random access procedures. If no configuration is included, GL interprets it as a failure due to e.g. lack of available relays in range. Besides, if the RUE associated with the GL fails or resigns from supporting the relaying functionality, the network redoes the relay selection and inform the GL about the new situation via a relay attachment notification, containing the most current relay information (or no configuration). As long as no failure occurs, the RUEs are only registered as relays, but do not reserve any resources for relaying, thus saving energy.

The signalling diagram of post-disaster message exchange between the GL and the RUE is presented in Figure 5.7. Using pre-set uplink radio resources, the victim device triggers the relaying functionality on the RUE by sending Relay Activation Message, optionally with early data piggybacked. Relay Activation Message Acknowledgement, being received back from the relay signifies the readiness for the direct communications. The user payloads are sent to the RUE in a form of unacknowledged packets, which are then sent to the EPC without any pre-processing. At the moment the VUEs restore the connectivity with the base station, Relay Deactivation Message is sent towards the relay, which responds with Relay Deactivation Message Acknowledgement and disables its relaying capabilities.

### 5.2.5 Implementation considerations

This section presents practical implementation considerations in the context of ns-3 open-source network simulator. The high-level description of the design and remarks regarding real-life deployment are included in this section. At the same time, finishing the model implementation in ns-3 and running a showcase simulation have been left for future work.

![Communication signalling at the time of a disaster](image-url)
The aim of the simulation is to showcase the proposed scheme and observe the communications between one or more Group Leader VUEs, which could not reach their eNB, and the RUE. Since the relay remains attached to its well-functioning eNB, the main part of the model design is to introduce the extended UE functionality of both the GL and the RUE. Figure 5.8 shows the RRC state machine of a GL. The white boxes and black arrows correspond to the original diagram of a LTE UE, implemented in ns-3. The proposed failure-safe scheme results in 3 new RRC states, marked in blue, together with the necessary transitions, visible in Figure 5.8:

1. `D2D_WAIT_ACK` is entered by the GL, when the failure of the eNB is detected⁴. Prior to the state transition, the device sends a RRC message `RelayActivationMessage`. If the RUE responds with a `RelayActivationMessageACK` Protocol Data Unit (PDU), which is received by the GL, then the GL transits to `D2D_CONNECTED` state. If no acknowledgement is received within a timeout, the procedure is abandoned and the GL returns to `CONNECTED_NORMALLY` state awaiting its eNB to recover.

2. `D2D_CONNECTED` can be entered from `D2D_WAIT_ACK` state if the VUE receives the relay acknowledgement. In this state, the actual D2D communications occurs for as long as necessary from the VUEs perspective.

3. `D2D_CLOSE_WAIT_ACK` state is entered, whenever the GL becomes aware of its eNB’s recovery, either directly or through another VUE in the deployment.

---

⁴The way the base station can be "failed" in ns-3 is to schedule an event setting the TX power of the eNB antenna to zero.

---

**Figure 5.8.** A modified RRC state machine of a GL VUE. The diagram is based on Figure 30 in [21].
In such a situation, the emergency direct communications is no longer necessary, thus the GL sends RelayDeactivationMessage RRC PDU and transits to \textit{D2D\_CLOSE\_WAIT\_ACK} state, awaiting RUE’s acknowledgement. Unless the RelayDeactivationMessageACK arrives, allowing the GL to come back to \textit{CONNECTED\_NORMALLY} state, the GL remains in \textit{D2D\_CLOSE\_WAIT\_ACK} state, re-sending the deactivation message. This behaviour ensures that the relay is deactivated as soon as possible, which minimises the energy consumption of the RUE device.

It is important to mention that in any of the new RRC states, the GLs are prepared for receiving uplink data from the corresponding VUEs-group members. In order to simplify the design and the implementation on the non-GL VUEs, there is no signalling procedure to establish UL data aggregation and instead, the VUEs automatically send their packets towards their GLs, using the pre-configured radio resources. The issues of synchronisation and scheduling within a VUE group are out of scope of this work.

The state machine of the RUE is only slightly modified (see Figure 5.9). A new state, \textit{CONNECTED\_NORMALLY\_D2D}, is entered by the device upon receiving the relay activation message from the GL. Once the acknowledgement is sent, the RUE becomes ready to serve the D2D relaying for the harmed IoT deployment. Note that the name of the new state includes \textit{CONNECTED\_NORMALLY}, meaning that the RRC connection conditions between the RUE and its eNB remain unchanged throughout the whole time the failure-safe scheme is triggered. Therefore, apart from forwarding the VUE uplink data, the RUE can also continue exchanging its application data. When the RUE receives a deactivation request, it sends an acknowledgement and

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{state_machine.png}
\caption{A modified RRC state machine of a RUE. The diagram is based on Figure 30 in [21].}
\end{figure}
immediately tears down the D2D state, returning to CONNECTED_NORMALLY state.

The communications between the GLs and the RUE is realised over the sidelink PC5 interface (see Figure 5.4). The IP packets from the affected devices can either be relayed "as-is" towards the EPC, or, in the case the relay has its own uplink data to send, the VUE payload can be used to create a combined packet with RUE payload included.

5.2.5.1 Implementation on real-life equipment

In addition to the design description above, this section contains remarks regarding the deployment of the failure-safe scheme on commercial CIoT hardware.

An important matter is to minimise the overhead that the proposed solution imposes on the end-devices. From the Relay UE perspective, the overall overhead to be minimised consists of 2 components:

1. the memory and CPU resources, needed to accommodate for the new behaviour, signalling and to keep track of the associated GLs,

2. radio resources to be used by the D2D communications, when the emergency situation occurs.

The footprint on the VUE devices refers to the GL logic, i.e. aggregating the VUE uplink data and communicating with the relay, whilst the non-GL features to be implemented by all the deployment include the association with the closest GL and the ability to send the pending UL packets directly to the GL at the time of the failure of the eNB.

Having that in mind, even though, as mentioned before, the direct communications between the relay and the GL occurs via PC5 interface and sidelink channels, the actual UE support of 3GPP sidelink features should be limited. For instance, the D2D data exchange at the time of the failure shall occur over pre-configured resources and is not expected to happen with any assistance from the EPC; therefore, the data structures and signalling concerning D2D capability advertisement and sidelink resource pool assignment can be excluded from the implementation.

In order to avoid privacy issues of the relay accessing the data of the VUEs, a lightweight end-to-end security solution ought to be applied on the user data (e.g. OSCORE, described in Chapter 8).

5.2.6 Discussion

It can be seen that the design of the proposed solution modifies the behaviour of the UEs to a moderate (GL) or minor extent (RUE). Nevertheless, all the UEs remain fully functional CIoT devices that can seamlessly co-exist with other UEs, which do not implement the scheme. The design is tailored to enable the communications
between the affected devices and the EPC only for the duration of the RAN failure; however, its low complexity compensates for such a specialised, limited applicability. Therefore, the proposed solution can be considered viable to implement in real-life.

An important difference between the idea presented here and the related works ([99,100]) is whether to rely on the cellular infrastructure. Both the literature examples and the proposed solution share a common goal of improving the availability of the affected sensors, but the mechanisms implemented in ([99,100]) require that the base stations and the core network are up and running, and assisting the process, so they cannot be fully effective in the scenario considered in this chapter. On the other hand, the approach presented in this chapter sustains the critical IoT application even though there is no involvement of the access and core network, due to the eNB failure. A drawback of the novel idea is its limited capability and flexibility, since direct communications cannot be used for regular operation (as in [100]), which could improve the energy-efficiency of the deployed sensors. However, it is believed that the simplicity of the solution (and at the same time, lower overhead it introduces) can make it attractive for those deployment owners that desires some extra reliability warranty in the extreme case only at minimal cost. The latter aspect becomes particularly important, when the issue of relay selection is brought about. In this light, the author’s failure-tolerant idea comes with great flexibility, because the role of the RUE may be realised by a powerful mobile terminal, but equally well by a off-the-shelf NB-IoT sensor board. Moreover, the probability of occurrence and the duration of the emergency situation are low enough for the relay candidate to only require a standard battery. All in all, the relaxed capability requirements of the relay candidate increases the likelihood of finding one, which must be considered crucial in the considered scenario.

5.2.6.1 A non-CIoT perspective

Sigfox and LoRaWAN do not employ cell attachment, i.e. the sensor does not establish an association with a particular gateway, thus any gateway that receives the uplink message automatically forwards it to the backend network. However, in the situation where the only gateway in the range of the sensors fails, the non-CIoT deployments face the same problem as the CIoT ones. As far as Sigfox is concerned, its design limitations, especially regarding the number of messages per day (up to 144 in the highest subscription option) and the payload constraint of 12 bytes effectively eliminate the possibility of using any Sigfox end-device as a relay. On the other hand, as discussed in [121], it is possible to incorporate D2D features in LoRaWAN networks in order to offload the NS, increase the data rate and optimise the energy consumption of the remote end-devices. The authors of [121] proposed a network-assisted D2D protocol, but they also argued that with the aid of the commercially available chips it is possible to implement other, reliable D2D communications flavours, provided that the aspects of synchronisation and security are properly addressed.
Every network will face a failure. Reliable IoT networks, required by smart communities, ought to deal with breakdowns and outages in such a way that critical services and applications can remain operable, at least to a limited extent.

One of the most harmful events that can occur to a LP-WAN deployment is RAN failure. This part of the dissertation studied the eNB breakdown scenario and examined the possibilities of providing failure-tolerance to the system. With a particular focus on CIoT technologies, and setting the main goal to sustaining the critical service even at the time no neighbouring base station can be reached, an idea of D2D-based enhancement was presented. The design assumes no network assistance during the failure and utilises an available relay device from a foreign IoT deployment to forward the uplink data from the victim devices towards the core network. The proposed solution is specialised for this sole purpose, therefore it introduces little overhead to the UEs taking part in the communications. Even a low-power NB-IoT sensor node is powerful enough to become a relay, which increases the chance of successful association prior to the failure event. Power consumption penalty is also minimal, as the enhancement is activated only after the eNB breakdown and disabled immediately after its recovery. In comparison with the identified related research efforts, it can be claimed that the proposed scheme is better suited to increase failure-tolerance of IoT infrastructures at minimal resource cost. An important future work step of performing system-level simulations should provide strong proofs for this statement.
Part III

Security
Contrary to coverage and reliability issues, which are often connected with deployment constraints, technical failures, environment properties or natural phenomena, the problem of security refers to well-planned illegal and/or unethical actions aimed at the victim system, which are meant to damage the network, steal the data and sabotage the services. In fact, some of the most severe threats to the networked computer systems are various types of hacker attacks, which may be targeted at the smart community itself and/or utilise its resources to perform another malicious activity elsewhere in the world (e.g. Mirai attack [122], described later in this section). Since smart communities of the future will be powered by ubiquitous wireless communication and IoT-based services, they will naturally be vulnerable to the hackers. Particularly dangerous attacks, considered in this thesis, are the following:

1. **Eavesdropping.** It is undoubtedly undesirable that a third-party captures sensitive traffic and collects data characterising the service, network devices and the people involved. In the reality of wireless communications, it is enough to apply a properly tuned radio receiver to be able to receive the packets, or, for instance, modify the firmware of the mobile phone to turn it into a remote microphone for the eavesdropper [123]. The eavesdropped information might be used for various purposes, such as: blackmailing, preparing further hacking or revealing private data. It is unfeasible to physically block radio packet capture, but one may encrypt the sensitive data in order to render the messages gibberish for anybody else than the legitimate service endpoints.

2. **Compromise of a network node.** Once the attacker gains control over a device, he/she can access any secret data stored on the hardware and utilise the captured equipment for malicious purposes, such as data leakage or internal network infiltration [124]. The actual severity of a particular attack depends on the type of the victim entity and its role in the network; compromising the sensor would allow for the stealth of the local credentials and configuration information, but possessing a base station might potentially give the possibility of controlling the whole local communications. Securing the devices physically,
employing tamper-proof mechanisms and avoiding suspicious software are examples of countermeasures for compromise attacks.

3. **Denial of Service (DoS).** This kind of attack relies on the fact that no network node is able to serve its original purpose if the capacity of the inbound communications exceeds the processing capabilities of the victim. Thus, by flooding the server with requests one may make it unable to provide the resources for the legitimate users. If the victim device is a part of the system’s security mechanism (e.g. a manager of the shared keying material), then its failure may lead to compromise of the whole network. In classical DoS, the attacker generates the messages from a single node or network interface, while the Distributed Denial of Service (DDoS) attacks refer to the case where multiple, possibly geographically distributed network devices participate in the attack. DDoS attacks have become even more powerful with the advent of IoT, due to the availability of thousands of weakly secured devices that can potentially provide enormous network capacity, e.g. 1.2 Tbit/s in the case of Mirai attack [122]. In order to prevent (D)DoS, the network needs to implement robust access control and anomaly detection to prevent unusually behaving devices from reaching out the protected servers.

4. **Man-in-the-middle (MitM) attacks.** In the case the hacker is able to control a device that acts as the intermediary node in the network, then with the aid of the compromised unit the attacker can monitor, analyse and, possibly, alter the communications by dropping and/or modifying some packets [125,126]. In IoT infrastructures, a typical MitM situation occurs when the base station or gateway is malicious.

The aforementioned attack examples has already been a problem in the era of Personal Computer (PC)s and World-Wide-Web, but nowadays, the IoT devices entering critical network infrastructure bring even more opportunities to the hackers and make some of the attacks even more powerful. First of all, the IoT devices are developed rapidly and sold in massive amounts as consumer electronics or as networking enablers for industrial and medical equipment. The market expectations and the complexity of solid security solutions often push the vendors to neglect the vulnerabilities and privacy issues. This was pointed out in [127], where the authors additionally claim, that the reason behind numerous security breaches of IoT devices is the lack of awareness and proper education about the severity and nature of security and privacy issues among the users. A prominent example of poor security culture is using weak, default or no passwords that made it possible for Mirai attack to compromise more than 500 000 devices trying only around 60 credentials.
6.1 Security goals

The aforementioned attacks and consequences thereof constitute a solid motivation for a solution that is able to protect smart communities and assure privacy of their inhabitants. Therefore, an effective network security mechanism provides countermeasures to the hacker efforts and ensures Confidentiality, Integrity and Availability of the communications (so called CIA triad, presented in Figure 6.1).

Confidentiality notion concerns preventing the data to be easily readable to third-parties by, for example, eavesdropping activity. Since in wireless communications the information exchanged over the air among the system nodes can be also captured by malicious entities, a way to maintain the data confidential is to apply encryption, so that only the sender and the receiver, sharing the secret key, can interpret the payload.

Integrity guarantees that the data are received in the same size and form (i.e., neither modified, nor deleted) as sent from the origin. Calculating and exchanging packet checksums or message authentication codes are the examples of providing integrity in a secure network and counteracting MitM attacks.

Availability ensures the service is reachable according to the specifications, but only for the legitimate users, in accordance with the system policies. Such functionality can help prevent DDoS attacks and at the same time enables user charging and subscription mechanisms. Authentication, attestation and access control lists may be used as availability enablers ([22], Chapter 7).

6.2 Contributions

In this part of the thesis, a selection of aspects under the umbrella of security in IoT communications is addressed. Chapter 7 contains an overview of LP-WAN security
mechanisms and experimental vulnerability analysis of the selected considered technologies. In Chapter 8, the security improvements of IoT systems are discussed. A new research direction of Sustainable Security for Internet of Things (SSIoT) is introduced (Section 8.1). The implementation and performance evaluation experiments of Group OSCORE protocol are described in Section 8.2. Figure 6.2 situates the aforementioned studies in a picture of an E2E IoT network architecture.
In this chapter, the aspect of security threats and standard LP-WAN solutions is addressed both from the device perspective (technology-agnostic) and from the communications perspective (each technology separately). Apart from introducing the current security mechanisms in the LP-WAN standards, several vulnerabilities and security weaknesses are identified and experimentally verified by means of PoC attacks. The work presented in Section 7.2 and the experiments from Section 7.3 have resulted in the following publication:


7.1 Related work

At the time of performing the security experiments, the studied issues in LP-WAN were mostly concerning LoRaWAN standards. Yang et al. [128] performed several attacks: replay, eavesdropping, packet modification, ACK spoofing and DoS by battery exhaustion, all on 1.0.x version of LoRaWAN. A thorough theoretical analysis of vulnerabilities in version 1.1 can be found in [129], however, lack of available hardware and network supporting the newest version did not allow for any PoC hacking¹. Unfortunately, publicly available vulnerability studies of Sigfox and/or cellular IoT (NB-IoT and LTE-M) were limited to a brief theoretical overview with no hands-on work [130,131]. All in all, even though the attacks presented in this chapter cover only a subset of security issues in LP-WAN, they enrich the collection of proven LoRaWAN, Sigfox and NB-IoT attacks, presented in the academia.

¹Even at the time of writing this thesis, LoRaWAN 1.1 devices remained unavailable in the market.
7.2 General LP-WAN security issues

From the perspective of an arbitrary IoT device (be it a sensor, or a gateway) one may identify several security threats that apply regardless of the particular LP-WAN technology used. First of all, if the hacker has physical access to the device and its interfaces, the hardware might get compromised in the case no hardware-level protection mechanisms are in place. The consequences of that may span from capturing the secret keys the compromised device uses for communications to controlling local access network, if the device is a gateway, or even the whole IoT deployment, should the victim be a crucial node in the network (e.g. an application server). As shown in Figure 7.1, the device can be potentially attacked via the exposed serial (such as Universal Serial Bus (USB)) or debugging (e.g. Universal Asynchronous Receiver/Transmitter (UART) or Joint Test Action Group (JTAG)) interfaces, which should not be present in the production grade equipment. In particular, diagnostic interfaces, such as JTAG, provide a means to find a mistake in the firmware or download the current image, but at the same time they open the doors for reverse-engineering, firmware modification, or even secret key extraction [132]. Hardware Security Module (HSM) typically stores security keys and includes a dedicated processor for cryptographic operations (hashing, encryption, signatures, random number generation). Moreover, it can provide tamper-proof functionality, so that the sensitive data (the keys) are erased upon the physical attack. Without HSM, the hacker can more easily read the secrets from the on-board memory (e.g. with the aid of malware). Nevertheless, HSM does not protect the IoT device from so called, side-channel attacks, where operation patterns, such as power consumption or wireless transmissions are analysed to extract the secret keys. In [133], it was proven that side-channel analysis could be applied to recover Universal Subscriber Identity Module (USIM) card secrets within several minutes. Ronen et al. [134] utilised a side-channel attack to obtain the global key used by Philips smart bulbs for secure firmware update, thus the authors could force

![Diagram of IoT device architecture](image-url)
an update of a malicious firmware and launch large-scale chain-reaction attack.

Even without any physical access to the victim device, the attacker can learn the behaviour of the device, the type of application and even the payload sent, despite the lack of knowledge of the encryption key. This can be possible in LP-WAN world, since the typical use-cases involve small amounts of data sent regularly, often with similar or even identical payload; moreover, the end-devices are very simple, so the number of possible message patterns is limited. All in all, by analysing the structure of the payload and the communications pattern one can deduce the application data even if the data transfer is encrypted2.

A natural vulnerability of any wireless network is signal jamming. Even though LP-WAN standards offer excellent coverage, coupling loss tolerance and spatial diversity in the uplink3, jamming can be done, as the bandwidth and transmission power of LP-WANs are small, so the equipment of the attacker can be less powerful as well. Furthermore, as downlink communication does not employ spatial diversity, the attack becomes easy as long as the hacker stays in proximity of the sensor. In the more sophisticated version of jamming, the attacker needs to jam specific fields of the packet (such as the checksum or message signature), so that the packet would be dropped by the receiver as a malformed packet. Instead, in the simplest jamming scenario it is sufficient to transmit a wideband signal of higher SNR than the victim device, turning the legitimate packet into noise.

Another security shortcoming of LP-WAN devices corresponds to their hardware and firmware limitations regarding the size of the security keys and the time to refresh them. The biggest problem appears in the situation when a sensor is supposed to reuse the same keys throughout its lifetime (possibly, 5-10 years). While the length of the key may allow for brute force protection at the time of deployment, this might not be the case after several years. If there is no mechanism to renew the keys, the device may become unusable.

The general security threats that are relevant to LP-WAN networks are illustrated in Figure 7.2.

### 7.3 Technology-specific security analysis

This section contains a description of the security functionality defined in the official documentation of LoRaWAN, Sigfox and NB-IoT/LTE-M technologies. Additionally, security vulnerabilities of these standards are identified and several experiments-attacks are presented to verify the selected security issues on the real setup. It has

2Example: the sensor sends a message when a door opens or closes. On one hand, the attacker can observe the deployment and associate the messages with the appropriate trigger. On the other hand, if the ciphertexts of "locked"/"unlocked" messages are of different sizes, the hacker can possibly discriminate the content without decryption.

3Spatial diversity does not apply to cellular IoT, as there is one-to-one association between the base station and the UE, and the eNBs do not receive the signal co-operatively.
to be mentioned that the network operators were informed about the discovered vulnerabilities, and no harm was made to the other wireless devices, as the tests were performed in a Faraday Cage and in an anechoic chamber.

7.3.1 LoRaWAN

LoRaWAN standard [24] secures data confidentiality by using encryption between sensors and Application Servers, and protects against data modification attacks by adding an integrity mechanism between sensors and the NS. As shown in Figure 7.3, both the header fields (MAC Header MHDR, device address DevAddr, flow control FCtrl, frame counter FCnt, frame options FOpts and port FPort) and the frame payload FRMPayload are integrity protected by means of Message Integrity Code (MIC), and the user data are additionally encrypted. From the security point of view, it is important to mention FCnt field, carrying 16 least significant bits of Frame Counter, which is a sequence number used in LoRaWAN to prevent replay attacks [23].

In LoRaWAN, 2 security keys are in use: Network Session Key (NwkSKey) and Application Session Key (AppSKey). The cryptographic algorithm applied is Advanced Encryption Standard (AES) with Counter (CTR) mode for encryption and Cipher-based Message Authentication Code (CMAC) for integrity protection. Since the encryption operates end-to-end, the AppSKey is shared between the sensor and the AS; the NwkSKey, used for integrity protection, is applied between the end-device and the NS.

Both the security keys NwkSKey and AppSKey, as well as the address of the device (DevAddr) must be generated for each end-device and stored in the hardware and in the AS and NS. This process may be conducted in two different ways, illustrated in Figure 7.4. One option, called Activation By Personalisation (ABP) assumes that the data are manually inserted in the device and the network (typically at the time of deployment). A more flexible approach allows for simultaneous generation of the keys on the servers and the sensor with the aid of wireless join procedure.
7.3 Technology-specific security analysis

Figure 7.3. Packet format and security mechanism in LoRaWAN (based on [23, 24]).

Figure 7.4. Key generation in LoRaWAN [23].
(Over The Air Activation (OTAA)). During the message exchange, the server and the end-device exchange random numbers used for key generation; the other necessary generation components are: a 128-bit Application Key (AppKey), a 64-bit Device Extended Unique Identifier (DevEUI) and a 64-bit Application Extended Unique Identifier (AppEUI), which are hard-coded on the sensor [24].

7.3.1.1 LoRaWAN packet forging

**Background:** this attack targets LoRaWAN integrity protection mechanism. The MIC, meant to prevent packet forging attacks, is calculated over the whole packet using NwkSKey. At the same time, the NS drops any message with invalid MIC. Thus, the packet cannot be easily forged by an attacker not knowing the key. The following pseudocode explains how the value of MIC can be derived for LoRaWAN version 1.0.x:

fields=MHDR|FHDR|FPort|FRMPayload
B0=0x49|0x00000000|Dir|DevAddr|FCntUp/FCntDown|0x00|length(fields)
MIC=AES128CMAC(NwkSKey, B0|fields) [0:3]

where | represents concatenation, 0x refers to hexadecimal numbers, Dir is a direction flag (0x00 for uplink, 0x01 for downlink), DevAddr is the 4-byte Device Address and FCntUp/FCntDown represents the Frame Counter\(^4\) for uplink or downlink transmission.

Even without the knowledge of the NwkSKey, a hacker may forge a packet by means of MIC bruteforcing. Considering 4-byte MIC field, there are \(2^{4\times8} = 4294967296\) possible values, so it would not be feasible to try all of those using a typical LoRaWAN setting - even at a rate of 1 packet per second the operation could take 136 years\(^5\). However, since the gateway is able to demodulate packets captured at different frequency channels and spreading factors at the same time, the attack can be accelerated. If multiple gateways are set to forward the data to the same NS, the bruteforcing becomes further parallellised. Assuming the rate 1 packet per second, 15 channels, 7 spreading factors and 5 gateways, exhausting all possible MIC values would take around 95 days. Considering several years of device lifetime, the attack might become reasonable in some situations.

Bypassing the access network and sending IP packets directly to the NS could make the attack orders of magnitude faster. By means of tools such as Zmap or Masscan one can scan the whole IPv4 address space of the same size as the LoRaWAN MIC in around 6 minutes (ca. 12 million packets per second), making this approach appealing for the attacker.

**Requirements:** packet forging attack can be conducted if either of the two following circumstances are met:

---

\(^4\)always 4 bytes; in the case of 2-byte Frame Counter, the remaining 2 bytes are zeros.

\(^5\)This is still an optimistic scenario, in which all the packets are correctly received and the session keys are not renewed throughout the trial
1. The link between the gateways and the NS is unauthenticated and unencrypted, which is the case for the gateways implementing the default Packet Forwarder by Semtech.

2. The hacker owns a gateway and connects it to the NS, which is facilitated by the fact that many NSes offer guidelines on how to attach a new gateway, and support roaming service.

Impact: if the attack is successful, the malicious party can forge any packet, send it and force the NS to pass it to the AS. From the application point of view, the forged messages will be decrypted by the AS most likely as rubbish, since the attacker does not know the AppSKey. Nonetheless, if replay protection is on and the sequence number of the malicious packet is high enough, the NS will accept the forged message and then block any victim’s packets until their Frame Counter reaches the value of the attacker plus one. The consequences of such DoS attack depend on the LoRaWAN version. In 1.0.x versions, the biggest acceptable difference between the sequence numbers MAX_FCNT_GAP is 16384; such limit is discarded in 1.1 release, effectively increasing the length of the DoS even up to 65535 \((2^{16} - 1)\) packets. The newest LoRaWAN standard additionally sets the Frame Counter size to 32-bits by default. As the packet field contains 16 least-significant bits of the Frame Counter, the NS "must infer the 16 most-significant bits of the frame counter from the observation of the traffic." [47] From the attacker’s viewpoint, it enables a situation where the values of the Frame Counter on the sensor and the server side do not match. Provided that the more than \(2^{16}\) LoRaWAN uplink packets are jammed and a new packet arrives to the NS, then the message will be accepted by the server assuming that the frames are in order, whereas in fact the actual value of the Frame Counter have changed!

The details of the NS implementation are out of scope of the LoRaWAN specification. Therefore, the NS may calculate the MIC of the incoming message in 2 ways:

1. Trying all permutations of the 16 most-significant bits of the Frame Counter, until the result matches the received MIC,

2. Using 16 most-significant bits of the Frame Counter of the most recently received valid packet (e.g. having received Frame Counter = 1, FCnt = 1, the packet with Frame Counter = 65537, FCnt = 2 will not be accepted).

If variant 1 is the case, the packet forging attack becomes easier, as the effective length of the MIC reduces to 16-bits. In the case of variant 2, if 65536 transmissions were jammed, the LoRaWAN device would be DoSed permanently. On the other hand, sending only 65536 LoRaWAN packets to achieve the forging can take just around 18 hours, sending 1 packet per second with no parallelisation. Such hack can be easily done at a very low cost (a single end-device).

The attack: the goal of this PoC attack was to verify if the selected NS would accept LoRaWAN packets with the same FCnt field (FCntUp) and different Frame
Counters (thus different MICs). In the case of success, it would mean that no more than 65535 tries are sufficient to forge an arbitrary packet; otherwise, a device would be DoSed after staying out of the gateways’ coverage for more than 65535 transmissions.

The infrastructure used supported LoRaWAN 1.0.x specifications, as 1.1 version was not implemented in the publicly available devices and operators. The end-device in use, Microchip RN2483 PICtail Daughter Board, operated with 32-bit Frame Counters, as well as The Things Network (TTN) server, the chosen NS. Kerlink Wirnet iFemtocell was applied as LoRaWAN gateway. Figure 7.5 presents the workflow of the attack. A LoRaWAN packet sent by the end-device is captured by the gateway and forwarded using Semtech Packet Forwarder (SPF) to the laptop (using User Datagram Protocol (UDP) on port 1700). A developed Python application extracts the payload from the SPF data (represented by base64 encoded bytes) and decodes the LoRaWAN header fields including the FCntUp. The only field the program modifies is the MIC, recalculated over the decoded values, for increasing Frame Counter. Specifically, since FCntUp represents 16 least-significant bits of the Frame Counter, the rest of the bits are selected from the range of $2^{16}$ possibilities (i.e. from 0x0000 to 0xFFFF). The application includes the newly computed MIC in the packet built using SPF protocol and sent to the TTN server. The messages are constructed using ascending values of the Frame Counter, thus it is sufficient to send several packets (and not 65536) to observe whether the NS would accept them.

**Results:** The outcome of the experiment is presented in Figure 7.6. The "victim" packet had FCnt=19 and carried letter 'C' in ASCII (43 hexadecimal value). To prevent the NS from interpreting the next packets as retransmissions (since their FCnt would still be 19), a new packet with FCnt=20 was sent. Afterwards, 10 LoRaWAN messages were sent, all with FCnt=19 and the same header fields, except the MIC, calculated for the increasing Frame Counter (i.e. 19, 19+65536=65555, 65555+65536=131091 etc.). The figure shows that only one of such packets was accepted by the NS (Frame Counter=65555). Consequently, any end-device being out of coverage for more than 65536 transmissions becomes permanently eliminated from the communications, unless manual intervention is done. Noteworthy, the decrypted payload value (0xA2) does not match the original 0x43, as the keystream used to encrypt and decrypt the user data depends on the Frame Counter value. Therefore, thus designed attack would not lead to inserting meaningful data to the application, but the attacker can still exploit the MitM situation and replay or selectively forward the packets using simple and affordable hardware.

### 7.3.2 Sigfox

The official Sigfox document describing the security approach can be found in [135]. It is important to underline that only integrity protection and authentication between the sensor and Sigfox Backend are enabled by default, while data encryption is disabled, unless explicitly requested by the subscriber, and provided that his devices
Listen for incoming UDP packets on port 1700

Kerlink iFemtocell Gateway

Personal Computer

Packet received?

no

yes

Is packet size larger than 200 bytes?

no

yes

Extract LoRaWAN frame from Semtech’s Packet Forwarder packet

Decode LoRaWAN fields from LoRaWAN frame

Calculate MICs based on the frame’s FCntUp for all possible Frame Counter values

Build new Semtech’s Packet Forwarder packets using the captured LoRaWAN frame (without the MIC) and the new MICs

Send the new SPF packets to the Network Server

Build packet using SPF’s protocol

Send packet using UDP on port 1700

LoRaWAN frame received?

no

yes

Figure 7.5. MIC bruteforcing workflow [4].
Figure 7.6. The results of the packet forging attack [4].

supports this feature. Moreover, the encryption terminates at the Sigfox Backend and not at the application server, which leaves Sigfox a possibility to access the user data in plaintext. The overview of security features in a Sigfox architecture can be seen in Figure 7.7.

Since the cryptographic algorithms used in Sigfox belong to AES128 group, each Sigfox device is assigned a unique 128-bit authentication and (possibly) encryption keys [25]. The authentication is achieved by means of AES in Cipher Block Chaining (CBC) mode, and the encryption uses AES in Counter (CTR) mode [16]. In order for a message to be authenticated a Message Authentication Code (MAC) is computed over the entire packet, except the checksum field, as presented in Figure 7.8. The length of the MAC field varies from 2 to 5 bytes, depending on the message size. Noteworthy, Sigfox uses a one-time number to register and bind a device to a specific Sigfox account. Such number, called Porting Authorisation Code (PAC) prevents the device from being accessed by any different operator or account. Protection against replay attacks is realised with the aid of 12-bit sequence numbers (sent within Frame Sync field).

As far as physical security is concerned, Sigfox considers 3 different options:

1. **Medium**, where the secrets are stored in the device memory without any protection,

2. **High**, which protects the keys inside a software-based protected area,
7.3 Technology-specific security analysis

Figure 7.7. Security in Sigfox network (inspired by [25]). Black - obligatory security features, Gray - optional features.

Figure 7.8. Sigfox packet format in uplink, based on Figure 2 from [25].
3. **Very High**, where the secrets reside in a tamper-proof Secure Element of the device.

Medium option corresponds to no physical security. In High case, stealing the keys might be more challenging, but tampering with the device or side-channel attack should accomplish the task. The only truly secure solution is the Very High option, but only if the implementation of the Secure Element is protected against side-channel analysis.

### 7.3.2.1 Sigfox replay attack + device DoS

**Background:** Sigfox replay protection relies on a 12-bit Sequence Number (SN), included in uplink messages and protected by MAC. Only a packet with a higher SN than the previously received one can be accepted by the Backend. At the time of performing the analysis, the algorithm used to calculate the MAC was not completely public; it was known that it uses AES in CMAC mode with 128-bit secret key and the SN as some of its inputs. There was no publicly available information regarding the SN in downlink scenarios.

Since the length of the SN make it overflow to 0 after 4096 messages and the secret key used to calculate the MAC remains unchanged throughout the device’s lifetime, replay attacks can be a real danger to Sigfox. In the case of Platinum subscription, allowing for 140 messages per day, it takes 30 days to reset the SN. However, one might exceed the subscription limitation, resulting from the 1% duty-cycle limit, and the extra messages would still be accepted by the Backend on a best-effort basis, thus accelerating the SN overflow.

**Impact:** after the attacker waits until the SN is set to 0, the hacker may replay an arbitrary message sent before the overflow, even multiple times, since the security key of the device cannot change and the MAC will always be valid at the Backend server. If the malicious party replays a message with larger SN, then the legitimate device will be effectively DoSed until its SN reaches the replayed one.

**Requirements:** Sigfox network drops any uplink message which SN difference towards the last received packet is higher than maximum, described as follows:

\[
\text{max\_sn\_gap} = \text{Max}(\text{daily\_transmission\_limit} \times 3, 20)
\]  \hspace{1cm} (7.1)

where \(\text{Max}\) denotes the maximum function. For Platinum subscription, the value of \(\text{max\_sn\_gap} = 420\). Therefore, in order to DoS the victim, the attacker needs to replay packets having SN increasing by no more than \(\text{max\_sn\_gap}\). In the Platinum case, 10 replays are enough to block the device throughout the whole SN range\(^6\).

**The attack:** The goal of the experiment was to observe if the replay of the message is possible in Sigfox and how it can affect the legitimate device. Thinxtra Xkit Sigfox device with a Platinum subscription was programmed to send Sigfox packets

\(^6\)In the extreme case, the device might be even DoSed without any hacker, e.g. due to lack of network coverage for a longer period of time (long enough to exceed the maximum gap).
every several seconds to force its SN reset. The board was placed in a controlled environment (Faraday Cage), so that most of the 4096 messages did not reach the Backend and the duty-cycle requirements were obeyed. Only occasionally was the device taken out of the cage to update the SN with the Backend within the allowed maximum gap. Several Sigfox frames were captured by HackRF One Software-Defined Radio (SDR) device at the time the SN value was around 200. Once the SN overflow occurred, the intercepted frame was replayed and the Sigfox Backend was observed to see the results of the attack.

**Results:** the effect of the attack on the Sigfox Backend can be observed in Figure 7.9. A packet having SN of 258 was captured. Once the Thinxtra Xkit SN overflowed to 0 and then it reached 93, the captured packet was replayed. It can be seen in the Figure that the replayed payload of 0000 reached the server at 16:30:04. The Backend warned about the discontinuity of the sequence number, as the packets of SN between 93 and 258 could have been lost. Afterwards the Thinxtra Xkit device sent new frames with SN equals 94 and 95, but the Backend did not accept them.

This attack shows that small sized Sequence Numbers (only 12 bits), together with the maximum gap rule and hardcoded security key make it possible for the attacker to replay previously captured Sigfox messages and perform a DoS on a legitimate device.

7.3.3 NB-IoT and LTE-M

Cellular IoT standards: NB-IoT and LTE-M are designed such that the compatibility with LTE is preserved as much as possible. This applies to the security architecture as well.

Figure 7.10 presents the protocol stack for Control Plane (CP) data exchange and

![Figure 7.9. The results of Sigfox replay attack [4].](image-url)
the security thereof. Both confidentiality and integrity services are provided both at
the NAS, connecting the UE with the core network (Mobility Management Entity
(MME)), and at the PDCP layer, used to communicate the UE with the eNB. Base
stations use regular IP to exchange information with the MME, thus IPSec suite is
applied to protect the integrity and encrypt the data.

As far as the User Plane (UP) security is
concerned, the eNB-MME link is protected
in the same way as for the CP, however
the Access Stratum connection at the PDCP
layer is only encrypted, but not integrity pro-
tected (see Figure 7.11).

As LTE, NB-IoT and LTE-M base their

Figure 7.10. Control Plane protocols and security (simplified) [26]. E - encryption, I - integrity
protection.

Figure 7.11. User Plane protocols and security (simplified) [26]. E - encryption, I - integrity
protection.

Figure 7.12. Control plane NAS packet for-
mat [27].
security on a 128-bit symmetric key that resides in a SIM card of the UE and is stored in the Authentication Centre (AuC) entity of the EPS. This key is also essential to derive a number of other security keying material, dedicated to protect specific aspects of the communication (e.g. encrypting the UP packets, integrity of the CP data etc.) [26]. These keys need to be recreated for each new security context. An important condition for such security mechanism to work is, that the secrets and private UE data (such as International Mobile Equipment Identifier (IMEI)) are exchanged securely between the end-device and the EPS core, for example by means of secure-enabled NAS.

The actual cryptographic algorithms for encryption and integrity protection are defined as families (EEA0,EEA1,EEA2,EEA3 and EIA0,EIA1,EIA2,EIA3, respectively) and the system may choose the desired standard, provided that the UE supports it [136]. Interestingly, one of the obligatory-to-implement schemes, EEA2 algorithm is a 128-bit AES in CTR mode, while EIA2 is a 128-bit AES in CMAC mode, which corresponds to the algorithms chosen by LoRaWAN and Sigfox. EEA0/EIA0 option corresponds to null ciphering/integrity algorithms and, if applied in the network, effectively disables the security. Every integrity protected packet includes a 4-byte MAC, computed by means of an EIA algorithm. Additionally 2 4-byte long incremental counters COUNT are defined (for AS and NAS) to serve replay protection. However, only the least Significant bits of the COUNT, called Sequence Number (SQN) are sent over the air (5, 7 or 12 bits, depending on the radio bearer), while the residual bits have to be updated locally by the network nodes. The value of COUNT is also one of the inputs for the encryption algorithms. Since the 8 Most-Significant Bits of the COUNT used in NAS are zeroes, the number of unique values drop to only $2^{24} = 16,777,216$. Albeit, the EPS core does not allow for the reuse of keying material, so when COUNT values approaches the maximum value, the security context renewal procedure is triggered [26, 27].

Figures 7.12, 7.13a and 7.13b show the format of PDCP and NAS packets, where differences in SQN length and the security features are highlighted.

![Figure 7.13. PDCP packet formats](image-url)
7.3.3.1 Attack on NB-IoT using malicious UE

**Background:** As NB-IoT UEs can send and receive data using IP, the attacks such as port scanning, ARP/DNS spoofing are a threat to NB-IoT networks. The UE needs to set up a Packet Data Protocol (PDP) context with an access point described by its Access Point Name (APN). Once established, the context allows the UE to connect to a PGW and grants it access to a private network. Typically, enterprises having large number of UEs would keep them as members of the same Local Access Network (LAN). In such a situation, the attacker could perform IP-based malicious actions on the UEs in the LAN, once he compromises or steals one of the UEs. The topology of such attack is depicted in Figure 7.14. One possibility of a harmful attack arises when a victim device opens a port to communicate with a IoT server. The hacker may then send malicious packets from the stolen UE, masqueraded as the server itself. Besides, the attacker could massively ping the other UEs, forcing them to respond draining their batteries; in a long run, this may lead to DoS by battery exhaustion, which may have severe consequences in critical applications.

**The attack:** the experiment aimed for exploring the potential of malicious UE presence in NB-IoT private network. The malicious device used was a LTE UE, so as to show whether it is possible to get access to the LAN also for a non-NB-IoT devices and to be able to launch an attack from a more powerful hardware implementing a full Transmission Control Protocol (TCP)/IP protocol stack. An Android phone was connected to a Danish MNO’s APN using NB-IoT SIMcard, that can be normally used to connect a NB-IoT device. The smartphone established a WiFi hotspot in order to allow a laptop to join the private APN network. Then, the PC tools and
processing power could be used for further steps. Since the attack aimed at a real network, the only activity performed on the accessed LAN was a ping scan. A tool called Zenmap was used to send Internet Control Message Protocol (ICMP) request to all network IP addresses to detect the connected UEs.

**Results:** as shown in Figure 7.15, the LTE UE could obtain a private IP address (10.X.X.12), becoming a member of a private network behind a router that does Network Address Translation (NAT). The whole 10.X.X.0-255 subnet space was scanned by means of the following command: `nmap -sn 10.X.X.12/24`. Two other UEs, represented by 10.X.X.5 and 10.X.X.7 addresses, were found. If this was a true hacker attack, the attacker could for example scan for open TCP/UDP ports on the two devices and send forged data on those ports.

### 7.4 Discussion

The analysis provided in this chapter reveals that although all the considered LP-WAN standards include well-thought security functionality in their system design, harmful hacks remain possible, and they do not require highly sophisticated equipment. The summary of the conducted LP-WAN attacks is presented in Table 7.1.

In the case of LoRaWAN, the integrity protection was targeted in the attack.

**Figure 7.15.** The results of the NB-IoT attack [4].
Table 7.1. Summary of LPWAN attacks [4].

<table>
<thead>
<tr>
<th>Target LPWAN</th>
<th>LoRaWAN</th>
<th>Sigfox</th>
<th>NB-IoT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Threat</td>
<td>Packet forging</td>
<td>Replay</td>
<td>Malicious UE</td>
</tr>
<tr>
<td>Impact</td>
<td>Message injection (gibberish payload), DoS (MIC bruteforce)</td>
<td>Message injection, DoS</td>
<td>Private NB-IoT network infiltrated</td>
</tr>
<tr>
<td>Cause</td>
<td>Short (4-byte) MIC</td>
<td>12-bit SN, max. SN gap</td>
<td>Poor protection from UEs private network</td>
</tr>
<tr>
<td>Exploitability</td>
<td>Easy (1.0.x) / Hard (1.1)</td>
<td>Easy</td>
<td>Medium/Hard</td>
</tr>
<tr>
<td>Prevalence</td>
<td>Common (1.0.x) / Rare (1.1)</td>
<td>Common</td>
<td>Common</td>
</tr>
</tbody>
</table>

The combination of several factors: relatively short MIC field (4 bytes), specifying maximum Frame Counter gap and letting the NS infer the actual Frame Counter value from the traffic observation allowed for a situation, in which the server accepted a forged packet and adopted its Frame Counter calculation to the hacker activity; the innocent device was blocked. To prevent such attacks, several actions may be considered. First of all, trying to brute-force the MIC requires sending large amounts of packets; thus, the servers ought to be able to detect and disallow such activity, preventing packet forging and DoS ing the end-devices and the servers themselves. Besides, future versions of the standard could revise the packet format and consider prolonging the MIC field to 5 or 6 bytes, significantly decreasing the feasibility of forging attempts. Furthermore, insecure packet forwarders, such as SPF, should be abandoned and it must always be ensured that the hacker cannot easily attach a malicious gateway to the core network.

Sigfox security weakness, exploited in the PoC attack was the length of the sequence number field (12 bits) and, consequently, relatively frequent rollovers. Similarly as in LoRaWAN, Sigfox specifies maximum SN gap above which the message cannot be accepted. Moreover, the security key used to authenticate the message is hard-coded in the device permanently. Jamming the victim sensor and replaying the message with higher SN allowed for both message injection and a temporary DoS of the device. Thus operating replay protection scheme is not suitable for critical applications, where sensitive data must arrive sharply on the expected time. On the other hand, enhancing the default security in Sigfox would either mean lowering the size of the payload to less than 12 bytes or enlarging the packets resulting in a smaller amount of daily transmissions (which results from the 1% duty cycle requirements in the Industrial, Scientific and Medical (ISM) band).

NB-IoT and LTE-M security architecture, inherited from LTE, represents the highest level of complexity among the analysed LP-WANs. Keying material can be renewed and strong encryption and integrity protection schemes are defined. However, by means of a smartphone using NB-IoT SIMcard, it was possible to enter a local

---

7 In particular, both sending too many packets and sending garbage data should be blocked.
IP network, where multiple devices (probably NB-IoT UEs) were registered. Such a scenario resulted rather from erroneous security policy implemented by a particular MNO, than from the design flaw of NB-IoT security. Therefore, best security practices must always be applied on each element of the system. Moreover transferring the data over NAS ensures Control Plane grade security, and additionally help minimise signalling overhead for battery-driven UEs [137].

7.4.1 The significance of the experiments

The contributions described in this chapter are meaningful for the research community, as they prove the security shortcomings of the considered LP-WANs, hence the results can be used to propose novel solutions. At the same time, the TTN LoRaWAN gateway and server provider and the operators of Sigfox and NB-IoT were informed about the outcome of the hacking experiments, and the vulnerabilities have been mitigated.

7.5 Summary

This chapter provides an investigation of low-power IoT security, taken from the device perspective and from LP-WAN communication perspective. An overview of LoRaWAN, Sigfox and NB-IoT/LTE-M security was complemented by 3 attacks on real networks, the results of which were discussed. It was shown that security design choices and/or lack of appropriate policies of the network administrator may facilitate malicious activities, such as enterprise network infiltration, message injection or DoS of legitimate end-devices. At the same time, the approach of supplementing the theoretical vulnerability analysis with experimental hacks can be seen as a successful means to raise the awareness about the need of continuous efforts toward security level improvement, both from the side of network operators, technology standardisation bodies and equipment vendors.
Network solutions powering smart communities must be properly protected from malicious actors. In Chapter 6, one could learn that there exist many dangerous assaults aiming at different network nodes and functionalities, which can lead to serious infrastructure damage, service disruptions and data leakage. Chapter 7 narrowed down the security considerations to the LP-WAN technologies under the interest in this work. Neither of the LP-WAN standards can be considered fully secure, as proven via the presented PoC attacks. This chapter comprises of studies that examine approaches and solutions that could improve the level of security in LP-WAN networks.

The contents of this chapter include theoretical and practical approaches that can contribute to a robust and environmentally-friendly security architecture of smart communities, also protecting user privacy to a greater extent than in the current releases of LP-WAN standards. Part of the research described here (Section 8.1 in particular) was also presented in the following conference paper:


8.1 Sustainable Security for IoT (SSIoT)

The theoretical considerations, presented in this section, address the problem that the massive IoT roll-out suffers from. On one hand, the aspects of security and sustainability are present and highly prioritised in IoT system requirements [138]; on the other hand, both the goals remain unreached as IoT devices are often vulnerable and can be easily hacked [139], and at the same time, together with the rest of Internet infrastructure, they contribute to high power consumption and environmental pollution [140]. It has to be mentioned that the level of security-awareness among IoT users is rather low, since they tend to choose the devices based on the price and with no regard to their security level [141]. Moreover, the availability of
free do-it-yourself IoT building automation platforms, e.g. OpenHAB [142] facilitates real-life deployments done by inexperienced users that do not follow good security practices, exposing their infrastructure to risks. There exists an intrinsic conflict between developing low-power, computationally-efficient IoT services, deployed on heavily-constrained hardware and including robust security mechanisms, that naturally bring additional resource usage and accelerate the drainage of the batteries. Therefore, the applications should be developed in a way that preserves a good balance between optimised power consumption and security guarantees. The necessity of combining the network security with energy-efficient communications was expressed in [143]. Addressing the limitations of IoT networks, the authors encourage energy harvesting and such mechanisms for security that do not introduce further overhead to the communications. For instance, one may use physically unclonable functions (PUF), where truly random material for key generation may be obtained through the properties of the device and its environment, or apply encryption only when it is really needed (avoiding encrypting payload-less error/alarm messages).

In response to the aforementioned issues with security and energy-efficiency in IoT, a need of a complete and comprehensive effort towards fully sustainable (more than only power-optimised) IoT security was identified.

### 8.1.1 SSIoT building blocks

This section describes the related work in the areas of Green/Sustainable Computing, Green/Sustainable IoT and Green/Sustainable Security, based on which the concept of SSIoT was derived.

#### 8.1.1.1 Green and Sustainable Computing

Saha defines Green Computing as "the practices and procedures of designing, manufacturing, using of computing resources in an environment friendly way while maintaining overall computing performance and finally disposing in a way that reduces their environmental impact" [144]. Originating from the problem of improving energy-efficiency of data centers, the direction of Green Computing has been attracting attention of the academia for more than 10 years [145,146]. At the same time, Sustainable Computing, according to the authors in [147], results from a clever combination of IoT and cloud computing; low-power devices widen the application portfolio of the system, while the cloud infrastructure powers the IoT with the necessary computational resources. For example, the power consumption and transmission delays of mobile applications can be optimised by carefully offloading the computations to the edge/cloud [148]. A security architecture detecting anomalies in smart grid deployments, which outperforms the compared proposals in terms of energy-efficiency, was presented in [109].
8.1 Sustainable Security for IoT (SSIoT)

8.1.1.2 Green and Sustainable Internet of Things

The notion of Green IoT was initially defined as:

- The application of IoT which reduce the greenhouse effect of the existing services
- The solutions tailored to diminish the environmental impact of the IoT itself [149,150].

Massive IoT use-cases require ultra-long battery-life of the devices; however, since the batteries are a source of heavily toxic substances and are hard to recycle [149], battery-less solutions are favoured. Energy harvesting is a promising technology of collecting the necessary energy from the surroundings; the main issue is to handle instability in the harvesting by advanced power management [151]. Zhang et al. [152] describe a method of efficient resource management, which optimises the data collection control and channel scheduling over short time intervals, while the power consumption is managed over longer periods of time.

The survey of green IoT, where a taxonomy of the notion is also included [153], shows that the energy-efficiency of the systems are typically improved, but often the level of security and privacy is lowered.

8.1.1.3 Green and Sustainable Security

Green security refers to "defining and investigating security solutions under an energy-aware perspective" [154]. Thus, it becomes insufficient to simply improve the protection of the system, if the side-effect is a substantial energy consumption penalty. This does not mean that optimising the power consumption and increasing the level of security are always contradictory. One can imagine a routing algorithm that turns off some of the nodes to save energy, but since the sleeping devices cannot be attacked from the network side, this behaviour has also positive implications towards the security [155]. Nonetheless, cryptographic operations, key generation and distribution require additional CPU and, possibly, networking efforts, thus increased power consumption must be expected. Although multiple research works have studied the challenge of achieving green security, focusing for example on security functions virtualisation [156,157], not many of the proposals set the same priority on security and energy efficiency [155].

8.1.2 SSIoT concept

The definition of SSIoT assumes that a sustainable security solution is both energy-efficient, environmentally-friendly (e.g. produces minimal amounts of waste, possibly avoiding dangerous, polluting waste) and economically feasible. Therefore, an idea under the umbrella of SSIoT fulfils 2 conditions:
1. Addressing IoT security issues from the broad perspective of the cost, power consumption and the environmental impact of the possible solution,

2. It effectively improves at least one of the SSIoT aspects (sustainability or security) and at the same does not negatively influence the other one.

The SSIoT research angle stems from the union of the 3 areas, discussed in Section 8.1.1: green/sustainable computing, green/sustainable security and green/sustainable IoT. As in green/sustainable IoT, SSIoT aims at improving the eco-friendliness of IoT deployments, however with the security goals taken into account with equal priority. One could find a close relation between the idea of SSIoT and green/sustainable computing, but the former one is focused entirely on low-power wireless network infrastructures. The new concept also relates to green/sustainable security, which in SSIoT becomes adapted to the specific nature of IoT and resource-constrained devices. Figure 8.1 visualise how all the aforementioned concepts contribute to SSIoT formulation.

Sustainable Security for Internet of Things (SSIoT) can be defined as "the collection practices, procedures and technologies aimed at improving the security level of IoT in an energy-efficient, affordable and eco-friendly manner" [6].

As stated before, all efforts within SSIoT research direction combine solid security with minimised environmental footprint, low cost and energy usage. The four components of SSIoT: security, energy efficiency, affordability and eco-friendliness are shown in Figure 8.2. The concept was conceived with the desire of raising the awareness about environmental and economical aspects of IoT security and to orient future investigations towards all those aspects.
8.2 Lightweight application-layer security: protecting CoAP communications

The introduction of LP-WAN security systems from Chapter 7, presenting how security is designed in LoRaWAN, Sigfox, NB-IoT and LTE-M standards revealed a common practice of allowing the network owner/operator to intercept the user data, if an optional encryption service is not active. Lack of trust towards the infrastructure and the MNO can be particularly understandable in sensitive use-cases of smart communities, where the payloads transmitted over the air (e.g. health parameters of the patient, critical system monitoring status) must not be disclosed to any unauthorised party. From the user perspective, a safe solution to protect privacy from both the malicious actors of the Internet and the network operator is to apply application-level security, established and managed beyond the control of the infrastructure owner. However, the complexity and additional overhead of the security overlay cannot hinder the economical feasibility of low-power IoT devices and applications. In this context, a novel and promising example of such robust and lightweight approach is Object Security for Constrained RESTful Environments (OSCORE), providing End-to-End confidentiality, integrity protection and non-repudiation properties to Constrained Application Protocol (CoAP), a popular application-layer protocol used in low-power IoT communications. Besides being more energy-efficient, the protocol’s advantage over the most prominent alternative, Datagram Transport Layer Security

Figure 8.2. SSIoT pillars (based on [6]).
Securing IoT communication

(DTLS) is selective encryption, leaving some of the packet fields in plaintext, so that advanced caching and proxying can be done without the need for intermediary nodes to decrypt the protected messages. While the original design of OSCORE security targets unicast client-server scenarios, this chapter focuses on a state-of-the-art extension of OSCORE, called Group OSCORE, enabling End-to-End secure group CoAP communications in one-to-many scenarios. The novel protocol was implemented in Contiki-NG operating system and performance evaluated on 2 Commercial Off-The-Shelf (COTS) hardware platforms: Zolertia Firefly Rev. A and Texas Instruments Simplelink CC1352R1 launchpad. To the best of the author’s knowledge, it is the first publicly available Group OSCORE implementation for constrained IoT devices and the first results on its performance. The preliminary results, presented in this dissertation, include memory overhead analysis and E2E communications delay measurements. The purpose of the performance evaluation study is to verify the feasibility of Group OSCORE in LP-WAN reality.

The remaining sections of this chapter firstly explain the operation of CoAP and its original security solution, Datagram Transport Layer Security (DTLS) and then introduce OSCORE and describe how Group OSCORE functionality was defined. Finally, the performance evaluation experiment and the initial results are presented and discussed.

8.2.1 CoAP overview

Many of the IoT stakeholders expect the devices to connect to the Internet and be integrated in the global network in a way that is compatible with well-established web services. These are often relying on HyperText Transfer Protocol (HTTP) protocol [158] and the REpresentational State Transfer (REST) [159], providing robust and convenient means of exchanging the information between clients and servers. With the aid of intuitive HTTP methods (e.g. GET, PUT, POST and DELETE) it has become easy to automate resource retrieval and modification via so called, REST Application Programming Interface (API). However, the HTTP architecture, developed with the thought of non-constrained devices, does not fit optimally in IoT realm, in terms of energy-efficiency and throughput [160]. Therefore, a new, more lightweight, yet still REST-complaint protocol for constrained devices was proposed under the name Constrained Application Protocol [29].

8.2.1.1 CoAP features

CoAP is a web protocol that exhibits similarity to HTTP (i.e., also uses methods such as GET, PUT, POST, DELETE), but is originally designed to operate over unreliable UDP transport (although the use of TCP is also considered [161]). Both synchronous and asynchronous data transfer is possible, and the resources are identified by Uniform Resource Identifiers (URIs), which corresponds to the web browsing use-case. CoAP also supports communications through network intermediaries: caches and proxies.
Such a feature is crucial in low-power IoT scenarios, such as LP-WAN, where the end-devices strive to achieve multiple years battery lifetime and the use of a proxy or a cache node may offload the constrained endpoints and postpone the downlink traffic, in order to adhere to the sensors’ deep-sleep cycles (e.g. PSM/eDRX periods in the case of NB-IoT/LTE-M devices).

Figure 8.3 shows the structure of CoAP messages. A minimal packet is a 4-byte header consisting of: Version field, Type, indicating a confirmable/non-confirmable request/response, a reset or an acknowledgement, Token Length, determining the presence and the size of the token (up to 8 bytes), Code (similarly as HTTP response code) and Message ID used for transaction binding and de-duplication purposes. The header may be followed by the token, one or more options organised in Type-Length-Value fashion (those further specify the communication parameters, analogically to HTTP options), a 1-byte payload delimiter and user data.

### 8.2.1.2 CoAP group communications

Although the original CoAP design focuses mostly on one-to-one client-server communications, group transmission is also supported and described in more details in [162]. The focus is put on one-to-many scenario, where a client controls and manages multiple servers, all devices belonging to one group. The client originates a CoAP request with the destination address being the IP multicast address of the group; all the group members that are servers may (but do not need to) respond by a unicast CoAP response, addressed to the client. To prevent traffic congestion and/or overwhelming the client device with the responses, the standard prohibits that the servers respond to the multicast request simultaneously. Instead, they should delay the transmission by a randomised value of leisure time, which must be no shorter than:

\[
leisure_{\text{min}}[\text{seconds}] = S \times G / R
\]  

(8.1)

where \( S \) denotes the message size, \( G \) corresponds to the group size and \( R \) is the transmission data rate. Furthermore, only non-confirmable messages may be exchanged in the group mode. Both the servers and the proxies might additionally
suppress some of the transmissions (e.g. sending error responses) to limit the interference, which could occur in large deployments. Note that the group responses can be related to their appropriate requests by the token value and not by the source/destination addresses\textsuperscript{1}.

Ishaq et al. proposed an alternative to the official IP multicast-based group communication [163,164]. In their unicast-based solution, the authors assign a role of Entity Manager (EM) to entities (groups). The EMs accept the group requests from the client and translate them into a series of unicast transactions with the servers [163] or a multicast message exchange with the server [164]. As a result, the system achieves higher level of reliability and can benefit from the security mechanisms defined for unicast communications (in the purely unicast example). However, performance issues of such a solution (e.g. delays, possible EM overload) may appear in some scenarios. On the other hand, a group communication scheme based on CoAP observe was introduced in [165] and further enhanced in [166]. In this approach, the client uses observe option to request group resources, which are sent upon value renewal, or after a timeout. The drawback of the observe-based idea is the significantly increased number of messaging (notifications) in large-scale networks.

\textbf{8.2.1.3 CoAP security with DTLS}

In the original specification of CoAP, it is recommended to use Datagram Transport Layer Security (DTLS) protocol [167] wherever security guarantees are required. DTLS is a transport security mechanism tailored for UDP [168] or other unreliable transport protocols; its goal is to provide \textit{secure channel}, i.e. secure exchange of arbitrary data, the structure of which is unknown to the protocol [169]. Applying DTLS results in data encryption, integrity protection and authentication. As visible in Figure 8.4, the design of the protocol is complex and consists of multiple building blocks, responsible for different stages of communications: The \textit{handshake} mechanism enables safe exchange of security parameters and session establishment between 2 entities, which afterwards are assigned client and server roles. The security keys in CoAP over DTLS can be pre-shared, based on raw public keys or derived from X.509 certificates [29]. By means of \textit{Change Cipher Spec} protocol, the cryptography method in use can be changed, and any errors can be sent with the aid of \textit{Alert} protocol. Finally, \textit{Application Data} scheme refers to actual secure communications based on the configuration agreed upon during the handshake. The above-mentioned protocols, together with application-layer payload, are encapsulated in the data structures of \textit{Record} layer, where basic information about the current sequence number, data type and length are kept [167].

\textsuperscript{1}The client sends the request targeting a group multicast address, but the servers responds setting the source address as their unicast ones and not the group one.
8.2 Lightweight application-layer security: protecting CoAP communications

![Diagram of DTLS Architecture](image)

Figure 8.4. DTLS Architecture (inspired by [30]).

### 8.2.1.4 Group CoAP security issues

Applying DTLS in CoAP applications can successfully secure unicast communications; however, the solution is not applicable to multicast scenarios. The main problem concerns the process of security keying distribution and management (handshake stage). A complex message exchange of the handshake protocol results in mutual agreement on the security parameters only for 1 client and 1 server. Applying that scheme to a large-scale group would result in substantial wireless communications overhead that cannot be tolerated in low-power IoT use-cases. This problem can be partially mitigated if a "proxy" device, aggregating server messages and handling client requests is used (see ref. [164]). However, the multicast scheme [162] is preferred in this thesis, due to lower number of messages sent over the network (faster performance, lower energy consumption).

Park [170] addresses the DTLS problem for multicast communication and proposes a security architecture in which the E2E protection is based on 2 keys: a group-wise key, securing the client’s requests and a pairwise key securing the responses of each of the servers. The task of key distribution and management is assigned to resource discovery/group controller node. The author introduces a security bootstraping protocol, allowing for key establishment without DTLS handshake; the DTLS record protocol is employed in the proposal in a modified form. Another DTLS-based approach, in which record sublayer is used and handshakes are suppressed is presented in [171]. All the members of the group share a security association and the group keying, whereas the individual keying is derived during the first multicast transaction (multicast request and multiple unicast responses). The scheme defined in [171] is more scalable and lightweight than the solution based on pure DTLS associations between each receiver with the sender.
8.2.2 Lightweight DTLS alternative: OSCORE

Protecting CoAP packets with DTLS may seem robust; however, one can argue that it is not fully suitable for low-power IoT use-cases. Firstly, encrypting the entirety of CoAP packet means that any proxying or caching functionality, normally enabled by special fields in the message, may not be used unless the intermediary network node decrypts the packet. This, in turn, introduces a severe security and privacy risk for the service, as the proxies and other nodes not owned by the users and/or the service owner cannot be trusted. Moreover, it becomes necessary to establish multiple DTLS channels (as many as many hops the E2E path comprises of), thus increasing the delay and power consumption of the network. Secondly, inability of using proxies and caches may lead to heavily sub-optimal energy consumption of the IoT end-devices, as the network cannot respect the sleeping patterns of the energy-efficient constrained nodes and postpone the downlink transmission until the deep-sleep timer expires. Consequently, the devices face the following alternative: to maintain the energy-saving policy and risk missing certain messages or remain ready for packet reception at the cost of highly increased power consumption. Addressing all those issues, a new security protocol for CoAP was specified [31].

Object Security for Constrained RESTful Environments (OSCORE) is an End-to-End security scheme that is proxy-aware, i.e. leaves the fields needed for the intermediaries unencrypted, thus the issues of DTLS can be avoided. Moreover, the design of OSCORE requires less resources than DTLS and does not impose a noticeable performance penalty on constrained devices [32].

Contrary to the channel security approach adopted by DTLS, OSCORE is based on object security, where the protocol is aware of how the data chunks are built, thus protecting each data object separately [169]. OSCORE requires the payload be efficiently encoded using Concise Binary Object Representation (CBOR), which allows for small message size and lightweight parser code [172]. With the aid of CBOR Object Signing and Encryption (COSE) [173], the CBOR data can be both encrypted (confidentiality) and integrity protected by a Message Authentication Code. A CoAP message protected by OSCORE is given a new option, shown in Figure 8.5 and its payload is in fact an encrypted COSE object.

Apart from supporting COSE authenticated encryption with additional data, in order to use OSCORE the endpoints must also be able to run algorithms for key derivation and authenticated encryption (HMAC-based Key Derivation Function (HKDF) and Authenticated Encryption with Associated Data (AEAD), respectively)\(^2\). Besides, the exact algorithms chosen and all other parameters necessary to perform the cryptography at both the communicating sides (i.e., the security context) must be agreed upon beforehand and stored on the end-devices. The structure of all types of security contexts: sender context, recipient context and common context is presented in Figure 8.6, where the sender is called client and the recipient is called server. Com-

\(^2\)An example of an AEAD algorithm, supported by currently available IoT devices can be AES-CCM with 128-bit key.
8.2 Lightweight application-layer security: protecting CoAP communications

1 byte n bytes

```
0 0 0 h k n n n Partial IV (optional)
```

1 byte s bytes remaining bytes

```
s (optional) kid context (optional) kid (optional)
```

**Figure 8.5.** The format of the OSCORE option [31]. The total length can be maximum 255 bytes.

*mon context* contains the identities of the sender-recipient pair (Context ID), the mutually agreed cryptographic algorithms (AEAD/HKDF identifier) and additional numbers used to derive the keying material and for the cryptography (Master Secret/Salt, Common Initialisation Vector (IV)). *Sender context* and *recipient context* are needed to secure the outgoing and incoming packets, respectively. They include the encryption/decryption key and the endpoint’s identity. The presence of sequence numbers and replay windows means that OSCORE protocol also protects the messages against replay attacks.

OSCORE classifies the available information from the CoAP packet and the security context into several types:

- Class E. This category corresponds to the CoAP Code field, user payload and those CoAP options that are not required for proxying. Class E data are en-

![OSCORE Client and Server security contexts](image_url)

**Figure 8.6.** OSCORE Client and Server security contexts [31, 32]
cryptected and integrity protected.

- Class I. Certain CoAP header fields and options may be accessed by the proxies, but may not be modified. An example can be the Version header field, which needs to be accessed by a node in order to process it correctly, but cannot be maliciously modified, e.g. downgrade to an insecure version. Thus, Class I data are integrity protected, but not encrypted.

- Class U. The remaining message contents can and should be modifiable by the intermediary nodes. For example, Token header values are to be refreshed with each network node the packet traverses. OSCORE takes no action on Class U data.

An unprotected CoAP message turns into a protected OSCORE message via the following procedure:

1. A new CBOR Object Signing and Encryption (COSE) data structure is formed to encapsulate all the confidential data (Class E fields and the payload).

2. Class I parts of the message to be authenticated and integrity-protected, together with certain security parameters, form the Additional Authenticated Data (AAD).

3. With the aid of the sender context components (Sender Key and Sender Sequence Number) the COSE object is encrypted and at the same time the AAD is integrity-protected, yielding COSE ciphertext and AEAD-tag, which are placed in the Message Content field of the COSE object.

4. The resultant ciphertext and tag from 3) become the payload of the protected message, and the remaining CoAP options remain unencrypted. Note that the COSE object, containing the Class E data in plaintext, is not transmitted.

A reverse procedure occurs at the recipient side, where the Recipient Key from recipient context is used by the server to process the incoming OSCORE message and decrypt the payload. Sequence number and Partial IV checks are performed to prevent message replays.

OSCORE has proven to outperform DTLS in terms of RTT, CPU usage and energy consumption, at the same time requiring less memory [32]. The idea is under development and field testing [174, 175], and has been included as the End-to-End security solution in the specification of Lightweight Machine-to-Machine (LwM2M) device management scheme [176].

8.2.3 Novel OSCORE flavour: Group OSCORE

Being currently under development, a new IETF draft [177] defines the adjustment and application of OSCORE protocol to group communication scenarios. The security guarantees remain the same, however group OSCORE additionally specifies how
the keying material can be derived in the network and how to achieve source authentication, confidentiality, integrity protection and message binding over IP multicast. In the standard, 2 modes are considered:

1. Group mode. This functionality is obligatory to implement and concerns CoAP requests addressed to multiple constrained servers, send over multicast and unicast CoAP responses.

2. Pairwise mode. In principle, this mode resembles the original unicast OSCORE scheme, but with some modifications regarding key derivation, the meaning of some fields of the OSCORE option and the fact that external AAD is used both for signing and encryption. Pairwise mode of operation is optional in group OSCORE systems and cannot be used for protecting the messages addressed for multiple servers. On the other hand, it can be applied to secure unicast server responses to a multicast (group) request, as pairwise mode is more lightweight and allow for even further energy-saving at the server side.

In this work, only group mode is investigated and described in details, since the scenario under interest involves multiple receivers.

8.2.3.1 Security context amendments
Similarly as in the unicast OSCORE, Group OSCORE requires that the communicating parties share the context data for security operations. The structure of the security context for the client and the sender has already been presented in Figure 8.6, but now includes the following changes:

- **Common context** also contains Counter Signature Algorithm and Counter Signature Parameters, in order to specify how to derive counter signatures, applied to protect group mode messages.

- To be able to sign group mode messages, **Sender context** is enriched by the private key of the sender.

- Since the receiver has to verify the message signatures with the aid of the sender’s public key, the other endpoint’s public key is added to the **Recipient context**.

8.2.3.2 The Group Manager role
A crucial entity in a Group OSCORE system is a Group Manager (GM), which supervises and is responsible for the group keying material. In particular, the official documentation defines many roles for the GM, the most important of them include:
• Creating and managing OSCORE groups; each group is given a unique Group ID (gid) and the GM additionally control how new devices can join the existing groups or when they can form a new one. Should an endpoint join a group, the Group Manager generates its Sender ID.

• (Re)generating the security context of a group. Any time a group is created or its membership status change (1 or more members join/leave), the gid, the sequence number and the keying material have to be reassigned. GM has to ensure that the group never reuses any of the previous values.

• Validating the public keys of the group members against the chosen countersignature scheme.

Note that in [177], only the responsibilities of the GM are defined; the actual implementation and the details about the communication between the GM and the group members are not specified.

8.2.3.3 Cryptography protocol differences

In this section, the details of operation that are different from OSCORE defined in RFC8613 are explained.

Group OSCORE also relies on CBOR Object Signing and Encryption, however, utilises different COSE object structure than OSCORE. Specifically, the use of countersignature is signalised and the computed signature is included in the data structure. Belonging to a group forces that the OSCORE option fields are interpreted in a different way. Specifically, kid and kid context (see Figure 8.5) are mandatory in Group OSCORE, and they hold the values of the Sender ID and the Group ID (Gid), respectively. Moreover, the Additional Authenticated Data in the group mode is needed both encryption and signing, and in each of the cases the structure of the AAD differs from the one used in the unicast OSCORE. In the encryption, the AAD structure additionally consists of the countersignature parameters and the gid, whilst in the signing, the AAD format is the same as in RFC8613, but also the entire OSCORE option value is included.

Group OSCORE protects the requests similarly as described in Section 8.2.2, taken into account the aforementioned modifications to the COSE object and the AAD structure; however, the final secured payload is also appended with the counter signature, calculated over the COSE object. The receiver of the request needs to verify the counter signature and additionally check, whether the gid included in the message is correct. The decryption and replay protection occurs in the same way as in OSCORE.

8.2.4 Performance evaluation experiment

In this work, Group OSCORE protocol has been implemented in C language as a part of Contiki-NG operating system for constrained devices. The design is an extension
of the OSCORE implementation in Contiki-NG [178] and is now available as an open-source [179]. The purpose of the performance evaluation experiment was to assess the memory and transmission time burden the protocol adds to a constrained IoT device. The setup of the trial consists of a laptop running Group OSCORE client implementation based on Californium [180] Java library, and 2 constrained devices: one of them, interfaced with the laptop and acting as a border router, enables the Java client running on the computer to communicate with the second constrained device, acting as the Group OSCORE server and running the newly implemented Contiki-NG code. The experimental setup is presented in Figure 8.7.

8.2.4.1 Memory occupancy analysis

In this part of Group OSCORE evaluation, different firmware images, containing the considered application-layer variants, were built and the resultant memory footprint was analysed both by Linux binary analysis tools: size and nm, as well as during runtime (observing peak stack occupancy). The cryptographic overhead in OSCORE and Group OSCORE scenarios was calculated by manually adding up the memory footprint of all cryptographic-related variables, constants, functions and data structures. On the other hand, for stack measurements, stack painting technique was used. Namely, the whole stack sector was initialised to a known value, and then during runtime, the amount of bytes where the value changed, could be monitored.

<table>
<thead>
<tr>
<th></th>
<th>Zoul Firefly Rev. A</th>
<th>TI Simplelink CC1352R1 launchpad</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RAM</td>
<td>ROM</td>
</tr>
<tr>
<td>OSCORE-SW</td>
<td>346</td>
<td>1956</td>
</tr>
<tr>
<td>OSCORE-HW</td>
<td>346</td>
<td>1926</td>
</tr>
<tr>
<td>GOSCORE-SW</td>
<td>1699</td>
<td>9861</td>
</tr>
<tr>
<td>GOSCORE-HW</td>
<td>2840</td>
<td>9133</td>
</tr>
</tbody>
</table>
Figure 8.8. Memory occupancy measured on Zoul Firefly Rev. A. Marked areas denote stack usage (in RAM usage) or memory occupied by the cryptography (in ROM usage).

Figure 8.9. Memory occupancy measured on TI Simplelink CC1352R1 launchpad. Marked areas denote stack usage (in RAM usage) or memory occupied by the cryptography (in ROM usage).
8.2 Lightweight application-layer security: protecting CoAP communications

The results of the analysis can be seen in Figure 8.8 (Zolertia Firefly Rev. A) and 8.9 (TI Simplelink CC1352R1 launchpad). The exact numbers regarding cryptographic overhead of OSCORE and Group OSCORE are included in Table 8.1. Note that CoAP and Group CoAP (GCOAP) scenarios correspond to no security scheme applied. In comparison to plain CoAP, Group OSCORE implementation brings noticeable increase of both Random Access Memory (RAM) and Read-Only Memory (ROM): Zoul boards experiences 34-37% Random Access Memory (RAM) usage increase and 35-38% Read-Only Memory (ROM) usage increase. In the case of the Simplelink device, the RAM is occupancy grows by 30% with Group OSCORE, whilst ROM requirements raise by 18% when HW acceleration is used, and by 23% otherwise. As a matter of fact, more memory is required, as the constrained endpoint needs to accommodate for elliptic curves definitions and hashing tables, as well as the mechanisms to sign and verify the messages. It can be observed that the overhead of Group OSCORE is less significant on Simplelink than on Zoul, due to different capabilities of those devices, as well as different implementations of security algorithms and dedicated hardware peripheral drivers. Nevertheless, one can see that hardware acceleration provides noticeable ROM occupancy decrease in both cases. It has to be mentioned that on all the platforms was it possible to build a fully-functional Contiki-NG operating system instance, including Group OSCORE, so that no more than half of the available RAM memory was required. This fact is critical for low-power operations, as for both tested devices, deep-sleep mode can only be enabled when 50% of the available RAM remains idle. Therefore, the complexity of Group OSCORE does not hinder the possibility of energy-efficient IoT sensor operation.

Noteworthy, OSCORE proves that end-to-end application layer security can be realised in lightweight mode. On both platforms, less than 10% RAM and no more than 17% of ROM occupancy was needed on top of plain CoAP implementation. Similarly as in Group OSCORE case, unicast OSCORE running with hardware acceleration of cryptography can further optimise the usage of ROM on the device.

8.2.4.2 Round-trip-time experiment

In this test, the client and the server, having their security context pre-configured, were exchanging Group OSCORE messages and the total E2E delay, or in other words, the time needed to complete a Group OSCORE transaction (request and response exchange) was the quantity under investigation.

During each of the scenarios, an image of the Contiki-NG Operating System (OS) was loaded on the tested IoT platform, and one of the following application layer solutions was included:

1. CoAP (no security),
2. Group CoAP (no security),
3. OSCORE with software-based cryptography implementation,
4. OSCORE with hardware acceleration for cryptography,
5. Group OSCORE with software-based cryptography implementation,
6. Group OSCORE with hardware acceleration for cryptography.

Each of the aforementioned options were tested on TI Simplelink CC1352R1 launchpad and Zolertia Firefly Rev. A boards, thus resulting in a total of 12 evaluation scenarios.

Figure 8.10. Round-trip-time measurements taken for various payload sizes.

Figure 8.10 shows the results of the RTT experiment. For each of the 17 selected payload sizes (from 1 to 128 bytes) 20 client requests were sent and the Java process captured how much time was needed to receive the response from the server. In general, the time needed to complete the transactions using Zoul Firefly Rev. A board are approximately 2 times longer as in the case of TI Simplelink CC1352R1 launchpad board. More interestingly though, certain observations could be made on both hardware platforms. Firstly, applying unicast OSCORE scheme does not result in noticeable increase of the E2E delay, regardless whether hardware cryptographic acceleration was applied. This agrees with the outcomes of the work presented in [32]. Secondly, it is important to note that Group OSCORE operation increases the RTT
significantly, which can be explained by the fact that apart from message encryption and decryption the countersignatures of the packets have to be processed (signed/verified). These additional operations were taking not less than 300 milliseconds on the Firefly board and 1 second on the TI board. Thirdly, a number of rather unexpected trends can be observed. For example, in Figure 8.10a, plain CoAP solution yielded higher delays than Group CoAP and unicast OSCORE variants. Moreover, looking at Figures 8.10a and 8.10b it is difficult to realise whether performing cryptographic operations in hardware in the Group OSCORE cases has positive or negative impact on the RTT. Finally, the total delay values do not seem to grow linearly with the increasing payload size, which might appear counter-intuitive. These unexpected observations may be explained by the fact, that the number of samples used to produce the RTT graphs (20 transactions for each solution) is too low for the results to stabilise and reflect the stable, long-term behaviour of the devices, accounting for hardware and wireless transmission uncertainties. Redoing the experiment for bigger amount of message exchanges has been left for future work.

8.2.5 Discussion

The experiments presented in this chapter included CoAP, Group CoAP, OSCORE with software/hardware cryptography and Group OSCORE with software/hardware acceleration of the cryptography. All of the variants could be built successfully, allowed for proper message exchange and fitted in the platforms’ RAM occupancy requirement for low-power operation. Thus, on one side the proposed implementations could be verified, and on the other side it could be proven on 2 real-life hardware pieces that both unicast OSCORE and Group OSCORE can be successfully applied in constrained environments. A general trend of ROM saving when choosing cryptography acceleration in hardware is noteworthy. Having tested the design on 2 types of constrained devices could reveal that both the absolute memory overhead of Group OSCORE (amount of bytes), as well as the differences between SW and HW acceleration versions can be different across hardware platforms.

The RTT evaluation case, though constrained by the number of samples involved, revealed that additional security operations, introduced with Group OSCORE, have a big impact on the E2E message exchange delay. Specifically, signing and verifying the messages is time-consuming; enough to say that 1-byte message transaction with Group OSCORE was significantly longer than 128-byte CoAP or OSCORE transactions. Therefore, only relaxed latency applications can tolerate such delay increase. Although the LP-WANs, considered in this thesis (NB-IoT, LTE-M, LoRaWAN and Sigfox) are to be deployed in latency-tolerant situations, one can identify extreme use-cases involving extended coverage and message repetitions (CIoT technologies), where adding Group OSCORE processing delays might cause the system to exceed the maximum expected E2E latency of 10 seconds.

Nonetheless, more experimental data (transactions per payload size) is required to provide more reliable picture of RTT performances of the investigated OSCORE
flavours. Future work includes repeating the RTT experiment and investigating CPU and radio transmitter’s energy consumption.

From the perspective of smart communities, Group OSCORE has a potential of becoming a default application-layer protection mechanism in those use-cases, where multitudes of resource-constrained servers are to be remotely managed, configured and supervised by a non-constrained client. For example, Group OSCORE transactions may enable secure and energy-efficient remote reconfiguration of all multicast group members (e.g. all remote metering devices in a given building). Similarly, new firmware might be updated on the sensors in a secure and private way. As originally stated in [177], Group OSCORE can also be suitable for light and, more generally, building automation scenarios, falling into a category of smart city applications.

8.3 Summary

This chapter presents a theoretical and a practical research effort concerning security improvement in low-power IoT. At first, a research angle of Sustainable Security for Internet of Things (SSIoT) was motivated and formulated, as to underline that future projects ought to be oriented towards more secure and privacy-protecting, yet also environmentally friendly and affordable IoT communications. Secondly, a performance evaluation study investigated a novel application-layer security proposal called Group OSCORE, designed to protect one-to-many data exchange of CoAP messages in the world of constrained devices. The protocol was implemented in-house as a part of an open-source operating system for IoT (Contiki-NG), and then tested on 2 commercially available hardware platforms: TI Simplelink CC1352R1 launchpad and Zolertia Firefly Rev. A. The results of memory occupancy analysis and the Round Trip Time experiment verify the correctness of the implementations and highlight that Group OSCORE can be used in real-life constrained IoT applications, such as integrated lighting and building control, secure bulk firmware or configuration updates. However, the additional time needed to sign and verify the messages is significant and must be taken into account at the time of the deployment.

In the future, the initial Group OSCORE experiments, presented in this dissertation, will be complemented by more thorough tests, also including energy consumption of the CPU and the radio module.
Part III Summary

Equipped with massive connectivity IoT technologies, smart communities become not only ready to embrace novel applications and services, but also more vulnerable to various kinds of malicious actions originated somewhere in the Internet. Thousands of IoT devices joining the global network are often insufficiently secured and can in fact amplify the effects of some classical distributed network attacks, such as Distributed Denial of Service (DDoS). This part of the thesis focused on evaluating the robustness of state-of-the-art security defined for LP-WAN networks and exploring theoretical and practical solutions that could positively impact the security guarantees of the smart community network infrastructure.

The analysis of LoRaWAN, Sigfox and NB-IoT security schemes showed that despite their complexity and maturity, effective and relatively cheap hacking attacks are possible. This has been experimentally proven for all these three standards, however, the immediate benefit was to inform the stakeholders (network operators and infrastructure vendors) and help them improve their security systems.

In this dissertation part, it was discovered that the recent academic efforts have targeted the areas of computing, IoT and network security from a green (or sustainable) perspective; albeit, the research on the topic combining all three was not clearly visible. Therefore, an angle of Sustainable Security for Internet of Things (SSIoT) was formulated in this dissertation with the hope that it will reorient future work onto not only secure, but also green and affordable IoT communications, and raise the level of concern about the environmental (as well as economical) footprint of IoT security.

With that in mind, Group OSCORE, a novel application-layer security solution for group communication using CoAP protocol, was investigated. Since the proposal offers confidentiality, integrity protection, message binding and source authentication for 1-to-many scenarios in a lightweight design, the experimental performance evaluation was conducted to discover the overhead Group OSCORE induces on a constrained device. It could be seen that real-life constrained IoT devices can accommodate for Group OSCORE functionality and still preserve an energy-efficient mode of operation. On the other hand, care must be taken when planning critical applications employing Group OSCORE security, as signing and verifying the messages increases the E2E delay noticeably.
Conclusion

This dissertation addresses three aspects recognised as the essential pillars of a reliable IoT architecture for smart communities: seamless coverage, failure-tolerance and robust security.

The research items included in this thesis unveil how the aforementioned reliability aspects have been realised in LP-WAN standards: NB-IoT, LTE-M, LoRaWAN and Sigfox, and introduce or explore solutions and practices that may be applied in a more reliable low-power IoT communications system.

Part I of the dissertation highlights the significance of empirical data collection in the process of understanding radio signal propagation. Apart from acquiring an immediate knowledge regarding the coverage performance in the measured area, high-quality field-trial datasets can be used to calibrate path-loss channel models, which increases their accuracy and facilitates successful IoT deployment planning. The acquired data enables further feature engineering and statistical investigations in terms of signal power and throughput, among others.

In Part II, it is shown that a CIoT failure-recovery enhancement that involves only constrained devices (i.e., with limited energy, memory and radio resources) and is activated only for the duration of the RAN failure, may be feasible to implement and increases the probability of finding a relay candidate in the neighbourhood. Such a targeted solution becomes an alternative to general D2D-based communications enhancements, due to its simplicity and minimal resource overhead. Further work is required so that the simulations evaluating the proposed scheme can be conducted.

Part III demonstrates experimentally that all the considered LP-WAN technologies could be hacked due to security design weaknesses or implementation issues. Finding and exploiting the vulnerabilities should be continued to ensure that the system has been made secure against the potential threats as soon as possible. Moreover, a new research angle of SSIoT invites the research community for development of more green-aware, economically feasible security solutions for IoT. Finally, through a performance evaluation experiment of in-house implementation, it was recommended that Group OSCORE protocol, offering strong security guarantees for application-layer group communications is incorporated in smart community network system.

The contributions of this document can be summed up as follows:

- Chapter 2: NB-IoT, LTE-M and LoRaWAN provide satisfactory connectivity in the considered harbour environment (Svanemollehavn, Denmark).
• Chapter 3: a commercial off-the-shelf NB-IoT device can successfully receive and demodulate downlink signals from the base station in deep-indoor environments (underground tunnel system) with more than 1km distance.

• Chapter 3: 3GPP 38.901 channel model can successfully predict indoor path-loss above-ground, but is not precise for underground scenarios.

• Chapter 3: a new, publicly available, large-scale underground measurement dataset containing UE radio statistics and environmental features was created. The empirical data are precisely localised with the aid of high-resolution LIDAR terrain map.

• Chapter 3: the underground behaviour of sub-GHz radio signal can be better understood by means of environment-related statistical features, engineered on the basis of tunnel geometry and transmitter-receiver straight-line distances. The deep-indoor predictor model proposals, derived with the aid of the engineered features, have proven higher accuracy that the current 3GPP 38.901 indoor component.

• Chapter 5: The newly proposed, lightweight failure-recovery scheme for CIoT utilises D2D link and relaying to maintain E2E communications under the event of eNB failure, and minimises the resource impact on the relay candidates.

• Chapter 7: security weaknesses, identified for NB-IoT, Sigfox and LoRaWAN were confirmed in the hacking experiments.

• Chapter 8: A newly-defined research angle, Sustainable Security for Internet of Things (SSIoT), is believed to attract industrial and academic attention and set the focus of future IoT security research on both security robustness, environmental impact reduction and affordability.

• Chapter 8: It was experimentally found that Group OSCORE, a novel application-layer security scheme for CoAP in group communication scenarios, introduce limited memory overhead on constrained IoT devices, thus deep-sleep mode can be used. On the other hand, Round Trip Time becomes significantly increased when message signing and verification is involved.

The remaining content of the thesis contains theoretical work introducing and explaining core knowledge, concepts and depicting the problem context. In Chapter 1, the LP-WAN standards (LoRaWAN, Sigfox, NB-IoT and LTE-M) are described and current coverage performances are discussed. Chapter 4 considers the problem of failure-tolerance in IoT networks in the context of RAN failure. An overview of the relevant hacking attacks, their impact and general mitigation methodology is included in Chapter 6.
Outlook

The findings from this PhD dissertation imply that the contemporary smart communities have been given several promising and robust communications technologies, enabling low-power, long-range and reliable use-cases. However, novel LP-WANs still require field testing and technological adjustments in order to secure seamless operation in edge-cases of smart communities’ deployment. For example, coverage performance of low-power IoT communications ought to be empirically evaluated in a wider spectrum of challenging scenarios and massive amount of radio samples are needed to produce generic, simple and powerful channel models, facilitating correct deployment decisions. The incorporation of lightweight D2D in LP-WAN technologies should be pursued, as soon-to-be-deployed billions of connected devices can benefit from one another’s proximity and achieve higher reliability, more optimised energy consumption and offloaded network core. Security breaches of LP-WANs should be intensively analysed and fixed, before the hackers utilise them for malicious purposes. Last, but not least, future security schemes can only be adopted in low-power IoT networks if they combine solid system protection with low resource and energy overhead.
Appendices
This chapter contains the following submitted research items:

D2D-enabled Failure-tolerance in Cellular IoT

1st Krzysztof Mateusz Malarski
Department of Photonics Engineering
Technical University of Denmark
Kongens Lyngby, Denmark
krmal@fotonik.dtu.dk

2nd Kalpit Dilip Ballal
Department of Photonics Engineering
Technical University of Denmark
Kongens Lyngby, Denmark
kdiba@fotonik.dtu.dk

3rd Sarah Ruepp
Department of Photonics Engineering
Technical University of Denmark
Kongens Lyngby, Denmark
sraru@fotonik.dtu.dk

Abstract—Failure-tolerance is an essential feature of any reliable system. In the context of Cellular IoT, mechanisms preventing the communications collapse at the event of failure are desired, if critical applications are to be deployed. In particular, this paper addresses the scenario, in which the end-devices, located in a remote and hard-to-reach area, experience a failure of the only base station in proximity. We propose a reliable and lightweight scheme that sustains the connectivity by means of a Device-to-Device link, established with a foreign device acting as a relay. Since the direct communications occurs only at the time of the failure, the solution introduces minimised overhead towards the relay device, which in our proposal can be realised by constrained hardware. We explain the assumptions, the behaviour and signalling patterns of the system. Finally, we discuss the applicability of our scheme and provide implementation considerations.

Index Terms—failure-safety, D2D, NB-IoT, LTE-M, relay

I. INTRODUCTION

Introduction of 5th generation mobile networks (5G), among other things, promises massive enhancements to communication speeds, growth in the machine type communication devices, and support for Ultra-Reliable Low Latency Communication (URLLC) for Mission-Critical Applications (MCAs) [1]. MCAs, such as industrial automation, e-health, smart cities and autonomous cars impose stringent network performance requirements from communication networks [2]. Critical applications such as e-health, remote monitoring of chronically ill patients require high security, low End-to-End (E2E) latency, high reliability and high availability for their expected functioning [3]. Failure to meet these network requirements can have severe implications in monitoring the patient safely and may in worst cases lead to casualties.

Narrowband-IoT (NB-IoT) and LTE for Machine Type Communication (LTE-M) are two Cellular IoT (CIoT) technologies introduced by 3rd Generation Partnership Project (3GPP) that are specially designed for applications that demand low power consumption, extended coverage with deep indoor signal penetration, and high reliability. Unlike most other Low-Power Wide-Area Network (LP-WAN) technologies, CIoT uses a licensed frequency spectrum for wireless communication. This allows them to avoid duty-cycle restrictions and limit the interference from other neighbouring devices [4]–[6].

CIoT has been developed on top of the existing Long-Term Evolution (LTE) standard. Therefore NB-IoT and LTE-M have similar characteristics and communication procedures as that of LTE. Both NB-IoT and LTE-M achieve extended coverage by message repetitions, apply energy-efficiency schemes: Power Saving Mode (PSM) and extended Discontinuous Reception (eDRX) and inherit LTE-grade security. Despite the similarities, LTE-M and NB-IoT differ in the operational bandwidth, peak-data rates and physical layer design. LTE-M uses a 1.4 MHz carrier bandwidth and PHY is compatible with LTE, making it possible to be deployed in an in-band configuration. NB-IoT is not fully compatible with LTE, but it occupies only 180 kHz of bandwidth and thus supports three deployment options, namely in-band, guard band, and stand-alone [7].

In order for CIoT devices to communicate with the core network, they need to attach to an evolved Node-B (eNB) in the area. This presents a potential risk of all the neighbouring base stations failing, resulting in a complete network outage, which is especially undesirable in hard-to-reach deployments. This is not acceptable for MCAs, as complete disconnection of the devices from the Evolved Packet Core (EPC) could lead to errors and disturbances in the system’s operation.

One solution to avoid complete disconnection of these CIoT devices is to use Device-to-Device (D2D) paradigm. D2D allows establishing a direct communication link between two CIoT devices in proximity where one of them is still attached to the eNB and relaying packets from the faded device to the network infrastructure [8]. A secondary use of D2D could be to offload traffic from the CIoT network. Offloading the device traffic from the network will help in the reduction of energy utilisation, and E2E delay [9], [10].

In this paper, we introduce an enhancement to CIoT operation that enables service continuity even when the eNB becomes unreachable. The User Equipments (UEs) communicate over a direct link with the relay UE, which forwards all uplink data towards the EPC. The proposed solution can be applied in NB-IoT and LTE-M deployments. Apart from network-assisted relay assignment prior to the emergency situation, the connectivity between the affected UEs and the relay relies on pre-configured resources and is maintained only until the eNB is recovered. We believe that the simplicity of our design and relaxed relay requirements increase the chance of finding a suitable relay in the neighbourhood. Our contributions can be reduced to the following items:

• we present the design of a failure-tolerant enhancement
for CIoT. Both the assumptions, the scope, the behaviour of the system and new signalling are included.
- we provide additional considerations on applying the solution on commercial off-the-shelf hardware.
- we discuss the applicability of the presented scheme and feasibility of applying it in non-cellular Internet of Things (IoT) systems.

The remainder of this paper is organised as follows. Section II presents the 3GPP perspective on D2D communications and discussed the related work. We define the considered scenario and state the assumptions in Section III. The description of our failure-tolerant enhancement for CIoT is included in Section IV. Section V contains implementation considerations and discusses the applicability of the solution. Finally, the work is concluded in Section VI.

II. EXISTING EFFORTS

The problem of Radio Access Network (RAN) unavailability has been addressed by 3GPP mainly from the public-safety perspective. The potential of applying D2D to restore the connectivity in emergency scenarios was first used in Release 12, where Proximity Service (ProSe) architecture was formulated for the first time (see Figure 1). Apart from introducing a new EPC entity (called ProSe function), the standard defines a number of new interfaces (PC1-PC5), connecting the network elements. In order to facilitate direct connectivity in terms of data flow, synchronisation and control, the D2D UEs are to use Sidelink (SL), bringing a number of new physical and logical channels.

As shown in Figure 2, 3GPP considers 3 deployment situations, where ProSe can be applied. Note that only in-coverage option can be used for the use-cases other than public-safety. The next releases brought about the functionality of relaying, implemented on Layer-3 (Rel. 13) and Layer-2 (Rel.14) [12]. In the latter solution, the remote UE can be better differentiated from the relay UE by the EPC, which allows for Quality of Service (QoS) guarantees and protects the remote UE payload from being accessed by the relay, being the case in Layer-3 scheme. However, as shown in [12], achieving QoS between the base station and the remote UE requires resolving many trade-offs regarding signalling overhead, implementation complexity and the overall delay.

Since the original design of ProSe and SL targeted classical LTE, a new work on applying the idea to the constrained CIoT equipment commenced in Release 14 [13]. The document considers Machine-Type Communication (MTC) applications, suggests simplifications to the channels and interfaces for NB-IoT and LTE-M and mentions that in the signalling procedures, the nature of power-saving schemes must be taken into account. Unfortunately, to the best of our knowledge, the initial report has not been transformed into a mature standard; instead, the most recent 3GPP efforts (e.g. in [14]) implies that the interest has been moved onto D2D and SL applications in vehicular networks.

A. Non-3GPP related work

Humukumbure et al. [15] achieves better reliability, minimises power consumption and limits the congestion of the network by incorporating D2D feature into the cellular infrastructure. Specifically, changes to random access procedure are proposed, where the base station announces the D2D mode of operation by means of broadcast messages, and then the UEs communicate with one another directly using Carrier Sense Multiple Access (CSMA) with collision avoidance. However, the work does not consider IoT devices and their different resource constraints.

A D2D solution enhancing the battery life and the availability of the CIoT deployment is described in [16]. The network supervises and maintains the role assignment (remote devices, relays) of the UEs with the aid of the collected environmental data (e.g. battery level and position). The authors derive new signalling behaviour, enabling UE attachment, (re)configuration of the transmission mode and uplink data transmission. The feasibility of the proposal is verified in system-level simulations; albeit, LTE radio is applied, so the effect of the behaviour on deep-sleep CIoT techniques cannot be observed. Furthermore, the application of core network controlled D2D solution is not suitable for the RAN failure problem, considered in this paper.

To the best of our knowledge, this study is the first to present a CIoT enhancement based on D2D, which targets eNB failure scenario.

III. PROBLEM DESCRIPTION

In this work, we take a closer look at a situation illustrated in Figure 3. A deployment of CIoT UEs realise that the eNB...
they have been attached to stops communicating (eNB1 in the Figure); this may be the consequence of hardware/software fault, or a hacker attack. Since the UEs are located in a remote and hard-to-reach area (e.g. coverage edge rural outdoor or deep-indoor), reattachment to another eNB can be impossible (see the Figure, where eNB2 is out of range). As a result, the affected UEs become disconnected from the network infrastructure and the corresponding service is discontinued.

One may identify 2 general approaches that could be applied to the system in order to avoid communication disruption: 1) to incorporate multiple radio technologies on the UE and use the non-CIoT one as a backup means of data transfer, which may be viable in practice, however, as shown in [17], the process of optimising the system in terms of technology integration, device size and battery requirement changes, as well as resource allocation may be difficult; 2) to enrich the CIoT standard with D2D capabilities, so that at the time of failure, other end-devices can form a connectivity chain leading to the closest operating eNB. In our work, approach 2) has been chosen, as we believe that the complexity of the design can be reduced and some problems avoided, if only a single Radio Access Technology (RAT) is considered.

**IV. OUR SCHEME**

We propose a D2D-based failure-recovery mechanism for NB-IoT and LTE-M networks, which addresses the problem depicted in Figure 3. Unlike in the case of alternative solutions, introduced in [15], [16], our scheme can be understood as a “targeted remedy”, since its purpose is to sustain the connectivity between the UEs and the EPC.

The operation of the mechanism is presented in Figure 4. The devices affected by their eNB breakdown suffer from the proximity of another UE, attached to a different eNB and willing to act as a relay at the time of emergency. The awareness of, so called Victim UEs (VUEs) about the Relay UEs (RUEs) and vice versa has been established in a network-assisted way prior to the time of the failure. The VUEs are organised into groups, led by special VUEs, Group Leaders (GLs). In order to limit the amount of packets sent and the number of D2D connections, the VUEs first forward the uplink data to appropriate GLs, which communicate via D2D link with the RUE and send the aggregated uplink packets from all the group members. Once the recovery of the eNB is perceived by the VUEs, the uplink traffic is immediately redirected towards the eNB and all the D2D connections with the RUE are torn down.

**A. Assumptions**

Our solution has been designed with a though of stationary deployment, which means that the assignment of GL roles among the VUEs may happen once and be preserved. Moreover, we assumed that the VUEs do not need to use signal repetitions to reach out their GLs. We estimate that the time to repair the eNB should not exceed 24 hours, though in some extreme cases, it can lasts for many days [18]. It was also assumed that it was possible to associate with at least 1 relay candidate prior to the failure, where the critical application, running on the VUEs required no more than 1 uplink message per 10 minutes, corresponding to 28.8 kB per day per device. Finally, we expect that even constrained, commercially available NB-IoT boards are powerful enough to implement our solution. This corresponds to an important requirement that the complexity and resource penalty of the scheme must be minimised, with the relay perspective being the first priority, and the VUEs being the second priority.

**B. Action sequence**

The following list explains how the system behaves at the time of RAN failure.

1) Thanks to the previously conducted configuration signalling (see Section IV-C), the VUEs keep the necessary information about the assigned RUE: Internet Protocol (IP) address, radio parameters, etc. On the other hand, the RUE stores the group IDs of the associated GLs.

2) The VUEs realise the eNB failure. This can be experienced as e.g. lack of Narrowband Reference Signal (NRS) for during the time the UE is not in a deep-sleep mode.

3) The GLs send a relay activation request message to the RUE. Non-GL devices transmit the application data to the GLs acting as aggregation nodes.

4) Once the relay acknowledgement is received, the GLs send the aggregated payload to the RUE in an unacknowledged way.
5) Without any intermediate processing, the relay forwards the received packets towards its eNB; since the RUE keeps serving its original service in the network, it may also need to send its data, which in such cases becomes combined with the relayed message.

6) As soon as the faulty eNB recovers (e.g., it resumes the downlink signalling), the VUEs notifies the closest GL about the event, being later on propagates among the remaining GLs. One of the GL devices sends a relay deactivation request to the RUE and all pending data are directed towards the reactivated eNB. In a similar way, any VUE that is yet to send its application packet, but is already aware of the eNB’s recovery, shall not transmit to the GL, but to the base station.

7) Upon receiving the relay deactivation request, the RUE replies with an acknowledgement and disables its relaying capabilities, thus releasing all the additional resources spent on being a RUE.

C. Signalling behaviour

The functionality described in the previous section is enabled by new signalling patterns, defining the initial relay assignment and (de)activation. Figure 5 presents a diagram showing how a GL can be assigned a RUE at the time of the first network attachment. It has to be mentioned that it is the EPC’s responsibility to store and maintain the list of available relay candidates, as well as valid associations between the RUEs and the VUE groups; moreover, current location and remaining battery lifetime information of the relays are continuously collected. It can be observed that the GL requests a relay while attaching to the EPC (relay attachment request is sent) and identifies itself by means of the group ID. In response to that, the network finds the most convenient (i.e., closest and having enough battery power) relay and originates a relay attachment response, containing the necessary configuration data, such as the relay IP address and pre-configured radio resources for the D2D communications. Should the relay fail or resign, the network repeats the relay finding process and sends a relay attachment notification, carrying the necessary configuration information. In case there was no suitable device that could act as a relay for a given GL (whether during the first attachment time, or after the previous RUE became unavailable), the GL can discover that by receiving empty configuration in the relay attachment response/notification message.

Note that the procedure of GL assignment within the VUE deployment is not presented, as it is assumed that the GL devices, possibly equipped with more powerful batteries, would include this functionality from the start of operation and never change this role throughout their lifetimes.

D. Implementation considerations

As stated in Section IV-A, the value of the proposed failure-proof scheme lays in the minimal overhead imposed on the end-devices. As long as the relay UE is concerned, the total overhead contains: 1) the CPU and memory resources required to implement the proposed behaviour, signalling patterns, and to maintain the awareness of the associated GLs, and 2) radio resources dedicated to the emergency D2D data exchange. On the other hand, the actual burden on the VUE side depends on the specific role in the deployment. All the non-GL VUEs should implement the behaviour of reacting on the downlink signalling loss by sending the uplink packets directly to the pre-configured GL. The GL VUEs need to carry the logic enabling relay attachment procedure and D2D communications with the RUE. Although in cellular networks the direct data transfer between the end-devices typically occurs by means of PC5 interface and sidelink channels, our solution does not require full ProSe functionality that would be anyway overwhelming to incorporate in constrained VUEs and RUEs. In particular, the emergency D2D communications during the eNB repair is expected to happen between pre-discovered devices using pre-configured radio resources and, for obvious reasons, the presence (and assistance) of the network is not possible. Therefore, the data structures and messaging regarding D2D capability advertisement, device discovery and sidelink resource pool assignment should be excluded from the implementation.

Since the relaying in our scheme is considered at Layer-3 and it was highlighted in [12] that Layer-3 relaying allows
the RUE to access the relayed IP payloads, we recommend that application-layer security mechanism, such as OSCORE, is applied as a countermeasure to the privacy problems.

V. Discussion

It can be observed that the proposed failure-tolerance enhancement exhibits a fundamental difference to other solutions found in the literature ([15], [16]). While the related studies, similarly to our work, apply D2D communications with the goal of increasing the availability of the system, they require that the network infrastructure continuously supervises the end-devices, which cannot be the case for the failure scenario investigated in this paper. Even though our solution is not capable of improving the energy-efficiency of the UEs (which is the benefit of using the proposal in [16]), only our design enables that the UEs continue communicating with the network even if, temporarily, no eNB operates in the range of the affected UEs. We foresee that limited complexity and resource penalty of our proposal can be appealing wherever it is desired to guarantee the availability in the extreme case at minimal cost. The minimised computational and battery requirements for the RUE makes it more probable to find a relay candidate in the neighbourhood of the VUEs, which really allows for D2D communications. All in all, we claim that our idea is viable to apply in real-life scenarios, and it is realistic to expect some neighbouring devices (even the constrained nodes) volunteer to act as relays. However, the design of a particular relay incentive mechanism is out of scope of this work.

A. Non-CIoT perspective

The considered RAN failure problem has been discussed in the context of NB-IoT and LTE-M. However, other IoT networks, such as Sigfox and Long-Range Wide Area Network (LoRaWAN), may be affected in the same way should the only gateway fail. The intrinsic payload size (12 bytes) and daily transmission limitations (144 messages) of Sigfox excludes the possibility for any Sigfox device to act as a relay. In the case of LoRaWAN, it has been discussed and proven in [19] that with the aid of currently available hardware chips it is feasible to enhance the LoRaWAN protocol with D2D communication and, as a result, improve the throughput, offload the backend network and optimise the energy usage of the remote sensors.

VI. Conclusions and Future Work

In this paper, we introduced a D2D enhancement for CIoT, which improves failure-tolerance of NB-IoT and LTE-M by sustaining the communication with the UEs, isolated by the eNB breakdown. Contrary to the examples found in the literature, our proposal requires that the direct link is only active during the eNB repair, and is deactivated otherwise. This, combined with low resource demand, enabling constrained relay devices is believed to increase the probability of being assigned a relay prior to the emergency situation.

In the future, we plan to provide system-level simulation results, verifying the operation of the proposed scheme. Furthermore, the details of the communications between the VUEs and the GLs, especially the aspects of synchronisation and scheduling, will be addressed.

Acknowledgment

This work was partially founded by Innovation Fund Denmark through the Eureka Turbo project IoT WatchLife.

References


[79] TSGR, “TR 138 901 - V14.3.0 - 5G; Study on channel model for frequencies from 0.5 to 100 GHz (3GPP TR 38.901 version 14.3.0 Release 14),”


[111] 3GPP, “Study on further enhancements to LTE Device to Device (D2D), User Equipment (UE) to network relays for Internet of Things (IoT) and wearables,” Technical Report (TR) 36.746, 3rd Generation Partnership Project (3GPP), 04 2018. Version 15.1.1.


