UNIVERSITÀ DEGLI STUDI DI PARMA

Dottorato di Ricerca in Tecnologie dell'Informazione

XXII Ciclo

PERFORMANCE OPTIMIZATION IN WIRELESS SENSOR NETWORKS

Coordinatore:

Chiar.mo Prof. Carlo Morandi

Tutor:

Chiar.mo Prof. Gianluigi Ferrari

Dottorando: *Paolo Medagliani*

Gennaio 2010

To my family and to Isotta
– simply thank you!

Marco Polo descrive un ponte, pietra per pietra.
– Ma qual è la pietra che sostiene il ponte? – chiede Kublai Kan.
– Il ponte non è sostenuto da questa o quella pietra, – risponde Marco, – ma dalla linea dell'arco che esse formano.
Kublai Kan rimane silenzioso, riflettendo.
Poi soggiunge: – Perché mi parli delle pietre?
È solo dell'arco che mi importa. Polo risponde: – Senza pietre non c'è arco.
(I. Calvino - Le città invisibili)

Contents

Introduction **1**

1 Literature Analysis and Motivations **3**

1.1 Introduction 3
1.2 Detection of Incoming Targets 4
1.3 Power Allocation Strategies 5
1.4 Data Fusion and Clustered Networks 6
1.5 Integrated Networks 8

2 The Zigbee Standard **9**

2.1 Introduction 9
2.2 Functional Description of the Zigbee Standard 10
2.3 Physical Layer of the Zigbee Standard 14
2.4 Access and Medium Control Layer of Zigbee Standard 17
2.4.1 Structure of the Superframe 17
2.4.2 Structure of the MAC Layer Packets 19
2.4.3 Data Transfer Modality 20
2.4.4 The CSMA/CA Protocol 25
2.5 Network Layer of Zigbee Standard 31
2.5.1 Network Creation and Management 32
2.5.2 Routing Algorithm 33
2.5.3 Network Topologies 34

2.6 Application Layer 37
2.7 Conclusions 39

3 Theoretical Performance Evaluation Framework 41
3.1 Introduction 41
3.2 Target Detection Model 42
3.2.1 Problem Statement 42
3.2.2 Physical Description of the Sensor 43
3.2.3 Preliminary Works on Random Deployment 44
3.2.4 Integration of Duty Cycles for Random Deployment 45
3.2.5 Latency after Detection 49
3.2.6 Energy Model 51
3.3 Adjacency Matrix-based Transmit Power Allocation Strategy 54
3.3.1 Definition of a Simplified Model for Zigbee WSNs 54
3.3.2 Network Lifetime 63
3.3.3 Optimal Transmit Power Allocation 66
3.3.4 Extension to Multi-hop Scenarios 70
3.4 Clustered Networks Performance Analysis 75
3.4.1 Network Tolerable Death Level 75
3.4.2 Possible Network Configurations 76
3.4.3 Sensing and Data Fusion 78
3.5 Concluding Remarks 80

4 Performance Evaluation 83
4.1 Introduction 83
4.2 The Opnet Simulator 84
4.3 Target Detection Performance 86
4.3.1 Validation 86
4.3.2 Latency after Detection 87
4.3.3 System Engineering 89
4.4 Adjacency Matrix-based Transmit Power Allocation Strategy 93
4.4.1 Considered Opnet Model 94

4.4.2 Performance Analysis 98
4.4.3 Validation of the Analytical Model 99
4.4.4 Impact of the Adjacency 102
4.4.5 Impact of Traffic 103
4.4.6 Impact of the Power Allocation Strategy 106
4.4.7 Network Lifetime Performance 112
4.5 Clustered Networks Performance 116
4.5.1 Simulation Setup 117
4.5.2 Performance Analysis 119
4.5.3 Impact of Tolerable Network Death Level 119
4.5.4 Impact of the Packet Generation Rate 126
4.5.5 Impact of Data Fusion Mechanisms 131
4.5.6 Design Guidelines 137
4.5.7 A Simple Analytical Model for the Zigbee Performance Surface 140
4.5.8 Analytical Approximation of the Performance Surfaces . . . 142
4.6 Concluding Remarks 144

5 RFID-Controlled Zigbee Networks **145**
5.1 Introduction 145
5.2 RFID Technology 146
5.3 Hybrid Zigbee-RFID Networks 147
5.3.1 System Model 147
5.3.2 Integrated Opnet Structure 149
5.3.3 Deep Sleep Algorithm 150
5.3.4 Deep Sleep Algorithm with Virtual Spatial Grid 154
5.4 Performance Analysis 155
5.4.1 Deep Sleep Algorithm 157
5.4.2 Impact of the Virtual Spatial Grid 160
5.5 Conclusions 163

6 Concluding Remarks and Future Work **165**

Bibliography **169**

Acknowledgments **181**

List of Figures

2.1 ISO/OSI stack of the Zigbee standard. 10

2.2 Network topologies allowed by the Zigbee standard. 13

2.3 Scheme of the modulator for the Zigbee standard at the frequencies of 868 and 916 MHz. 15

2.4 Scheme of the modulator for the Zigbee standard at the frequency of 2.4 GHz. 16

2.5 Structure of a superframe. 17

2.6 General structure of a MAC layer data packet. 19

2.7 Data transfer from an end device to a PAN coordinator. 21

2.8 Data transfer from a PAN coordinator to an end device. 24

2.9 CSMA/CA mechanism in the IEEE 802.15.4 protocol. 28

2.10 Logical scheme of the application layer of the Zigbee protocol. . . . 38

3.1 (a) Logical scheme of the sensing duty cycle and (b) model for the sensing range of a node. 46

3.2 Integration domain for the evaluation of $P\{\mathscr{E}_1, \mathscr{E}_2\}$. 47

3.3 Pairwise connections in a scenario with $N = 10$ nodes. Two values for the total network transmit power are considered: (a) $P_{\text{tot}} = 5 \cdot 10^{-5}$ W and (b) $P_{\text{tot}} = 2.5 \cdot 10^{-5}$ W. In both cases, the proposed optimized power allocation strategy is used. 70

3.4 Graphical notation for communication links: (a) A and B communicate with each other (bidirectional communication); (b) only A can transmit to B (monodirectional communication). 70

3.5 Illustrative clustered topologies: (a) uniform with 4 clusters and 4 relays, (b) non-uniform with a 10-2-2-2 configuration with 4 relays, (c) non-uniform with a 14-1-1 configuration and 3 relays, and (d) non-uniform with a 14-1-1 configuration and 1 relay. The nodes denoted with "R" are only dedicated to relay packets from the RFDs to the AP, whereas the nodes denoted with "FC" are also used to fuse data received from the RFDs. 77

4.1 Simulation (solid lines) and analytical (dashed lines) P_{md} results as functions of the duty cycle β_{sens} considering different sensing ranges. The target enters with speed $v = 15$ m/s a monitored area of $d_{\text{s}} = 1000$ m side, where $N = 50$ sensors are randomly deployed. 87

4.2 Simulation (solid lines) and analytical (dashed lines) P_{md} results as functions of the duty cycle β_{sens}, considering different period lengths T_{sens}. The size of the area is $d_{\text{s}} = 1000$ m, in which $N = 50$ sensors, with sensing range equal to $r_{\text{s}} = 50$ m, are randomly placed. The speed of the target is $v = 15$ m/s. 88

4.3 Latency as a function of the number of hops traversed by a packet. Both (i) experimental (dashed lines) and (ii) theoretical results (solid lines) are presented. 89

4.4 Joint optimization of D, P_{md}, and L. No constraints are imposed. . . 91

4.5 Lifetime maximization under probability of missed detection (P^*_{md}) and latency (D^*) constraints. 92

4.6 Considered network topologies with $N = 20$ nodes. 94

4.7 Considered network topologies with $N = 10$ nodes. 95

4.8 Network topologies with $N = 10$ nodes randomly deployed over a 10 m^2 square surface, used for (a) PER comparison and (b) evaluation of the network lifetime. (c) Network topology with $N = 10$ nodes randomly deployed over a 50 m^2 square surface, used for the evaluation of the network lifetime. 97

4.9 PER as a function of the total offered traffic load. The simulation results are obtained considering (i) default backoff exponent, i.e. $BE_{\min} = 3$ and $BE_{\max} = 5$, and (ii) modified backoff exponent, i.e. $BE_{\min} = BE_{\max} = 7$. Different values of per-node transmit powers are considered. The allocated transmit power is the same at each node. 100

4.10 (a) Network transmission rate and (b) delay, as functions of the total offered traffic load (Ng), in the presence of (i) default backoff exponent, i.e. $BE_{\min} = 3$ and $BE_{\max} = 5$, and (ii) modified backoff exponent, i.e. $BE_{\min} = BE_{\max} = 7$. Different values of transmit powers are considered. The allocated transmit power is the same at each node. 101

4.11 PER as a function of the sparsity index of the adjacency matrix. In the simulations, for scenarios with $N = 20$ nodes, the packet generation rate is set to $g = 0.1$ pck/s, whereas for scenarios with $N = 10$ nodes, the packet generation rate is set to $g = 0.2$ pck/s. 102

4.12 PER as a function of the sparsity index of the adjacency matrix. In the simulations, for scenarios with $N = 20$ nodes, the packet generation rate is set to $g = 1$ pck/s, whereas for scenarios with $N = 10$ nodes, the packet generation rate is set to $g = 2$ pck/s. 104

4.13 PER as a function of the aggregated offered traffic load Ng. Different topologies for scenarios with $N = 20$ nodes are considered. In all cases, the sparsity index is set to 0.87. 105

4.14 (a) Network transmission rate and (b) delay as functions of the aggregated offered traffic load Ng. Various topologies with $N = 20$ nodes are considered. The sparsity index is set at 0.87. 106

4.15 PER as a function of the aggregated offered traffic load Ng. Different transmit power allocation strategies are considered for the topology with $N = 20$ nodes presented in Figure 4.6 (b): (i) fixed per-node transmit power (unif) and (ii) optimized transmit power (opt). Both simulation (solid lines) and analytical (dashed lines) results are presented. The dotted lines refer to scenarios where no optimized power allocation strategy is used. 107

4.16 P_{er} as a function of the aggregated offered traffic load (Ng). Different power allocation configurations are considered for the topology with $N = 10$ nodes presented in Figure 4.7 (a). Both simulative (solid lines) and analytical (dashed lines) results are presented. The dotted lines refer to scenarios where no optimized power allocation strategy is used. 109

4.17 P_{er} comparison between the proposed power allocation scheme and the RSSI-based power allocation strategy. 110

4.18 P_{er} comparison between the proposed power allocation scheme and the uniform transmit power allocation scheme. 111

4.19 Residual energy per node. Both RSSI-based and adjacency-based transmit power allocation schemes are considered. The given PER is $P_{\text{er}} = 5 \cdot 10^{-2}$. 112

4.20 (i) Per-node (solid and dotted lines) and (ii) average network (dashed with symbols) residual energies as functions of time. Both RSSI-based (dotted lines and dashed with squares) and adjacency matrix-based (solid lines and dashed with circles) transmit power allocation schemes are considered. The target PER is $P_{\text{er}} = 5 \cdot 10^{-2}$. 114

4.21 Residual energy with adjacency matrix-based transmission power allocation strategy proposed here, in a scenario with $N = 10$ nodes after (a) 0 days, (b) 45 days, and (c) 90 days. 115

4.22 Residual energy with RSSI-based power allocation strategy, in a scenario with $N = 10$ nodes after (a) 0 days, (b) 45 days, and (c) 90 days. 116

4.23 Performance evaluation in a scenario with $N = 64$ RFDs and various uniform clustering configurations: (a) network transmission rate and (b) delay as functions of χ_{net}; (c) delay as a function (parameterized with χ_{net}) of the network transmission rate. The packet generation rate is equal to $g = 2$ pck/s. 120

4.24 Performance evaluation in a scenario with $N = 64$ RFDs and various non-uniform clustering configurations: (a) network transmission rate and (b) delay as functions of χ_{net}; (c) delay as a function (parameterized with χ_{net}) of the network transmission rate. 122

4.25 Performance evaluation in a scenario with $N = 16$ RFDs and various *uniform* clustering configurations: (a) network transmission rate and (b) delay as functions of χ_{net}; (c) delay as a function (parameterized with χ_{net}) of the network transmission rate. The packet generation rate is equal to $g = 2$ pck/s. 124

4.26 Performance evaluation in a scenario with $N = 16$ RFDs and various *non-uniform* clustering configurations: (a) network transmission rate and (b) delay as functions of χ_{net}; (c) delay as a function (parameterized with χ_{net}) of the network transmission rate. The packet generation rate is equal to $g = 2$ pck/s. 125

4.27 Performance evaluation in a scenario with $N = 16$ RFDs and various *uniform* clustering configurations: (a) network transmission rate and (b) delay as functions of χ_{net}; (c) delay as a function (parameterized with χ_{net}) of the network transmission rate. The packet generation rate is equal to $g = 10$ pck/s. 127

4.28 Performance evaluation in a scenario with $N = 16$ RFDs and various *non-uniform* clustering configurations: (a) network transmission rate and (b) delay as functions of χ_{net}; (c) delay as a function (parameterized with χ_{net}) of the network transmission rate. The packet generation rate is equal to $g = 10$ pck/s. 128

4.29 Performance evaluation in a scenario with $N = 16$ RFDs and various *uniform* clustering configurations: (a) network transmission rate and (b) delay as functions of the packet generation rate. The network tolerable death level χ_{net} is fixed to 50%. 129

4.30 Performance evaluation in a scenario with $N = 16$ RFDs and various *non-uniform* clustering configurations: (a) network transmission rate and (b) delay as functions of the packet generation rate. The network tolerable death level χ_{net} is fixed to 50%. 131

4.31 Performance analysis in a scenario without clustering: probability of decision error performance as a function of the SNR at the RFDs. Both simulations (solid lines) and theoretical (dashed lines) results are shown. 132

4.32 Performance analysis in a scenario without clustering: (a) probability of decision error at the AP as a function of the SNR at the RFDs, and (b) throughput and delay as functions of the number N of transmitting RFDs. 133

4.33 Probability of decision error in scenarios with $N = 5$ and $N = 30$ RFDs, respectively, in the absence of clustering. Both the presence (solid lines) and the absence (dotted lines) of ACK messages are considered. 136

4.34 Probability of decision error at the AP in scenarios with $N = 16$ RFDs both in uniform and non-uniform clustering configurations. The considered topologies are the following: (i) no clustering, (ii) 2 8-RFD clusters and 2 FCs, (iii) 4 4-RFD clusters and 4 FCs, (iv) 14-1-1 clustering configuration with 3 FCs, (v) 10-2-2-2 clustering configuration with 4 FCs, and (vi) 8-2-2-2-2 clustering configuration with 5 FCs. 137

4.35 Performance evaluation in a scenario with $N = 16$ RFDs and various uniform and non-uniform clustering configurations; (a) delay as a function of network transmission rate and χ_{net}, (b) network transmission rate as a function of delay and χ_{net}, and (c) χ_{net} as a function of network transmission rate and delay. The considered packet generation rate is $g = 2$ pck/s. 138

4.36 Performance evaluation in a scenario with $N = 64$ RFDs and various uniform and non-uniform clustering configurations; (a) delay as a function of network transmission rate and χ_{net}, (b) network transmission rate as a function of delay and χ_{net}, and (c) χ_{net} as a function of network transmission rate and delay. The considered packet generation rate is $g = 2$ pck/s. 140

4.37 Performance evaluation in a scenario with (a) $N = 16$ RFDs and (b) $N = 64$ RFDs. Both the realistic surface (transparent) and its approximation (filled) are shown. 141

5.1 Network scenario with $N = 14$ nodes, out of which and $N_{\text{min}} = 9$ are active and $N - N_{\text{min}} = 5$ have been turned off by the deep sleep algorithm. 148

5.2 Logical scheme of the integrated network: (a) the Zigbee and RFID networks are combined together; (b) the integrated nodes and AP are obtained from the integration of Zigbee and RFID devices. 150

5.3 Initial steps of the deep sleep algorithm for selective wake-up of remote nodes: (A) N_{min} out of N RFDs are selected by the AP; (B) the selected RFDs are notified through the associated RFID tags; (C) the activated RFDs start communicating. 151

5.4 Steps of the replacement of an active node with residual energy under threshold and a sleeping node with higher residual energy: (A) the AP identifies an RFD with low residual energy; (B) a sleeping node with high residual energy is turned on and (C) the selected active RFD with low residual energy is turned off; (D) the set of active nodes communicate. 152
5.5 Network scenario with $N = 27$ nodes, out of which and $N_{\min} = 19$ are active and $N - N_{\min} = 8$ have been turned off by the deep sleep algorithm. The monitored surface is not divided into cells.. 153
5.6 Network scenario with $N = 27$ nodes, out of which and $N_{\min} = 19$ are active and $N - N_{\min} = 8$ have been turned off by the deep sleep algorithm. The monitored surface is divided into cells and only one node per cell is active, according to the spatial grid procedure. . . . 155
5.7 Average residual energy as a function of the days of simulation, in scenario with $N = 40$ nodes. Various values (25, 30, 40) of $N_{\min}$ are considered for the application of the deep sleep algorithm. In the case with $N_{\min} < N$, various values of the energy step are used. 158
5.8 Average residual energy, as a function of $N_{\min}$, considering various values for the number of days of simulation. In all cases, $N = 40$. . . 159
5.9 Average residual energy, as a function of the number of days of simulation, for various packet generation distributions: (i) constant and (ii) Poisson. In all cases, $N = 40$, whereas $N_{\min}$ is set to either 25 or 40.160
5.10 Average cells coverage with and without the virtual spatial grid. In all cases, the deep sleep algorithm is used ($N_{\min} = 25$) and $N = 40$. . 161
5.11 Average energy consumption with energy step equal to 1% of the initial energy level. The number of nodes is $N = 40$. The deep sleep algorithm is applied considering $N_{\min} = 20$, $N_{\min} = 25$, respectively. The performance with the virtual spatial grid is also shown. 162

List of Tables

2.1 Channel frequencies of the Zigbee standard. 14

3.1 Current consumption in each state for a generic CC2420 radio module. 66
3.2 Computation times for networks of different sizes. Results obtained with Mosek 5 (64-bit version) with a Core 2 Duo CPU at 3.16 GHz and with 4 GB RAM. 69

4.1 Parameters of the Zigbee standard. 98
4.2 Sparsity indexes of the adjacency matrix for the scenario presented in Figure 4.6 (b). Different values of overall network transmit power are considered. 108
4.3 Sparsity indexes of the adjacency matrix for the scenario presented in Figure 4.7 (a). Different values of overall network transmit power are considered. 110
4.4 Aggregate throughput, in a scenario without clustering, as a function of the number N of transmitting RFDs. 135
4.5 Coefficients of the linear combination in (4.1) and corresponding normalized MMSEs in the cases with $N = 16$ and $N = 64$, respectively. 142
4.6 Coefficients used for the computation of the approximating function in the case with $N = 16$ and $N = 64$, respectively. 144

Introduction

During recent years, we have witnessed a large diffusion of wireless devices with sensing capabilities, especially due to the improvements in manufacturing techniques, which have lead to the production of cost-effective and energy-efficients components. For these reasons, there is an increasing interest for both civilian and military applications based on wireless sensing devices. Even if the first monitoring applications were related to surveillance in military applications, nowadays a large number of applications concern industrial or environmental monitoring in civilian scenarios.

Usually, wireless sensor networks (WSNs) are supposed to be deployed in harsh environments, where the human intervention of external operators is difficult or, in extreme cases, impossible. It is therefore of paramount importance that a WSN can operate properly for a sufficiently long period of time, e.g., on the order of months. On the other side, the maximization of the lifetime should not affect largely the network performance, which should remain above a minimum required Quality of Service (QoS), depending on the specific application at hand. In order to properly deal with this unavoidable trade-off, an optimization strategy must be adopted.

In this thesis, we investigate some strategies for the configuration of WSNs, all based on the optimal tuning of key node parameters. After an accurate literature survey (Chapter 1), this thesis will be structured in the following way.

- Chapter 2 is dedicated to the description of the Zigbee standard, which is based for the first two layers of the ISO/OSI stack on the IEEE 802.15.4 standard. We first introduce an overall view of the standard, followed by a description of the characteristics of the physical layer of the ISO/OSI stack. In particular, the

main emphasis is on the Medium Access Control (MAC) layer, which regulates the policy for the accesses to the channel. Finally, we quickly provide some details on the network and application layers of the Zigbee standard.

- Chapter 3 introduces some analytical frameworks for the derivation of optimization strategies. Three main scenarios are considered: (i) target detection in the presence of random node deployment and duty cycles of sensing and communication interfaces; (ii) power optimization in star/tree networks for the minimization of the probability of error at the Access Point (AP); (iii) data fusion and QoS evaluation in the presence and in the absence of clustering.

- Chapter 4 focuses on performance evaluation of the three considered scenarios. In this case, for each scenario we provide insights for possible optimization strategies which allow to maximize the required performance indicators and/or satisfy the given performance constraints.

- Chapter 5 is dedicated to the integration of the Zigbee and RFID technologies. We present a possible approach in which the network in charge of exchanging data (formed by Zigbee devices) is controlled by a logical controlling network (formed by RFID devices). The latter selects which Zigbee nodes must be activated in the former network and which can remain in sleep mode in order to save energy.

In all cases, we shed light on the aspects which characterize the energy consumption of WSNs and, consequently, their lifetime. Moreover, the presented simulation and experimental results tend to validate analytical frameworks and, therefore, make the proposed optimization strategies reliable. Finally, Chapter 6 is dedicated to concluding remarks and future works.

Chapter 1

Literature Analysis and Motivations

1.1 Introduction

Wireless sensor networks (WSNs) are an interesting research topic, both in military [1–3] and civilian scenarios [4]. In particular, remote/environmental monitoring [5], surveillance of reserved areas, etc., are important fields of application of WSNs. These applications often require very low power consumption and low-cost hardware [6]. One of the most common standards for wireless networking with low transmission rate and high energy efficiency has been proposed by the Zigbee Alliance [7]. The increasing interest for sensor networks has, therefore, pushed a significant research activity on the design of efficient fault-tolerant networks with the longest possible lifetime [8]. In particular, since the WSNs are generically used in harsh environments, optimization strategies are getting more and more importance in order to extend the lifetime of the WSNs and, at the same time, increase the other performance indicators, such as network transmission rate, delay, and minimize the probability of transmission errors.

This chapter is structured as follows. In Section 1.2, we introduce the problem of incoming target detection in a given monitored surface. In Section 1.3, we present

the possible power allocation strategies present in the literature. In Section 1.4, we focus on the problems of clustered networks, wither in the presence or in the absence of data fusion at the intermediate nodes. Finally, in Section 1.5, we review the literature concerning optimization strategies which aim at the minimization of the power consumption of IEEE 802.15.4 networks. In addition, we also provide some details for the RFID technology and for proposed integration strategies between Zigbee and RFID devices.

1.2 Detection of Incoming Targets

WSNs are formed by battery-powered devices commonly used for environmental monitoring, military surveillance, and industrial automation. These devices are typically composed of an embedded microcontroller with some memory, a radio transceiver, physical transducers that sense the environment, and a battery. Recent advances in hardware miniaturization, low-power radio communications, and battery lifetime, together with the increasing affordability of such devices, are paving the road for a widespread usage of WSNs in a vast array of applications.

In particular, WSNs are expected to create a major shift in future public safety and military surveillance systems. When integrated to heterogeneous surveillance systems in complement to traditional high-power and bulky observation sub-systems (e.g., mounted optronic systems and radars), hundreds of tiny sensor nodes can help secure and protect people and assets in remote or inaccessible areas. Through the use of embedded transducers such as acoustic, seismic or infrared sensors, they can perform local or collaborative target signature detection and classification as well as trigger actuators (e.g., flash lights, sirens). These nodes can be easily deployed and recovered, are lightweight, and provide cost-effective complements to existing surveillance systems. A WSN can be exposed as a sub-system to the rest of the information system through gateway nodes which can offer backhaul connectivity and have more capabilities in terms of storage and processing. Most of the sensors are inherently resource-constrained because of their size and cost, which result in limits in terms of energy, computational speed, storage capacity, and communication band-

width.

In the literature, a few papers address in detail the problem of target detection and decision reporting. In [9], the authors present the design and the implementation of a monitoring system, referred to as VigilNet, based on a WSN. The authors derive an energy-efficient adaptable surveillance strategy and validate it through experimental tests. In [10], under the assumptions that the road network map is known and the target movement is confined into roads, the authors describe an algorithm, referred to as Virtual Scanning Algorithm, which ensures that the incoming target will be detected before reaching a given protection point. However, the above approaches do not provide a global analytical framework for the optimal tuning of system parameters, such as sensing and communication duty cycles. In [11], the authors derive an analytical model for the probability of detection and the average detection delay under the assumption that the deployment of sensors guarantees a complete coverage of the monitored area. In their models, the authors take into account tunable parameters of the system such as node density and overall duty cycle. In [12], the authors formulate the target detection problem as a line-set intersection problem and use integral geometry to analytically characterize the probability of target detection for both stochastic and deterministic deployments. Compared to [11], they analyze WSNs where sensors have heterogeneous sensing capabilities.

1.3 Power Allocation Strategies

An interesting research direction for WSNs is the design of network architectures that can guarantee high energy efficiency. In particular, since the overall energy available in a WSN is typically limited (all nodes are battery-equipped), the research community has focused on the derivation of transmit power allocation strategies that maximize a specific performance indicator yet still guarantee high energy savings.

In [13], the authors compare three power control schemes by analyzing the received signal-to-noise ratio in dense relay networks. In particular, one of these opportunistic schemes aims at extending the lifetime of the relays, in order to maximize the lifetime of the entire network. In [14], the authors introduce a power allocation

scheme that minimizes the estimation mean-square error at the fusion center of a network where sensors transmit to the fusion center over noisy wireless links. In [15], the authors jointly optimize the data source quantization at each sensor, the routing scheme and the power control strategy in a WSN in order to derive an efficient solution for the problem of overall network optimization. Finally, in [16] the authors present an opportunistic power allocation strategy based on local and decentralized estimation of the links' quality. In this scenario, only the nodes that experience channel conditions above a specific quality threshold are allowed to transmit in order to avoid waste of energy. In [17], the authors introduce a dynamic power allocation scheme for WSNs which relates the received signal strength indicator (RSSI) to the received signal-to-interference plus noise ratio (SINR). In particular, they propose two possible approaches: (i) a first approach based on a Markov chain system characterization and (ii) a second approach based on the minimization of the average packet error rate (PER).

1.4 Data Fusion and Clustered Networks

In several applicative scenarios, the nodes are not able to communicate directly with the AP. In such a case, the sensors are grouped in small subsets of nodes, referred to as clusters, depending on how they are placed and the environmental characteristics (some sensors might not communicate directly with the AP) or in order to reduce their transmission range (and, consequently, to save battery energy). Several approaches have been proposed in order to evaluate the performance of clustered WSNs. In [18], the authors present a system level design methodology for clustered wireless sensor networks based on a semi-random communication protocol. In [19], the authors analyze the impact of energy consumption in wireless sensor networks and provide a model for describing the energy consumption behavior and derive the optimal transmission range. In [20], the authors investigate how the energy efficiency of a clustered wireless sensor network is affected by the transmit power distribution, the total number of sensors in a cluster, the required end-to-end packet error rate, and the relative magnitudes of intra-cluster and inter-cluster distances.

Another important research topic in the field of wireless sensor networking is the embedded data decision strategy, often involving *data fusion*. In [21], the authors present theoretical results on the optimal decision rule and its application to data fusion. In [22], the authors follow a Bayesian approach for the minimization of the probability of decision error at the access point (AP). The data fusion mechanism has also a strong impact on practical applications. In [23], the authors analyze several methods of multi-sensor data fusion, such as Bayesian estimation, Kalman filtering, and Dempster-Shafer evidence theoretical methods, in order to design a move-in-mud robot. In [24], the impact of source-destination placement and communication network density on the energy costs and delay associated with data aggregation are evaluated.

The problem of extending the sensor network lifetime has been studied extensively. In particular, the derivation of upper bounds for the sensor network lifetime has been exploited. In [25–33], various analyses are carried out according to the particular sensor network architecture and the definition of sensor network lifetime. In [34], a simple formula, independent of these parameters, is provided for the computation of the sensor network lifetime and a medium access control (MAC) protocol is proposed to maximize the sensor network lifetime. In [35], a distributed MAC protocol is designed in order to maximize the network lifetime. In [36], network lifetime maximization is considered as the main criterion for the design of sensor networks with data gathering. In [37], the authors consider a realistic sensor network with nodes equipped with TinyOS, an event-based operating system for networked sensor motes. In this scenario, the network lifetime is evaluated as a function of the average distance of the sensors from the central data collector. In [38], an analytical framework, based on the Chen-Stein method of Poisson approximation, is proposed in order to find the critical time at which isolated nodes, i.e., nodes without neighbors in the network, begin to appear, due to the deaths of other nodes. Although this method is derived for generic networks where nodes are randomly deployed and can die in a random manner, this can also be applied to sensor networks. Finally, an important area of application of wireless sensor networking is the medical field. In [39], an analysis of network lifetime using IEEE 802.15.4 sensor networks [40] is derived for this kind

of applications.

1.5 Integrated Networks

One of the newest standards for wireless networking with low transmission rate and high energy efficiency has been proposed by the Zigbee Alliance [7, 41]. An experimental analysis of Zigbee WSNs, taking into account the impact of the most important system parameters (e.g., the Received Signal Strength Indication (RSSI), throughput, network transmission rate, and delay) is presented in [41–43]. One of the most interesting research directions for WSNs is the design of network architectures with high energy efficiency. In [44], the authors analyze different approaches and possible optimization strategies in order to reduce the power consumption of IEEE 802.15.4 networks. In [45], instead, a mechanism for shutting down the radio frequency interface of wireless sensors, in order to reduce power consumption, is presented.

On the other side, Radio Frequency IDentification (RFID) devices are also receiving more and more attention, by both industrial and scientific communities. In particular, they can be used for luggage identification in airports, biological materials identification in hospitals, monitoring of post parcels, tracking of livestocks, efficient monitoring of objects in supply chains, etc. [46–49]. One of the newest areas of interest for the RFID technology is pervasive computing, typically carried out integrating different technologies, such as RFID devices and WSNs [50]. In [51, 52], the authors propose and evaluate three different system architectures in order to combine WSNs with RFID systems.

Chapter 2

The Zigbee Standard

2.1 Introduction

In this chapter, we introduce the Zigbee standard which has been largely adopted in both the analytical and the simulation frameworks. The increasing requirements of monitoring applications, in both civil and military scenarios, yielded, in 2004, to the creation of a standard able to guarantee wireless data communication together with high energy efficiency. The Zigbee standard is based, for the first two layers of the ISO/OSI stack, on the IEEE 802.15.4 standard, which is expressively designed for low power consumption, at the price of a reduced transmission rate of $R = 250$ kbps. The remaining levels of the ISO/OSI stack, instead, are designed by the Zigbee consortium. More precisely, two versions of the IEEE 802.15.4 standard have been released: the first in 2003 and the second in 2006. The latter introduces supplementary specifications, especially for the PHYsical (PHY) and *Medium Access Control* (MAC). However, since the IEEE 802.15.4 implementation included in the Opnet simulator and used in the larger part of this work refers to the 2003 model, in the following of this chapter we will present the 2003 version of this standard.

The structure of the chapter is the following: in Section 2.2 is a functional description of the Zigbee standard, whereas in Section 2.3 we introduce the physical layer of the Zigbee standard; in Section 2.4 we present the MAC layer of the ISO/OSI

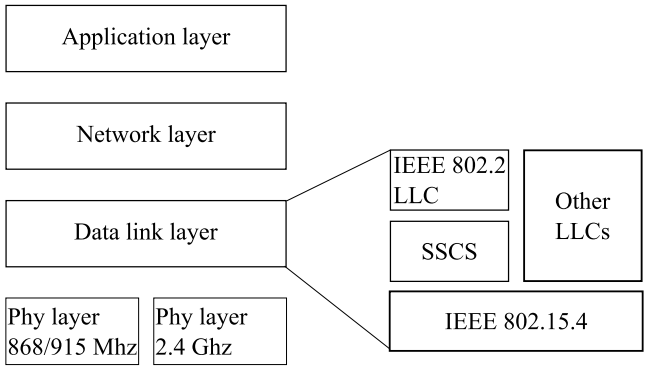


Figure 2.1: ISO/OSI stack of the Zigbee standard.

stack; in Section 2.5 we describe the network layer, whereas in Section 2.6 we shortly present the application layer of the Zigbee standard. Finally, Section 2.7 concludes this chapter.

2.2 Functional Description of the Zigbee Standard

The logical subdivision of the layers in the Zigbee protocol, shown in Figure 2.1, is perfectly compliant with that of the ISO/OSI stack [53]. Since the Zigbee standard is based on the IEEE 802.15.4, the following description will hold for both the two standards.

The data link layer is composed by some subcomponents such as: (i) a MAC sublayer, (ii) a Logical Link Control (LLC) sublayer, and (iii) a Service-Specific Convergence Sublayer (SSCS). The roles of LLC sublayer are flow control, data retransmission in case of collisions, and data aggregation and disaggregation of upper layers packets. The adopted LLC, described by the IEEE 802.2 standard [54], provides for an unreliable service. The allowed communication paradigms between LLC entities are “one-to-one,” “one-to-many,” and “one-to-all.” This standard is not connected oriented, therefore the arrival sequence of the packets may be different from the transmission one. In addition, the LLC does not provide for error and flow control.

The SSCS plays an intermediate role between the IEEE 802.15.4 and the LLC

sublayers and allows to separate them, making them independent one from each other. The use of a SSCS is not mandatory in the Zigbee standard.

The main functionalities of the IEEE 802.15.4 sublayer, which will be presented in details in Section 2.4, are the following [40]:

- channel access management, through the Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) technique;
- reliable communication between two distinct MAC entities;
- security management through a 128-bit symmetrical Advanced Encryption Standard (AES) key [55];
- generation of synchronization packets, referred to as beacons;
- synchronization between the nodes through the beacon messages;
- support of association and disassociation of nodes from a Personal Area Network (PAN);
- management of Guaranteed Time Slot (GTS) management.

The nodes can operate in either beacon or beaconless modality. Of course these modalities are mutually exclusive.

In beacon mode, the use of specific beacon packets allows to synchronize the nodes in the network. The time is discretized in intervals referred to as superframes, which are, in their turns, divided into shorter time intervals. This fact introduces a higher complexity of the devices but, on the other side, allows a better duty cycle and channel access management, including the priority reservation of GTS. On the contrary, the use of a beaconless modality allows to simplify the structure of nodes, since there is no synchronization between the nodes and the superframe structure is not used. However, the performance of beaconless networks is lower than that of beacon enabled networks. Obviously, the MAC protocol used by the beacon and beaconless networks are different, especially in the way they manage the timing, the eventual synchronization, and the retransmission of packets.

The IEEE 802.15.4 devices are grouped in PAN networks, which differ one from each other for their identification number (PAN ID). Each PAN is composed by at least one node, which has the role of the PAN coordinator. In addition to this type of node, there exist nodes with other roles in the network: coordinators and end devices. The PAN coordinator has additional tasks than standard coordinators, such as network initialization through the determination of the configuration parameters, external nodes association, and periodical beacon transmission. The coordinators are not allowed to set up the PAN ID and to configure the network parameters. The end devices, instead, do not have functionalities. They are only responsible for controlling eventual sensors/actuators connected with them. The devices can logically divided into two main classes: Remote Function Devices (RFDs) and Full Function Devices (FFDs). The difference among these classes is that the FFDs implement all the functionalities provided by the Zigbee standard, whereas the RFDs implement only a subset of them. In addition, the FFDs are able to communicate with all the other nodes in the network, if in the transmission range, whereas the RFDs are able to communicate only with a FFD.

Focusing on the network topology, the MAC layer provides for two different network configurations: (i) star topology and (ii) peer-to-peer topology. In the former case all the nodes transmit directly to the PAN coordinator, which is responsible for eventually forwarding the packets to the destination. This configuration is suitable for small networks where no relay of packets is required. In the latter case, instead, if the nodes are FFDs, they allowed to communicate directly one with each other if in the transmission range. On the other hand, the RFDs are required to communicate with an associated FFD. In the peer-to-peer mode, it is possible to create networks with a complex structure. The layer which is in charge of managing the network and the routing of the packets is the NetWorK (NWK) layer.

The NWK layer provides for the multihop functionalities but is not responsible for inter-PAN communications. Since it differs at logical level from the MAC level, a different node nomenclature is adopted. In particular, three classes of nodes are considered:

- Zigbee router: it is an optional node in the network, it acts as a coordinator at

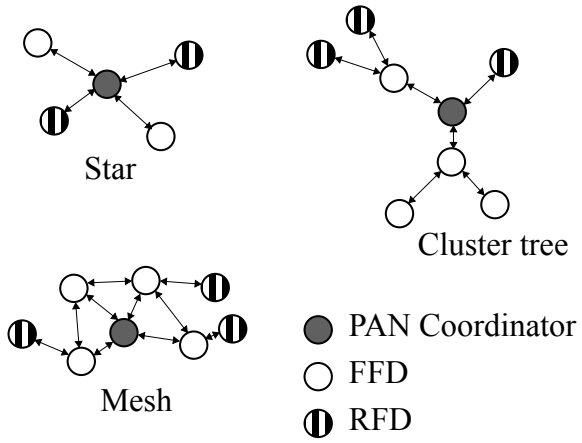


Figure 2.2: Network topologies allowed by the Zigbee standard.

MAC layer, it supports the association of other nodes in the network and it is able to relay the packets in the network according to a routing protocol;

- Zigbee coordinator: it acts as the PAN coordinator at MAC layer, it has the same functionalities of Zigbee routers but it is also able to set up the network by configuring the network parameters;
- Zigbee End Device: it corresponds to the End Device at MAC layer. it has neither routing nor association management capabilities.

A group of Zigbee nodes associated with and connected, either directly or indirectly, to a Zigbee coordinator forms a Zigbee network. These nodes correspond to those which form the underlying PAN network. The NWK layer defines three routing strategies: (i) neighborhood routing, (ii) tree routing, and (iii) ad-hoc routing, as will be presented in Section 2.5. According to these routing strategies, it is possible to define three networks topologies, based on the two presented for the MAC layer and shown in Figure 2.2: (i) star, (ii) cluster tree. and (iii) mesh topology.

The star topology corresponds exactly to the star topology defined at the MAC layer, since only direct communications from remote nodes to the Zigbee coordinator

Table 2.1: Channel frequencies of the Zigbee standard.

Channel number	Frequency (MHz)
$k = 0$	868.3
$k = 1, \ldots, 10$	$906 + 2(k-1)$
$k = 11, \ldots, 26$	$2405 + 5(k-11)$

are allowed. The other two topologies, instead, rely on the peer-to-peer topology at MAC layer and support multi-hop communications. The mesh topologies uses the ad-hoc routing strategy based on dynamic routing tables and does not support the beacon modality. On the other side, the cluster tree topology does not use routing table since packets are relayed according to tree routing.

2.3 Physical Layer of the Zigbee Standard

The 2003 version of the IEEE 802.15.4 standard defines three operating frequencies: the first is around 2.4 GHz, the second is around 916 MHz, and the third is around 868 MHz. In particular, the frequencies around 868 MHz are used in the EU, those centered at 916 MHz are used in the U.S.A., whereas those around 2.4 GHz are worldwide available. The frequency allocation of the 27 channels available for the Zigbee standard are presented in Table 2.1.

The transmission rates at physical level are 20 and 40 kbps for the frequency bands at 868 and 916 MHz, respectively, whereas at 2.4 GHz the transmission rate is 250 kbps. These values are low if compared with the performance obtained by other IEEE 802.1x standards. In particular, it is possible to range from 1 Mbps for the Bluetooth and Wibree standards to 250 Mbps for the IEEE 802.11n standard. For sake of completeness, some modulation changes in the 2006 version of the IEEE 802.15.4 protocol allow the data rate of each frequency band to reach the value of 250 kbps.

All the frequency bands used by the Zigbee standard are unlicensed, that is they

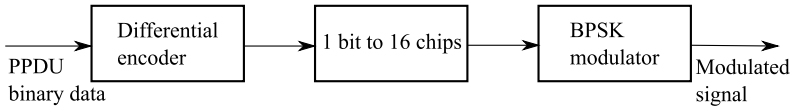


Figure 2.3: Scheme of the modulator for the Zigbee standard at the frequencies of 868 and 916 MHz.

don't require Government authorization for their exploitation. Neglecting all the benefits that this choice can bring, the negative aspect is that Zigbee network may suffer from co-channel interference from other devices operating at the same frequency. In particular, in addition to Bluetooth, IEEE 802.11b, and IEEE 802.11g, several electronic devices exploit the 2.4 GHz band, making critical the coexistence of devices operating in this interval of the spectrum [56]. According to simulation data provided in [40], the most critical coexistence is between the IEEE 802.15.4 and the IEEE 802.11b standards; a transmission of an IEEE 802.11b device, with a frequency shift of 3 MHz and at a distance of 80 m from an IEEE 802.15.4 device, is able to prevent the transmission of the latter (the impact of a shift of 47 MHz is perceptible only at a few meters). On the contrary, the impact of IEEE 802.15.4 on IEEE 802.11b transmissions is scaled by a factor of 4. Concluding, Zigbee and IEEE 802.11x standards can coexist in the same frequency band only if the IEEE 802.11x standards have a low utilization factor. Otherwise, the two standards must operate in frequency bands sufficiently separated (at least 7 MHz) one from the other, in order to reduce the impact of the interference and guarantee the coexistence. The situation is better concerning the Bluetooth standard, because the mutual interference has no impact when the distance between two devices is larger then 10 m.

The modulation scheme adopted by the Zigbee standard depends on the chosen frequency band. Concerning the bands at 868 and 916 MHz, the logical scheme of the modulator is shown in Figure 2.3. The modulator can be split in three separated blocks. The first block makes a differential encoding of the information bit, the second makes a spectral expansion by associating to each bit, generated from the previous block, a 16-chips pseudo-random (PN) sequence, in order to realize a direct sequence spread spectrum (DSSS) modulation [57]. Finally, the bit generated by

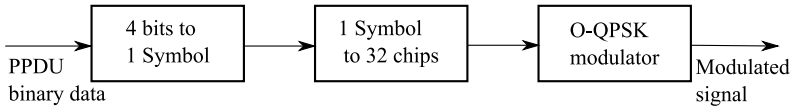


Figure 2.4: Scheme of the modulator for the Zigbee standard at the frequency of 2.4 GHz.

the second block are modulated according to a Binary Phase Shift Keying (BPSK) modulation and sent to the transmitter. The main difference between the 868 and 916 MHz frequency bands is the chip rate which is 20 and 40 Kchips/s, respectively.

Similarly to the modulator for the bands at 868 and 916 MHz, also the modulator for the 2.4 GHz, shown in Figure 2.4, can be split into three blocks. However, in this case in the first block 4 information bits are encoded with one of the 16 available symbols with whom, in the second block, one of the 16 32-chips PN expansion sequence is associated, in order to realize the DSSS expansion. Finally, in the third block an Offset-Quadrature Phase Shift Keying (O-QPSK) modulation is applied in order to transmit the signal. On the in-phase component of the modulator the odd indexes of the bits of the expansion sequence are used, whereas on the quadrature-phase the even indexes of the bits of the expansion sequence are used. Due to the presence of the offset, this modulation can be seen as a Minimum Shift Keying (MSK) with half-sine shaping pulse. According to the chosen solution, it is possible to reach lower Bit Error Rate (BER) compared with the other IEEE 802.1x standards. Some works have proved that under the assumption of low signal-to-noise (SNR) ratio, the IEEE 802.15.4 modulations have better performance than that of IEEE 802.11 and IEEE 802.15.1 [58].

From a protocol point of view, the PHY layer defines a Physical Protocol Data Unit (PPDU) which is formed by the following fields:

- Synchronization Header (SHR): it has a fixed dimension of 5 bytes and it allows the receiver the synchronization with incoming bit stream;
- Physical Header (PHR): it has a fixed dimension of 1 byte and it contains the information about the frame length;

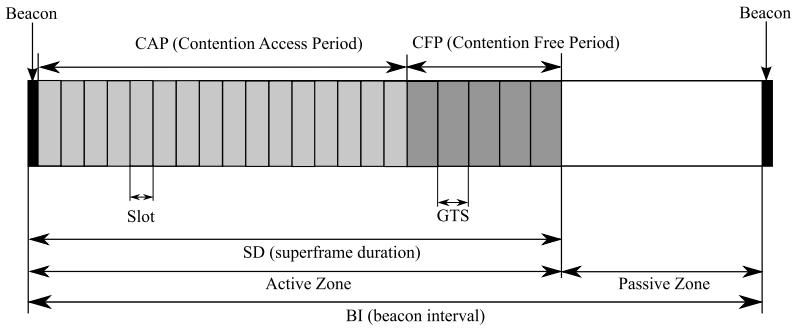


Figure 2.5: Structure of a superframe.

- Physical Service Data Unit (PSDU): since it contains the payload carried by the packet, it has a variable length. The maximum length allowed by the Zigbee standard is 127 byte, therefore this is the maximum length also for the MPDU (MAC Protocol Data Unit).

2.4 Access and Medium Control Layer of Zigbee Standard

2.4.1 Structure of the Superframe

As introduced in Section 2.2, in the beacon modality, the time axis is discretized in time intervals referred to as superframes, which are set by the beacon messages. The structure of a superframe is shown in Figure 2.5. A superframe is made of a multiple of the length of a physical modulation symbol, denoted as $T_{\text{sym}} = 1.6\mu s$ for the frequency band around 2.4 GHz. The duration of the superframe, including the length of the beacon and denoted as Beacon Interval (*BI*), is decided by the coordinator and it is equal to

$$BI = d_{\text{SF}} \cdot 2^{BO}$$

where the constant *BO* (Beacon Order) ranges between 0 and 15. The case with $B0 = 15$ corresponds to beaconless modality; the meaning of the constant d_{SF} will be shortly introduced. The superframe is composed by an active zone, during which the communications between remote nodes and coordinator are allowed, and a pas-

sive zone, during which, instead, data communications are forbidden. This strategy allows the nodes to use a low-power consumption modality for a part of the superframe, reducing their duty cycle and, consequently, their energy consumption. The duration of the active zone, referred to as Superframe Duration (*SD*), is lower then *BI* and is defined as:

$$SD = d_{\mathrm{SF}} \cdot 2^{SO}$$

where the constant *SO* (Superframe Order) ranges in values between 0 and *BO*. Intuitively, when $SO = BO$, there is no passive zone. The active zone is divided in aNumSuperframeSlots slots, that is 16 in the standard configuration. Each slot is made of a given number of physical symbols (d_{sl}) which depends on the value of *SO*. In particular, we have that:

$$d_{\mathrm{sl}} = \mathrm{aBaseSlotDuration} \cdot 2^{SO}$$

where aBaseSlotDuration represents the length of a slot in the case of $SO = 0$. At this point it is possible to introduce the constant d_{SF}, which represents the number of symbols in the active zone when $SO = 0$ and which corresponds to the constant aBaseSuperframeDuration defined by the standard:

$$d_{\mathrm{SF}} = \mathrm{aBaseSuperframeDuration} = \mathrm{aNumSuperframeSlots} \cdot \mathrm{aBaseSlotDuration}.$$

Finally, the active zone is divided into two temporal intervals. In the former, referred to as Contention Access Period (CAP), the access to the channel is regulated by the CSMA/CA protocol; in the latter, instead, referred to as Contention Free Period (CFP), the access to the channel is deterministic, therefore there is no contention. The CFP is made of a given number of GTS, which may differ from one node to the other. Obviously, a deterministic protocol does not fully exploit the capacity of the channel but, on the other side, it allows to guarantee a required Quality of Service (QoS) at each node. In some applications, this requirement is more important than the optimal channel exploitation.

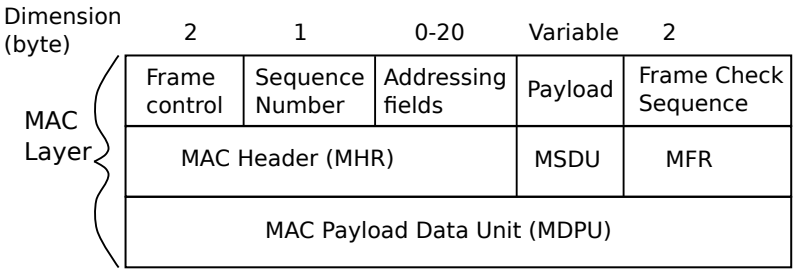


Figure 2.6: General structure of a MAC layer data packet.

2.4.2 Structure of the MAC Layer Packets

A generic MAC layer packet, in the following referred to as frame, is called MAC Protocol Data Unit (MPDU). As shown in Figure 2.6, the frame is composed by three fields: (i) MAC header (MHR), (ii) Service Data Unit (MSDU), and (iii) MAC footer (MFR). Depending on the type of MSDU, it is possible to identify 4 classes of frames:

- beacon, used by the PAN coordinators to send the homonym packets;
- data, used by the higher levels for data transfer;
- acknowledgment, used to confirm the correct reception of a frame;
- MAC Command, used to manage all the control information between the devices.

In the simplest case, that is in the case of a data frame, the MSDU is the Protocol Data Unit received by the higher level and corresponds to the payload of the transmitted message. The MFR is formed by a single field referred to as Frame Check Sequence (FCS), which contains a 16-bit Cyclic Redundancy Check (CRC), computed according to the MHR and MSDU fields. Finally, the MHR is formed by 3 fields:

- Frame control, which has a fixed size of 2 bytes, defines the frame typology, and contains information about the addressing type and other specific control flags.

- Sequence Number, which contains an 8-bit unambiguous frame identifier. Whether the frame is a beacon, this field contains the Beacon Sequence Number (BSN), randomly initialized by the coordinator and incremented at each beacon transmission. In all the other cases, this field contains the Data Sequence Number (DSN), which is randomly initialized by the source node and incremented at each new packet transmission. In the case of a retransmission, the DSN is not incremented. The ACK frame contains the same DSN of the packet which needs for reception confirmation.

- Addressing Fields, which is formed, at its turn, by four sub-fields, containing the source and destination nodes addresses and the relative PAN identifiers. The latter are optional and have a fixed dimension of 2 bytes. The former, instead, even if still optional, may have two different formats, that is a 16-bit short format and a 64-bit long format. Depending on the selected configuration, the Addressing Fields size may range between 0 and 20 bytes.

The maximum size of a MPDU corresponds to the maximum size of the Service Data Unit at the PHY layer, that is 127 bytes. Considering the presence if the MHR and FHR fields, the maximum size of the MSDU is related to the addressing type and ranges between 102 and 122 bytes. The ACK frame, instead, uses neither the Addressing Fields nor the MSDU and has a size of 5 bytes.

2.4.3 Data Transfer Modality

The IEEE 802.15.4 standard allows for several way to transmit a data frame. Basically, it is possible to classify these modalities in two families: (i) direct and (ii) indirect transmissions. In the former family, the source node transmits directly to the destination node without any preventive request. In the latter family, instead, the effective data frame transmission begins after that the source node has signaled the destination node of the incoming transmission.

Indirect transmissions are useful in two cases:

- in star topologies, in communications from the PAN coordinator to a generic node in the network;

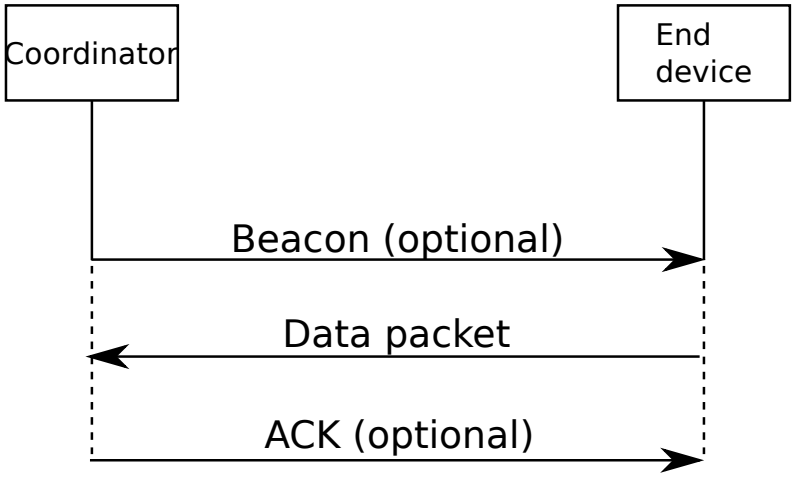


Figure 2.7: Data transfer from an end device to a PAN coordinator.

- in tree topologies (at the NWK layer), in communications from the parents to the children, when the children are end devices.

It is not possible to perform indirect transmissions during GTS periods. The modality of transmissions between to generic nodes, in a peer-to-peer network in beacon mode, has not been defined by the MAC layer, with the exception of the above mentioned tree topology. In all the other cases, direct transmissions are used, especially in communications to the PAN coordinator (uplink) and beaconless peer-to-peer data exchange.

In the case of direct data frame transmission, the frame exchange sequence, shown in Figure 2.7, is composed by the following operations:

- optional transmission of the beacon;
- medium access;
- data frame transmission;
- transmission of the optional ACK by the destination node;

- processing of the received packet [1]

In beaconless modality, a node can access the communication channel only through an unslotted CSMA/CA. On the other side, in beacon modality, there are two distinct methodologies to access the communication channel: (i) the slotted CSMA/CA mechanism inside the CAP and (ii) the use of a guaranteed access mechanism, such as the GTS, which may happen only during the CFP. In beacon modality, the transmission procedure begins only in the correspondence of the boundary between two consecutive slots, referred to as backoff boundary and after that the beacon which determines the beginning of the superframe has been received and correctly processed.

An additional constraint is due to the fact the whether the remaining time in the current superframe is not sufficient to complete the data exchange sequence, the transmission and the medium access procedure are delayed to the following superframe. Once that the medium access has been completed, the status of the radio interface (RF) of the device must be switched from reception to transmission in a time interval strictly lower than the constant aTurnAroundTime, referred to as T_{TAT} and set by the standard equal to 12 T_{sym}. Once concluded the operation, the data frame transmission may begin. In beacon modality, the transmission must begin in the correspondence of a backoff boundary, whereas in the beaconless modality it can begin as soon as the transmitting node has changed the status of its RF, that is after a T_{TAT}.

Once that the data frame transmission is correctly concluded, the receiving node transmits the ACK packet, whether required by the single packet. The use of ACK packets allows to maximize the reliability of the transmitted packets but, on the other side, it largely degrades the network performance, especially in terms of throughput, delay, and lifetime. However, while in classic data networks, the content of each data frame is important since it is normally unique, in physical phenomena monitoring applications, which are typical of WSNs, the content of a data frame is often not unique. Therefore, in such scenarios maximizing the reliability of a single packet may not necessarily lead to the maximizing the overall network reliability and it may also be counterproductive. For this reason, the use of ACK is totally application

[1] In transmissions without the ACK, this phase is immediately subsequent to the reception of the data frame and the ACK frame transmission does not require the medium access strategy.

dependent.

If the source node does not receive the ACK message in a given time interval, the packet is declared collided and retransmitted reinitializing the parameters of the CSMA/CA mechanism. This procedure can be repeated a maximum number of times, whose maximum is determined by the parameter MaxCSMARetry. When this value is exceeded, the packet is definitively discarded. The limit time interval for the reception of the first bit of an ACK frame is set by the constant AckWaitDuration, which can be defined as:

$$T_{\text{wait}}^{\max} = T_{\text{TAT}} + T_{\text{bu}} + T_{\text{ACK}}$$

dove T_{ACK} is the number of symbols necessary for the transmission of an ACK frame, whereas T_{bu} is the duration of a backoff unit of the CSMA/CA mechanism, as will be shown in Subsection 2.4.4. This duration is referred to as aUnitBackoffPeriod and it is equal to $20T_{\text{sym}}$. Since the dimension of an ACK frame is 11 bytes, in the case of the 2.4 GHz transmissions, $T_{\text{ACK}} = 22T_{\text{sym}}$. From this it can be derived that $T_{\text{wait}}^{\max} = 54T_{\text{sym}}$. In the case of transmissions without the acknowledgment, the source node cannot be detect if the packet is lost or correctly delivered, therefore each packet is sent only once.

Finally, the reception of an eventual ACK message or the end of the data frame reception is followed by a period referred to as InterFrame Space (IFS), which is necessary for the receiving node to process the exchanged frame. In addition, during the IFS the source node cannot transmit another packet to the same destination node, but it has to wait for the end of the IFS in order to begin the transmission of the following data frame. The value of the IFS depends on the length of the transmitted packet. For packets whose dimension is lower than the threshold value aMaxSIFSFrameSize = 18 bytes, the IFS is referred to as Short InterFrame Spacing (SIFS) and its duration must be larger than aMinSIFSPeriod $= 12T_{\text{sym}}$; vice versa for longer data frames the IFS, referred to as Long InterFrame Spacing (LIFS) and its duration must be larger than aMinSIFSPeriod $= 40T_{\text{sym}}$. In the following, we will denote the length of the IFS with T_{IFS}.

In the case of indirect data transfer, that is downlink communication, the message exchange scheme is slightly more complex, as shown in Figure 2.8, and is composed

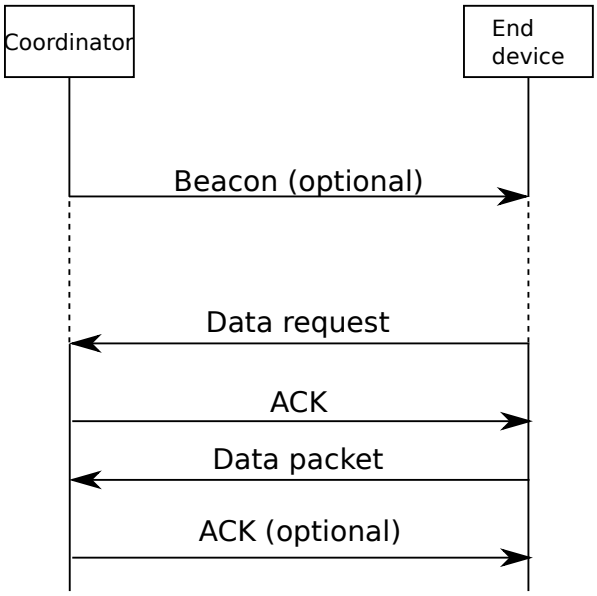


Figure 2.8: Data transfer from a PAN coordinator to an end device.

by the following steps:

- the destination node accesses the communication channel;
- the destination node sends a data request command frame to the PAN coordinator;
- the coordinator replies with an ACK frame;
- the coordinator sends the data frame;
- eventually, the destination node sends the ACK message;
- the received frame is processed.[2]

In beacon modality, a node become aware of having a pending data frame, which needs to be received, through the beacon packets sent by the coordinator at the beginning of each superframe. Therefore, a node will send the data request only when there

[2] In transmissions without ACK, this phase is immediately subsequent to data frame reception phase.

are data pending which are directed to that node. Vice versa, in beaconless networks, the nodes, which do no have information about pending packets, need to cyclically poll the coordinator through data request frames. In both cases, the destination node will need to obtain the access to the channel before sending the data request message. The coordinator, once received the data request, will reply with an ACK frame containing information about the eventual presence of pending packets addressed to the node which is sending the data request. If there are the conditions to complete the data exchange procedure in the current CAP, the coordinator will send the data frame immediately after the ACK without performing the medium access operations. This strategy is referred to as piggybacking. Vice versa, the coordinator will transmit the data frame in the following CAP according to the CSMA/CA strategy. After the reception of the data frame, the destination node will eventually reply with the ACK frame, if required, without performing the medium access strategy. Finally, an IFS time interval, necessary for data processing, follows the packet reception.

The remaining types of frame, that is those different from the data frame, are transmitted according to the following procedure:

- a command frame is transmitted in the way of a data frame; the only difference is that the command frame can be transmitted only during the CAP and not during the CFP;
- the beacon frames are transmitted outside both the CAP and CFP, require neither the confirmation through the ACK nor the use of CSMA/CA strategy and are followed by an IFS interval;
- the ACK frame are transmitted only after the reception of a data frame or a command frame. They are followed by an IFS time interval and are transmitted without performing the CSMA/CA strategy.

2.4.4 The CSMA/CA Protocol

In random medium access mechanism, like the CSMA, there is no authorities or controlling nodes which are able to determine which node has the right to access

the channel in a certain instant. In fact, the nodes manage the channel access in a distributed fashion. The peculiarity of the CSMA/CA is given by the fact that each node, before the channel access, sense the channel in order to determine whether the channel is busy or idle Given the result of the sensing operation, the behavior of the node is not univocally defined. This characteristic allows to classify the CSMA/CA-based protocols according to the concept of "persistence" which will be explained in the following. During the normal CSMA/CA operations, the time is discretized in time slots, referred to as backoff units, of constant duration equal to T_{bu}.

The parameters, which characterize the CSMA/CA protocol and are defined as multiple of a backoff unit, are the following:

- Number of Backoffs (NB): it is the number of backoff cycles elapsed for the transmission of the current packet.
- m: it corresponds to the macMaxCSMABackoffs and it is the maximum number of backoff cycles, referred to as backoff stage, allowed for the transmission of the current packet;
- Backoff Exponent (BE_i): it is the exponent used in the computation of the waiting time which elapses between two channel access attempts in the i-th cycle of backoff;
- BE_{min}: it corresponds to macMinBE and it is the minimum value to which BE can be set, that is the default value used in the initialization phase;
- BE_{max}: it corresponds to aMaxBE and it is the maximum value to which BE can be set;
- CW: it is the dimension of the contention window and it is equal to 1 slot in the unslotted case and 2 slots in the slotted case.

The differences between the slotted and the unslotted versions are not so considerable to justify a separate treatment, therefore in the following we will present only

the simplest version, that is the unslotted case, highlighting when required the differences between the two versions. In order to help the comprehension, in Figure 2.9, the flow charts of both the versions of the CSMA/CA protocol are presented

The operations of the unslotted CSMA/CA protocol can be presented schematically in the following way.

1. The source node selects a discrete random number uniformly distributed in the interval $[0, 2^{BE_i} - 1]$ and waits for a number of slots equal to the previously generated value. This waiting is elapsed also at the beginning of the backoff cycle, that is when the *NB* and *BE* are equal to the default values. This fact, which could be interpreted as an inefficiency, belongs to the collision avoidance characteristic of the IEEE 802.15.4 protocol and it is useful especially in periodical monitoring applications in which may happen that several nodes transmit at the same instant.

2. Once elapsed the waiting interval, the node verifies the status of the channel through a Clear Channel Assessment (CCA) operation, which allows to measure the energy level received by the RF interface over a time interval equal to $8T_{\text{sym}}$. If the result of this measurement is lower than the energy detection threshold, the channel is assumed idle and the CCA is successful. Vice versa the channel is assumed busy and the CCA fails.

3. If the channel is idle, a node can immediately transmit the packet which is stored in the transmission queue. Instead, if the CCA fails, it is necessary to compare *m* and *NB*. If $NB = m$, the CSMA/CA is declared failed and the packet is discarded; vice versa, the node increments of one the value of *BE*, given that *BE* is lower than the maximum allowed value BE_{max}, and the procedure is repeated from point 1. Incrementing *BE* allows to double the time interval in which it is possible to chose the number of slots to be waited for and, consequently, to double the average waiting time; for such a reason, this protocol is referred to as binary exponential backoff. This behavior is a peculiarity of the collision avoidance mechanism of the IEEE 802.15.4 protocol, since it decreases the probability that two nodes perform their following transmission

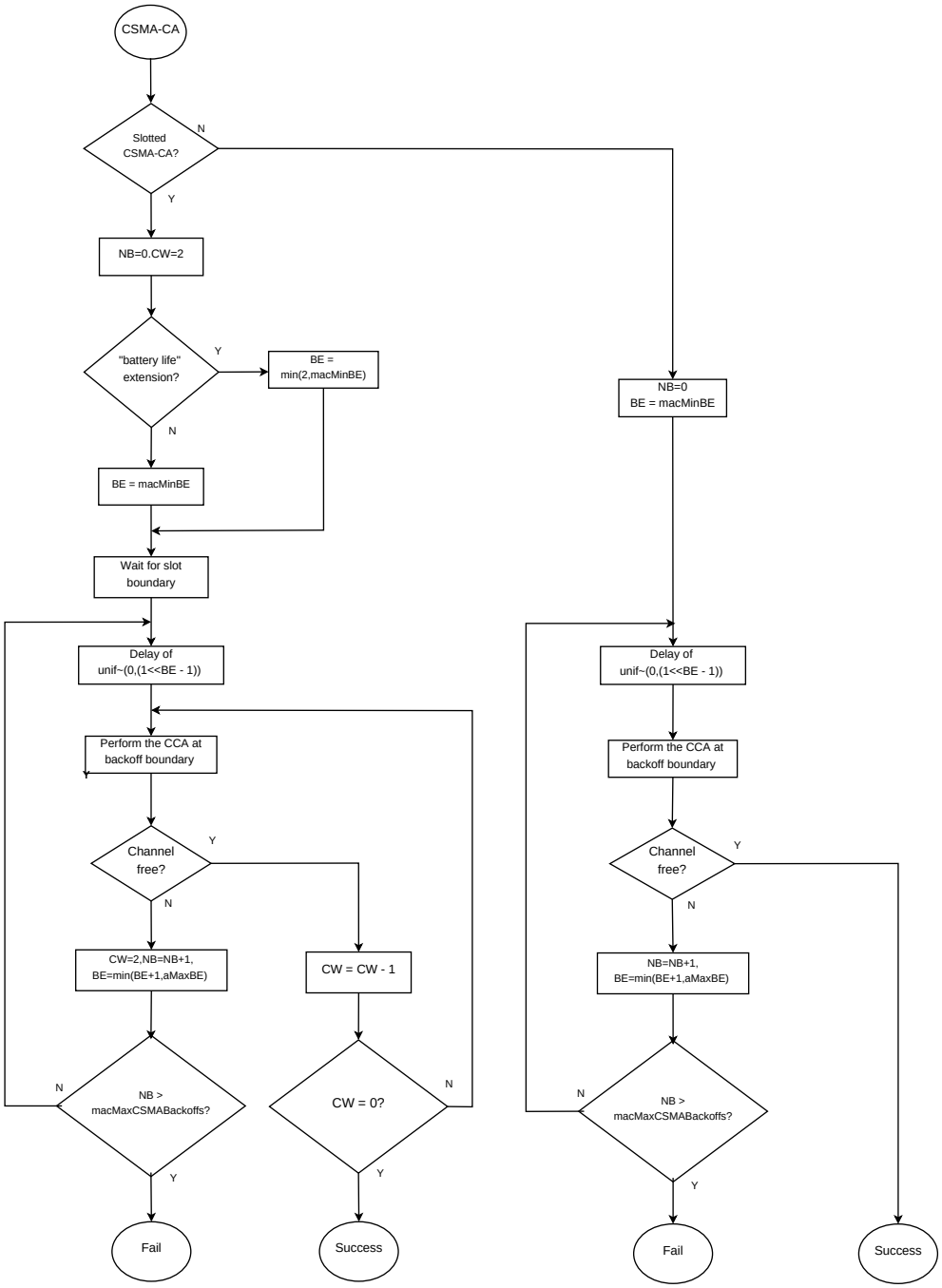


Figure 2.9: CSMA/CA mechanism in the IEEE 802.15.4 protocol.

attempts in the same slot.

The main difference between the slotted and the unslotted versions is that the all the CSMA/CA operations and the frame transmissions must be performed in the correspondence of a backoff boundary. Therefore, a node cannot begin the transmission as soon as a packet is put in the transmission queue, but it must wait for the following backoff unit or for the following beacon if still not synchronized. For the same reason, the ACK frame cannot be sent at the end of the transition of the status of the RF interface (radio switching), but it has to wait for the beginning of the following slot. Due to this behavior, depending on the transmission time of the current packet and the time elapsed to perform the radio switching, in same circumstances there can be an unused time slot between the reception of a packet and the transmission of the related ACK. Therefore, an eventual CCA operation performed in the same slot by another node would not be successful, determining a collision between the incoming packet and the ACK frame to be transmitted.

In order to solve this problem, the standard states that, in the slotted version, a node must perform two successful CCA operations in a row, that is the contention window has a duration equal to 2 slots ($CW = 2$), differently from the unslotted version in which the CW is equal to 1 slot. This expedient prevents from possible collisions but, on the other side, leads to a performance degradation due to the non optimal channel utilization. In particular, this phenomenon is really stressed in the presence of communications without the ACK. [45, 59]

Finally, in the slotted scenario it is foreseen a low energy consumption modality, which can be activated by setting the parameter macBattLifeExt. In this modality, independently of the duration of the active zone, the activity period of the coordinator in the CAP is reduced to a number of slot between 6 and 8. In this case, the initial value of BE is decreased, in order to increase the probability of transmitting during the active zone of the node supposed to receive a message from the coordinator.

Although the efforts to prevent this problem, frame collisions are possible due to several causes. First of all, when two or more nodes are not in radio visibility, the status of the channel inside the PAN is no longer the same for all the nodes, but groups of nodes see channel states slightly correlated or, at most, uncorrelated. The

fact that the nodes do not have a global view of the channel unhinges the basis of the CSMA/CA protocol and makes the channel sensing operations useless.

The classical example for this phenomenon is when two nodes, that is A and B, which are not in radio visibility, want to transmit to the same node C. In this case, the node B is not able to detect whether A is transmitting to the node C but, vice versa, is able to detect when the node C is transmitting to the node A. Consequently, it may happen that when the node A is transmitting to the node C, the node B starts transmitting to C, leading to a packet collision. This problem is known as hidden terminal problem and afflicts both the CSMA/CA versions of the IEEE 802.15.4 standard. There are some particular expedients which allow to overcome this problem, such as the use of RTS/CTS in the IEEE 802.11 standard or the use of a dedicated channel, as for the protocols based on the principle of the Busy Tone Multiple Access (BTMA). [60]

Also in an ideal scenario with perfect synchronization between the nodes, the absence of hidden terminal problem, idle channel and without propagation delay, the collisions are anyway possible. In fact, it is sufficient that two or more nodes perform the CCA operations at the same instant to incur in a packet collision. Also this problem affect both the versions of the CSMA/CA. The only difference is given that in the slotted case, the vulnerability interval is equal to the duration of a slot, whereas in the unslotted version the vulnerability interval corresponds to the duration of the CCA, that is 8 T_{sym}. In the slotted version the problem is worsened by the fact that the standard requires that all the transmissions, which cannot be completed in the current superframe, must be delayed to the following superframe. This behavior is problematic; in fact, as proved by [61], when two or more nodes are forced to delay the transmission in the same superframe, inevitably all the nodes involved will perform the CCA in the first useful slot, inducing a collision. In the above-mentioned ideal scenario, the contemporaneous CCA is the only cause of collision for beacon configuration. In the beaconless case, instead, the lower dimension of the vulnerability interval and the absence of following superframe transmission delay, reduce the probability of having contemporaneous CCA. On the other side, there exists another situation that may generate collisions: the standard defines that the maximum time in-

terval during which the device must perform the radio switching is equal to the T_{TAT} constant and is set to 12 T_{sym}, whereas the CCA operation has a duration of 8 T_{sym}. Therefore, if a node performs the CCA in the time interval between the reception of a packet and the transmission of the corresponding ACK frame, the node would see the channel idle and this would lead to a collision between the ACK and the data frame sent by the node.

2.5 Network Layer of Zigbee Standard

The NWK layer performs the following tasks:

- creation and management of a network;
- choice of the network topology;
- routing of the packet both in unicast (from one source to one destination) and in multicast modalities;
- security management;
- management of the logical associations between the applications resident on different nodes in the network.

The last point is referred to as binding by the Zigbee standard and will be presented in details in Section 2.6. Considering the categories of the nodes presented in Section 2.2, only the Zigbee coordinators and the routers have an active role in network maintenance, whereas the end devices do not perform any operation and, consequently, they will have lower requirements and a lower set of functionalities. The main difference is that a router and a coordinator must belong to the FFD class, whereas an end device can also belong to the RFD class. In addition, coordinator and router nodes can create routing tables, necessary for the correct ad-hoc packet routing provided by the Zigbee standard. On the other side, end devices do not have a routing table, therefore they can communicate only with the parent node to which they are connected. There is also another table, different from the routing table and referred to

as neighborhood table, which contains information about all the surrounding nodes which are in the transmission range of a given node [3].

2.5.1 Network Creation and Management

The Zigbee coordinator is the node which initializes the network setting the fundamental parameters, both and MAC and NWK layer and notifying the associated nodes of the selected parameters. In addition, the Zigbee coordinator is responsible for the periodical beacon transmission to associated nodes. The mentioned parameters concern the creation and the assignment of network addresses, the binding association, the choice of the network topology and the choice of the routing strategy used in the network.

The choice of which node will become the network coordinator either can be a-priori defined by opportunely programming the higher levels of the stack or can be random. The default Zigbee policy is that each FFD device, which is not able to associate itself to any existing network, starts the creation of a new network, setting itself as the network coordinator. Once that the network is working, the coordinator can receive association requests from other nodes, both router and end devices. If there are available addresses and the compatibility and security requirements are satisfied, the coordinator allows the applying node to access the network. The coordinator will become the “parent” of the associated node which, at its turn, will become the “child.” The children of the coordinator can receive association requests from other nodes in the network, becoming parents of other nodes. Iterating this process, it is possible to create very complex network topologies. We remark that end devices cannot become parents of other nodes, therefore they cannot accept incoming association requests.

As soon as the association process ends, the children receive from the parent node an address which is unambiguous in the network; in the case of routers, they will implicitly receive a group of addresses which could be assigned to their eventual children. The coordinator can assign the addresses in two different ways: (i) delegating to the higher levels the choice of the address assigned to each node, (ii) using a

[3] This table could be created also in the end devices, but the standard is not clear about this point.

distributed (tree-based) addressing mechanism, which allows to automatically generate the addresses to be assigned to the nodes in the network. The last addressing mode is efficient since it allows to use a hierarchical routing algorithm which does not require touring tables.

2.5.2 Routing Algorithm

In the Zigbee standard, the routing is presented from a global point of view [62], without distinguishing between the different implementation of the algorithms. In order to help the reader in better understanding, it is better to interpret the routing through a classification of the several routing algorithms. At this proposal, we can state that the Zigbee standard provides for three different routing algorithm: (i) an hierarchical algorithm, (ii) an ad-hoc routing algorithm based on Ad hoc On Demand Distance Vector (AODV, [63]), and (iii) an algorithm referred to as neighborhood routing. In the following, the ad-hoc routing will be indicated through the acronym AODV, even if there exists a slight difference among the two.

The neighborhood routing is the simplest routing strategy, It consists of verifying whether the destination address is in the routing tables of the neighboring nodes; in such a case, the source node transmits directly to the node which this information, using the direct data transfer mechanism. The standard does not specify whether this approach can be implemented also by the end devices or only by the router nodes.

The hierarchical routing algorithm is based on the fact that each Zigbee node has a direct parent-child degree of kinship with at least another node in the network, therefore exploiting in a recursive way these links, it is possible to communicate with all the other nodes in the network through multi-hop paths. This algorithm, when used jointly to the distributed addressing mechanism, does not require routing tables, therefore its implementation in the nodes is really simple. In fact, starting from the destination address of a frame, through a simple equation it is possible to understand in which direction the packet must be forwarded, that is towards either the parent or the child node. At the same time, this approach has two problems: first of all, due to its nature, it exploits only sub-optimal paths (which in exceptional cases can be optimal) increasing the delay; the second problem is that, due to the absence of

routing tables, it is not possible to dynamically modify the paths, therefore a failure in a path may harm the communication.

The AODV routing algorithm allows to obtain unicast paths, especially in scenarios characterized by dynamic links such those for mobile ad-hoc WSNs, using a limited amount of computational resources. Differently from the tree routing, AODV is a routing algorithm based on routing tables which must be stored in the routers of the network. The correct functioning of AODV is based on the transmission of Route Request (RREQ) packets in broadcast. These packets are forwarded by intermediate nodes in order to reach either the destination node or an intermediate node which has a valid and updated route for the destination node. The nodes which have information will reply through the transmission of Route Response (RREP) unicast packets which will be forwarded till the source node. The route discovery process begins with the transmission of a RREQ packet from the source node and ends with either the reception of a RREP packet or the expiration of a timer in the source node, which indicates that the route discovery process has failed. The information necessary for the correct routing of the packet are distributed among the network, in the routing and discovery tables of each routing with forwarding capabilities.

2.5.3 Network Topologies

One of the tasks of the Zigbee coordinator is the choice of the network topology. As described in Section 2.2, the NWK layer networks may present three different topologies: (i) star, (ii) tree, and (iii) mesh. Also in this case, the standard does not provide a clear distinction between the different network topologies.

The star topology is characterized by the fact that there are no routers in the network, since the communications are only between the end devices and the Zigbee coordinator, that is there is only a parent-child communication type and the maximum number of hops traversed by a packet is two, one from the end device to the coordinator and one from the coordinator to the destination node belonging to the same PAN. In this configuration, no NWK routing functionalities are exploited. Due to this simple structure, the star topology gives the following advantages:

- synchronization between the nodes;
- control on the latency in the network;
- low-power operation support due to the use of active zones;
- QoS support due to the use of GTS defined by the IEEE 802.15.4.

On the other side, there are also the following contraindications:

- at most two-hops communications are possible;
- the absence of redundancy between the paths yields a lack of robustness and requires that all the links towards the coordinator are reliable.

The tree topology is formed in the following way: the PAN coordinator is the root of the tree; starting from this node, it unravels a branch structure which is composed by routers, which are used to create the branches of the tree, and by end nodes, which act as the leaves of the tree and which do not have any role at NWK layer. Let's note that whether a RFD is connected to the branch of the tree, this node automatically becomes a leaf, since no additional node can be connected through the RFD. Vice versa, if a FFD is connected to a branch, it will behave like a leaf, till another node will be connected to it and it will start behaving like a branch. The tree topology can be seen as an extension of the star topology. The standard configuration of a tree topology requires the use of the distributed addressing mechanism and the tree routing algorithm. The peculiar tree structure, the possibility of using the beacon packets, together with the tree routing algorithm, make this topology energy efficient. In fact, by opportunely shifting the active zones of the branches connected to the same router, it is possible to use really short duty cycles, allowing only small subsets of nodes to be active at the same time, with all the other nodes are in low-power consumption state. Obviously, this expedient allows to significantly reduce the overall energy consumption and to increase the lifetime of the nodes, especially in the case of symmetrical trees. Finally, the advantages of tree topology are the following:

- for each path it is possible to evaluate the latency with a required accuracy;

- in beacon modality, it is possible to operate in low-power mode, making possible the use of battery powered routers;
- multi-hop communications are feasible.

On the other side, the main disadvantages of the tree topology are the following:

- the maximum network diameter is fixed;
- the latency, even if predictable, is high;
- due to the absence of redundancy, the path must be reliable to avoid failures.

Finally, the mesh topology is characterized by the fact that a node can communicate with all the other nodes in the network, under the limitation given by its transmission range and its routing capability (an end device can transmit only to its parent). The advantages and disadvantage of mesh networks are the opposite of those of tree networks. In fact, mesh networks are, usually, quite energy consuming since they do not support the beacon modality, therefore the nodes cannot be placed in low-power mode. On the other side, the strength of mesh networks is the resistance to node/link failures. In fact, the capability of communicating with all the nodes in the network allows to implicitly introduce redundancy in the paths, introducing the possibility of exploiting alternative paths in the case of failures. The advantages of mesh topology are the following:

- the network diameter is not necessarily fixed;
- the latency can be lower than in the case of tree topology;
- the network is robust due to the presence of multiple paths between a source and a destination node.

On the other hand, the disadvantages are the following:

- it is hard to predict the latency;
- the router must be always on, therefore battery powered routers cannot be used;

- the presence of routing tables requires higher memory capacity inside the routers;
- network discovery capabilities are required.

2.6 Application Layer

Each node, according to the Zigbee nomenclature, is referred to as device. This definition includes not only the hardware characteristics of the node, but also its ISO/OSI stack. On each Zigbee device it is possible to configure up to 240 Application Object, that is software applications which exploit the underlying hardware to perform some tasks. As for TCP/IP-based networks, where the applications are univocally identified through the used port, each Application Object is associated to an unambiguous numerical identifier, referred to as end-point. A set of Application Objects implementing a Zigbee profile, even if operating on distinct peripherals, is referred to as Zigbee Application. A profile is a set of guidelines which allow the interoperability between Application Objects respecting these guidelines. A profile is said public if the specifications are public available, allowing to have interoperability between devices built by different producers. For example, the Zigbee alliance has defined profiles for the nowadays most common applications, such as domotic automation, light control, etc. Vice versa, it is also possible to define private profiles which do not have the interoperability property.

Data exchanged in the network are defined by the attributes, that is they are associated with a couple (identifier, numerical format): for example, the identifier “temperature” is associated with the data type “16-bit integer number,” or the identifier “switch” is associated with the boolean data “on/off.” The input and output attributes of an Application Object are denoted with the term Device Descriptor. Since the descriptor is associated with the applications, depending on the considered application field, the same device may present different characteristics. In addition, a set of attributes associated with a given Application Object are identified univocally inside a profile by a Cluster ID because they form a cluster. We point out that this definition of cluster is different from the one that will be presented in Chapter 3.

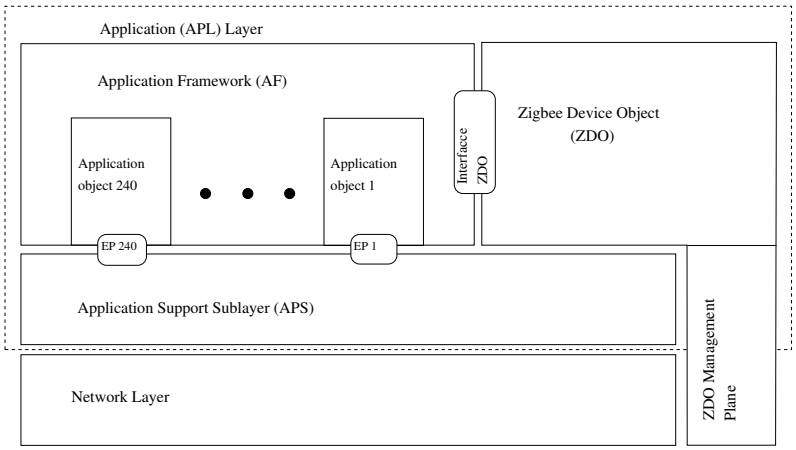


Figure 2.10: Logical scheme of the application layer of the Zigbee protocol.

The operation of mutual association between two or more couple of values (end-point, cluster ID), belonging to two or more different Application Objects, is referred to as binding and allows the involved nodes to communicate one with the other. Let's note that the introduction of this cluster-based communication mechanism is the main modification in 2006 version of the Zigbee standard with respect to the previous version which was based on two specific services: (i) Message (MSG) and (ii) Key Value Pair (KVP). Not only the Application Layer (APL) but also the NWK and MAC layers take part into the binding creation. The associations can be of different types: (i) one-to-one, (ii) one-to-many, (iii) many-to-one, and (iv) many-to-many, that is one or more end-points can be associated with one or more end-points on the remote peripherals. All the existing bindings in the network are stored by the coordinator in a binding table.

Finally, referring to Figure 2.10, we present the protocol structure of the Zigbee APL. There are three main sub-levels: (i) Application Support Layer (APS), (ii) Application Framework (AF), and (iii) Zigbee Device Object (ZDO). The APS sublayer provides a data transfer service to the higher sublayers, taking care of the creation of APL level packets, referred to as Application Protocol Data Units (APDUs), compliant with the requirements of the lower layers, since they do not support packet

fragmentation; in addition, the APS takes part in the binding operations. The AF sublayer, instead, is responsible for the preparation of the APDU frame, which can be of either KVP or MSG type and which will be forwarded to the lower layers. Finally, the ZDO sublayer is a special application which is associated with the end-point 0 and which manages the automatic operations of the network. The user-defined applications notify the ZDO the parameters that settle their behavior. In particular, the specified parameters concern the input and output clusters which are used to transmit from and to each end-point. If an output cluster of a local end-point corresponds to the input cluster of a remote end-point, the devices are said to be "corresponding." The ZDO provides also for an automatic corresponding-nodes research functionality. Once the nodes have been identified, the ZDO automatically contacts them and sets up the connection.

2.7 Conclusions

In this chapter we have presented the Zigbee standard. We have pointed out how this standard looks for the highest possible energy saving, while trying to send data at a sufficiently high transmission rate. Another evident aspect is the simplicity of this standard at the APL. The goal of this standard is to allow a programmer of Zigbee-based applications to easily manage the required operations. The network creation and management, in fact, are automatic operations which must be "transparent" to the APL. Finally, a lot of attention has been given to the MAC layer, because large part of the work presented in the following will involve MAC characteristics.

Chapter 3

Theoretical Performance Evaluation Framework

3.1 Introduction

In this chapter, we present the analytical tools which allow to evaluate and optimize the performance of WSNs. One of the most common applications of WSNs is the border surveillance, especially for military purposes. In such a scenario, a set of nodes, equipped with sensors with detection capabilities, can be placed in order to protect and monitor the accesses to a given area. Since the nodes of a WSN are traditionally equipped with batteries, the energy consumption becomes an important parameter for optimal network configuration. However, maximizing the network energy saving can be sometimes counterproductive, because the use of a limited amount of transmission power at the nodes may cause an increment of the number of packets lost at the AP due to collisions. In addition, in order to structure the network and decrease the number of exchanged packets, a clustered approach can be used. In such a way, the nodes in charge of the management of the clusters (cluster-heads) can fuse the information transmitted by the nodes belonging to each cluster. This operation allows the cluster-heads to retransmit only a subset of the received packets, representing the result of the local decision carried out on the basis of the received information.

The structure of the chapter is the following: in Section 3.2, we introduce an analytical model for the derivation of the probability of missing a target, whereas in Section 3.3, we provide for a power allocation strategy based on the minimization of the collisions between the packets received by the AP. In Section 3.4, we present the analytical model for the data fusion at the cluster-heads and at the AP. Finally, in Section 3.5, we draw some conclusions about the analytical models presented.

3.2 Target Detection Model

In this section, we introduce the problem of target detection in sparsely deployed WSNs [64]. In particular, we focus on random node deployment in the considered monitored area. After presenting the problem that we are facing, we describe the physical model for the sensing interface of the nodes in the presence of duty cycles. In addition, we also provide for a simplified model for the latency of the transmission of an alert message to the AP of the network and for the energy consumption. The goal of this section is the derivation of a framework which allows to evaluate the performance of the systems in terms of (i) probability of detecting a target entering the monitored area (P_{d}) or, conversely, probability of missing it (P_{md}), (ii) latency after detection, and (ii) energy consumption.

3.2.1 Problem Statement

The surveillance of a given area is a sensible aspect in many military and civilian applications. In particular, it is often required that a set of nodes detects an incoming target which crosses the monitored area. The simplest solution would be to place the sensors in the proximity of all the entrance points of the area. However, in many realistic scenarios it is not possible to cover all the possible entrance points, since a number of sensors, large enough to ensure the required detection, is often not available. Therefore, it is necessary to set an appropriate node deployment that guarantees the highest P_{d} of any incoming target.

In addition, due to the physical morphology of the monitored area, it is often not possible to deal with cables placement, so that a wireless solution is the winning

choice for ensuring the required P_{d}. In a WSN there is a node which acts as the access point (AP), for the other nodes, to reach control centers outside of the network. A generic node can also relay messages for the others.

In these networks, battery energy is a precious resource and nodes are cyclically "switched off," according to proper duty cycles, generally at both sensing and the communication levels. However, these operations can affect both the probability of detection, since a node may be off when the target crosses its sensed area, and the transmission latency of an "alert" message form the detecting node to the AP. On the opposite, in this way it is possible to extend the network lifetime.

3.2.2 Physical Description of the Sensor

A wireless sensor devices embeds two main sub-units, i.e., (i) the sensing sub-unit and (ii) the communication sub-unit. The former is equipped with a seismic sensor, whose sensing range r_{s} is maximum over a rocky surface, in which the vibrations due to an incoming target propagate with low attenuation. Since the seismic sensors can be placed also over different surfaces, such as sandy or clayey terrains, where the propagation model is different and the attenuation is higher, we consider different values of r_{s}. In order to reduce the energy consumption of the system, the sensing part can be periodically switched off, according to a normalized duty cycle $\beta_{\mathrm{sens}} \in [0,1]$ over a period T_{sens} (dimension: [s]). More precisely, nodes sense the surrounding environment for the interval of length $\beta_{\mathrm{sens}} T_{\mathrm{sens}}$ and sleep for the interval of duration $(1-\beta_{\mathrm{sens}})T_{\mathrm{sens}}$. The power consumption associated with the sensing operations is denoted as Ω_{sens}. We assume that all the sensors have the same r_{s}, β_{sens}, and T_{sens}.

The communication interface of the nodes has a transmission range r_{T} (dimension: [m]), under the constrain that $r_{\mathrm{T}} \gg r_{\mathrm{s}}$. Generally, r_{T} ranges between 100 m and 1000 m (in line-of-sight scenarios).

We have assumed that only the node detecting the target transmits an alert message to the AP and that there can be only one target in the monitored area at a time. As a result of these assumptions, there are no collisions between the packets transmitted by the wireless devices. In such a scenario, the Medium Access Control (MAC) protocol must guarantee the lowest energy consumption and the lowest latency in the

network. In order to keep the MAC layer as simple as possible, we have chosen the X-MAC protocol [65], since it does not require synchronization between the nodes (reducing energy consumption) and it keeps the latency low, at the opposite of scheduled MAC protocols, especially for multi-hop networks. The X-MAC protocol and the corresponding parameters, i.e., S_{d}, S_{p}, S_{al}, $\Omega_{\mathrm{T_x}}$, $\Omega_{\mathrm{R_x}}$, and Ω_{s}, will be described in Subsection 3.2.5. Since it is likely that a packet has to be relayed in order to reach the AP, the average number of hops that a packet has to traverse will be denoted with N_{hop}.

To make the derivation of P_{d} (or, equivalently, of the probability of missed detection P_{md}) feasible, we assume the monitored area to be a square with sides of length d_{s} (dimension: [m]). In this area, N sensors are identically and independently deployed in a random fashion under the constrain that their sensing ranges do not overlap. We also assume that the potential targets penetrate the monitored area following a linear and uniform trajectory. Trajectories are characterized by an angle of arrival θ and a target constant speed v (dimension: [m/s]). Since there is no information about the entrance point, we also assume that the target enters the monitored area from a random point along the perimeter of the monitored surface.

3.2.3 Preliminary Works on Random Deployment

Our analytical framework for the evaluation of the probability of target missed detection, considering moving targets, extends the results presented in [12], which we rapidly recall in the following. For more details about the following derivation, the interested reader is referred to [12, 66, 67].

In order to detect a target in a squared area with perimeter L_0, N sensors are randomly placed over the field of interest. Sensors s_i, $i = 1, \ldots, N$ have a sensing area of perimeter L_i. Assuming that there is no priori knowledge about the direction and the entrance point of the target, the probability $P_{\mathrm{d}}(k)$ that at least $k \geq 1$ sensors detect the target crossing the field of interest is

$$P_{\mathrm{d}}(k) = 1 - \sum_{w=0}^{k-1} \sum_{j=1}^{|Z_{N,w}|} \prod_{i=1}^{|z_j|} q_{z_j(i)} \prod_{v=1}^{|\bar{z}_j|} (1 - q_{\bar{z}_j(v)}) \tag{3.1}$$

where $|Z_{N,w}|$ denotes the possible w-tuples z_j of vector $[1,...,N]$, i.e., the possible groups of w sensors that can detect the target at a time, $\overline{z}_j$ denotes the complement $(N-w)$-tuple of z_j, $q_i = q_{z_j(i)} = q_{\overline{z}_j(i)} = L_i/L_0$, and q_i can be interpreted as $P\{\text{Sensor on the trajectory}\}$.

Assuming that $L_i = L\ \forall i \in \{1,\cdot,N\}$, equation (3.1) can be rewritten as

$$P_{\mathrm{d}}(k) = 1 - \sum_{i=0}^{k-1} \binom{N}{i} \frac{L^i (L_0 - L)^{N-i}}{L_0^N}.$$

Eq. (3.1) can also be used to derive the probability of missed detection of a target. In this case, observing that $P_{\mathrm{md}} = P(Z_{N,0})$, i.e., the probability of not detecting any target, and $z_j = 0, \overline{z}_j = \{1,\cdots,N\}$, the probability P_{md} of missing a target can be written as

$$P_{\mathrm{md}} = \prod_{i=1}^{N} \left(1 - \frac{L_i}{L_0}\right). \tag{3.2}$$

According to the model introduced in Subsection 3.2.2 and recalling that all the sensors have the same sensing range, Eq. (3.2) can be rewritten as follows

$$P_{\mathrm{md}} = \left(1 - \frac{2\pi r_{\mathrm{s}}}{4 d_{\mathrm{s}}}\right)^N. \tag{3.3}$$

3.2.4 Integration of Duty Cycles for Random Deployment

To integrate sleeping duty cycles at sensing level, we extend the previous model to express P_{d} or, equivalently, P_{md}, as a function of both the duty cycle β_{sens} and the geometrical configuration of the WSN. The probability of detecting a target is the probability that there is a sensor on the target's trajectory (event denoted as $\mathscr{E}_{\mathrm{SoT}}$) *and* the sensor is active when the target is crossing the sensed area (event denoted as $\mathscr{E}_{\mathrm{det}}$). Therefore, the probability that a single sensor detects a target is

$$P_{\mathrm{d-1}} = P\{\mathscr{E}_{\mathrm{SoT}}, \mathscr{E}_{\mathrm{det}}\} = P\{\mathscr{E}_{\mathrm{det}} | \mathscr{E}_{\mathrm{SoT}}\} P\{\mathscr{E}_{\mathrm{SoT}}\}. \tag{3.4}$$

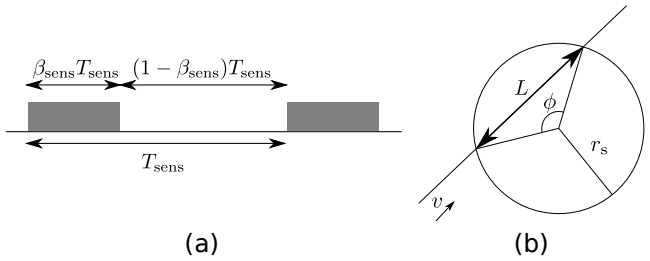


Figure 3.1: (a) Logical scheme of the sensing duty cycle and (b) model for the sensing range of a node.

According to [12], $P\{\mathscr{E}_{\text{SoT}}\}$ can be expressed as $2\pi r_{\text{s}}/(4d_{\text{s}})$.

In order to evaluate $P\{\mathscr{E}_{\text{det}}|\mathscr{E}_{\text{SoT}}\}$, we consider the scheme for the sleeping duty cycle presented in Fig. 3.1 (a). Since the target arrives with a finite speed v, the crossing time is $T_{\text{cross}} = L/v$, where L is a random variable which expresses the length of the intersection between the target's trajectory and the area sensed by a sensor, as shown in Fig. 3.1 (b). Since there is no information about the arrival of the target, its arrival has been assumed uniformly distributed over the period T_{sens}.

When the sensor is on, any incoming target will be detected. In the case that the sensor is off, i.e., during the interval of duration $\beta_{\text{sens}}T_{\text{sens}}$, the analysis has to be refined. Let $\mathscr{E}_{\text{target}}$ be the event {The sensor is on at the instant at which the target enters the sensed area}. Applying the total probability theorem [68], $P\{\mathscr{E}_{\text{det}}|\mathscr{E}_{\text{SoT}}\}$ can then be expressed as

$$\begin{aligned} P\{\mathscr{E}_{\text{det}}|\mathscr{E}_{\text{SoT}}\} &= P\{\mathscr{E}_{\text{det}}|\mathscr{E}_{\text{target}},\mathscr{E}_{\text{SoT}}\}P\{\mathscr{E}_{\text{target}}|\mathscr{E}_{\text{SoT}}\} \\ &+ P\{\mathscr{E}_{\text{det}}|\bar{\mathscr{E}}_{\text{target}},\mathscr{E}_{\text{SoT}}\}P\{\bar{\mathscr{E}}_{\text{target}}|\mathscr{E}_{\text{SoT}}\} \end{aligned} \tag{3.5}$$

where $P\{\mathscr{E}_{\text{det}}|\mathscr{E}_{\text{target}},\mathscr{E}_{\text{SoT}}\} = 1$. Since $\mathscr{E}_{\text{target}}$ and $\mathscr{E}_{\text{SoT}}$ are independent—in fact, the activity cycle of a sensor does not depend on the target—one can write,

$$P\{\mathscr{E}_{\text{target}}|\mathscr{E}_{\text{SoT}}\} = \mathrm{P}\{\mathscr{E}_{\text{target}}\} = \int_0^{\beta_{\text{sens}}\mathrm{T}_{\text{sens}}} \frac{1}{\mathrm{T}_{\text{sens}}}\mathrm{dt} = \beta_{\text{sens}}$$

and

$$P\{\bar{\mathscr{E}}_{\text{target}}|\mathscr{E}_{\text{SoT}}\} = 1-\beta_{\text{sens}}.$$

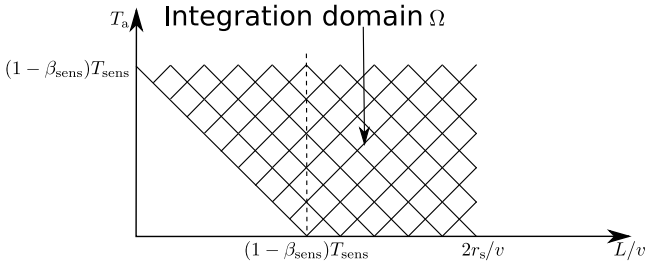


Figure 3.2: Integration domain for the evaluation of $P\{\mathscr{E}_1, \mathscr{E}_2\}$.

We are now going to evaluate the last term of Eq. (3.5). According to the conditioning on $\bar{\mathscr{E}}_{\text{target}}$ and $\mathscr{E}_{\text{SoT}}$, the target arrival time, denoted as T_a, is a uniformly distributed random variable over an interval of length $(1-\beta_{\text{sens}})T_{\text{sens}}$. To have successful detection, the target must remain in the sensed area until the sensor turns on its sensing device in the following active period. In this case as well, one must distinguish between two cases: (i) $T_{\text{cross}} > (1-\beta_{\text{sens}})T_{\text{sens}}$ and (ii) $T_{\text{cross}} < (1-\beta_{\text{sens}})T_{\text{sens}}$. In the former case, each target will be detected, since it remains in the sensed area for a time interval longer than the sleep period. Therefore, in this case $P_d = P\{\mathscr{E}_{\text{SoT}}\}$. In the latter case, instead, the target will be detected if it enters the sensed area in the last part of the sleep period, so that it will be detected in the following active period.

Letting $\mathscr{E}_1 = \{T_a + L/v > (1-\beta_{\text{sens}})T_{\text{sens}}\}$ and $\mathscr{E}_2 = \{L/v < (1-\beta_{\text{sens}})T_{\text{sens}}\}$, the probability that the target is detected, given that the sensor is in the sleep state when the target enters the sensed area, can be expressed as

$$P\{\mathscr{E}_{\text{det}} | \bar{\mathscr{E}}_{\text{target}}, \mathscr{E}_{\text{SoT}}\} = P\{\mathscr{E}_1, \mathscr{E}_2\} = \iint_{\Omega} f(T_a, L)\, \mathrm{d}T_a\, \mathrm{d}L.$$

where the integration domain Ω is shown in Fig. 3.2. Since T_a and L are independent, the joint probability density function (pdf) can be expressed as the product of the marginal pdfs, i.e., $f(T_a, L) = f(T_a)\, f(L)$. Since T_a is a uniformly distributed random variable, it follows that $f(T_a) = 1/[(1-\beta_{\text{sens}})T_{\text{sens}}]$. In order to express $f(L)$, it is necessary to make some further considerations. In particular, the pdf of the angle ϕ, shown in Fig. 3.1 (b), is needed. The length l of the chord, i.e., of a particular

realization of L, can be expressed as the following function of ϕ

$$l = 2r_{\mathrm{s}} \sin\left(\frac{\phi}{2}\right)$$

with $\phi \in [0, 2\pi]$.

According to the fundamental theorem at page 93 of [68], $f(L)$ becomes

$$f(L) = \begin{cases} \dfrac{1}{\pi\sqrt{r_{\mathrm{s}}^2 - \left(\frac{L}{2}\right)^2}} & \text{if } 0 < L < 2r_{\mathrm{s}} \\ 0 & \text{else.} \end{cases} \tag{3.6}$$

Finally, taking into account that one needs to evaluate L/v in order to correctly compute $P\{\mathscr{E}_1, \mathscr{E}_2\}$, the following joint pdf can be used as

$$f(T_{\mathrm{a}}, Y) = \begin{cases} \dfrac{v}{\pi c\sqrt{r_{\mathrm{s}}^2 - \left(\frac{vY}{2}\right)^2}} & \text{if } 0 < Y < \dfrac{2r_{\mathrm{s}}}{v}, \\ & 0 < T_{\mathrm{a}} < c \\ 0 & \text{else.} \end{cases} \tag{3.7}$$

where $Y = L/v$ and $c = (1 - \beta_{\mathrm{sens}})T_{\mathrm{sens}}$.

Considering the integration domain, two possible cases can be distinguished: (i) $\frac{2r_{\mathrm{s}}}{v} < (1-\beta_{\mathrm{sens}})T_{\mathrm{sens}}$ and (ii) $\frac{2r_{\mathrm{s}}}{v} > (1-\beta_{\mathrm{sens}})T_{\mathrm{sens}}$. In the former case, the integration domain, referring to Fig. 3.2, reduces to the triangle on the left. $P\{\mathscr{E}_1, \mathscr{E}_2\}$ can thus be expressed as

$$P\{\mathscr{E}_1, \mathscr{E}_2\} = \int_0^{\frac{2r_{\mathrm{s}}}{v}} \int_{c-y}^{c} f(T_{\mathrm{a}}, Y)\, \mathrm{d}T_{\mathrm{a}}\, \mathrm{d}Y = \frac{4r_{\mathrm{s}}}{\pi c v}. \tag{3.8}$$

In the latter case, i.e., when $2r_{\mathrm{s}}/v > (1-\beta_{\mathrm{sens}})T_{\mathrm{sens}}$, the integration domain is that shown in Fig. 3.2, therefore the expression of $P\{\mathscr{E}_1, \mathscr{E}_2\}$ can be rewritten as

$$\begin{aligned} P\{\mathscr{E}_1, \mathscr{E}_2\} &= \int_0^{c} \int_{c-y}^{c} f(T_{\mathrm{a}}, Y)\, \mathrm{d}T_{\mathrm{a}}\, \mathrm{d}Y \\ &+ \int_c^{\frac{2r_{\mathrm{s}}}{v}} \int_0^{c} f(T_{\mathrm{a}}, Y)\, \mathrm{d}T_{\mathrm{a}}\, \mathrm{d}Y \\ &= \frac{4r_{\mathrm{s}} - 2\sqrt{4r_{\mathrm{s}}^2 - c^2v^2}}{\pi c v} + 1 - \frac{2\mathrm{asin}\left(\dfrac{cv}{2r_{\mathrm{s}}}\right)}{\pi} \end{aligned} \tag{3.9}$$

Combining equations (3.8) and (3.9) into equation (3.5), the expression for $P\{\mathscr{E}_{\text{det}}|\mathscr{E}_{\text{SoT}}\}$ can be obtained

$$P\{\mathscr{E}_{\text{det}}|\mathscr{E}_{\text{SoT}}\} = \beta_{\text{sens}} + (1-\beta_{\text{sens}})\text{P}\{\mathscr{E}_1, \mathscr{E}_2\}. \tag{3.10}$$

Finally, extending the model in order to take into account that N independent sensors can detect the target, the probability of missed detection becomes

$$P_{\text{md}} = (1-P_{\text{d}-1})^N = \left(1 - P\{\mathscr{E}_{\text{det}}|\mathscr{E}_{\text{SoT}}\}\frac{2\pi \text{r}_\text{s}}{4\text{d}_\text{s}}\right)^N. \tag{3.11}$$

In the case of heterogeneous sensing model, the derivation is almost identical. The only difference is that we should start the derivation from Eq. (3.2) instead of from Eq. (3.3).

3.2.5 Latency after Detection

In this section, we propose an analytical model for the alert transmission latency, i.e., the delay between the detection instant of the presence of an intruder by a sensing node and its notification to the sink. In the following, we first derive the per hop latency, denoted as $D_{1\,\text{hop}}$, and then the latency over a multi-hop path. We then verify the validity of this model through experimental measurements on a Crossbow MicaZ testbed.

As mentioned in Subsection 3.2.2, we use X-MAC [65], a low-power asynchronous MAC-layer protocol for duty-cycled WSNs. X-MAC uses Low-Power Listening (LPL), or preamble sampling, to enable low-power communications between a sender and a receiver which do not synchronize their wake-up and sleep schedules. Indeed, a sender with data sends a preamble at least as long as the sleep interval of the receiver. This guarantees that the receiver will wake up, detect the preamble, and stay awake for the reception of the data. X-MAC uses a *strobed preambling* approach in which the sender quickly alternates between sending the packet destination address and a short wait time so that the receiver could potentially abort this process to receive data. This approach allows to further reduce energy consumption and the per-hop latency in comparison to protocols using long-preambles such as B-MAC [69].

The average per hop transmission latency can be expressed as

$$D_{1\,\mathrm{hop}} = \frac{(1-\beta_{\mathrm{comm}})^2 T_{\mathrm{comm}}}{2} + S_{\mathrm{p}} + S_{\mathrm{al}} + S_{\mathrm{d}} \tag{3.12}$$

where β_{comm} is the (normalized) communication duty cycle over the period T_{comm}, and S_{p}, S_{al}, S_{d} durations (dimension: [s]) of the strobed preamble, the acknowledgement of the preamble, and the alert packet, respectively. Considering the status of the receiving node, i.e., if the communication subsystem is either turned on or off, the probability that a node begins its transmission when the receiving node is on is β_{comm} and the associated latency is $S_{\mathrm{p}} + S_{\mathrm{al}} + S_{\mathrm{d}}$. On the opposite, the probability that a node is off is $1-\beta_{\mathrm{comm}}$. We evaluate $D_{1\,\mathrm{hop}}$ simply as the average between worst and best cases. The best case is when the a node starts transmitting exactly when the receiving node starts its LPL operations, so that the packet is transmitted after $S_{\mathrm{p}} + S_{\mathrm{al}} + S_{\mathrm{d}}$. In the worst case, the transmitting node waits for the entire duration of the sleep interval. In addition, since the receiving node must receive an entire preamble before sending the acknowledgment message, the worst case takes into account that two transmissions of the preamble may be required in order to start the communication. In this case, the latency introduced by the transmission is $(1-\beta_{\mathrm{comm}})T_{\mathrm{comm}} + (S_{\mathrm{p}} + S_{\mathrm{al}}) + S_{\mathrm{d}}$. Scaling this term by the probability that the receiver is off, averaging the best and the worst cases and adding the latency, related to the case with the receiver on, scaled by its probability of being on, equation (3.12) is obtained.

Considering a multi-hop path, the average global latency can be expressed as follows

$$D = D_{1\,\mathrm{hop}} N_{\mathrm{hop}} \tag{3.13}$$

where N_{hop} denotes the average number of nodes that the alert message traverses to reach the sink.

Note that the model relies on a collision-free transmission of the alert as we assume a single target detection scenario with low arrival rate. Therefore, there is only one sending node at a time that does not experience packet losses and retransmissions. We also neglect the processing time of the packets.

3.2.6 Energy Model

As nodes operate on batteries, the way they consume energy directly impacts the lifetime of the surveillance system. To take this into account, we now propose a simple energy model for the engineering toolbox.

The energy consumption of nodes can be roughly given by the sum of the energies consumed by its hardware components. For the sake of simplicity, we only integrate in the energy model contributions from the sensing sub-unit and the radio transceiver. We also define the network lifetime as the time needed for the average residual energy E_{r} to be lower then a threshold value E_{th}.

To derive an expression for the network lifetime, we evaluate the energy consumed after a given interval t. The average residual energy E_{r} at a generic instant t can be expressed as

$$E_{\mathrm{r}}(t) = NE_{\mathrm{i}} - N\Omega_{\mathrm{tot}}t \tag{3.14}$$

where E_{i} is the initial energy of a node and Ω_{tot} is the power consumption given by sensing and communication operations.

According to the description of the X-MAC protocol in Subsection 3.2.5, there are four possible states for a node: (i) transmission, (ii) reception, (iii) sleep, and (iv) LPL, with corresponding power consumptions denoted as $\Omega_{\mathrm{T_x}}$, $\Omega_{\mathrm{R_x}}$, Ω_{s}, and Ω_{LPL}, respectively.

Ω_{tot} can then be computed as follows

$$\Omega_{\mathrm{tot}} = \Omega_{\mathrm{sensing}} + \Omega_{\mathrm{LPL}} + (\Omega_{\mathrm{R}} + \Omega_{\mathrm{T}})P_{\mathrm{d}}N_{\mathrm{target}} \tag{3.15}$$

where: $\Omega_{\mathrm{sensing}}$ is the power consumption associated with the sensing device in the activity period of duration T_{sens}; Ω_{LPL} is the power required when performing the LPL operations (over a period of duration T_{comm}); Ω_{R} is the power used by a node to receive a packet; Ω_{T} is the power used to transmit an alert packet; P_{d} is the target detection probability; and N_{target} is the number of times that a target appears during a reference period. The expected power to send a packet can be expressed as

$$\Omega_{\mathrm{T}} = \left[\Omega_{\mathrm{T_x}}S_{\mathrm{d}} + \frac{(1-\beta_{\mathrm{comm}})T_{\mathrm{comm}}}{2(S_{\mathrm{p}}+S_{\mathrm{al}})} \cdot (\Omega_{\mathrm{T_x}}S_{\mathrm{p}} + \Omega_{\mathrm{R_x}}S_{\mathrm{al}})\right]\frac{N_{\mathrm{hop}}}{NT_{\mathrm{comm}}} \tag{3.16}$$

where the first additive term, i.e., $\Omega_{T_x} S_d$, is the energy spent to transmit a packet, whereas the last additive term expresses the energy consumption due to the periodic preamble transmission in order to notify the receiving node of a packet arrival. As in Subsection 3.2.5, since node are not synchronized, we assume the average number of preamble transmissions to be the mean between best and worst cases. The term N_{hop}/N is introduced owing to the fact that only a set of the N sensors is used to relay the alert message to the AP.

The expected power to receive a packet can be expressed as

$$\Omega_R = \frac{[\Omega_{R_x} S_d + (\Omega_{R_x} S_p + \Omega_{T_x} S_{al})] N_{hop}}{N T_{comm}} \tag{3.17}$$

where $\Omega_{R_x} S_d$ is the energy spent to receive a packet and $\Omega_{R_x} S_p + \Omega_{T_x} S_{al}$ is the energy spent for the reception of the preamble and the transmission of the acknowledgment message. Since in the considered scenario there is no information about the position of the AP, the average number of hops has been determined by calculating, through the Dijkstra algorithm, the shortest path for each node to reach any other node in the network. Given that r_T is around 250 m, the value of N_{hop} is around 2.7. However, since it is reasonable to consider a integer value of N_{hop}, the value has been rounded to 3.

The power associated with the LPL operations can be expressed as

$$\Omega_{LPL} = \Omega_{R_x} \beta_{comm} + \Omega_s (1 - \beta_{comm}) - \Gamma_{T_x} - \Gamma_{R_x} \tag{3.18}$$

where Γ_{T_x} and Γ_{R_x} are two corrective terms. In particular, equations (3.16), (3.17), and (3.18) are evaluated over a period T_{comm}. However, during normal operations the node either performs LPL operations or transmits/receives a packet. Γ_{T_x} and Γ_{R_x} are used to refine the power consumption due to LPL operations. In fact, the LPL and the transmission and reception intervals overlap for short intervals, so that without these two terms the power consumption budget would be higher than the correct one. In particular, Γ_{T_x} can be expressed as

$$\Gamma_{T_x} = \left[\Omega_s \left[\frac{(1 - \beta_{comm}) T_{comm}}{2} + S_d + S_{al} \right] + \Omega_{R_x} S_p \right] \frac{P_d N_{target} N_{hop}}{N T_{comm}} \tag{3.19}$$

whereas Γ_{R_x} can be expressed as

$$\Gamma_{R_x} = \left[(S_{al} + S_d)\,\Omega_s + \Omega_{R_x} S_p\right] \cdot \frac{P_d N_{target} N_{hop}}{N T_{comm}}. \tag{3.20}$$

The Γ_{T_x} term must take into account the fact that, during the transmission operations, such as periodical preamble sending over an interval of duration $(1-\beta_{comm})T_{comm}/2$, packet transmission and acknowledgment reception, the node would normally be in the sleep state, whereas during the transmission of the preamble, that will be acknowledged by the receiving node, the node would normally be in the reception state. This is basically a correction factor, since otherwise the energy consumed by the node with this model would be higher than the real value because the reception and transmission operations overlap over a period with normal LPL operations. Similar considerations can be carried out for the Γ_{R_x} term. In fact, when the node is waiting for the acknowledgment window to transmit the acknowledgment message, receiving the preamble and transmitting a packet, according to the LPL operations, it would normally be in the sleep state. The same considerations can be carried out when the node is receiving the preamble, since it would normally be in the receiving state for the LPL operations.

Finally, the power consumed during sensing operations can be expressed as

$$\Omega_{sensing} = \beta_{sens}\Omega_{sens}. \tag{3.21}$$

Introducing the expressions in (3.19) and (3.20) into (3.18) and the expressions in (3.16), (3.17), (3.18), and (3.21) into (3.14), it is possible to derive an expression for the energy consumption which depends on both sensing and communication parameters.

In order to derive an explicit expression for the network lifetime L, equation (3.14) can be rewritten as

$$L = \frac{NE_i - E_{th}}{N\Omega_{tot}} \tag{3.22}$$

where E_{th} is a given residual energy threshold, which can be used to model the physical behavior of a node.

3.3 Adjacency Matrix-based Transmit Power Allocation Strategy

One of the main constraints of the WSNs is the limited amount of energy available at the nodes. In order to maximize the lifetime and minimize the collisions between transmitted packets, a power allocation strategy can be used [70]. In the following, we first introduce some key parameters of a Zigbee WSN. Then, we present a simplified version of its MAC protocol and under the assumption of low traffic load, we propose a simplified analytical model for the estimation of the following main network performance indicators: (i) the Packet Error Rate (PER) at the AP, i.e., the ratio between the number of packets lost at the AP and the total number of transmitted packets in the network, and (ii) the average delay. Finally, we describe the power allocation strategy which allows to minimize the PER and, at the same time, to extend the lifetime of the network.

3.3.1 Definition of a Simplified Model for Zigbee WSNs

Each sensor node is characterized by two main parameters: (i) its position on a two-dimensional plane and (ii) its transmit power, as stated in the following definition (for the sake of simplicity, we will simply use the term "sensor" to refer to a wireless node with sensing capabilities).

Definition 1 *A* sensor *is represented by a couple* $s = (x, P)$*, where* $x \in \mathbb{R}^2$ *is the sensor position and* $P \in \mathbb{R}$ *is its transmit power.*

We remark that the previous definition is based on the assumption that the positions of the nodes are known. This is realistic in several practical applications, such as industrial or home monitoring, where the spatial distribution of the nodes is *a priori* determined. In more general scenarios, the positions of the nodes could be unknown. In such case, one should also consider proper localization algorithms. However, once the positions of the nodes are estimated, our framework for optimized transmit power control can be directly applied.

We assume that the detection operation is described by an ideal threshold model, as stated in the following assumption.

Assumption 1 (Threshold reception) *Given two sensors* $s_1 = (x_1, P_1)$ *and* $s_2 = (x_2, P_2)$, *there exists a* minimum power function $\Pi(x_1, x_2)$ *such that sensor* s_2 *receives the transmission of sensor* s_1 *if and only if*

$$P_1 \geq \Pi(x_1, x_2).$$

This assumption holds because of propagation loss (according to the Friis formula) and assumes that a threshold detector is used at the receiver [71]. In fact, in this case the power P_r received by sensor s_2 can be expressed as:

$$P_\mathrm{r} = P_1 G_\mathrm{t} G_\mathrm{r} \left(\frac{\lambda}{4\pi r} \right)^{\alpha} \tag{3.23}$$

where G_t and G_r are the gains of the transmit and receive antennas, r is the distance, λ is the wavelength, and α is the path loss exponent. According to the ideal threshold detector model, sensor s_2 receives a transmission from sensor s_1 if and only if $P_r > P_{\min}$, where $P_{\min}$ is the (pre-defined) receiver reception threshold. In this case,

$$\Pi(x_1, x_2) = \frac{P_{\min}}{G_\mathrm{t} G_\mathrm{r}} \left(\frac{4\pi r}{\lambda} \right)^{\alpha}.$$

A sensor network can be introduced as a set of sensors, characterized by their positions and their transmit powers, together with an associated minimum power function.

Definition 2 *A* sensor network *of N elements is an ordered set* $\mathscr{S} = (c, \Pi, s_1, s_2, \ldots, s_N)$, *where* $s_1, s_2, \ldots, s_N$ *are sensors,* $c \in \mathbb{R}^2$ *is the position of the AP, and* $\Pi : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ *is the associated minimum power function.*

Definition 3 (Adjacency matrix) *Given a sensor network* $\mathscr{S} = (c, \Pi, s_1, s_2, \ldots, s_N)$, *with* $s_i = (x_i, P_i)$, $i = 1, \ldots, N$, *its associated* adjacency matrix *is given by*

$$A(\mathscr{S}) \in \mathbb{R}^{N \times N}$$

where

$$A_{ij} = A(\mathscr{S})_{ij} = \begin{cases} 1 & \text{if } P_i \geq \Pi(x_i, x_j) \\ 0 & \text{otherwise}. \end{cases}$$

The complement of $A(\mathscr{S})$ *corresponds to*

$$\overline{A}(\mathscr{S}) = \begin{cases} 1 & \text{if } A_{ij} = 0 \\ 0 & \text{otherwise}. \end{cases}$$

The number of ones in the adjacency matrix is given by the adjacency *of sensor network* $\mathscr{S}$ *and denoted by* $|A(\mathscr{S})|$*. The complementary adjacency is given by the number of zeros in the adjacency matrix and denoted by* $|\overline{A}(\mathscr{S})|$*.*

For each $i = 1, \ldots, N$ *,we define the following two sets:*

$$\begin{aligned} \mathscr{R}_i &\triangleq \{j = 1, \ldots, N | A_{ji} = 1\} \\ \mathscr{T}_i &\triangleq \{j = 1, \ldots, N | A_{ij} = 1\} \end{aligned}$$

which represent the sets of indices of the sensors that s_i *can receive from and transmit to, respectively. We denote by* $\overline{\mathscr{R}}_i$ *and* $\overline{\mathscr{T}}_i$ *the complements of these two sets.*

In order to make the theoretical analysis feasible, a Zigbee WSN is described by the following simplified model.

Assumption 2 (Simplified model)

1. *Poisson generation: the traffic generated by each sensor in the network is modeled as a homogeneous Poisson process [72]. The processes associated with different sensors are independent of each other and have intensity g (dimension: [pck/s]) [73].*

2. *Limited CCA: before transmission, the i-th sensor waits for a random backoff time, with average* T_{B_1} *(dimension: [s]), and then checks if the channel is clear. This clear channel assessment (CCA) is limited only to those sensors whose indices lie in the set* $\mathscr{R}_i$*. In other words, the sensing is limited only to those sensors that can effectively (i.e., with sufficiently high received power) transmit to the i-th sensor. The CCA has a duration equal to* T_{CCA}*.*

3. *Infinite number of backoffs: if the channel is found busy, the current sensor transmission is delayed by a random backoff time with average T_{B_2} (dimension: [s]). During the backoff period the traffic generation at the transmitting sensor does not stop. There is no limit on the total number of subsequent backoffs that a single packet transmission can incur.*

4. *Constant transmission length: each transmission has the same length $T_{\text{trans}} = L/R$, where L is the packet length (dimension: [b/pck]) and R is the transmission data rate (dimension: [b/s]).*

5. *Transmission turnaround time: after sensing, if the channel is found idle, each sensor waits a turnaround time, denoted as T_{TAT} (dimension: [s]), before starting its transmission.*

For each sensor s_i ($i = 1, \ldots, N$) the following counting processes can be defined.

- $G_i(t)$: the number of times that sensor s_i has checked if the channel is clear in the time interval $[0,t]$.
- $B_i(t)$: the number of times that a packet transmission of sensor s_i has been delayed, through the backoff mechanism, in $[0,t]$.
- $T_i(t)$: the number of times that a sensor s_i has transmitted in $[0,t]$ (counting both successful and unsuccessful transmissions).
- $E_i(t)$: the number of transmission errors incurred by sensor S_i in $[0,t]$.

For a counting process $P(t)$, define the steady state intensity as follows:

$$F[P] \triangleq \lim_{t\to\infty,\tau\to 0} \frac{\mathbb{E}[P(t+\tau) - P(t)]}{\tau} \tag{3.24}$$

where $\mathbb{E}[\cdot]$ denotes expectation. We recall that for the stationary Poisson traffic generation processes, the steady state intensity is constant and denoted by g. In the following, we assume that the limit at the righthand side of Equation 3.24 exists for all previously defined counting processes: this is equivalent to assuming that the network

reaches a "steady state." Under this hypothesis, the following equilibrium conditions must be satisfied:

$$F[T_i] = g \tag{3.25}$$

$$F[G_i] = g + F[B_i]. \tag{3.26}$$

Equation 3.25 states that, at steady state, the intensity of transmissions must be equal to the intensity of traffic generation. Equation 3.26 states that, at steady state, the intensity of channel sensing has to be equal to the sum of the intensities of packet generations and backoffs.

The backoff traffic intensity can be expressed as follows:

$$F[B_i] = F[G_i]\chi_i$$

where χ_i represents the ratio between the numbers of backoffs and transmission attempts. In this way, the processes $\{T_i(t)\}$ and $\{B_i(t)\}$ satisfy the following relations:

$$F[G_i] = F[B_i] + F[T_i] = \frac{g}{1-\chi_i} \tag{3.27}$$

$$F[B_i] = \frac{\chi_i}{1-\chi_i} g. \tag{3.28}$$

The term χ_i can be equivalently interpreted as the probability, for the ith sensor, to assess that the channel is busy during the CCA. In order to derive a simple expression for χ_i, it is assumed that the processes $\{T_i(t)\}$ are uncorrelated and Poisson. This simplification is appropriate under low traffic conditions. In fact, in this case, $F[B_i] << g$ and the processes $\{T_i(t)\}$ are statistically very similar to Poisson traffic generation processes. However, as it will be shown in Chapter 4, the estimated PER obtained with these simplifications is close to that predicted by (realistic) simulations also under relatively high traffic conditions.

Under the above simplifications, χ_i equals the probability of finding at least one packet transmission event, during a time interval equal to the transmission length T_{trans}, in the set of independent Poisson processes $\{G_j(t)\}_{j\in\mathscr{R}_i}$. In other words, one

can write:

$$\chi_i = \lim_{t\to\infty} \mathscr{P}\left\{ \max_{j\in\mathscr{R}_i} \{T_j[t+T_{\text{trans}}] - T_j[T_{\text{trans}}]\} > 0 \right\}.$$

In order to compute χ_i, it is worth remarking that the probability of finding no packet transmissions from the ith sensor in a time interval of length T_{trans} is given by $e^{-F[T_i]T_{\text{trans}}}$. Since the process T_i is assumed to be Poisson and uncorrelated from the other $\{T_j\}_{j\neq i}$, the probability of finding no transmission events from the sensors belonging to $\mathscr{R}_i$ (i.e., those sensors that can be received by the ith sensor) in T_{trans} is given by

$$\prod_{j\in\mathscr{R}_i} (e^{-F[T_j]T_{\text{trans}}}).$$

In conclusion, the probability of finding at least one packet transmission event in a time interval equal to T_{trans} from any of the sensors that can be received by the ith sensor is given by

$$1 - \prod_{j\in\mathscr{R}_i} (e^{-F[T_j]T_{\text{trans}}}).$$

Therefore, χ_i can finally be expressed as follows:

$$\begin{aligned} \chi_i &= 1 - \prod_{j\in\mathscr{R}_i} (1 - e^{-F[T_j]T_{\text{trans}}}) \\ &\simeq \sum_{j\in\mathscr{R}_i} gT_{\text{trans}} \\ &= g|\mathscr{R}_i|T_{\text{trans}} \end{aligned} \tag{3.29}$$

where we have used Equation 3.25 and approximated $\prod_{j\in\mathscr{R}_i}(1-e^{-gT_{\text{trans}}})$ with $\sum_{j\in\mathscr{R}_i} gT_{\text{trans}}$. The latter simplification holds under low traffic conditions, where $gT_{\text{trans}} << 1$. The notation $|\mathscr{R}_i|$ stands for the number of elements of the set $\mathscr{R}_i$. From Equation 3.29, using the approximations $1/(1-\chi_i) \simeq 1+\chi_i$ and $\chi_i/(1-\chi_i) \simeq \chi_i$, that hold for small values of χ_i, the following simplified expressions for network sensing and backoff intensities can then be obtained:

$$\begin{aligned} F[G_i] &\simeq (1+\textstyle\sum_{j\in\mathscr{R}_i} gT_{\text{trans}})g \\ F[B_i] &\simeq (\textstyle\sum_{j\in\mathscr{R}_i} gT_{\text{trans}})g. \end{aligned}$$

In general, the number of transmission errors accumulated by sensor s_i can be written in the following form:

$$F[E_i] = \gamma_i F[G_i] + \lambda_i F[T_i] + \eta_i F[T_i] + \kappa_i F[T_i] \tag{3.30}$$

where the four terms at the righthand side can be characterized as follows. The term $\gamma_i F[G_i]$ represents the intensity of transmission errors occurred due to the occupation of channel by a packet transmission that could not be detected by the ith sensor during a CCA interval. The term $\lambda_i F[T_i]$ represents the intensity of transmission errors due to interference from other sensors that cannot receive s_i. The term $\eta_i F[T_i]$ represents the intensity of transmission errors resulted from another sensor beginning to transmit when s_i is waiting the turnaround time between the CCA and the transmission act. Finally, the term $\kappa_i F[T_i]$ represents the intensity of transmission errors due to the fact that other sensors can begin transmission in the first subinterval, of length T_{TAT}, of a transmission act from sensor s_i. In fact, if some other sensor begins transmission during the turnaround time, it cannot detect the previous starting instant of a transmission by s_i. The last two terms appearing in Equation 3.48 take into account the transmission errors independent of the network connectivity and are significant in the overall network error analysis.

Under the assumption of low traffic load and with the simplification that all relevant processes are Poisson and independent, the coefficient γ_i in Equation 3.48 can be approximated as follows:

$$\begin{aligned}
\gamma_i &= \lim_{t\to\infty} \mathscr{P}\{\max_{j\in\overline{\mathscr{R}}_i}\{T_j[t+T_{\text{trans}}] - T_j[T_{\text{trans}}]\} > 0\} \\
&= 1 - \prod_{j\in\overline{\mathscr{R}}_i} (1 - e^{-F[T_j]T_{\text{trans}}}) \\
&\simeq \sum_{j\in\overline{\mathscr{R}}_i} F[T_j]T_{\text{trans}} \simeq |\overline{\mathscr{R}}_i| T_{\text{trans}} g.
\end{aligned}$$

Similarly, the coefficient λ_i in Equation 3.48 can be approximated as

$$\begin{aligned}\lambda_i &= \lim_{t\to\infty} \mathscr{P}\{\max_{j\in\overline{\mathscr{T}}_i}\{G_j[t+T_{\text{trans}}]-G_j[T_{\text{trans}}]\}>0\}\\ &= 1-\prod_{j\in\overline{\mathscr{T}}_j}(1-e^{-F[G_j]T_{\text{trans}}})\\ &\simeq \sum_{j\in\overline{\mathscr{T}}_j} F[G_j]T_{\text{trans}} \simeq |\overline{\mathscr{T}}_j|T_{\text{trans}}g.\end{aligned}$$

The coefficient η_i in Equation 3.48 can be approximated as

$$\begin{aligned}\eta_i &= \lim_{t\to\infty} \mathscr{P}\{\max_{j=1,\ldots,N}\{T_j[t+T_{\text{TAT}}]-T_j[t]\}>0\}\\ &= 1-\prod_{j=1,\ldots,N}(1-e^{-F[G_i]T_{\text{TAT}}}) \simeq NgT_{\text{TAT}}.\end{aligned}$$

Finally, the coefficient k_i in Equation 3.48 is given by

$$\begin{aligned}\kappa_i &= \lim_{t\to\infty} \mathscr{P}\{\max_{j=1,\ldots,N}\{T_j[t+T_{\text{TAT}}]-T_j[t]\}>0\}\\ &= 1-\prod_{j=1,\ldots,N}(1-e^{-F[T_i]T_{\text{TAT}}}) \simeq NgT_{\text{TAT}}.\end{aligned}$$

Using the expressions found above for the coefficients γ_i, λ_i, η_i and κ_i in Equation 3.48, the transmission error intensity can be approximated as

$$F[E_i] \simeq \left[(|\overline{\mathscr{T}}_i|+|\overline{\mathscr{R}}_i|)T_{\text{trans}}+2NT_{\text{TAT}}\right]g^2.$$

Therefore, the overall network error intensity can be estimated as follows:

$$\sum_{i=1}^{N} F[E_i] \simeq (2|\bar{A}(\mathscr{S})|T_{\text{trans}}+2N^2T_{\text{TAT}})g^2$$

and the error probability, i.e., the ratio between the overall network error intensity and the generation intensity (given by Ng), becomes

$$\begin{aligned}P_{\text{er}} &= \frac{\sum_{i=1}^{N}F[E_i]}{Ng}\\ &\simeq \left(2\frac{|\bar{A}(\mathscr{S})|}{N^2}T_{\text{trans}}+2T_{\text{TAT}}\right)Ng. \qquad (3.31)\end{aligned}$$

Equation 3.31 shows that, under the considered simplifying assumptions, the error probability grows linearly with the network complementary adjacency $|\bar{A}(\mathscr{S})|$.

In the following, we find an estimate of the average network delay. First of all, we remark that if, after the first backoff, the channel is found idle, the total delay is given by

$$D_{\min} = T_{\mathrm{B}_1} + T_{\mathrm{CCA}} + T_{\mathrm{TAT}} + T_{\mathrm{trans}}.$$

This is the minimum average delay that a packet incurs if the channel is found idle at the first transmission attempt. If the channel is found busy, the sensor waits for a backoff time with average T_{B_2}, then senses the channel again. If, taking into account the second transmission attempt, the channel is found idle for the second time, the overall delay can be expressed as $D_{\min} + D_{\mathrm{BO}}$, where

$$D_{\mathrm{BO}} = T_{\mathrm{B}_2} + T_{\mathrm{CCA}}.$$

This delay model is slightly different from that presented in Section 3.2.5 since, in this case, we are considering higher packet generation rate and unslotted CSMA/CA.

Under the low traffic load assumption, the probability of having more than one backoff during a single transmission act is negligible. Therefore, the average transmission delay becomes

$$\begin{aligned} D_i &= D_{\min} + D_{\mathrm{BO}}\gamma_i = D_{\min} + D_{\mathrm{BO}} \sum_{j \in \overline{\mathscr{R}}_i} F[T_j] T_{\mathrm{trans}} \\ &\simeq D_{\min} + D_{\mathrm{BO}} |\overline{\mathscr{R}}_i| g T_{\mathrm{trans}} \end{aligned}$$

where D_{BO} is the average backoff time. The average network delay can then be expressed as

$$\begin{aligned} D &= \frac{\sum_{i=1}^{N} D_i}{N} \\ &\simeq D_{\min} + D_{\mathrm{BO}} \frac{|A(\mathscr{S})|}{N} T_{\mathrm{trans}} g. \end{aligned} \tag{3.32}$$

Equation 3.32 for the delay shows that the network delay depends linearly on the network adjacency. We remark that, since we considered star topologies, the PER

and delay statistics collected at each node are less significant than those calculated at the AP, which instead provide a better description of the network behavior. Should more complicated topologies be considered, the proper metrics need to be taken into account.

In conclusion, under the low traffic load assumption, the PER and the delay at the AP of a WSN can be estimated as follows:

$$\begin{aligned} P_{\text{er}} &\simeq (2\frac{|\bar{A}(\mathscr{S})|}{N^2}T_{\text{trans}} + 2T_{\text{TAT}})Ng \\ D &= T_{\text{B}_1} + T_{\text{CCA}} + T_{\text{TAT}} + (T_{\text{B}_2} + T_{\text{CCA}})(T_{\text{trans}}\frac{|A(\mathscr{S})|}{N}g). \end{aligned} \tag{3.33}$$

3.3.2 Network Lifetime

An important parameter for a WSN is the network lifetime. This performance indicator can be interpreted in several ways. For example, in [17] the network lifetime is defined as the time interval at the end of which the probability of outage falls below a maximum value than can be tolerated, on average, over the transmission links before the network is declared dead. In particular, the network degradation (i.e., the increase of the probability of outage) is assumed to be caused by fading and battery depletion. In [74], the network lifetime is related to the minimum number of sensors that need to be active before declaring the network dead. More precisely, when the number of active nodes drops to below this minimum number due to battery depletion, the network dies.

In this section, the network lifetime is defined similarly to that proposed in [74]. More precisely, since we are focusing on power control, i.e. minimization of the total transmit power for a given PER, we consider a definition of network lifetime based on the overall residual energy in the network. If the overall residual energy at time t, denoted as $E_{\text{res}-\text{net}}(t)$ is higher than a pre-defined threshold, which may depend on a required network operational quality of service (QoS), then the network is declared alive. On the other hand, if the residual energy becomes lower than this threshold, then the network is declared dead. We point out that this definition of residual energy is slightly different from that presented in Section 3.2.6 since here we are not focusing on the impact of the communication and sensing duty cycles. In this section, we are interested only in the impact of the transmit power on the network residual energy.

The network residual energy at time t can be expressed as:

$$E_{\text{res}-\text{net}}(t) = NE_{\text{I}-\text{node}} - E_{\text{cons}-\text{net}}(t)$$

where $E_{\text{I}-\text{node}}$ is the initial per-node energy and $E_{\text{cons}-\text{net}}(t)$ is the average energy consumed, at network level, up to time t. In order to evaluate $E_{\text{cons}-\text{net}}(t)$, one can write:

$$E_{\text{cons}-\text{net}}(t) = P_{\text{cons}-\text{net}}t = NP_{\text{cons}-\text{node}}t \tag{3.34}$$

where $P_{\text{cons}-\text{net}}$ is the average network-level consumed power and $P_{\text{cons}-\text{node}}$ is the average consumed power at each node. When the proposed power allocation strategy is used, the consumed power at each sensor is different. However, in order to simplify the analytical model, we consider the average network-wide power consumed, then we derive the average power consumed at each node. At this point, the evaluation of the average network residual energy at any instant reduces to the evaluation of $P_{\text{cons}-\text{node}}$.

In order to evaluate $P_{\text{cons}-\text{node}}$, one can observe that it depends on the average powers consumed by the nodes in each of the following possible states: (i) transmission (tx), (ii) reception (rx), (iii) CCA, (iv) BO, and (v) idle. We denote the average percentages of time, in 1 s, spent by the nodes in each of the previous states as (i) $\tau_{\text{tx\_state}}$, (ii) $\tau_{\text{rx\_state}}$, (iii) $\tau_{\text{CCA\_state}}$, (iv) $\tau_{\text{Boff\_state}}$, and (v) $\tau_{\text{idle\_state}}$, respectively. The average power consumed at each node can then be evaluated as follows:

$$\begin{aligned} P_{\text{cons}-\text{node}} \quad = \quad & \tau_{\text{tx\_state}}P_{\text{tx\_state}} + \tau_{\text{rx\_state}}P_{\text{rx\_state}} + \tau_{\text{CCA\_state}}P_{\text{CCA\_state}} \\ & + \tau_{\text{Boff\_state}}P_{\text{Boff\_state}} + \tau_{\text{idle\_state}}P_{\text{idle\_state}}. \end{aligned} \tag{3.35}$$

At this point, we simply need to evaluate (i) the percentages of time and (ii) the powers appearing at the right-hand side of Equation 3.35. We start with the percentages of time. As stated, we refer to the percentages of time spent in the various states within a 1 s interval. We remark that the assumption of a reference time equal to 1 s holds since $gT_{\text{trans}} \ll 1$. In fact, if $gT_{\text{trans}} \geq 1$, each node would always be in the transmission state, and the network would not function. Likewise, the other percentages of times are all lower than 1 under the assumed low traffic load conditions. The

percentage of time spent by a node in the tx state can be computed as

$$\tau_{\text{tx\_state}} = gT_{\text{trans}}.$$

The percentage of time spent in the rx state for a generic node i can be computed as the sum of the transmission time percentages of the nodes which are within the transmission range of node i, i.e., from the nodes belonging to $\mathscr{R}_i$. Owing to the previous derivations, this percentage of time does not depend on the particular node and, exploiting the results in Equations 3.27 and 3.28, can be expressed as follows:

$$\tau_{\text{rx\_state\_i}} = \sum_{j \in \mathscr{R}_i} gT_{\text{trans}} = gT_{\text{trans}}|\mathscr{R}_i| = gT_{\text{trans}} \sum_{j=1}^{N} |A_{ji}|.$$

The percentage of time spent in the CCA phase can be evaluated as

$$\tau_{\text{CCA\_state}} = F[G_i]T_{\text{CCA}} \simeq \left(1 + \sum_{j \in \mathscr{R}_i} gT_{\text{trans}}\right) gT_{\text{CCA}} = \left(1 + gT_{\text{trans}} \sum_{j=1}^{N} |A_{ji}|\right) gT_{\text{CCA}}.$$

The fraction of time spent in the BO state by a generic node i, under the assumption that the node experiences only a single BO before transmitting a packet, can be written as

$$\tau_{\text{Boff\_state}} = F[B_i]T_{\text{B}_1} \simeq \left(\sum_{j \in \mathscr{R}_i} gT_{\text{trans}}\right) gT_{\text{B}_1} = \left(gT_{\text{trans}} \sum_{j=1}^{N} |A_{ji}|\right) gT_{\text{B}_1}.$$

Finally, since the previous percentages of time have been evaluated with respect to a reference interval that equals to 1 s, the percentage of time spent by a node in the idle state can be expressed as

$$\tau_{\text{idle\_state}} = 1 - (\tau_{\text{tx\_state}} + \tau_{\text{rx\_state}} + \tau_{\text{CCA\_state}} + \tau_{\text{Boff\_state}}).$$

In order to evaluate the average power consumption in each state, we refer to the results presented in [44], where the authors evaluate the power consumption of a generic node equipped with a CC2420 radio. In particular, the current consumption in the tx state depends linearly on the transmit power. The power consumption in each state is shown in Table 3.1. These terms have been obtained as a linear in-

$I_{\text{idle\_state}}$	396 uA
$I_{\text{rx\_state}}$	19.6 mA
$I_{\text{CCA\_state}}$	19.6 mA
$I_{\text{Boff\_state}}$	396 uA
$I_{\text{tx\_state\_}i}$	$7.886P_i + 0.009711$ mA

Table 3.1: Current consumption in each state for a generic CC2420 radio module.

terpolation of the values presented in [44]. We point out that the dimension of the coefficient P_i is 1/V. The voltage reference for the evaluation of the consumed power is $V_{\text{DD}} = 3$ V. Given the current consumption, it is possible to derive the associated power consumption by simply multiplying the current consumption by the reference voltage. In this way, the values of the powers in the various states (excluding the tx state) become:

$$P_{\text{idle\_state}} = V_{\text{DD}} I_{\text{idle\_state}} = 1.188\,\text{mW} \tag{3.36}$$

$$P_{\text{rx\_state}} = V_{\text{DD}} I_{\text{rx\_state}} = 58.8\,\text{mW} \tag{3.37}$$

$$P_{\text{CCA\_state}} = V_{\text{DD}} I_{\text{CCA\_state}} = 58.8\,\text{mW} \tag{3.38}$$

$$P_{\text{Boff\_state}} = V_{\text{DD}} I_{\text{Boff\_state}} = 1.188\,\text{mW}. \tag{3.39}$$

The power consumed in the tx state can be expressed as the arithmetic average of the specific transmit powers used by all nodes in the network:

$$P_{\text{tx\_state}} = \frac{\sum_{i=1}^{N} I_{\text{tx\_state\_}i} V_{\text{DD}}}{N} = \frac{\sum_{i=1}^{N} (23.598 P_i + 0.029133)}{N}. \tag{3.40}$$

Note that the values of $\{P_i\}$ will be determined by the proposed power allocation strategy. Obviously, if a uniform power allocation strategy is used, i.e. $\{P_i\}$ are all equal, the proposed derivation still holds.

3.3.3 Optimal Transmit Power Allocation

In this subsection, we discuss the following problem:

Problem 1 (Transmission error optimization) *Upon the assignment of a total available transmit power P_{tot} for the sensor network $\mathscr{S}$, distribute it among the sensors in the network in order to minimize the PER at the AP.*

This problem is equivalent to minimizing the overall transmit power to guarantee a desired PER at the AP.

Under low traffic load assumption, using Equation 3.33 on the PER, the solution of Problem 1 is equivalent to the maximization of the adjacency $|A(\mathscr{S})|$ of sensor network $\mathscr{S}$. This fact allows to recast Problem 1 in the following form.

Problem 2 (Network adjacency maximization) *Upon the assignment of a total available transmit power P_{tot} for the sensor network $\mathscr{S}$, distribute it among the sensors in the network in order to maximize the network adjacency $|A(\mathscr{S})|$.*

Assign to each sensor s_i a transmission power $P_i > 0$, $i = 1,\ldots,N$. Then, the network adjacency is given by the following function:

$$|A(\mathscr{S})| = Q_{\text{tot}}(P_1,P_2,\ldots,P_N) = \sum_{i=1,\ldots,N,\ j=1,\ldots,N} H(P_i - \Pi(x_i,x_j)) \tag{3.41}$$

where

$$H(x) = \begin{cases} 1 \text{ if } x \geq 0 \\ 0 \text{ otherwise} \end{cases}$$

is the Heaviside function. We remark that the Heaviside function $H(P_i - \Pi(x_i,x_j))$ appearing in Equation 3.41 is 1 if the ith sensor transmits with a power sufficient to reach the jth sensor, and 0 if otherwise.

For each sensor s_i, define $\mathscr{P}_i$ as the following set of transmit power values:

$$\mathscr{P}_i \triangleq \{\Pi(x_i,x_j),\ j = 1,\ldots,N,\ \text{ with } j \neq i : \Pi(x_i,x_j) \geq P_{\min,i}\} \cup \{P_{\min,i}\} \tag{3.42}$$

where $\Pi(x_i,x_j)$ is the transmit power with which sensor s_i can reach sensor s_j and $P_{\min,i}$ is the minimum power that allows the ith sensor to reach the AP. According to this definition, the set $\mathscr{P}_i$ contains the value of the minimum transmit power required by the ith sensor to reach the AP, together with the values of the transmit powers that allow s_i to reach the other sensors of the network and are higher than $P_{\min,i}$.

The following property leads to the possibility of limiting the search of possible transmit powers for a sensor s_i to the set $\mathscr{P}_i$.

Proposition 1 *For any set of transmit powers $P_i > 0$, $i = 1,\ldots,N$, there exists a set of values $\bar{P}_i \in \mathscr{P}_i$, such that*

$$Q_i(\bar{P}_1,\bar{P}_2,\ldots,\bar{P}_N) = Q_i(P_1,P_2,\ldots,P_N), \tag{3.43}$$

$$\bar{P}_i \leq P_i,\ \forall i = 1,\ldots,N. \tag{3.44}$$

Proof.
Define

$$\bar{P}_i \triangleq \max\{P \in \mathscr{P}_i : P \leq P_i\}. \tag{3.45}$$

Equation 3.44 follows immediately from Equation 3.45. Moreover, the function at the righthand side of Equation 3.41 is piecewise constant with respect to any argument P_i, discontinuous on those values in which $P_i = \Pi(x_i,x_j)$ for any $j = 1,\ldots,N$. From Equations 3.42 and 3.45 it follows that function Q is continuous in the set $[\bar{P}_1,P_1] \times [\bar{P}_2,P_2]\cdots[\bar{P}_N,P_N]$ and therefore constant in this set. Hence, Equation 3.43 holds. □

Proposition 1 simply means that, in the ideal threshold detection hypothesis, it is not convenient to allocate to sensor s_i a transmit power that does not belong to the set $\mathscr{P}_i$, since it would employ extra power without gaining extra connectivity. For instance, in a network composed of 4 sensors, suppose that sensor 1 can reach the AP using a transmit power of 0.5 mW, whereas it needs 1 mW to reach sensor 2, 2 mW to reach sensor 3, and 0.2 mW to reach sensor 4, respectively. In this case, $\mathscr{P}_i = \{0.5\,\text{mW}, 1\,\text{mW}, 2\text{mW}\}$ contains the transmit powers that allow to reach the AP and sensors 2 and 3. The optimal transmit power for the first sensor should be chosen in this set. In fact, for example, it would be inconvenient to choose a transmit power of 1.5 mW instead of 1 mW, because the connectivity would be the same despite the increased transmit power (sensor 1 would still reach the AP and sensors 2 and 4).

The power allocation problem may be written in the following form.

Problem 3 (Discrete optimization problem) *For each sensor $i = 1,\ldots,N$ choose a transmit power $P_i \in \mathscr{P}_i$ such that the function $T(P_1,P_2,\ldots,P_N)$ (defined by Equa-*

Number of sensors	Mean [s]	Std. Dev. [s]
10	0.040829	0.0064042
20	0.053216	0.00795104
50	0.17279	0.0730227
100	0.80042	0.336531
200	4.38015	1.46177

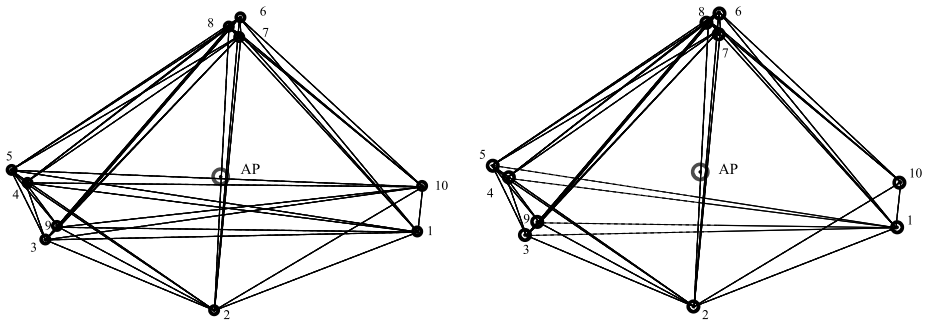
Table 3.2: Computation times for networks of different sizes. Results obtained with Mosek 5 (64-bit version) with a Core 2 Duo CPU at 3.16 GHz and with 4 GB RAM.

tion 3.41) is maximized while satisfying the constraint

$$\sum_{i=1}^{N} P_i \leq P_{\text{tot}}.$$

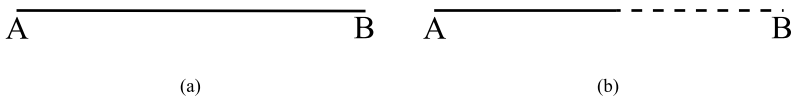
This problem corresponds to a *multiple choice knapsack problem*, which has been extensively studied in the literature [75] and can be solved by standard computational libraries, such as MOSEK [76]. It is well known that this problem is NP-complete and the computation time increases very quickly as the number of sensors in the network grows. However, this is a standard optimization problem and some recent tools allow finding the exact solution in a reasonable time, in many cases of practical interest. Table 3.2 shows the computation time (namely the mean value and the standard deviation), in relation to the size of the sensor network, obtained with MOSEK 5 (64-bit version) running over a Core 2 Duo CPU with a clock frequency of 3.16 GHz and a 4 GB RAM. Furthermore, it is worth noting that accurate suboptimal solutions to problems of larger size (i.e. considering larger networks) could be obtained through heuristic methods.

An illustration of how the proposed approach works is depicted in Figure 3.3, whose legend is shown in Figure 3.4. When the transmit power budget is large enough to allow each node to communicate with any other node (Figure 3.3a), all bidirectional connections are active (solid lines, as shown in Figure 3.3a). When the power budget is not large enough (Figure 3.3b), the proposed optimized transmit



(a) (b)

Figure 3.3: Pairwise connections in a scenario with $N = 10$ nodes. Two values for the total network transmit power are considered: (a) $P_{\text{tot}} = 5 \cdot 10^{-5}$ W and (b) $P_{\text{tot}} = 2.5 \cdot 10^{-5}$ W. In both cases, the proposed optimized power allocation strategy is used.



(a) (b)

Figure 3.4: Graphical notation for communication links: (a) A and B communicate with each other (bidirectional communication); (b) only A can transmit to B (monodirectional communication).

power allocation strategy allocates the transmit power to the nodes in a way that the number of 1's in the adjacency matrix is maximized. This means that some connections may be missing (absence of connecting lines between the nodes) or become monodirectional (half solid and half dashed lines, as shown in Figure 3.3b).

3.3.4 Extension to Multi-hop Scenarios

In this subsection, we will extend the results obtained for the single-hop case to scenarios with multi-hop communications. We assume that the network has a tree topology and the root corresponds to the AP. The leaves of the tree are sensors which send the generated packets to their respective parent nodes. The (intermediate) nodes which

are not leaves may originate their own packets and send them to their respective parent nodes together with the data received from their children nodes. All packets generated by the nodes are transmitted towards the AP.

Let N be the total number of sensors in the network (including the access point, which is denoted by s_N). For each sensor s_i ($i = 1, \ldots, N-1$), denote by d_i the index of the sensor to which s_i sends its packets.

Define the routing matrix $R \in \mathbb{R}^n \times \mathbb{R}^n = (r_{ij})$ as follows:

$$r_{ij} = \begin{cases} 1 \text{ if } i = d_j \\ 0 \text{ otherwise.} \end{cases}$$

Define also recursively, starting from the leaves, the following quantity at each node:

$$\phi_i = 1 + \sum_{j|i=d_j} \phi_j$$

where the term ϕ_i represents the number of nodes (including node i) in the sub-tree which has node i as root.

For each sensor s_i, define, as in the single-hop case, the counting processes $G_i(t)$, $B_i(t)$, $T_i(t)$, and $E_i(t)$. We also assume that the packets are generated at each node as a Poisson process with intensity g.

Under the above assumptions, the following equilibrium conditions must be satisfied:

$$F[T_i] = g + \sum_{k \in s_i} (T_k - E_k) \tag{3.46}$$

$$F[G_i] = F[T_i] + F[B_i]. \tag{3.47}$$

Condition (3.46) states that, at steady state, the intensity of transmission must be equal to the sum of the intensities of generated and propagated traffic. Condition (3.47) states that at steady state the intensity of channel sensing is equal to the sum of intensities of traffic and backoffs.

As in the single-hop case, all processes $G_i(t)$, $B_i(t)$, $T_i(t)$ and $E_i(t)$ are assumed to be uncorrelated and Poisson. This simplification is clearly plausible only in low

traffic conditions. In the following, we will denote with g_i, b_i, t_i, e_i the corresponding intensities of these processes.

As in the single-hop case, the number of transmission errors of sensor s_i can be written in the following form:

$$F[E_i] = \gamma_i F[G_i] + \lambda_i F[T_i] + \eta_i F[T_i] + \kappa_i F[T_i]. \tag{3.48}$$

However, in this case $\gamma_i F[G_i]$ represents the intensity of transmission errors due to the fact then, when s_i transmits, a sensor, which can reach s_{d_i} but not s_i, begins transmitting, thus generating a collision. The term $\lambda_i F[T_i]$, instead, represents the intensity of transmission errors due to interference from other sensors that can communicate with sensor d_i but cannot receive from s_i. The term $\eta_i F[T_i]$ represents the intensity of transmission errors made because any sensor which can reach the destination sensor d_i begins any transmission when s_i is waiting the TAT between CCA and transmission. Finally, the term $\kappa_i F[T_i]$ represents the intensity of transmission errors due to the fact that any sensor which can reach the destination node d_i begins transmitting in the first subinterval of length T_{TAT} of a transmission of sensor s_i. In fact, due to the TAT, if some sensor begins transmitting in this time intervals, it cannot detect the preceding starting instant of a transmission by s_i. The last two terms appearing in (3.48) take into account transmission errors which "penalize" the use of a high transmit power.

Under the assumption of low traffic load and with the simplification that all relevant processes are Poisson and independent, the term γ_i in (3.48) can be computed as follows:

$$\gamma_i = \lim_{t\to\infty} \mathscr{P}\left\{ \max_{j\in\overline{\mathscr{R}}_i\cap\mathscr{R}_{d_i}} \{T_j[t+T_{\text{trans}}] - T_j[T_{\text{trans}}]\} > 0 \right\}$$

that is

$$\begin{aligned} \gamma_i &= 1 - \Pi_{j\in\overline{\mathscr{R}}_i\cap\mathscr{R}_{d_i}} (1 - e^{-F[T_j]T_{\text{trans}}}) \\ &\simeq \sum_{j\in\overline{\mathscr{R}}_i\cap\mathscr{R}_{d_i}} F[T_j]T_{\text{trans}} \simeq \sum_{j\in\overline{\mathscr{R}}_i\cap\mathscr{R}_{d_i}} t_j T_{\text{trans}} \,. \end{aligned}$$

Similarly, the term λ_i in (3.48) can be expressed as

$$\lambda_i = \lim_{t\to\infty} \mathscr{P}\left\{ \max_{j\in\overline{\mathscr{T}}_i\cap\mathscr{R}_{d_i}} \{G_j[t+T_{\text{trans}}] - G_j[T_{\text{trans}}]\} > 0 \right\}$$

that is

$$\begin{aligned} \lambda_i &= 1 - \Pi_{j\in\overline{\mathscr{T}}_j\cap\mathscr{R}_{d_i}}(1-e^{-F[G_j]T_{\text{trans}}}) \\ &\simeq \sum_{j\in\overline{\mathscr{T}}_i\cap\mathscr{R}_{d_i}} F[G_j]T_{\text{trans}} \simeq \sum_{j\in\overline{\mathscr{T}}_i\cap\mathscr{R}_{d_i}} g_j T_{\text{trans}}. \end{aligned} \tag{3.49}$$

The error due to other transmissions that can reach sensor s_{d_i} beginning during the turnaround time T_{TAT} can be estimated as

$$\eta_i = \lim_{t\to\infty} \mathscr{P}\left\{ \max_{j\in\mathscr{R}_{d_i}} \{T_j[t+T_{\text{TAT}}] - T_j[t]\} > 0 \right\}$$

so that

$$\eta_i = 1 - \Pi_{j\in\mathscr{R}_{d_i}}(1-e^{-F[G_i]T_{\text{TAT}}}) \simeq \sum_{j\in\mathscr{R}_{d_i}} t_j T_{\text{TAT}} \,.$$

Finally, the error contribution due to other transmitters that begin transmitting in the first subinterval of length T_{TAT} is given by

$$\kappa_i = \lim_{t\to\infty} \mathscr{P}\left\{ \max_{j\in\mathscr{R}_{d_i}} \{T_j[t+T_{\text{TAT}}] - T_j[t]\} > 0 \right\}$$

and, therefore,

$$\kappa_i = 1 - \Pi_{j\in\mathscr{R}_{d_i}}(1-e^{-F[T_i]T_{\text{TAT}}}) \simeq \sum_{j\in\mathscr{R}_{d_i}} t_j T_{\text{TAT}} \,.$$

Remark that terms κ_i and η_i penalize the use of high transmission power, since they grow when the network adjacency increases.

Under the low traffic load hypothesis, the number of backoffs and transmission errors is low and $g_i \simeq t_i \simeq \phi_i$.

Using these approximations, together with the expressions for the coefficients γ_i, λ_i, η_i, and κ_i in (3.48), we obtain the following estimate for transmission errors intensity:

$$F[E_i] \simeq \left[\left(\sum_{j \in \overline{\mathscr{R}}_i \cap \mathscr{R}_{d_i}} \phi_j + \sum_{j \in \overline{\mathscr{T}}_i \cap \mathscr{R}_{d_i}} \phi_j \right) T_{\text{trans}} + 2 \sum_{j \in \mathscr{R}_{d_i}} \phi_j T_{\text{TAT}} \right] \phi_i \, .$$

Therefore, the overall network error intensity can be estimated as follows:

$$\sum_{i=1}^{N-1} F[E_i] \simeq \sum_{i=1}^{N-1} \left[\left(\sum_{j \in \overline{\mathscr{R}}_i \cap \mathscr{R}_{d_i}} \phi_j + \sum_{j \in \overline{\mathscr{T}}_i \cap \mathscr{R}_{d_i}} \phi_j \right) T_{\text{trans}} + 2 \sum_{j \in \mathscr{R}_{d_i}} \phi_j T_{\text{TAT}} \right] \phi_i. \tag{3.50}$$

Finally, the error probability is given by the ratio between the error intensity and the generation intensity (given by Ng).

In the multi-hop case, the expression of the error probability is more complicated than the single-hop counterpart (3.31). Let

$$I_i(p_i) = \sum_{k | k \notin \mathscr{T}_i, d_k \in \mathscr{T}_i} \phi_k \phi_i T_{\text{trans}}.$$

The term I_i depends only on set $\mathscr{T}_i$, which is a function of the transmission power p_i.

Moreover, define:

$$\begin{cases} J_i(p_i, p_j) = \phi_i \phi_j T_{\text{TAT}} \text{ if } i \neq j,\ j \in \mathscr{T}_i \text{ and } d_i \in \mathscr{T}_j \\ J_i(p_i, p_j) = 0 \ \text{ otherwise} \, . \end{cases}$$

In this way, minimizing the error probability is equivalent to finding the minimum of function

$$T(p) = \sum_{i=1}^{N-1} I_i(p_i) + \sum_{i,j=1}^{N-1} J_i(p_i, p_j) \, .$$

This becomes a quadratic optimization problem with discrete variables, since the transmit powers can assume only discrete values and the quadratic terms are positive (see Subsection 3.3.3, where a linear programming problem with discrete variables has similarly been obtained in the single-hop case). This means that the optimal solution can be always found my means of branch-and-bound techniques. We have solved this problem using the Matlab toolbox Yalmip [77], together with Mosek [76].

3.4 Clustered Networks Performance Analysis

In this section, we present a model for the performance analysis of clustered networks. In particular, we focus on the impact on network transmission rate and delay of the required Quality of Service (QoS), expressed here as the percentage of RFDs' deaths which can be tolerated by the network before being declared dead [78]. We consider for the analysis both clustered and unclustered topologies. In addition, we introduce an analytical model for the derivation of the probability of decision error P_e in the presence of data fusion. In Chapter 4, this model will be confirmed through the use of the Opnet simulator.

3.4.1 Network Tolerable Death Level

A critical issue in wireless sensor networking is the network lifetime, since nodes are typically equipped with a limited-energy battery and may be subject to failures. First, one has to define when the network has to be considered "alive," and several definitions have been proposed in the literature. In general, the network can be considered alive until a proper QoS condition is satisfied. Obviously, the more stringent this QoS, the shorter the network lifetime. In this section, we consider, as network lifetime QoS, the percentage of RFDs' deaths at which the overall network is assumed to be dead. This percentage is defined as *network tolerable death level* and we denote it with χ_{net}. This choice is motivated by the fact that χ_{net} quantifies the intuitive idea that a minimum number of observations (or a minimum spatial density of observations) may be required for proper network operations. In other words, if the lifetime QoS condition is stringent, the network is considered dead just after few RFDs' deaths.

According to results in the field of reliability theory [79], we model the lifetime of a single RFD as an exponentially distributed random variable with mean value equal to 300 s. We point out that our approach can be extended to account for any RFD lifetime distribution. In [80] the authors show the same results, considering arbitrary instead of specific time units, e.g., seconds, hours or days. In particular they confirm that the analysis, carried out here for a mean value equal to 300 s, is valid also for different values of average lifetime. We also point out that this definition is compliant with the lifetime definitions presented in Section 3.2.6 and Section 3.3.2. In fact, in this section we are dealing with arbitrary time units and exponential distribution of the lifetime of each RFD, but we could remove these assumptions and consider as energy depletion model the above presented lifetime analytical models.

3.4.2 Possible Network Configurations

In order to allow the deployment of large-scale networks, the sensors may be grouped into *clusters*, i.e., they transmit their data to intermediate nodes (denoted as *cluster-heads*), which may properly modify these data and relay them to the coordinator [80].

An illustrative representation of some of the network configurations of interest with $N = 16$ RFDs is given in Fig. 3.5. In particular, the presented schemes can be grouped into two main classes: (i) networks with *uniform* clusters (all clusters have the same number of sensors, as in Fig. 3.5 (a)) and (ii) networks with *non-uniform* clusters (the dimensions of the clusters may vary, as in the cases in Figs. 3.5 (b), (c), and (d)). In the scenarios with uniform clustering, the considered network configurations include: (i) one 16-node cluster and no relay (i.e., direct transmission from each RFD to the AP); (ii) one 16-node cluster with one relay (i.e., transmission from each RFD to the AP through a relay); (iii) two 8-node clusters with two relays (one relay per cluster); (iv) 4 4-node clusters with 4 relays (one relay per cluster); and (v) 8 2-node clusters with 8 relays (one relay per cluster). In the simulations with non-uniform clustering, instead, the adopted network configurations are: (i) 8-2-2-2-2 with 5 relays (i.e., the network is divided into 5 clusters, one formed by 8 RFDs and each of the other four formed by 2 RFDs, and each cluster is connected to the AP through a relay); (ii) 10-2-2-2 with 4 relays (i.e., the network is divided into 4

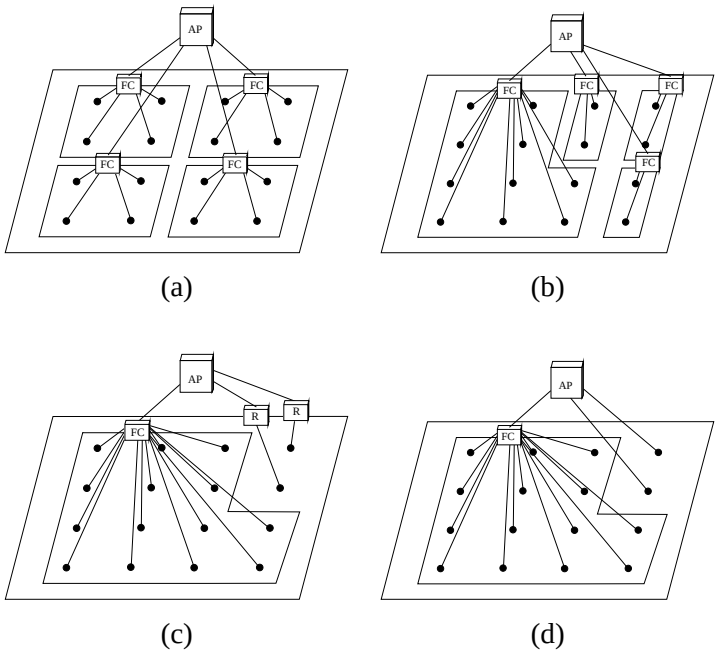


Figure 3.5: Illustrative clustered topologies: (a) uniform with 4 clusters and 4 relays, (b) non-uniform with a 10-2-2-2 configuration with 4 relays, (c) non-uniform with a 14-1-1 configuration and 3 relays, and (d) non-uniform with a 14-1-1 configuration and 1 relay. The nodes denoted with "R" are only dedicated to relay packets from the RFDs to the AP, whereas the nodes denoted with "FC" are also used to fuse data received from the RFDs.

clusters, one formed by 10 RFDs and other formed by 2 RFDs each, connected to the AP through a relay), as shown in Fig. 3.5 (b); (iii) 14-1-1 with 3 relays (i.e., one cluster is composed by 14 RFDs and two clusters are composed by one node each), as shown in Fig. 3.5 (c); and (iv) 14-1-1 with 1 relay (i.e., only one cluster composed by 14 nodes and connected to the AP through a relay, while the other two RFDs communicate directly to the AP), as shown in Fig. 3.5 (d). The network configurations adopted in the scenarios with 64 RFDs are similar to those in the case with 16 RFDs, except for the fact that the number of nodes in each cluster is quadruplicated (e.g.,

the configuration 14-1-1 with 3 relays is replaced by the configuration 56-4-4 with 3 relays).

The delay in the presence of relays is given by the sum of two terms: (i) the average[1] transmission delay between the RFDs and the relay and (ii) the average transmission delay between the relay and the AP. In particular, the first delay is calculated as the arithmetic average of the delays introduced by all relays. For example, in the 14-1-1 scenario with only one relay node (shown in Fig. 3.5 (d)), the delay is computed as follows:

$$D \simeq \frac{1}{16}\sum_{i=1}^{16}\overline{D}_i = \frac{1}{16}[\overline{D}_{\text{AP}} + \overline{D}_{\text{AP}} + 14\cdot(\overline{D}_{\text{AP}} + \overline{D}_{\text{relay}})] = \overline{D}_{\text{AP}} + \frac{14}{16}\cdot\overline{D}_{\text{relay}} \quad (3.51)$$

where $\overline{D}_{\text{AP}}$ is the average delay of a direct transmission to the AP (either from an RFD or the relay) and $\overline{D}_{\text{relay}}$ is the delay introduced by the communication from an RFD to the relay.

3.4.3 Sensing and Data Fusion

In this subsection, we summarize the data fusion mechanism proposed in [81], which will be used in our simulator. The reader is referred to [81] for more details. In general, N remote sensors (namely, the RFDs) observe (in a *noisy* manner) a *common* binary phenomenon H, defined as

$$H = \begin{cases} H_0 & \text{with probability } p_0 \\ H_1 & \text{with probability } (1-p_0) \end{cases} \quad (3.52)$$

where $p_0 \triangleq P(H = H_0)$. Throughout this section, we will consider equally distributed phenomena, i.e., $p_0 = 1/2$. As previously introduced, the RFDs may be clustered into $n_c < N$ groups, and each RFD communicates only with its corresponding relay, which acts as a fusion center (FC). The FCs collect data from the RFDs in their corresponding clusters and make local decisions on the status of the binary phenomenon. Then, each local FC transmits over an ideal wireless channel to the coordinator,

[1] In the following, the expectation is computed over the simulation time.

which makes the final decision. Being the observed signal the same across the RFDs and assuming that the observation noises are Gaussian and independent with the same distribution $\mathcal{N}(0,\sigma^2)$, the common signal-to-noise ratio (SNR) at the sensors can be defined as follows [81]:

$$\text{SNR} = \frac{s^2}{\sigma^2} \tag{3.53}$$

where s is the intensity of the observed signal. Each sensor makes a decision comparing its observation r_i with a threshold value τ_i and computes a local decision $u_i = U(r_i - \tau_i)$, where $U(\cdot)$ is the unit step function. In order to optimize the system performance, the thresholds $\{\tau_i\}$ need to be optimized. Even though, in general, a common value of the decision threshold for all sensors might not be the best choice, in the following we assume a fixed (optimized) threshold value τ equal to $s/2$ [81].

Uniform Clustering

In a scenario with *uniform* clustering, the sensors are grouped into identical clusters, i.e., each of the n_c clusters contains d_c sensors, with $n_c \cdot d_c = N$. An illustrative example, in a scenario with $N = 16$ and $n_c = 4$, is given in Fig. 3.5 (a).

The fusion rules at both the FCs and the AP are majority-like fusion rules with the fusion thresholds set, respectively, to k and k_f. Using a combinatorial approach (based on the repeated trials formula [68]) and taking into account the majority-like fusion rules, the probability of decision error at the coordinator can be expressed as follows [81]:

$$\begin{aligned} P_e &= P(\widehat{H} = H_1|H_0)P(H_0) + P(\widehat{H} = H_0|H_1)P(H_1) \\ &= p_0\,\text{bin}(k_f, n_c, n_c, \text{bin}(k, d_c, d_c, 1-\Phi(\tau))) \\ &\quad + (1-p_0)\,\text{bin}(0, k_f - 1, n_c, \text{bin}(k, d_c, d_c, 1-\Phi(\tau - s))) \end{aligned} \tag{3.54}$$

where $\Phi(x) \triangleq \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp(-y^2/2)\,\mathrm{d}y$ and

$$\text{bin}(a, b, n, z) \triangleq \sum_{i=a}^{b} \binom{n}{i} z^i (1-z)^{(n-i)}\,.$$

Non-Uniform Clustering

In this case, the RFDs do not aggregate regularly. In particular, the i-th cluster has size $d_{\mathrm{c}}^{(i)}$ ($i = 1, \ldots, n_{\mathrm{c}}$) and it hods that $\sum_{i=1}^{n_{\mathrm{c}}} d_{\mathrm{c}}^{(i)} = N$. Since the RFDs are not equally distributed among the clusters, the optimized decision thresholds at the RFDs could depend on the considered cluster. We remark that the fusion rules at the FCs and the AP are majority-like fusion rules with proper values of the fusion thresholds, as in Subsection 3.4.3.

Let us define $p_\ell^{1|1}$ ($p_\ell^{1|0}$, respectively) as the probability that the ℓ-th FC decides for H_1 when H_1 (H_0, respectively). After a few manipulations, it can be shown that the probability of decision error can be expressed as follows [81]:

$$\begin{aligned} P_{\mathrm{e}} &= p_0 \sum_{i=k_{\mathrm{f}}}^{n_{\mathrm{c}}} \sum_{j=1}^{\binom{n_{\mathrm{c}}}{i}} \prod_{\ell=1}^{n_{\mathrm{c}}} \{\mathrm{string}(i,j,\ell) p_\ell^{1|0} + (1 - \mathrm{string}(i,j,\ell))(1 - p_\ell^{1|0})\} \\ &\quad + (1-p_0) \sum_{i=0}^{k_{\mathrm{f}}-1} \sum_{j=1}^{\binom{n_{\mathrm{c}}}{i}} \prod_{\ell=1}^{n_{\mathrm{c}}} \{\mathrm{string}(i,j,\ell) p_\ell^{1|1} + (1 - \mathrm{string}(i,j,\ell))(1 - p_\ell^{1|1})\} \end{aligned} \tag{3.55}$$

where $\mathrm{string}(i,j,\ell) = 1$ if there is a "success" (corresponding to a decision, at an FC or at the coordinator, in favor of H_1), whereas it is 0 if there is a "failure" (corresponding to a decision, at an FC or at the coordinator, in favor of H_0). In other words, $\mathrm{string}(i,j,\ell)$ is an auxiliary binary function used to distinguish, in the repeated trials formula [68], between a success and a failure. It can be shown that (3.55) reduces to (3.54) in the presence of uniform clustering [81].

3.5 Concluding Remarks

In this chapter, we have presented a global analytical framework which allows to optimally tune the node parameters in order to: (i) reduce the probability of missed detection of an incoming target and minimize the latency of alert notification after a target detection, (ii) minimize the collisions between the packets at the AP of the network, and (iii) maximize the lifetime of the network according to a given QoS. In

the following, we will confirm the validity of our models, showing some performance results obtained applying the analytical models presented in this section.

Chapter 4

Performance Evaluation

4.1 Introduction

In this chapter, we will present the performance results of the analytical framework presented in Chapter 3. Whenever possible, the theoretical results will be compared to those obtained through either the simulator or the experimental testbed, in order to effectively check the validity of the described framework. In particular, the use of a simulator is useful to simplify the analysis of the performance of a WSN because it is possible to (i) simulate the behavior of the network even in the presence of a large number of nodes (which may not be available in the reality) and (ii) reduce the duration of the analysis since it is possible to "virtualize" the time. The analytical framework allows also to derive a strategy to improve the performance of the required indicators according to some given constraints.

The structure of this chapter is the following. In Section 4.2 we introduce the structure of the simulator and we will introduce the main characteristics of the implemented models. In Section 4.3, we present the performance of the target detection model, together with an approach for optimally tuning the node's physical parameters. In Section 4.4, we show the performance of the adjacency-based power allocation strategy. In Section 4.5, we describe the results of both the data fusion mechanism and the cluster approach. Finally, Section 4.6 concludes this chapter.

4.2 The Opnet Simulator

Since the experimental validation of the analytical results sometimes requires a large number of nodes, which often are not available, a possible solution is the use of a simulator, which models the implementation of the layers of the ISO/OSI standard. The simulator that we have used is the Opnet Modeler [82], since it provides for both the slotted and the unslotted models of a IEEE 802.15.4 device.

The Opnet model has a hierarchical structure. In particular, it is possible to distinguish between three different tiers: (i) network, (ii) node, and (iii) process tier. The node model specifies object in network domain, whereas the process model specifies object in node domain. In particular, in the network domain it is possible to physically place the nodes, setting the distances between them, their speed and the communication links. In the node model, instead, it is possible to configure the internal queues of the node and the transceivers interfaces, whereas in the process domain, there is the C implementation of the algorithms which govern the layers of the ISO/OSI stack. This third layer is modeled as a finite state machine, whose transitions are due to the occurrence of events, that is particular activities which occur at a certain time. For example, an event can be associated to the reception of a packet from a remote node or from a different layer of the ISO/OSI stack, or to the expiration of an internal timer.

The Opnet modeler is a discrete event simulator. This means that simulation time progresses only when an event occurs. We point out that the simulation time is different from the real time. Since the simulation time is not continuous, when an event occurs, the simulator , after executing the instructions associated to the event, “jumps” to the following event or, more precisely, to the time instant associated to the next scheduled event. Therefore, the real time interval between two subsequent events is skipped by the simulator. Conversely, when no events are scheduled, the simulation time does not progress.

The Opnet simulator allows to manage different types of packets. Normally, the packets carry unformatted data, that is no information is transmitted within the packets. However, it is also possible to transmit formatted data, which will be processed by the destination node.

In this chapter, we introduce only the model which refers to the unslotted implementation of the IEEE 802.15.4 standard, since the performance results presented here refers to the unslotted version of the CSMA/CA protocol. Conversely, the integrated RFID-controlled model is based on the slotted version of the IEEE 802.15.4 standard, which is presented in Chapter 5.

The basic unslotted model of the IEEE 802.15.4 is a built-in Opnet model created at the National Institute of Standardization and Technologies (NIST) [83]. The main characteristics of the unslotted version of the IEEE 802.15.4 implementation of the protocol are the following:

- no difference between FFD and RFD nodes;
- no implementation of router nodes;
- standard implementation of the CSMA/CA protocol, as described in Section 2.4;
- optional transmission of acknowledgment (ACK) messages;
- there is no attenuation on the transmission of the packet, that is two packets are lost only in the case of collisions;
- the channel is modeled as an infinite buffer;
- no modelization of the energy consumption of the node.

Since no router nodes are implemented in the basic model of the IEEE 802.15.4 standard, we have developed a node which “relays” the received packets towards the AP node. As soon as the relay node receives the packet, it tries to retransmit it towards the AP using the CSMA/CA MAC algorithm. When required, this nodes can also act as cluster-heads of the cluster. In addition, this node can also fuse the information received by the RFDs associated to it.

4.3 Target Detection Performance

In this section, we present the performance related to the theoretical model presented in Section 3.2. We first validate, through simulations, the analytical model for the determination of the P_{md} of a target crossing the monitored area[1]. Then, we compare the latency performance of the proposed model with the results carried out through an experimental testbed. Finally, after providing some results related to the simplified energy model, we introduce the optimization framework, which allows to tune the node's parameters in order to maximize/minimize the required performance indicators.

4.3.1 Validation

In order to validate the analytical framework, we analyze the probability of missed detection P_{md} through simulations. The reference model for the simulation set-up has been described in Sec. 3.2.2. In order to reduce possible statistical fluctuations due to random node placement and random trajectory of the target, the simulations have been repeated 1000 times over 1000 different scenarios. In Fig. 4.1, we show P_{md} as a function of β_{sens}, considering different values of r_{s}. The number of nodes is $N = 50$ and the speed of the target is $v = 15$ m/s. As the intuition may suggest, the longer the sensing range, the higher the probability of detecting any incoming target. When the value of β_{sens} becomes small, the target can cross the sensed area during the sleep period of the sensor without being detected, thus increasing the probability of missed detection. In Fig. 4.1, the simulation results (solid lines) are compared with the theoretical ones (dashed lines). The results show a good agreement between the two models, especially for large values of r_{s}, confirming the validity of the analytical model.

In Fig. 4.2, we evaluate P_{md} as a function of β_{sens}. In all cases, the target speed v is set to 15 m/s. In this case as well, the same considerations carried out for Fig. 4.1 hold. When T_{sens} is sufficiently small, β_{sens} has a limited impact on P_{md}. On the opposite,

[1] We point out that the simulations carried out in this section are obtained through a specifically designed simulator and not through the use of the Opnet simulator.

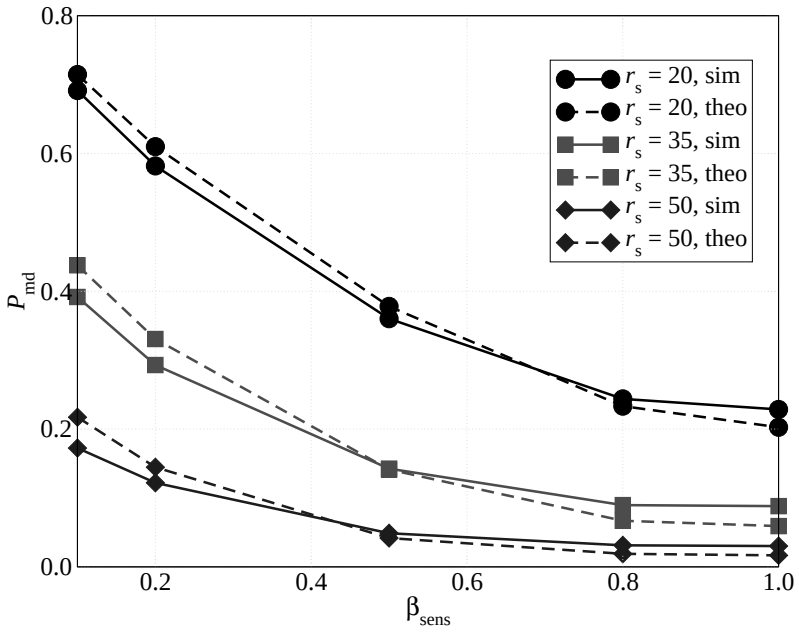


Figure 4.1: Simulation (solid lines) and analytical (dashed lines) P_{md} results as functions of the duty cycle β_{sens} considering different sensing ranges. The target enters with speed $v = 15$ m/s a monitored area of $d_{\text{s}} = 1000$ m side, where $N = 50$ sensors are randomly deployed.

the larger T_{sens}, the larger the impact of β_{sens}. When $\beta_{\text{sens}} = 1$, the sensing period has no impact on the performance on the system. Instead, when $\beta_{\text{sens}} < 1$, a larger T_{sens} leads to a larger P_{md}, because it is more likely that the target cross the sensed area during the sleep phase of the sensor, without being detected. Also in this case, there is a good agreement between the simulation (solid lines) and theoretical (dashed lines) curves.

4.3.2 Latency after Detection

In order to verify the analytical model of the multi-hop alert transmission latency, we have run a set of experimental tests with a testbed of 4 Crossbow MicaZ nodes

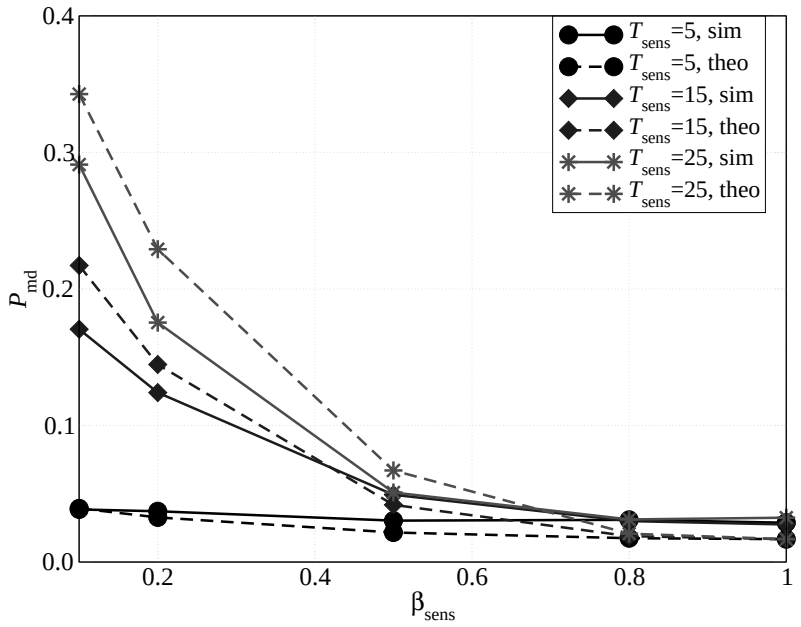


Figure 4.2: Simulation (solid lines) and analytical (dashed lines) P_{md} results as functions of the duty cycle β_{sens}, considering different period lengths T_{sens}. The size of the area is $d_{\text{s}} = 1000$ m, in which $N = 50$ sensors, with sensing range equal to $r_{\text{s}} = 50$ m, are randomly placed. The speed of the target is $v = 15$ m/s.

deployed in chain topology. The first node injects a packet every 2 s. Each subsequent node forwards the packet to its neighbor until the packet reaches the last node, which reverses the transmission's direction. We have measured the round trip time ranging N_{hop} from 2 to 6, and β_{comm} to either 0.067, 0.1 and 0.2.[2] For each pair of values, we compute the average latency as the mean round-trip time of 100 samples, along with the 95% confidence interval. In Fig. 4.3, a comparison between the theoretical and experimental results is shown. The curves are quite close, even though a gap appears when β_{comm} becomes lower and the duration of the sleep interval increases. This may

[2]These values of β_{comm} correspond to a static active period of 8ms and T_{comm} respectively equal to 120ms, 80ms and 40ms.

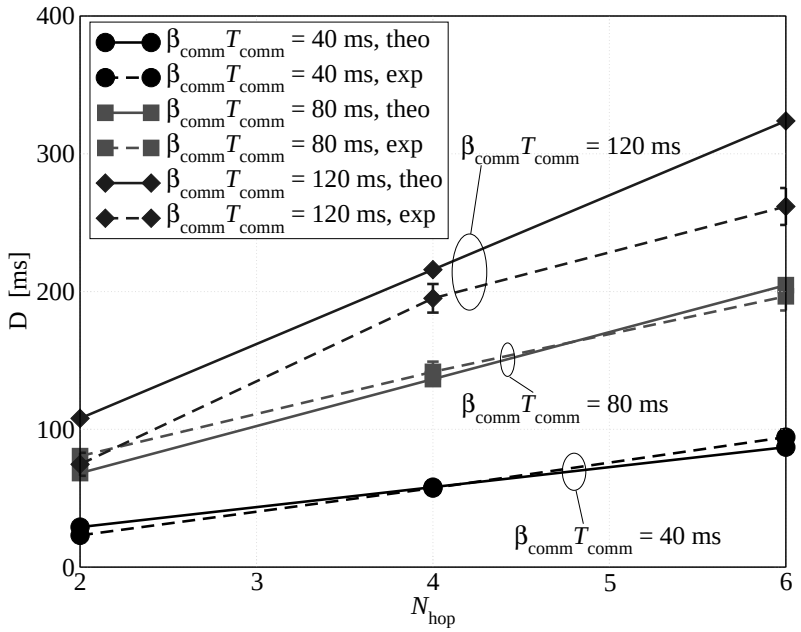


Figure 4.3: Latency as a function of the number of hops traversed by a packet. Both (i) experimental (dashed lines) and (ii) theoretical results (solid lines) are presented.

come from the unavoidable alignment of sleep schedules on the packet's way back, which tends to reduce the latency for large values of sleep interval durations, i.e., small values of β_{comm}. In the context of the reference scenario, the experiments show the validity of the analytical model, so that the latency can be roughly approximated as $D \simeq N_{\text{hop}}(1-\beta_{\text{comm}})T_{\text{comm}}/2$.

4.3.3 System Engineering

This subsection illustrates the use of an engineering toolbox for optimally configuring the network. First, we investigate the space of optimum configuration solutions under the perspective of the trade-offs faced by the kind of WSN under consideration. Then, we present an application case in which a given surveillance system is optimally configured to maximize the sustainability of the network, i.e., it lifetime L, under

certain quality of service requirements in terms of probability of missed detection (P_{md}) and latency for alert transmission to the AP (D).

The engineering toolbox consists of the three expressions derived in previous sections for P_{md}, L, and D. As equations (3.11) and (3.22) are not linear, standard optimization techniques for linear programming cannot be used. However, the three equations identify a convex set, which makes gradient-based optimization techniques feasible [84]. When two or more functions need to be optimized, we used multi-objective optimization techniques, which allow to simultaneously optimize conflicting objective functions subject to certain (if any) constraints. Throughout this section, the target arrival rate N_{target} is fixed to 10 targets per day.

System Trade-offs

This subsection investigates the space of optimum configuration parameters (number of nodes N, duty cycles β_{comm} and β_{sens}), using unconstrained optimization, to highlight the trade-offs faced by the considered WSN. In order to solve the optimization problem, we choose a Pareto-compliant ranking method based on evolutionary techniques, namely the Non-dominated Sorting Genetic Algorithm-II (NSGA-II) [85]. Fig. 4.4 shows the result of the joint optimization of the three objective functions, namely latency, lifetime, and probability of missed detection. These results show that the longer the latency or the higher the probability of missed detection, the longer the lifetime. In particular, when the latency is short, i.e., the nodes must have the communication interfaces on for a large portion of the interval T_{comm}, the lifetime is affected mainly by the duty cycle β_{comm}, so that the impact of β_{sens} is negligible. On the other hand, for long latency values the nodes can keep β_{sens} low, and, consequently the impact of P_{md} is far more pronounced.

The important outcome of the results in Fig. 4.4 is that it provides details about the optimal network configuration, since tuning the network with the parameters derived from the execution of the NSGA-II algorithm guarantees that the network performance is not biased toward one of the performance indicators.[3]

[3]The D axis ends at 0.2 s because the maximum latency value, which is associated to the smallest value of β_{comm} (i.e., 0.0025) allowed by the X-MAC protocol, is 0.15 s over the considered average

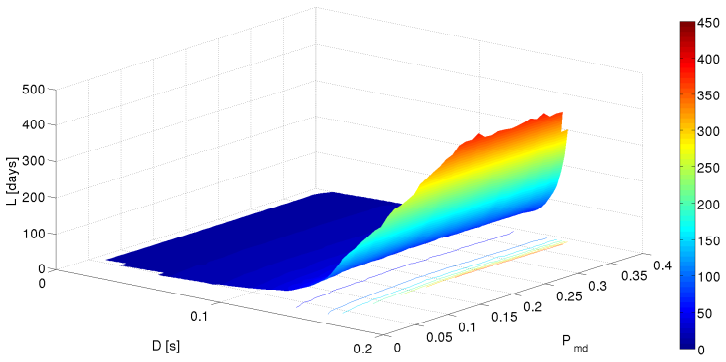


Figure 4.4: Joint optimization of D, P_{md}, and L. No constraints are imposed.

An Application Case

This subsection investigates a realistic use case where one wants to maximize the lifetime of a given WSN according to given constrained maximum values of D and P_{md}, denoted as P^*_{md} and D^*, respectively. This case consists of the optimization of a single-objective function, given constraints on the two other functions. The optimal parameters can be evaluated using single-objective convex programming techniques. The `fmincon` Matlab toolbox or the optimization approach presented in Subsection 4.3.3 can be used. In Fig. 4.5, the lifetime of the network has been maximized under the constraints $P_{\text{md}} < P^*_{\text{md}}$ and $D < D^*$. The optimization has been carried out considering several values of P^*_{md} and D^*. In particular, for each considered pair of values, we have evaluated the maximum allowed lifetime that can be obtained. Of course, the more stringent the requirements on D and P_{md}, the shorter the longest achievable lifetime. In fact, in the case of small values of P^*_{md} and D^*, a sensor must keep on its sensing interface (to minimize P_{md} and D) and its communication interface (to minimize D) for a large portion of the period, so that the energy consumption increases and the lifetime reduces. On the other hand, when the requirements are less

3-hop path.

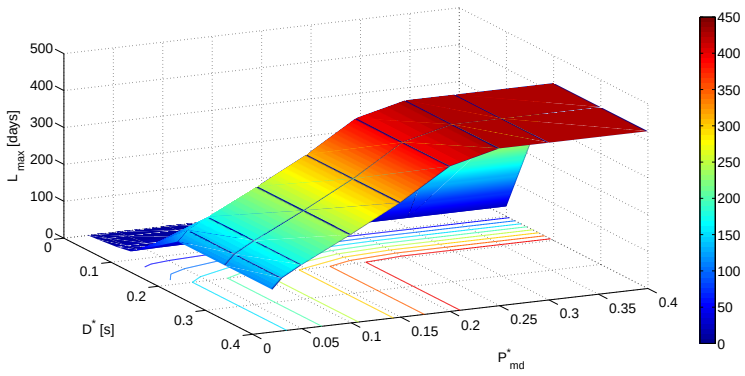


Figure 4.5: Lifetime maximization under probability of missed detection (P^*_{md}) and latency (D^*) constraints.

stringent, the sensing and communication interfaces can be switched off for a longer portion of the period and, consequently, the lifetime increases.

Focusing on the shape of the surface, generated by interpolation of the simulation results, it is possible to understand the contribution of both sensing and communication operations. For a given value of P^*_{md}, one can observe that the shape of the projection of the network lifetime over the $D^* - L_{\text{max}}$ plane remains the same. In the case of the joint optimization of the three metrics, the maximum latency is limited to 0.15 s, and this is why the D axis ranges between 0 s and 0.2 s. In this case, instead, the D^* and P^*_{md} axes refer to the constraints required by a given application and, therefore, they can range in a wider interval. When the latency requirements are stringent, the lifetime is short. Obviously, if if the latency requirements are relaxed, the lifetime increases till saturation to a value which depends on P^*_{md}. A short latency corresponds to a short network lifetime, since (i) the power consumed during reception is several orders of magnitude larger than the power consumed in the sleep phase and (ii) the communication interface of each node must be on for a large portion of the communication period T_{comm} in order to ensure that a packet is delivered to the AP in a short time. The impact of the latency on L_{max} becomes negligible, i.e., the

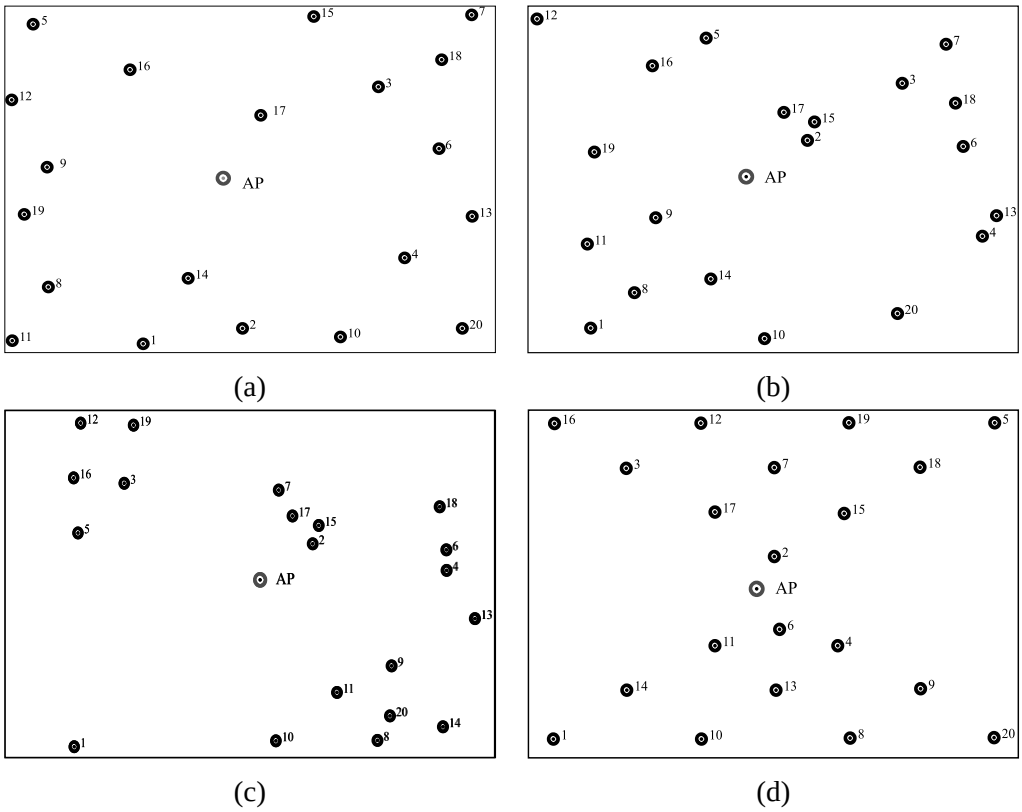
lifetime saturates, with respect to D^*, when $D^* \simeq 0.15$ s. In fact, this value corresponds to the minimum allowed value of β_{comm} that guarantees that a preamble is correctly received in order for a communication to start. Therefore, larger values of D^* has no impact on the latency, since lower values of β_{comm} cannot be selected.

Focusing on the $P^*_{\text{md}} - L_{\max}$ plane, it turns out that the power consumption due to sensing operations is basically negligible when the latency requirements are limited. In fact, the power consumption associated with the sensing interface is one order of magnitude lower than that related to the reception operations. When the maximum tolerable latency becomes longer than 0.15 s, the power consumption related to the communication interface remains constant and the lifetime is affected only by the sensing power consumption. In this case, the lower P^*_{md} (i.e., the more stringent the constraint on the probability of missed detection), the higher the energy consumption, since a seismic sensor must be on for a longer interval.

Unlike the case with joint optimization (with three objective functions) considered in Fig. 4.4, in this case the parameters associated with the given values of D^* and P^*_{md} may not lead to optimum network configuration, even if the lifetime is maximized. In fact, the solution obtained with a single objective minimization may lead to a "dominated" network configuration, i.e., a solution where one performance indicator is dominating over the others. This approach, however, guarantees that the network parameters are correctly configured and that no energy is wasted by the nodes during their communication/sensing operations.

4.4 Adjacency Matrix-based Transmit Power Allocation Strategy

In this section, we present the results of the power allocation strategy presented in Section 3.3. We first present the Opnet model that we have considered for our simulations. Then, we show the performance of the Zigbee networks, in terms of PER, delay, and network lifetime, focusing on the impact of the adjacency matrix structure, the traffic load, and the used power allocation strategy.



(a) (b)

(c) (d)

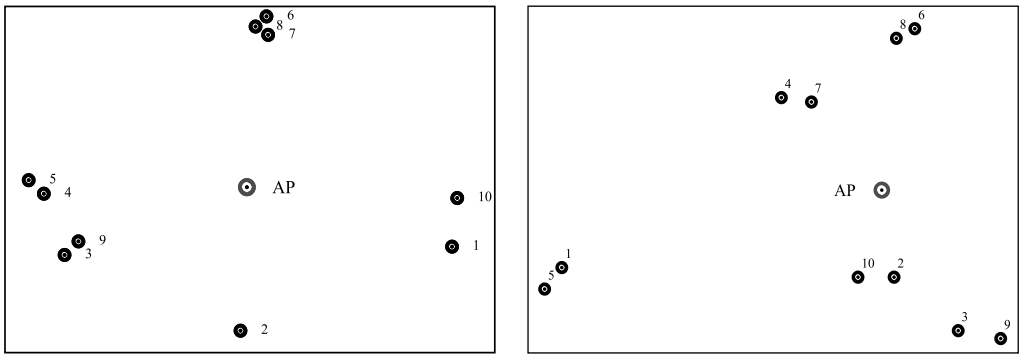
Figure 4.6: Considered network topologies with $N = 20$ nodes.

4.4.1 Considered Opnet Model

The simulations have been carried out with the Modeler package of the Opnet simulator [82] and a built-in Zigbee network model designed at the National Institute of Standards and Technologies (NIST) [83]. We have considered a scenario where N nodes transmit directly to the AP. In particular, the considered topologies for $N = 20$ are shown in Figure 4.6, whereas those for $N = 10$ are shown in Figure 4.7.

More precisely:

- in Figure 4.6 (a), $N = 20$ nodes are randomly deployed over a 100 m^2 square



(a) (b)

Figure 4.7: Considered network topologies with $N = 10$ nodes.

area (the width of the side of the surface will become meaningful for the typical values of the transmit power considered in the following. Moreover, the maximum transmission range allowed by the Zigbee standard is 100 m) and are approximately concentrated towards the external perimeter of the surface (we point out that the considered surface for $N = 10$ sensors is smaller than that for $N = 20$ sensors);

- in Figure 4.6 (b), $N = 20$ nodes are deployed over the same surface as before, but present a few cluster and isolated nodes;
- in Figure 4.6 (c), $N = 20$ nodes are placed in order to form four small groups and only one node is isolated from the others;
- in Figure 4.6 (d), $N = 20$ nodes are placed over a regular grid and form two "triangular" grids which converge at the AP;
- in Figure 4.7 (a), $N = 10$ nodes are approximately at the same distance from the AP and form small groups isolated from each other;
- in Figure 4.7 (b), $N = 10$ nodes are clustered in groups of two. In particular, four pairs of nodes are placed near the AP, whereas the remaining pair is far

from the AP.

We believe that the considered topologies are representative of a large set of possible WSN topologies. However, we remark that the proposed framework can be applied to a WSN with a generic topology.

Since the proposed power allocation strategy aims at PER minimization, we have considered the network topology presented in Figure 4.8 (a) to highlight the performance gain given by the proposed adjacency-based power allocation scheme.

In order to highlight the impact of the proposed power allocation strategy on the network lifetime, we have considered two scenarios with $N = 10$ nodes randomly deployed over a 10 m^2 square surface and over 50 m^2 square surface. These topologies are shown in Figure 4.8 (b) and Figure 4.8 (c), respectively.

Since the NIST Zigbee network Opnet model was developed to analyze the coexistence between IEEE 802.15.4 and IEEE 802.11 standards in small environments, it did not take into account signal attenuation [86]. In our simulations, instead, we have neglected the impact of co-existing IEEE 802.11 networks and we have introduced the channel attenuation according to the Friis propagation model. In particular, the Friis formula is given by Equation 3.23 and we assume $G_{\mathrm{r}} = G_{\mathrm{t}} = 1$ (omnidirectional antennas), $\lambda = 0.125$ m ($f_{\mathrm{c}} = 2.4$ GHz), and $\alpha = 2.1$. In all cases, r is shorter than 100 m, which is the maximum transmission range allowed by the Zigbee standard. If the received power is higher than a pre-defined threshold, fixed to -90 dBm, the nodes can exchange packets.

For each of the considered topologies, the distance between the nodes and consequently the power attenuation is computed offline on the basis of the coordinates of the nodes. These values are then used to fill the adjacency matrix. In particular, consider a pair of nodes (s_i, s_j) with $i \neq j$: if s_i is sufficiently close to transmit to s_j, we insert a "1" in the corresponding entry of the adjacency matrix (i.e. the ith row and the jth column); otherwise, we mark the absence of communication with a "0". We remark that the communication links may be asymmetric: even if s_i can communicate with s_j, the opposite may not hold. The distances between the nodes are also used to determine (i) the minimum (per-node) transmit power which allows each node to reach the AP and (ii) the maximum transmit power which guarantees

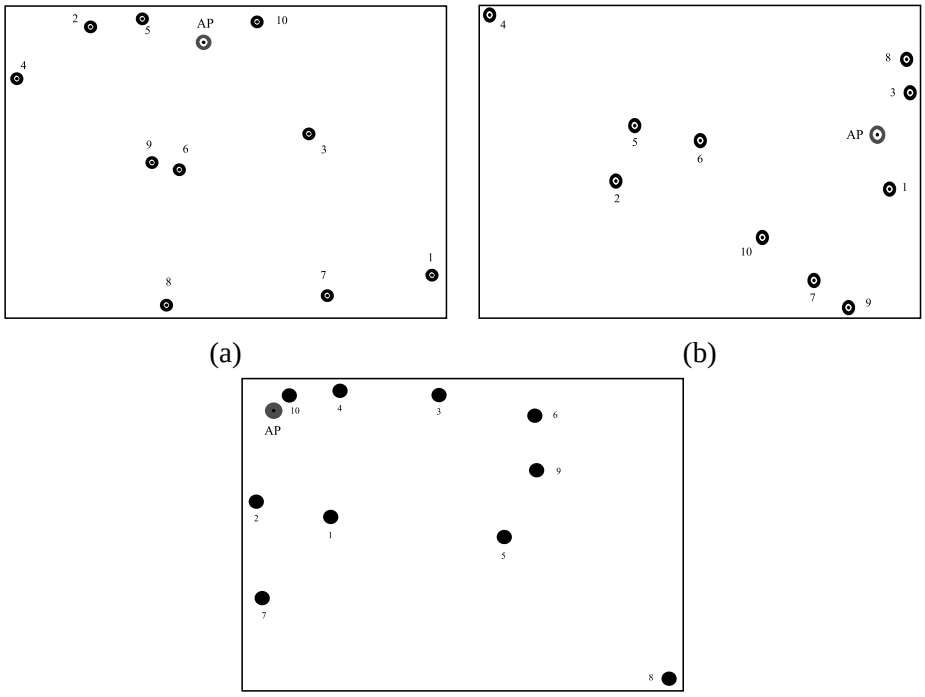


Figure 4.8: Network topologies with $N = 10$ nodes randomly deployed over a 10 m^2 square surface, used for (a) PER comparison and (b) evaluation of the network lifetime. (c) Network topology with $N = 10$ nodes randomly deployed over a 50 m^2 square surface, used for the evaluation of the network lifetime.

that each node can reach any other node in the network.

The Zigbee standard provides indications about the values of the main network parameters introduced in Chapter 2. The values of the relevant parameters for our simulations are shown in Table 4.1. We remark that the Opnet simulator expresses all time-related parameters as multiples of the fundamental time unit, which corresponds to the inverse of the transmission data rate R. The simulations have been repeated several times with different seed initialization parameters in order to ensure that possible

Table 4.1: Parameters of the Zigbee standard.

Fundamental time unit	$4\,\mu$s
LIFS	$640\,\mu$s
T_{CCA}	$128\,\mu$s
ACK window duration	$864\,\mu$s
T_{TAT}	$192\,\mu$s
T_{B}	$320\,\mu$s
L (packet length)	512 (payload) + 120 (header) bits

statistical fluctuations are avoided. The Opnet simulator also stores into log files the values of important metrics related to (i) packet transmission, such as the numbers of correctly received packets and noisy packets, and (ii) packet generation, such as the numbers of sent packets and dropped packets.

We remark that the simplified theoretical model presented in Section 3.3.1 is compliant with the simulation model just described.

4.4.2 Performance Analysis

In this subsection, we present the performance results in the presence of transmit power control. In particular, we focus on the following key performance indicators: (i) PER, (ii) delay D (dimension: s), and (iii) network transmission rate S (dimension: bit/s). The delay is defined as the average time interval between transmission and correct reception instants of a data packet. The network transmission rate is defined as the number of bits correctly received by the AP per unit of time. In addition, we present the performance results, in terms of residual energy, of the proposed power allocation strategy by comparing them with those obtained by the power allocation strategy proposed in [17].

The simulations have been carried out using different values of the overall network transmit power and, consequently, different values of the transmit powers allocated to the sensors. In particular, we have considered two possible transmit power

allocation strategies: (i) each node has the same transmit power (uniform power allocation); (ii) the transmit power varies from node to node and is allocated using the strategy presented in Subsection 3.3.3; (iii) the transmit powers are allocated according to the strategy proposed in [17]. In all cases, the obtained simulation results are directly compared with the results predicted by the theoretical model. In fact, referring to the scenarios shown in Figure 4.6 and Figure 4.7, we have first set the same transmission power at each node in order to allow (a) each couple of nodes to communicate with each other (the used transmit power is denoted as $P_{\mathrm{t}}^{\max}$) and (b) each node to reach at most the AP (the per-node transmit power is denoted as $P_{\mathrm{t}}^{\min}$). In the following, we will denote as $\{P_{\mathrm{i}}\}$, $i = 1,\ldots,N$ the transmit powers assigned to the nodes using the proposed power allocation strategies, and denote the overall available power as P_{tot}. In particular, we will denote as $P_{\mathrm{tot}}^{\max} = P_{\mathrm{tot}}^{\max}/N$ the overall transmit power that guarantees that each node, using the same transmit power of $P_{\mathrm{t}}^{\max} = P_{\mathrm{tot}}^{\min}/N$, can transmit to any other node. Similarly, we will denote as $P_{\mathrm{tot}}^{\min}$ the overall power that guarantees that each node, using a transmit power of $P_{\mathrm{t}}^{\min}$, can reach at most the AP. We also remark that, in any case, the following condition will hold:

$$\sum_{i=1}^{N} P_i \leq P_{\mathrm{tot}}.$$

4.4.3 Validation of the Analytical Model

Through simulations, we first validate the assumptions, behind the analytical model, of neglecting the impact of the backoff exponents on the network performance. We have compared the PER in a scenario with $N = 20$ nodes, different values of the backoff exponent and without the use of ACK messages. The results are shown in Figure 4.9. The solid lines refer to scenarios where each node can communicate with any other node in the network (the common per-node transmit power is $P_{\mathrm{t}}^{\max}$), whereas the dashed lines refer to scenarios where minimum common per-node transmit power (equal to $P_{\mathrm{t}}^{\min}$) is used. In both cases, the line with circles and the line with squares, which refer to scenarios with default and modified backoff exponents, basically overlap. This fact confirms the analytical assumption that the backoff exponent

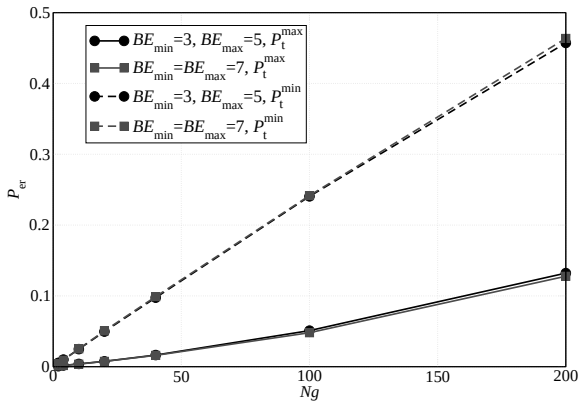
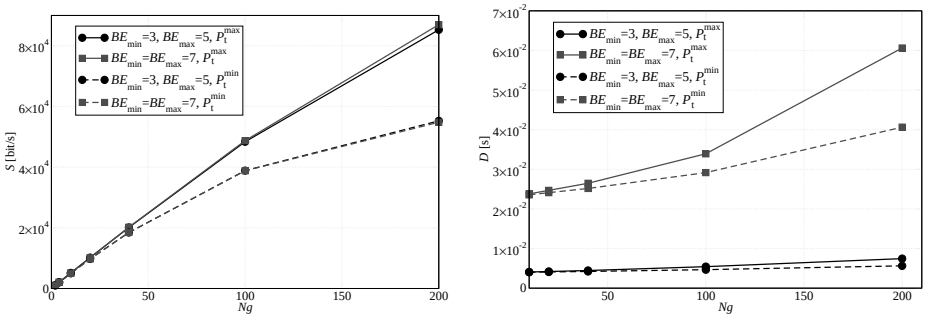


Figure 4.9: PER as a function of the total offered traffic load. The simulation results are obtained considering (i) default backoff exponent, i.e. $BE_{\min} = 3$ and $BE_{\max} = 5$, and (ii) modified backoff exponent, i.e. $BE_{\min} = BE_{\max} = 7$. Different values of per-node transmit powers are considered. The allocated transmit power is the same at each node.

has a very limited impact on the network performance and it can be neglected, thus simplifying the theoretical model. A larger backoff exponent does not affect the performance in terms of PER, because the default value of the backoff exponent is large enough to decorrelate the backoff intervals of two nodes that could not transmit. In fact, according to the CSMA/CA MAC protocol described in Chapter 2, a node must double the range of the backoff interval and retry to transmit the packet after sensing the channel and finding it busy. If another node performs the same operations at the same time, some sort of correlation between the two transmitting nodes may emerge. However, the use of the random backoff interval guarantees that the two nodes will not collide at the subsequent transmission attempt.

The validity of the analytical model has also been verified in terms of delay and network transmission rate. The corresponding results are shown in Figure 4.10. As for the PER results in Figure 4.9, the network transmission rate is not affected by the use of different backoff exponent values. In fact, this performance indicator is strictly



(a) (b)

Figure 4.10: (a) Network transmission rate and (b) delay, as functions of the total offered traffic load (Ng), in the presence of (i) default backoff exponent, i.e. $BE_{\min} = 3$ and $BE_{\max} = 5$, and (ii) modified backoff exponent, i.e. $BE_{\min} = BE_{\max} = 7$. Different values of transmit powers are considered. The allocated transmit power is the same at each node.

related to the PER, therefore, recalling the results shown in Figure 4.9, the network transmission rate remains basically the same for different values of $BE_{\max}$. Considering Figure 4.10 (a), the solid lines, which refer to the case with default backoff exponent, and the dashed lines, which refer to the case with modified backoff exponent, basically overlap. On the other hand, the use of different backoff exponents affects the delay performance. Since the backoff window is larger in the case with $BE_{\min} = BE_{\max} = 7$, a node may wait for a longer period before sending the packet on the channel. Considering Figure 4.10 (b), it can be observed that the delay is longer in the case with the modified backoff exponent. In fact, in this case a node has to wait, on average, for a longer period before transmitting a packet. In addition, using the transmit power $P_{\mathrm{t}}^{\max}$, the delay is higher because a node is more likely to sense other transmitting nodes during its CCA and, in this case, waits for a longer period before transmitting. On the other hand, if the common transmit power is set to $P_{\mathrm{t}}^{\min}$ and no ACK mechanism is used, it is less likely that a transmitting node will sense another simultaneously transmitting (and thus colliding) node.

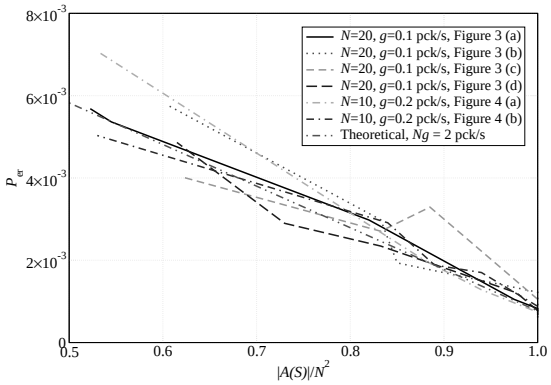


Figure 4.11: PER as a function of the sparsity index of the adjacency matrix. In the simulations, for scenarios with $N = 20$ nodes, the packet generation rate is set to $g = 0.1$ pck/s, whereas for scenarios with $N = 10$ nodes, the packet generation rate is set to $g = 0.2$ pck/s.

4.4.4 Impact of the Adjacency

According to the analytical results in Section 3.3.1, the performance, in terms of PER, depends only on the adjacency, regardless of their specific positions. This emerges clearly from the results shown in Figure 4.11, where the PER is shown as a function of $|A(\mathscr{S})|/N^2$, i.e. the sparsity index of the adjacency matrix. In this figure, the performance with the network topologies in Figure 4.6 and Figure 4.7 is evaluated. For the simulation results, in the scenarios with $N = 20$ nodes, the packet generation rate is set to $g = 0.1$ pck/s, whereas in scenarios with $N = 10$ nodes, the packet generation rate is set to $g = 0.2$ pck/s, in order to keep the product Ng (i.e. the overall traffic load) constant and make the comparison between different topologies meaningful. In the same figure, the PER predicted by the analytical model, given by the expression in Equation 3.31, is also shown. As one can see, the simulation curves are very close to the corresponding analytical curves and this is more pronounced for values of $|A(\mathscr{S})|/N^2$ in the proximity of 1. In fact, when the value of $|A(\mathscr{S})|/N^2$ is close to 1, the network is strongly connected and a node can sense any other node. Observing

the Opnet log files (not reported here for lack of space) stored by the nodes during the simulations, when $|A(\mathscr{S})|/N^2$ approaches 1 (i.e. the network is fully connected and during a CCA operation each sensor can detect the transmissions of all other sensors), the packets are dropped only during the TAT, when a node cannot sense other active nodes in the network during the transmission of its packets. On the other hand, when the value of $|A(\mathscr{S})|/N^2$ decreases, some nodes may become isolated from the other nodes (except for the AP) and may no longer be able to sense them, so that those packets may collide at the AP, leading to a PER increase.

Referring to Figure 4.11, the topology of the nodes in the the network has a very limited impact on the PER when $|\bar{A}(\mathscr{S})|/N^2$ is close to 1 (the curves basically overlap). When $|\bar{A}(\mathscr{S})|/N^2$ becomes lower, instead, the PER is higher in the scenarios relative to the topologies in Figure 4.6 (b) and Figure 4.7 (a). For instance, considering node 12 in Figure 4.6 (b), when the transmit power is set to the minimum allowed value $P_{\mathrm{t}}^{\mathrm{min}}$ (equal at each node) to reach the AP, this node is isolated from most of the remaining nodes in the network. The packets transmitted by this node are likely to collide with those transmitted by nodes that are out of its transmission range, thus degrading the performance in terms of PER.

Similar considerations can be carried out for the scenarios where the aggregate traffic is set to $Ng = 20$ pck/s. The performance of these scenarios is shown in Figure 4.12. In this case, the statistical fluctuations are reduced since the number of packets transmitted by each node during the simulation is larger. However, the behavior in this case is also similar to that presented in Figure 4.11. There is a little dependence on the topology of the network, especially for values of $|\bar{A}(\mathscr{S})|/N^2$ close to 1. When the number of ones in the adjacency matrix reduces, the PER increases and the impact of isolated nodes heavily affects the network performance.

4.4.5 Impact of Traffic

As shown in Section 3.3.1, the performance of a WSN, under the assumption of low traffic load, depends only on the number of ones in the adjacency matrix. In Figure 4.13, the sparsity index $|\bar{A}(\mathscr{S})|/N^2$ is set to 0.87 and the PER is shown as a function of the aggregated offered traffic Ng. For all scenarios, we have considered

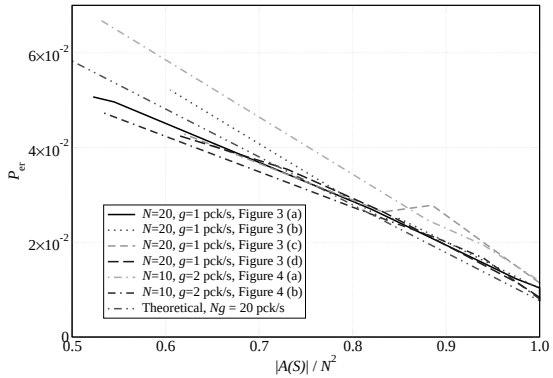


Figure 4.12: PER as a function of the sparsity index of the adjacency matrix. In the simulations, for scenarios with $N = 20$ nodes, the packet generation rate is set to $g = 1$ pck/s, whereas for scenarios with $N = 10$ nodes, the packet generation rate is set to $g = 2$ pck/s.

at most $g = 10$ pck/s, since for larger values of g the assumption of low traffic load is no longer satisfied. The considered topologies are those in Figure 4.6. From our analysis it turns out that, for small values of Ng, all lines basically overlap, regardless of the network dimension and the number of nodes. In the inset of Figure 4.13, a closeup of the curves for low values of Ng is shown. The overlap of the curves is due to the proposed transmit power allocation strategy, which minimizes the number of collisions between the nodes. On the other hand, when Ng increases, the number of collisions increases as well, and the approximations behind the analytical model no longer hold. In this figure, the curve relative to the PER predicted by the analytical model and given by Equation 3.31 is shown. The PER predicted by the analytical model, for a given sparsity index, is lower than that obtained through simulations, especially for low offered traffic load. In our analytical model, in fact, we have assumed that under the assumption of low traffic load, the number of retransmission attempts due to the backoff algorithm is negligible. Through this transmit power allocation scheme it is then possible to set the transmit power at each node, in order to reach a desired sparsity index and consequently improve the performance.

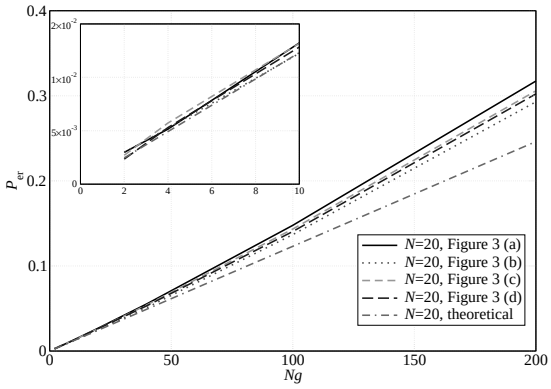
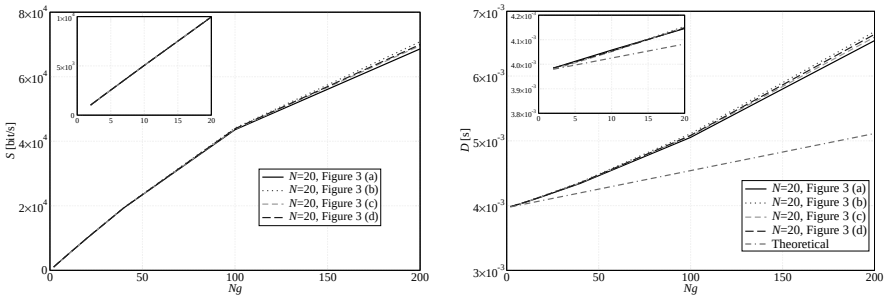


Figure 4.13: PER as a function of the aggregated offered traffic load Ng. Different topologies for scenarios with $N = 20$ nodes are considered. In all cases, the sparsity index is set to 0.87.

The same network configurations have been used in order to verify the impact of the sparsity index on the network transmission rate and delay. In Figure 4.14 a,b, these performance indicators are presented as functions of the overall traffic load. Considering Figure 4.14 (a), all presented curves basically overlap. This fact underlines once more that the number of ones in the adjacency matrix does not affect the performance. In particular, for small values of Ng the overlap is almost perfect and one can say that the performance depends only on the adjacency matrix and the related sparsity index. On the other hand, these results suggest that a given performance can be obtained with any network, provided that the transmit power is correctly allocated among the nodes.

Similar considerations can be carried out for the delay performance, analyzed in Figure 4.14 (b). In this case, there is also a good overlap between the simulation curves relative to different topologies and scenarios with different numbers of remote nodes. In particular, the number of ones in the adjacency matrix, i.e. the number of active connections, is the only characteristic that affects, for small values of the aggregated offered traffic load, the network performance. When the traffic load increases, however, the assumptions made in Section 3.3.1 do not hold anymore. In this figure,



(a) (b)

Figure 4.14: (a) Network transmission rate and (b) delay as functions of the aggregated offered traffic load Ng. Various topologies with $N = 20$ nodes are considered. The sparsity index is set at 0.87.

the analytical curve given by the expression in Equation 3.32 is also shown. In this case as well, the delay predicted by the analytical model is lower than that obtained with simulations. As in the previous case, the impact of the backoff procedure, due to the packet retransmission, has not been taken into account, thus the average delay predicted by our analytical model is lower.

4.4.6 Impact of the Power Allocation Strategy

In this subsection, we present the impact of the proposed transmit power allocation strategy on the performance of WSNs. In particular, as anticipated at the beginning of Subsection 4.4.2, we consider three possible transmit power allocation strategies.

1. In the former case, the transmit power is set in order to allow each sensor either to communicate with any other sensor in the network or to communicate at most with the AP.

2. In the proposed adjacency matrix-based power allocation strategy, the power is different at each sensor and is set according to optimization strategy presented in Section 3.3.1, where the total amount of available transmit power is assigned to each sensor in order to minimize the PER at the AP. This rule

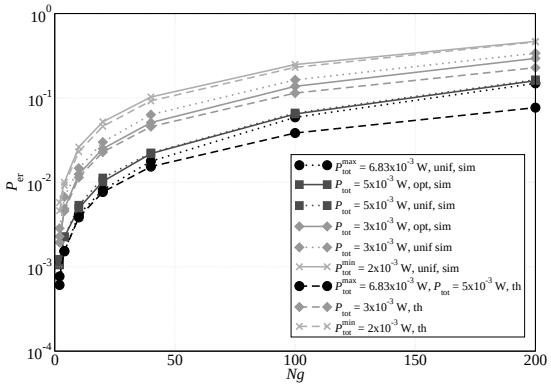


Figure 4.15: PER as a function of the aggregated offered traffic load Ng. Different transmit power allocation strategies are considered for the topology with $N = 20$ nodes presented in Figure 4.6 (b): (i) fixed per-node transmit power (unif) and (ii) optimized transmit power (opt). Both simulation (solid lines) and analytical (dashed lines) results are presented. The dotted lines refer to scenarios where no optimized power allocation strategy is used.

leads to allocate small transmit powers to nodes which are isolated and large transmit powers to nodes which can be connected with a large number of remote nodes. In this way, it is possible to minimize the collisions at the AP. Of course, there will still be nodes which cannot sense each other (leading to possible collisions), but this is due to the limited amount of overall network transmit power.

3. The power allocation strategy proposed in [17], which will be discussed in the following.

In Figure 4.15, the PER is shown as a function of the offered traffic load Ng in the network, for the topology presented in Figure 4.6 (b). Different values of overall network transmit power, with the corresponding sparsity indexes of the adjacency matrix shown in Table 4.2, are considered. Under the transmit power allocation strategy presented in Section 3.3.1, in Figure 4.15, a performance comparison between

Table 4.2: Sparsity indexes of the adjacency matrix for the scenario presented in Figure 4.6 (b). Different values of overall network transmit power are considered.

Available power (P_{tot})	Sparsity index
$6.83 \cdot 10^{-3}$ W ($P_{\text{tot}}^{\max}$)	1
$5 \cdot 10^{-3}$ W	1
$3 \cdot 10^{-3}$ W	0.85
$2.9 \cdot 10^{-3}$ W	0.8325
$2 \cdot 10^{-3}$ W ($P_{\text{tot}}^{\min}$)	0.605

scenarios with and without the use of the proposed transmission power allocation strategy is presented. The overall transmission power available in the network is allocated either assigning a common transmit power to all nodes or using the proposed transmit power allocation strategy. Of course, in the latter case, the sparsity index of the adjacency matrix is maximized according to the available power and, referring to the results presented in Figure 4.11 and Figure 4.12, the higher is the sparsity index, the lower is the PER.

Fixing the sparsity index to 1, from the results in Figure 4.15 it can be observed that the performance is almost the same, regardless of the chosen transmit power allocation strategy. This confirms that the PER performance depends only on the number of ones in the adjacency matrix. In the other cases, when the number of connections between the nodes decreases, the probability of collisions at the AP increases, since it is likely that one transmitting node cannot sense another transmitting node out of its transmission range. However, the curves have the same trend for all values of offered traffic load. In particular, when Ng is low, it is likely that the number of collisions at the AP is low. Instead, when the traffic load is larger, the probability that two nodes transmit at the same time increases and the PER increases as well. In Figure 4.15, the analytical results are shown as dashed lines. These curves are close to those associated with simulation results, especially for scenarios in which the sparsity index is small. Once more, the good agreement between the analytical results and the simulation results is confirmed, thus further validating the analytical model. For the sake

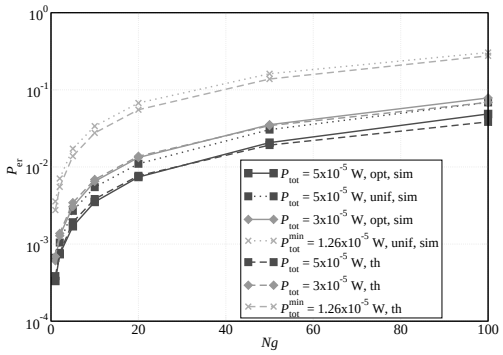


Figure 4.16: P_{er} as a function of the aggregated offered traffic load (Ng). Different power allocation configurations are considered for the topology with $N = 10$ nodes presented in Figure 4.7 (a). Both simulative (solid lines) and analytical (dashed lines) results are presented. The dotted lines refer to scenarios where no optimized power allocation strategy is used.

of comparison, in Figure 4.15 we also show the PER in scenarios where no transmit power allocation strategy is used (dotted lines). In these cases, the performance is worse than in the case with the optimized transmit power allocation strategy. In fact, given a value of overall network available power, the proposed power allocation strategy allows to maximize the sparsity index of the adjacency matrix and, therefore, reduce the PER.

In Figure 4.16 the PER is shown as a function of the offered traffic load Ng for the scenario with $N = 10$ nodes presented in Figure 4.7 (a). The corresponding values of the sparsity indexes of the adjacency matrix are shown in Table 4.3. In this case, since the offered traffic load is lower, there is a better agreement between analytical and simulation models. In fact, the assumption of low traffic holds almost for all considered values of Ng. For small values of Ng, the curves are almost overlapped. Instead, when Ng increases, the backoff procedure leads to a gap between the simulation and the analytical curves. This gap, however, remains smaller than that in Figure 4.15. This confirms that the proposed analytical model can predict almost perfectly the

Table 4.3: Sparsity indexes of the adjacency matrix for the scenario presented in Figure 4.7 (a). Different values of overall network transmit power are considered.

Available power (P_{tot})	Sparsity index
$5 \cdot 10^{-5}$ W	1
$3 \cdot 10^{-5}$ W	0.94
$1.26 \cdot 10^{-5}$ W ($P_{\text{tot}}^{\text{min}}$)	0.53

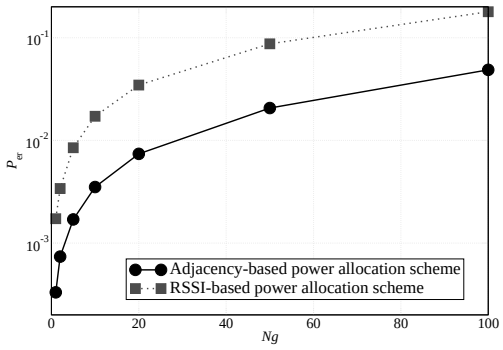


Figure 4.17: P_{er} comparison between the proposed power allocation scheme and the RSSI-based power allocation strategy.

performance of WSNs, especially for low values of Ng. Similarly to Figure 4.15, a comparison between the scenarios with (solid lines) and without (dotted lines) the use of the transmission power allocation strategy is also shown in Figure 4.16. In particular, given a pre-defined value of overall network available transmit power, the proposed approach maximizes the number of ones in the adjacency matrix and, consequently, improves the network performance.

For the sake of completeness, a performance comparison between the PER guaranteed by the proposed power allocation scheme and that guaranteed by the power allocation scheme derived in [17] is considered in Figure 4.17. In [17], the authors aim at dynamically allocating the power at each node, in order to minimize the PER, under the assumption of additive white Gaussian noise (AWGN) channels and ne-

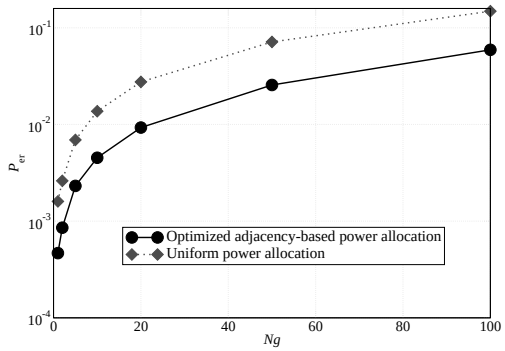


Figure 4.18: P_{er} comparison between the proposed power allocation scheme and the uniform transmit power allocation scheme.

glecting the impact of interference due to other transmitting nodes. In particular, they assume that the packet generation rate is low enough to prevent packet collisions. The power is allocated to each node according to the quality of the link that it experiences in order to reach the AP. The link quality is measured in terms of RSSI. In order to make the comparison fair, we have applied the power allocation strategy of [17] in a scenario without channel noise but in the presence of multiple access interference, which has been modeled as a Gaussian random variable [87]. As one can see from the results in Figure 4.17, our approach, based on the maximization of the sparsity of the adjacency matrix, tends to reduce as much as possible the number of collisions at the AP. Therefore, it is more efficient, especially for networks where the offered traffic load starts to become significant.

In order to highlight the performance gain, in terms of PER, due to the proposed transmit power allocation scheme, with respect to a uniform power allocation strategy, the simulations have been carried out for a fixed overall network transmission power, considering the network topology presented in Figure 4.8 (a). The corresponding results, shown in Figure 4.18, underline that the proposed power allocation strategy effectively lowers the PER. The results in Figure 4.18 also suggest that, for a fixed target PER, the proposed power allocation scheme allows to support an almost

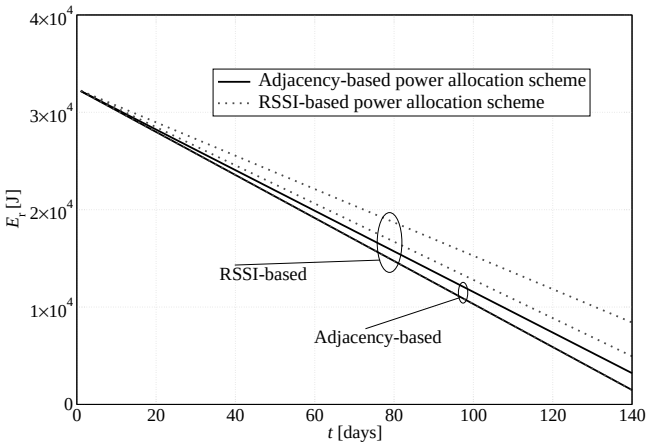


Figure 4.19: Residual energy per node. Both RSSI-based and adjacency-based transmit power allocation schemes are considered. The given PER is $P_{\text{er}} = 5 \cdot 10^{-2}$.

double aggregated offered traffic load, with respect to that supported by a uniform power allocation strategy. In fact, through the proposed approach, a node, before transmitting, can sense a larger number of neighboring nodes (in a relatively dense partition of the network), and therefore prevent packet collisions.

4.4.7 Network Lifetime Performance

We now evaluate the network lifetime in the scenario with $N = 10$ nodes shown in Figure 4.8 (b), setting the average packet generation rate to $g = 1$ pck/s, and considering both the proposed adjacency matrix-based power control approach and the RSSI-based power control strategy presented in [17]. Since no battery models are provided in Opnet, the residual energy performance analysis has been carried out through Matlab. In particular, the battery depletion model refers to Equations 3.36–3.40. Each node is equipped with a 3 V battery with an initial energy of 32.4 kJ. In Figure 4.19, the residual energy per node is shown considering a target PER equal to $P_{\text{er}} = 5 \cdot 10^{-2}$. Since the distances between the nodes are small, the transmission powers of the nodes is low (between 0.374 μW and 2.2 μW). Since the current consumed

during the transmission phase is $I_{\text{tx\_state\_}i} = 7.886P_i + 0.009711$, the second term of the right-hand side of the current expression dominates, i.e., $I_{\text{tx\_state\_}i} \simeq 0.009711$. In other words, the effect of the transmit power on the energy consumed in the tx state is negligible. In fact, this is confirmed by the fact that the two families of curves, referring to adjacency-based (solid lines) and RSSI-based (dotted lines), are quite close. On the other hand, since the proposed approach aims at maximizing the connections between nodes, the power consumption is slightly higher in our approach than in [17].

In order to highlight the energy saving improvement introduced by our power allocation technique, we have considered the network, shown in Figure 4.8 (c), with $N = 10$ nodes randomly deployed over a 50 m^2 square surface, where, unlike all previous scenarios, the path loss exponent α is 3 — recall that all previous results refer to scenarios with $\alpha = 2.1$. In this case, the transmit power at each node is larger, so that the impact of the transmit power on the energy consumed in the tx state is more evident. In Figure 4.20, the residual energy in each node is shown, considering both the adjacency matrix-based transmit power allocation strategy proposed here and the RSSI-based power allocation scheme presented in [17]. Considering the proposed power allocation strategy (solid lines), the use of lower transmit powers allows to drastically reduce the nodes' deaths. In particular, excluding the case of a specific node which is far from the AP and uses a high transmit power, the other nodes die later than in the case with RSSI-based power allocation scheme (dotted lines). In fact, according to the proposed model, a node delays its packet transmission if it senses that other neighboring nodes are transmitting. In this way, since a larger number of nodes can sense each other, the number of transmissions (successful or not) reduce and the nodes waste less power to process incoming packets. In the same figure, the average residual energy for both proposed power allocation schemes is also shown. These curves confirm that the adjacency-based power allocation scheme allows to extend the network lifetime because it increases the network residual energy.

Since each node transmits with a different power and receives a different number of packets from neighboring nodes, it will experience different power consumptions according to its spatial position and the number of surrounding nodes. In Figure 4.21, the energy distribution in the network considering the use of the proposed power

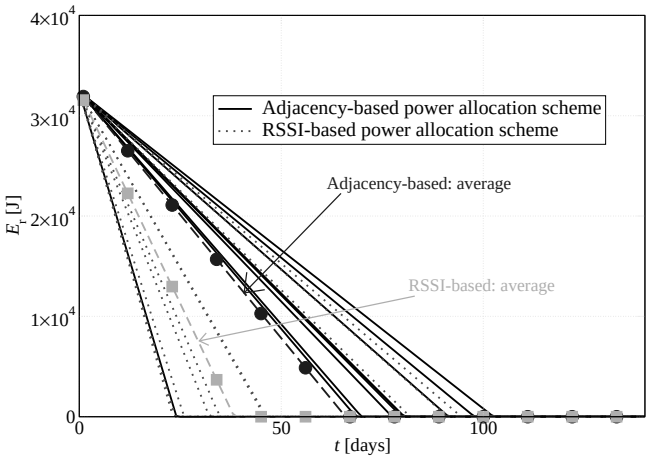


Figure 4.20: (i) Per-node (solid and dotted lines) and (ii) average network (dashed with symbols) residual energies as functions of time. Both RSSI-based (dotted lines and dashed with squares) and adjacency matrix-based (solid lines and dashed with circles) transmit power allocation schemes are considered. The target PER is $P_{\text{er}} = 5 \cdot 10^{-2}$.

allocation strategy is shown. The reference topology is the same as that considered in Figure 4.20 (i.e. the network topology of Figure 4.8 (c)). In Figure 4.21a, the initial energy in the network, i.e. at all nodes, is shown. As one can see, since all nodes have the same battery energy (3.24 kJ), the initial "surface" lies on a plane. In Figure 4.21 (b), a snapshot of the residual energy in the network after 45 days is presented. The residual energy in the node far from the AP is lower than that in the other nodes, because this node transmits with much higher power. Finally, in Figure 4.21 (c), the residual energy in the network is presented after 100 days. The farthest node from the AP has run out of battery, but the other nodes are still able to communicate, since the use of low (on average) transmission powers prevents rapid battery depletion.

In Figure 4.22, the residual energy performance is shown considering the RSSI-based power allocation scheme presented in [17]. As in the previous case, in Fi-

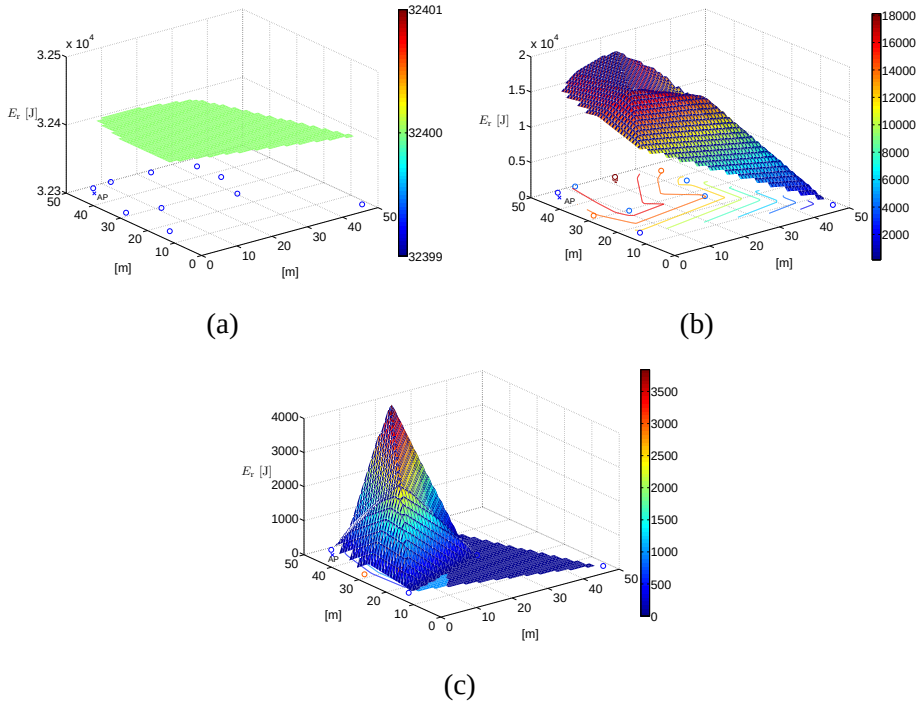
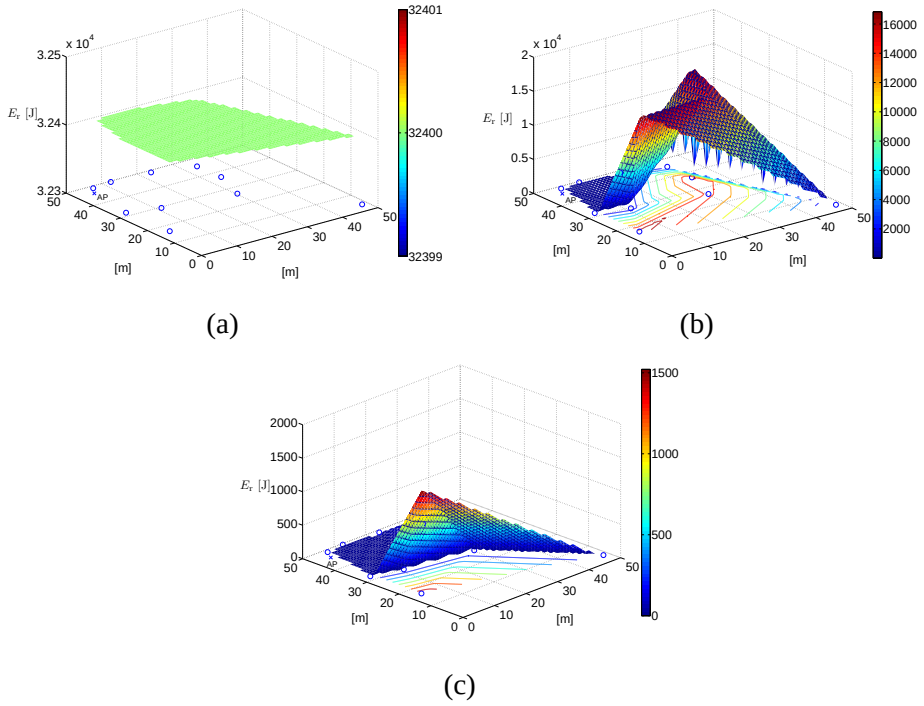


Figure 4.21: Residual energy with adjacency matrix-based transmission power allocation strategy proposed here, in a scenario with $N = 10$ nodes after (a) 0 days, (b) 45 days, and (c) 90 days.

gure 4.22 (a) the initial energy in the network is shown (as for the previous figure, in this case as well all nodes have the same initial energy). In Figure 4.22 (b), the residual energy after 45 days is shown. As observed in Figure 4.21 (b), the farthest node from the AP has the lowest residual energy. However, in this case the nodes near the AP also have low residual energies. According to the RSSI-based power allocation scheme, in fact, these nodes experience low attenuation and are therefore assigned high values of transmit power. In this way, they consume a significant amount of energy to transmit a packet. Finally, in Figure 4.22 (c), the residual energy in the network after 100 days is shown. In this case, only one node is still alive, whereas the



(a) (b)

(c)

Figure 4.22: Residual energy with RSSI-based power allocation strategy, in a scenario with $N = 10$ nodes after (a) 0 days, (b) 45 days, and (c) 90 days.

batteries of the remaining nodes have run out of energy.

4.5 Clustered Networks Performance

In this section, we introduce the results related to clustered networks. In particular, we show the impact of the transmission rate and of the tolerable network death level on network transmission rate, delay, and throughput. In addition, we present the impact of the data fusion mechanism, focusing on the impact of the MAC protocol on the probability of error at the AP. Then, we provide some design guidelines useful to configure network parameters (such as clustering configuration, tolerable network death

level, and packet generation rate) in order to meet given performance constraints.

4.5.1 Simulation Setup

The simulations have been carried out with the Opnet Modeler simulator and a built-in Zigbee network model designed at the National Institute of Standards and Technologies (NIST). Since this model refers only to the first two layers of the ISO/OSI stack, we have extended it by deriving an Opnet model for a FC, which, in addition to providing relay functionalities, implements the intermediate data fusion mechanisms described in Subsection 3.4.3. We assume (ideal) wireless communications between the RFDs and the FCs, and between the FCs and the coordinator.

Each simulation result is obtained by averaging over ten consecutive runs in order to make possible statistical fluctuations negligible. The duration of each simulation has been chosen so that the simulation ends as soon as the condition related to the minimum number of alive RFDs in the network (i.e., the percentage of RFDs which are still operative) is no longer respected. The packet length, referring to the network tolerable death level performance analysis, is 632 bits (i.e., 512 bits of data payload and 120 bits introduced by MAC and physical layers).

The RFDs carry out noisy observations of a randomly generated binary phenomenon H and make local decisions on the status of this phenomenon. The observation noise is modeled as an Additive White Gaussian Noise (AWGN). Subsequently, the RFDs embed their decisions into data packets, which are sent either to the coordinator (in the absence of clustering) or to the FCs (in the presence of clustering). The decisions are assumed to be either 0 (no phenomenon) or 1 (presence of the phenomenon).

- In the absence of clustering, the coordinator makes its final decision through a majority-like rule directly on the messages received from the RFDs. Obviously, if some packets are lost due to medium access collisions, decisions are made only on the received packets (this leads to a reduced reliability of the final decision). If all the packets related to a set of observations of the same phenomenon are lost, the coordinator decides randomly (i.e., with probability 0.5) for one

of the two possible values. Finally, if half of the decisions are in favor of one phenomenon status and the other half are in favor of the other, the coordinator decides for the presence of the phenomenon.

- In the presence of clustering, each FC makes an intermediate decision on the basis of the messages received from the RFDs associated with its cluster and forwards this decision to the coordinator, which makes the final decision on the basis of the received messages (from either an FC or an RFD). The decision rule used by the FCs is the same of that used by the coordinator (i.e., majority decision).

In both scenarios, it is possible to evaluate, by simulation, the probability of decision error of the coordinator, by comparing, in each simulation run, the final decision made by the coordinator with the true status of the phenomenon. Together with the probability of decision error, the simulator allows to evaluate: the network transmission rate S, defined as the ratio between the number of bits correctly received by the coordinator in the simulation time; the throughput S_{th}, defined as the ratio between the number of packets correctly delivered at the coordinator and the number of packets sent by the RFDs; and the delay D (dimension: [s]), defined as the time interval between the transmission and the reception instants of a generic packet. The last performance indicator of interest is the aggregate throughput (dimension: [pck/s]), defined as $S_{\text{agg}} = N \cdot g \cdot S_{\text{th}}$, where N is the number of transmitting RFDs and g is the packet generation rate (dimension: [pck/s]).

In each simulation run, the status of the phenomenon (either 0 or 1) is chosen randomly in consecutive phenomenon realizations. In the case of data fusion performance analysis, the local decisions are then inserted into data packets which are 96 bit long. In order to estimate a probability of decision error at the AP of the order of 10^{-6}, approximately 115200 local decisions on the phenomenon status need to be transmitted. Since in the Opnet implementation each decision must be coded into a char, the memory occupation is therefore equal to 8 bits per observation. In addition, in each packet a null char terminator (8 bits) and a 32-bit sensor identifier must be added. Considering a packet length of 96 bits, therefore each packet can contain 7

local decisions. From this simple analysis, it follows that 16458 packets need to be transmitted. Since the packet generation rate g is equal to 2 pck/s, the simulation duration is set to 8229 s, in order to guarantee that all required local decisions are sent to the coordinator. We have chosen this value as a compromise between simulation duration and achievable probability of decision error at the AP. The effective packet length is equal to 216 bits, since, besides a payload of 96 bits containing the local decisions, a header of 120 bits is introduced by physical and MAC layers. The RFDs send the packets over an ideal wireless channel which introduces neither attenuation nor fading. At the end of the simulation, the probability of decision error is computed by a direct comparison between the sequence of decisions at the AP and the sequence of true phenomenon realizations.

4.5.2 Performance Analysis

In the first two subsections, we evaluate the performance from a networking perspective, considering scenarios both with and without clustering. More precisely, in Subsection 4.5.3, the impact of the network tolerable death level on the throughput and delay is analyzed; in Subsection 4.5.4, the same network performance indicators are evaluated, for a fixed value of χ_{net}, considering various values of the packet generation rate. In Subsection 4.5.5, instead, we consider the presence of data fusion and its impact on the probability of decision error, the throughput, the aggregate throughput, and the delay is analyzed. In the figures in Subsection 4.5.3 and Subsection 4.5.4, we indicate, for each performance curve, the confidence interval 2σ, where σ is the standard deviation, over consecutive simulation runs, with respect to their average value.

4.5.3 Impact of Tolerable Network Death Level

$N = 64$ RFDs

In this case, we consider scenarios with no cluster (no relay), and a number of clusters ranging from 1 to 8 (with a relay per cluster). In Fig. 4.23, (a) the network transmission rate and (b) the delay are shown, as functions of the required network tolerable

death level χ_{net}, in scenarios with $N = 64$ RFDs. In Fig. 4.23 (c), instead, the curves in the previous two subfigures are combined, obtaining transmission rate-delay curves (parameterized with respect to the network tolerable death level χ_{net}).

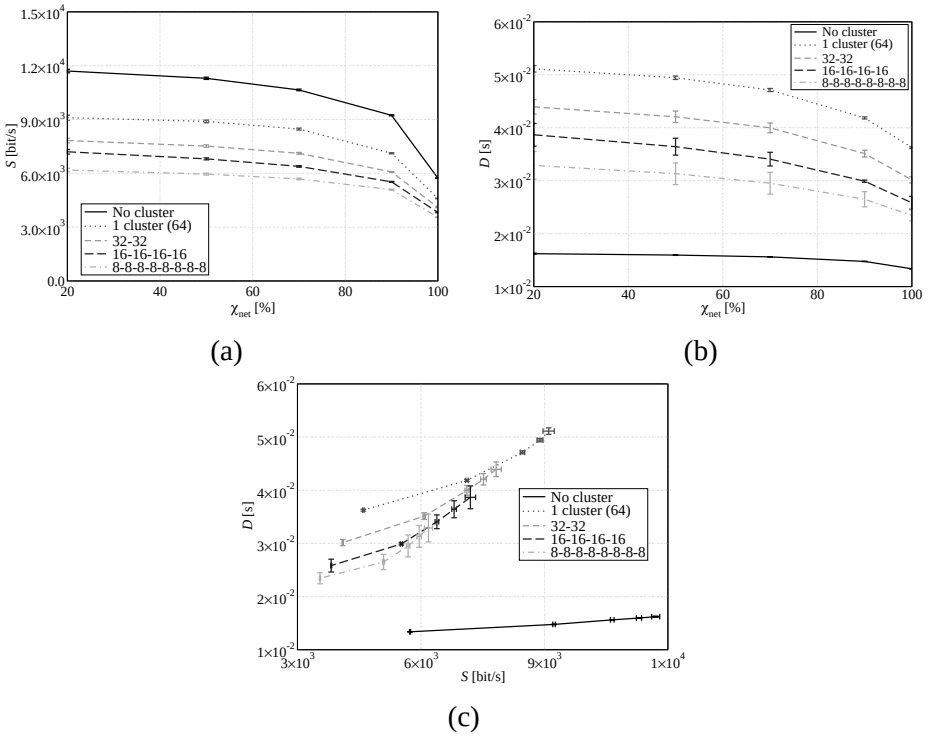


Figure 4.23: Performance evaluation in a scenario with $N = 64$ RFDs and various uniform clustering configurations: (a) network transmission rate and (b) delay as functions of χ_{net}; (c) delay as a function (parameterized with χ_{net}) of the network transmission rate. The packet generation rate is equal to $g = 2$ pck/s.

Looking at Fig. 4.23 (a), the network transmission rate is first slightly decreasing, for small values of the network tolerable death level, and then suddenly drops in correspondence to $\chi_{\text{net}} \simeq 90$ %. In the scenario with many small clusters, the network transmission rate remains low since there is a larger number of transmitting nodes, and this increases the number of collisions.

Considering the results in Fig. 4.23 (b), unlike what the intuition may suggest, the delay in the scenario with one large cluster (with $N = 64$ RFDs) and a single relay is higher than the delay in the scenario with 8 relays (8-8-8-8-8-8-8-8). The delay reduces when the number of clusters increases, since in this case the relay of each cluster has to manage a smaller number of packets. As expected, the delay is lowest in the absence of clustering, i.e., when all RFDs transmit directly to the AP. Obviously, in all cases the delay decreases for increasing values of percentage of nodes' deaths required to kill the network (i.e., for increasing values of χ_{net}). In fact, the more RFDs die, the more efficiently the surviving RFDs can be served: therefore, the time-averaged delay reduces. Comparing Fig. 4.23 (a) with Fig. 4.23 (b), it can be noted that the network configuration which guarantees the best performance, in terms of network transmission rate and delay, is the one without relays (lines with circles).

Finally, in Fig. 4.23 (c) the delay is shown as a function of the network transmission rate (the points of each curve correspond to different values of network tolerable death level). These curves give a concise (instantaneous) picture of the network operating status: the best operating conditions would correspond, obviously, to the bottom right (low delay and high transmission rate).

We now analyze the network performance in the presence of *non-uniform* clustering. In Fig. 4.24 (a), the network transmission rate is shown as a function of χ_{net}, for various clustering configurations. As in the case with uniform clustering, the network transmission rate is first slightly decreasing (for small values of χ_{net}) and then decreases rapidly (for values of χ_{net} higher than 90%). In fact, for small values of χ_{net} the RFDs' deaths are balanced by the reduced number of collisions in the common wireless channel, so that the network transmission rate remains high. When the number of RFDs' deaths becomes large, instead, the overall traffic generated by the RFDs reduces and, consequently, the network transmission rate necessarily reduces. The best performance is obtained in the 56-4-4 clustering configuration with 1 relay. In this case, in fact, the presence of the eight directly-connected RFDs has a beneficial impact on the network transmission rate, since the packets can be transmitted without the need of being relayed. The worst performance, instead, is obtained in the scenario with 32-8-8-8-8 clustering configuration and 5 relays: in this case, the num-

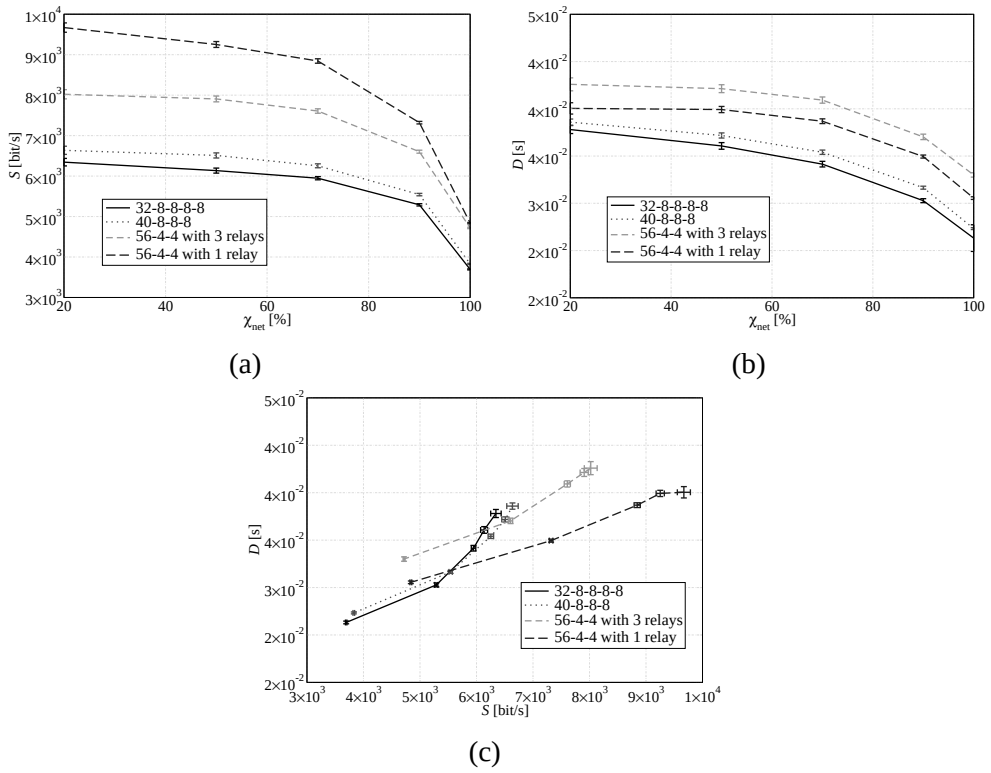


Figure 4.24: Performance evaluation in a scenario with $N = 64$ RFDs and various non-uniform clustering configurations: (a) network transmission rate and (b) delay as functions of χ_{net}; (c) delay as a function (parameterized with χ_{net}) of the network transmission rate.

ber of transmitted packets is larger due to the larger number of relays and, thus, the higher probability of finding the channel busy. In Fig. 4.24 (b), the delay performance is shown. The performance, in the presence of non-uniform clustering, is similar in all cases where the RFDs are connected to the AP through a relay. In these cases, in fact, when the number of RFDs'deaths increases, it is more likely that a node finds the channel free and, therefore, can transmit its data. The network performance in the scenario with 56-4-4 clustering configuration and 1 relay, instead, shows a different

behavior; in this case, both throughput and delay (the latter is obtained according to equation (3.51)) are affected by the 8 RFDs directly connected to the coordinator, which transmit more frequently, since they are not affected by the presence of a relay node acting as a bottle-neck. In Fig. 4.24 (c), the delay is shown as a function of the network transmission rate. Similarly to Fig. 4.23 (c), this figure provides a concise characterization of the network operating point.

$N = 16$ RFDs

We first consider a scenario with *uniform* clustering. The performance results, in terms of network transmission rate and delay, are shown in Fig. 4.25, where various network configurations are compared. Looking at the results in Fig. 4.25 (a), where the network transmission rate is shown as a function of the network tolerable death level, one can observe that the best performance is obtained, as in the case with 64 RFDs, in the absence of clustering. Similarly to the case with 64 RFDs, the transmission rate is monotonically decreasing and the *shape* of the curves is the same regardless of the number of clusters. From the results in Fig. 4.25 (b), where the delay is shown as a function of χ_{net}, one can notice that there is a substantial agreement with the delay performance in the presence of 64 RFDs (compare, for instance, Fig. 4.25 (b) with Fig. 4.23 (b)). As in the case with 64 RFDs, the lowest possible delay is obtained without clustering. In the presence of clustering, instead, the best performance is obtained in the case with 8 2-sensor clusters. This can intuitively be explained considering that when a relay is connected to a large number of RFDs (i.e., a large cluster), it is likely that the RFDs find the channel "busy" and, therefore, have to wait longer in order to transmit their packets. In the case with only a few RFDs connected to a relay, instead, the latter has to manage a limited number of packets.

Let us now turn our attention to a scenario with *non-uniform* clustering. The performance results, in terms of network transmission rate and delay, are shown in Fig. 4.26.

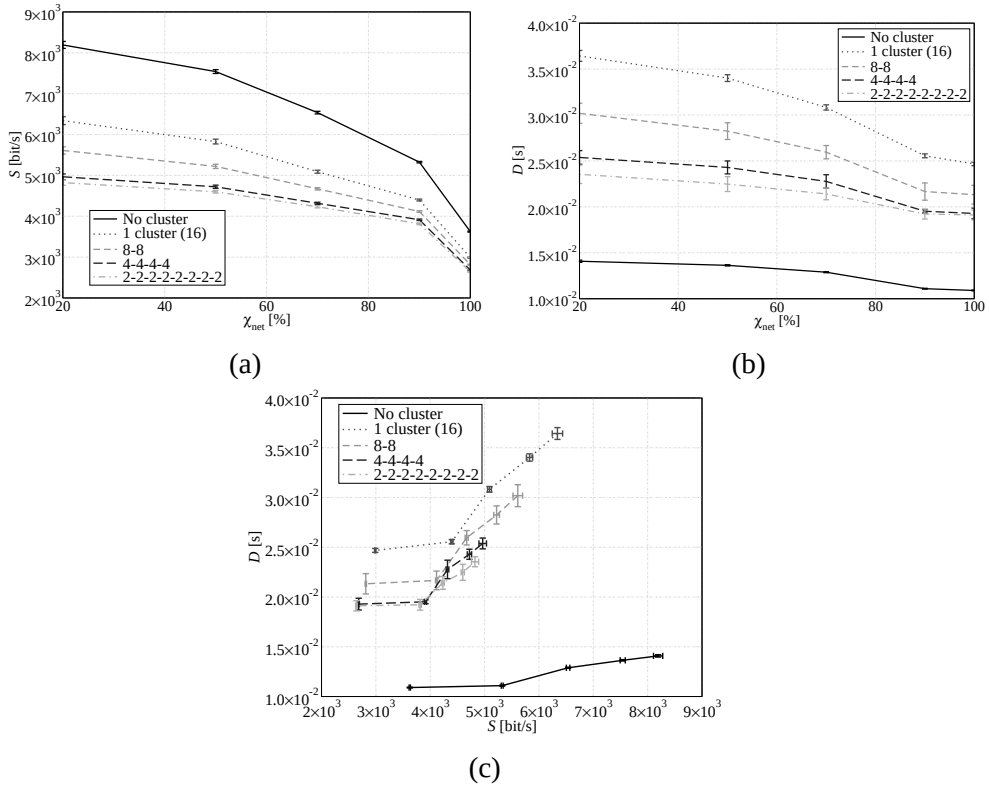


Figure 4.25: Performance evaluation in a scenario with $N = 16$ RFDs and various *uniform* clustering configurations: (a) network transmission rate and (b) delay as functions of χ_{net}; (c) delay as a function (parameterized with χ_{net}) of the network transmission rate. The packet generation rate is equal to $g = 2$ pck/s.

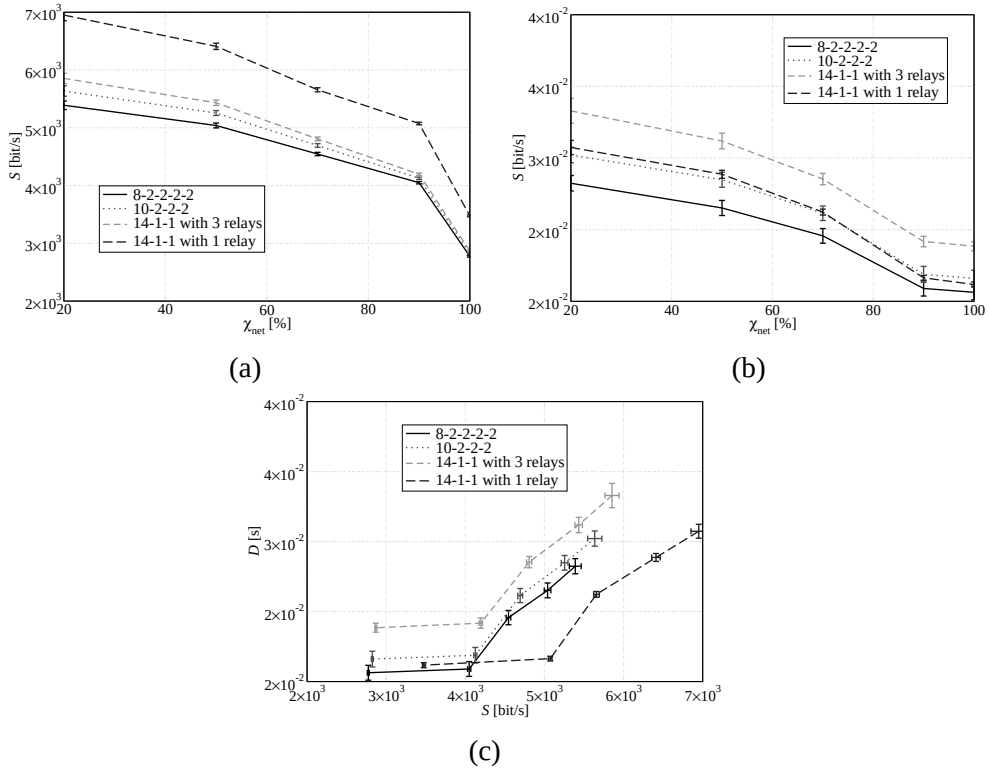


Figure 4.26: Performance evaluation in a scenario with $N = 16$ RFDs and various *non-uniform* clustering configurations: (a) network transmission rate and (b) delay as functions of χ_{net}; (c) delay as a function (parameterized with χ_{net}) of the network transmission rate. The packet generation rate is equal to $g = 2$ pck/s.

Considering the performance results in Fig. 4.26 (a), where the network transmission rate is shown as a function of χ_{net}, one can observe that the highest network transmission rate is obtained, regardless of the value of χ_{net}, with the configuration 14-1-1 (with 1 relay), formed by one big cluster (connected to the AP through a relay) and two single RFDs connected directly to the AP. This can be explained considering that, unlike the scenarios where an RFD is connected to the AP through a relay, in this case the two RFDs connected to the AP can send their data packets directly, avoiding

a “bottle-neck” relay. For all the other non-uniform clustered configurations, one can observe that there is no significant difference. In particular, this difference becomes negligible at large values of χ_{net}. Observing the delay performance in Fig. 4.26 (b), the lowest delay is guaranteed almost everywhere by the 8-2-2-2-2 configuration. However, the delay associated with 14-1-1 (with 1 relay) configuration, i.e., the one which guarantees highest network transmission rate, has a peculiar behavior. More precisely, the delay curve of the 14-1-1 (with 1 relay) configuration is a steep function of χ_{net}: for small values of χ_{net}, the delay is approximately equal to that associated with the 10-2-2-2 clustering configuration; when χ_{net} becomes 100 %, however, the corresponding delay becomes lowest. Finally, in Fig. 4.26 (c) the delay is shown as a function of the network transmission rate, for various values of the network tolerable death level χ_{net}. As one can see, the curves are closer to each other than in a scenario with uniform clustering. In this case as well, the choice of the network operating point depends on the specific user needs.

In Fig. 4.27 and Fig. 4.28, the performance of the networking schemes considered in Fig. 4.25 and Fig. 4.26 is analyzed by increasing the packet generation rate g from 2 pck/s to 10 pck/s, considering both uniform and non-uniform clustering. The network transmission rate in the case with $g = 10$ pck/s has the same shape than in the case with $g = 2$ pck/s, the only difference being the fact that it is five times higher. In fact, even if the traffic load is 5 times higher, this load is not critical, so that the number of collisions is limited and each generated packet can be immediately transmitted.

4.5.4 Impact of the Packet Generation Rate

In this subsection, we investigate the impact of the packet generation rate only in scenarios with $N = 16$ RFDs. In particular, we consider the same network topologies described in Subsection 3.4.2, i.e., with both uniform and non-uniform clustering. The network tolerable death level is set to 50%, i.e., the simulation stops when half of the nodes die. This analysis can be straightforwardly extended to scenarios with 64 RFDs.

We first analyze the network performance in the case with the following uniform

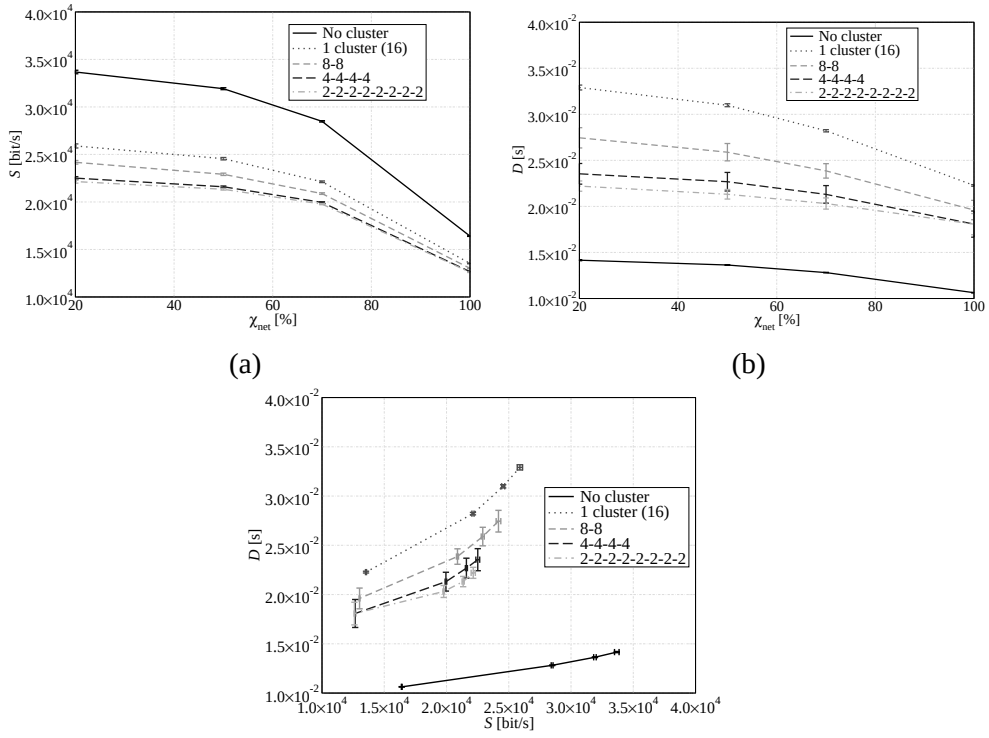


Figure 4.27: Performance evaluation in a scenario with $N = 16$ RFDs and various *uniform* clustering configurations: (a) network transmission rate and (b) delay as functions of χ_{net}; (c) delay as a function (parameterized with χ_{net}) of the network transmission rate. The packet generation rate is equal to $g = 10$ pck/s.

clustering configurations: (i) no cluster (i.e., direct transmission from the RFDs to the AP), (ii) 1 16-node cluster and 1 relay, (iii) 2 8-node clusters and 2 relays, (iv) 4 4-node clusters and 4 relays, and (v) 8 2-node clusters and 2 relays. In Fig. 4.29 (a) and Fig. 4.29 (b), we present performance results in terms of network transmission rate and delay, respectively. The network transmission rate has a similar behavior in all considered network scenarios, except for variations of the confidence interval due

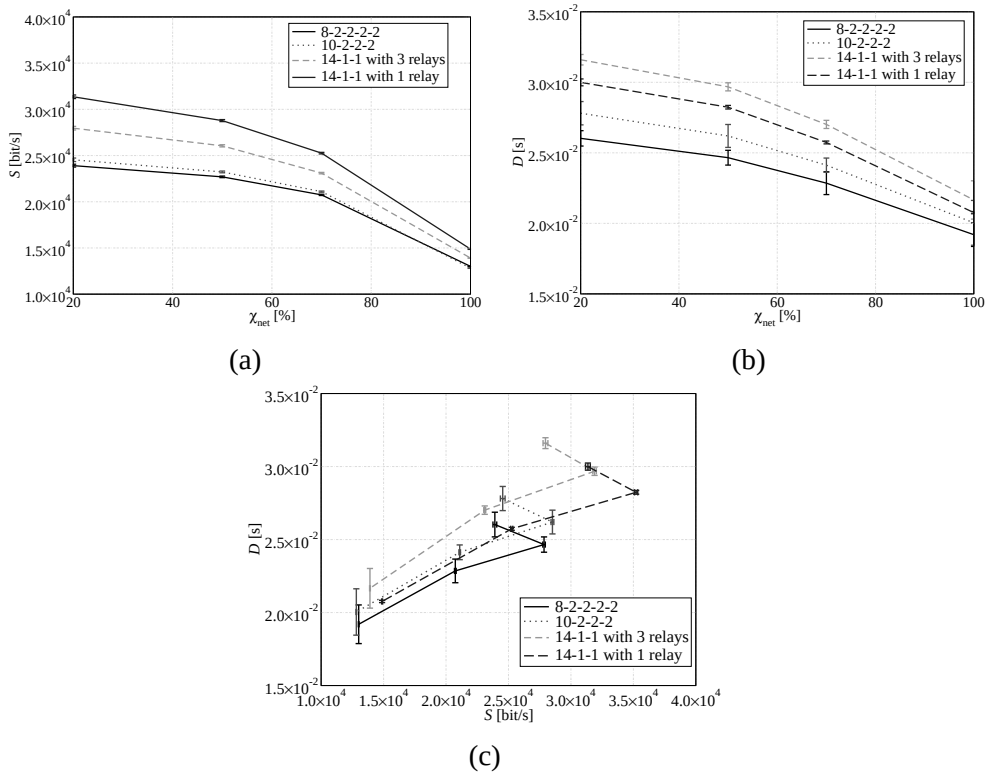
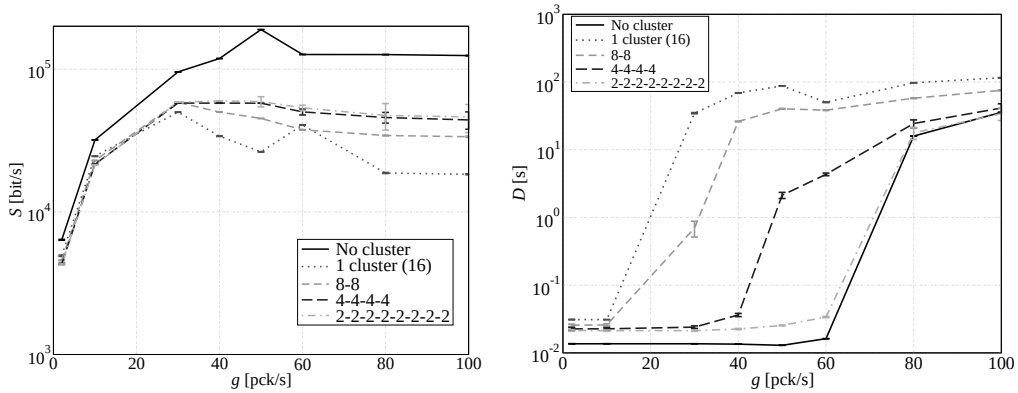


Figure 4.28: Performance evaluation in a scenario with $N = 16$ RFDs and various *non-uniform* clustering configurations: (a) network transmission rate and (b) delay as functions of χ_{net}; (c) delay as a function (parameterized with χ_{net}) of the network transmission rate. The packet generation rate is equal to $g = 10$ pck/s.

to the limited duration of simulations.[4] All curves shown in Fig. 4.29 (a) first increase till a maximum value, after which they slowly decrease. In this case, as in the other

[4]The Opnet simulator is based on a pseudo-random value generator, which can be initialized through a user-defined seed. In our simulations, we have experienced that there exist some values (not predictable) of this seed, which lead to a worse performance. If a larger number of simulations were considered, i.e., a larger number of seeds (e.g., 100 different seeds instead of 10) were used, the confidence interval would shrink significantly.



(a) (b)

Figure 4.29: Performance evaluation in a scenario with $N = 16$ RFDs and various *uniform* clustering configurations: (a) network transmission rate and (b) delay as functions of the packet generation rate. The network tolerable death level χ_{net} is fixed to 50%.

previously described cases, the best network performance is obtained in the absence of the relay node. In the presence of clustering, instead, the network transmission rate curves change their behavior, approximately in correspondence to a packet generation rate equal to 20 pck/s, due to the increasing traffic load. In a scenario with low traffic load, the best clustered configuration is that with one 16-node cluster and one relay, because the relay node can efficiently manage all packets sent by the nodes. When the traffic load increases, instead, the best network configuration is that with 8 2-node clusters and 8 relays. In fact, in this case it is more likely to find the relay busy, so that the probability of finding a relay idle increases in the presence of a larger number of relay increases.

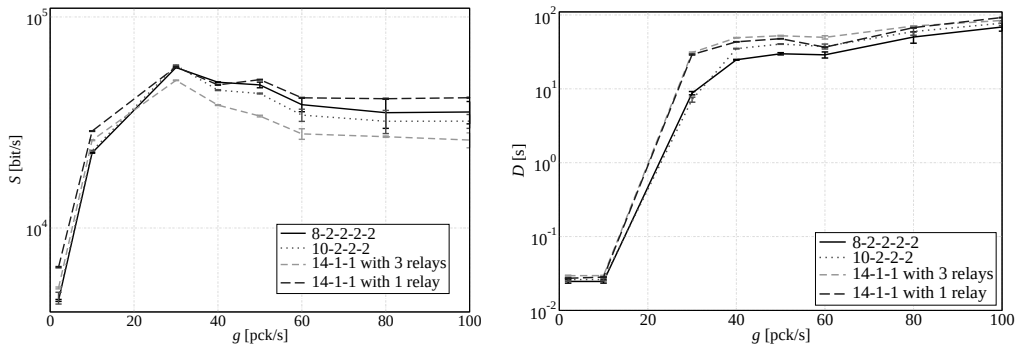
In Fig.4.29 (b), the delay is investigated as a function of the packet generation rate. All curves present a floor for small values of the packet generation rate, then the delay quickly increases and reaches a maximum value which depends on the specific clustering configuration. Considering clustered scenarios, the best configuration is that with 8 2-node clusters and 8 relays, whereas the worst performance is obtained

in the scenario with one 16-node cluster and one relay. Intuitively, in the presence of a few small clusters, it is likely that the relays are ready to receive new incoming packets. On the opposite, when there is only one 16-node cluster with one relay, it is likely that a node, in need of sending a data packet, finds the relay node occupied by another transmitting node.

In Fig. 4.30 (a) and Fig. 4.30 (b), instead, we present the same performance results (i.e., network transmission rate and delay) obtained in the presence of non-uniform clustering. The considered network configurations are the following: (i) 8-2-2-2-2 with 5 relays (i.e., the network is divided into 5 clusters, one formed by 8 RFDs and the other four formed by 2 RFDs each, and each cluster is connected to the AP through a relay); (ii) 10-2-2-2 with 4 relays (i.e., the network is divided into 4 clusters, one formed by 10 RFDs and other formed by 2 RFDs each, connected to the AP through a relay), as shown in Fig. 3.5 (b); (iii) 14-1-1 with 3 relays (i.e., one cluster is composed by 14 RFDs and two clusters are composed by one node each), as shown in Fig. 3.5 (c); and (iv) 14-1-1 with 1 relay (i.e., only one cluster composed by 14 nodes and connected to the AP through a relay, while the other two RFDs communicate directly to the AP), as shown in Fig. 3.5 (d).

In Fig. 4.30 (a), the network transmission rate is shown as a function of the packet generation rate. As one can see, all curves present a very similar behavior: for low values of the packet generation rate, the network transmission rate quickly increases, then reaches a maximum, and finally decreases to a saturation value which depends on the network configuration. When the packet generation rate is high, the best performance is obtained in the scenario with 14-1-1 clustering configuration with one relay. On the opposite, the worst performance is obtained with the 14-1-1 configuration with 3 relays. The presence of two RFDs connected directly to the AP has a beneficial effect on the network transmission rate, since it is likely that these two directly connected RFDs transmit without waiting. If we consider scenarios where all RFDs are connected to the AP through a relay, according to the results in Fig. 4.29 (a) the best performance is obtained in the scenario with many small clusters.

In Fig. 4.30 (b), the delay is shown as a function of the packet generation rate. Unlike the case with uniform clustering (see Fig. 4.29 (b)), in this case the curves are



(a) (b)

Figure 4.30: Performance evaluation in a scenario with $N = 16$ RFDs and various *non-uniform* clustering configurations: (a) network transmission rate and (b) delay as functions of the packet generation rate. The network tolerable death level χ_{net} is fixed to 50%.

close to each other. For small values of the packet generation rate the delay is low, but it rapidly increases for values of the packet generation rate between 20 pck/s and 30 pck/s. Finally, when the packet generation rate is high, the delay saturates to a maximum value, which is approximately the same for all network configurations. In this case as well, the best performance, in terms of delay, is obtained in the scenario with many small clusters. In the scenario with 14-1-1 clustering configuration and one relay, unlike the scenarios where all RFDs are connected to the AP through one relay, the overall delay is affected by the presence of the two directly connected RFDs, which are likely to reserve communications with the AP.

4.5.5 Impact of Data Fusion Mechanisms

In *non-clustered* scenarios (i.e., star topology), various values for the number N of RFDs are considered. Moreover, the impact of the presence/absence of ACK messages is taken into account. In *clustered* scenarios, instead, the number of RFDs is set to $N = 16$, and various clustering configurations are investigated: (i) 8-8 (i.e., 2 8-

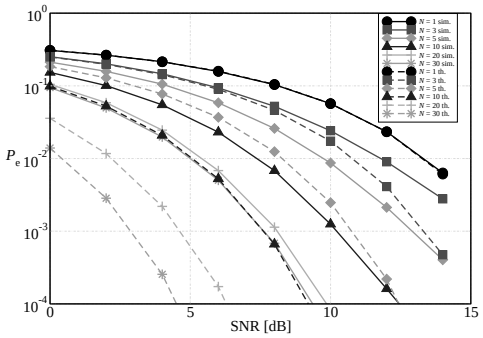
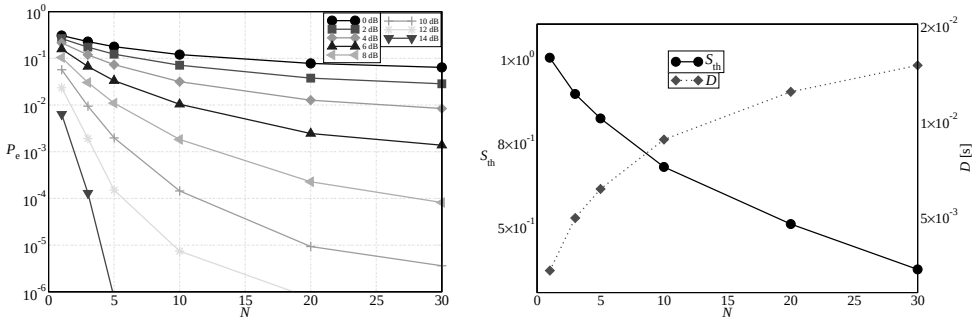


Figure 4.31: Performance analysis in a scenario without clustering: probability of decision error performance as a function of the SNR at the RFDs. Both simulations (solid lines) and theoretical (dashed lines) results are shown.

RFD clusters with 2 FCs), (ii) 4-4-4-4 (i.e., 4 4-RFD clusters with 4 FCs, (iii) 14-1-1 with 3 FCs (i.e., one cluster is composed by 14 RFDs and two clusters are composed by one RFD each), and (iv) 10-2-2-2 with 4 FCs (i.e., the network is divided into 4 clusters, one formed by 10 RFDs and others formed by 2 RFDs each), and (v) 8-2-2-2-2 with 5 FCs (i.e., the network is divided into 5 clusters, one formed by 8 RFDs and others formed by 2 RFDs each).

In Fig. 4.31, the performance in scenarios without clustering is analyzed: more precisely, the probability of decision error at the AP is shown as a function of the observation SNR. In the configuration with $N = 1$ RFD (solid line with circles), i.e., point-to-point communication between an RFD and the coordinator, the probability of decision error has the typical trend of On Off Keying (OOK)—this is a "sanity check" for our simulator[5]. The coordinator, in fact, may decide only for either 0 or 1 using the (possibly erroneous) decisions received from the RFDs. As one can see, the probability of decision error reduces for increasing numbers of RFDs. This can be explained recalling the data fusion mechanism described in Section 3.4.3. In fact, since communication links are modeled as ideal, for a fixed observation SNR, the larger

[5]Note that the modulation schemes foreseen by the Zigbee standard have no effect on the probability of decision error, since the channel is considered ideal.



(a) (b)

Figure 4.32: Performance analysis in a scenario without clustering: (a) probability of decision error at the AP as a function of the SNR at the RFDs, and (b) throughput and delay as functions of the number N of transmitting RFDs.

the number of received decisions, the lower the probability of decision error with a majority decision rule. Note that for increasing numbers of RFDs, the relative reduction of the probability of decision error is negligible (e.g., the improvement is higher when the number of RFDs increases from 1 to 10 than when the number of RFDs increases from 20 to 30). In the same figure, we also show theoretical results (dashed lines) obtained with the analytical framework summarized in Subsection 3.4.3. Since this analytical framework does not take into account the medium access policies, the predicted performance is in agreement with the simulation results only in a scenario with $N = 1$ RFD. In the other scenarios, the probability of decision error predicted by the analytical framework is better than that obtained through simulations. This is due to the fact that our simulator takes also into account the presence of collisions. In this case, since some packets may be lost or dropped, the probability of decision error worsens. The negative impact of the collisions exacerbates when the number of transmitting nodes increases.

In Fig. 4.32, the performance in scenarios without clustering is analyzed: more precisely, (a) the probability of decision error is shown as function of the number of transmitting RFDs, and (b) the throughput and the delay are shown as functions of

the number of nodes. In Fig. 4.32 (a), the probability of decision error is a monotonically decreasing function of N for all considered values of the observation SNR at the RFDs. For each network configuration, the probability of decision error reaches a floor which is strictly related to the collisions and, therefore, to the maximum achievable throughput in each scenario—in the cases with high SNR (12 dB and 14 dB), instead, the floor is not visible since it appears for very low (out of scale) values of the probability of decision error.

In Fig. 4.32 (b), the throughput and the delay are shown as functions of the number of transmitting RFDs. These curves are obtained by considering a fixed SNR (equal to 0 dB) at the RFDs. Our results, however, show that the throughput and the delay are not affected by the value of the observation SNR at the RFDs. We consider, in fact, ideal communication channels, so that the noise affects only the observations at the RFDs and not the packets transmitted from either the RFDs or the FCs. Consequently, throughput and delay do not depend on the observation SNR. The throughput curve (solid line with circles) decreases monotonically. In particular, for small values of N, it remains close to 1. When the number of transmitting nodes increases, instead, the number of collisions in the channel increases as well and the throughput reduces. Comparing the results in Fig. 4.32 (a) with those in Fig. 4.32 (b), the negative impact that a larger number of RFDs has on the throughput is compensated, in terms of probability of decision error, by the data fusion mechanism. In Fig. 4.32 (b), we also show the delay (dotted line with diamonds). As the intuition may suggest, the delay is small for small values of N. When the traffic increases, instead, due to a larger number of collisions, the delay is higher, since the channel is busy for a longer period of time and the probability of finding the channel idle reduces. Finally, for large values of N, the delay seems to start saturating to a maximum value. In this case, in fact, due to the increased offered traffic, it is likely that there is at least an RFD ready to send its packet as soon as the channel is idle.

In Table 4.4, we show the aggregate throughput predicted by previous analysis. When the number of transmitting nodes is small, the aggregate throughput is high (close to the maximum possible for each network configuration). When the number of transmitting RFDs increases, instead, the aggregate throughput tends to saturate.

Table 4.4: Aggregate throughput, in a scenario without clustering, as a function of the number N of transmitting RFDs.

N	S_{agg} [pck/s]
1	2
3	5.35156158
5	8.1906672
10	13.4860858
20	20.1639324
30	22.1357394

Once the saturation is reached, the number of collisions is so large than an increase of the traffic load has no longer effect on the aggregate throughput.

In Fig. 4.33, we present a comparison, in terms of probability of decision error in no-clustered scenarios, between scenarios with and without the use of ACK messages. As expected, the presence of ACK messages guarantees better performance, because each message sent by a RFD to the coordinator is confirmed in the case of correct reception. In particular, the presence of ACK messages allows the coordinator to make its final decision on the basis of a larger number of observations (especially in the case with $N = 30$ RFDs, where the gain introduced by the presence of ACK messages is higher), increasing the reliability of the decisions. On the opposite, the use of ACK messages, in the presence of a larger number N of transmitting RFDs, increases the traffic, so that the performance, in terms of throughput, worsens.

In Fig. 4.34, we analyze the impact of non-uniform clustering on the probability of decision error—as a performance reference, the probability of decision error in the case with uniform clustering is also shown. We consider scenarios with $N = 16$ RFDs and the following network configurations: (i) no clustering, (ii) 2 8-RFD clusters with 2 FCs, (iii) 4 4-RFD clusters with 4 FCs, (iv) 14-1-1 with 3 FCs, (v) 10-2-2-2 with 4 FCs, and (vi) 8-2-2-2-2 with 5 FCs. According to the analytical results presented in [81] and the previously shown simulation results, the best network performance is obtained in the absence of clustering. The worst performance, instead, is obtained

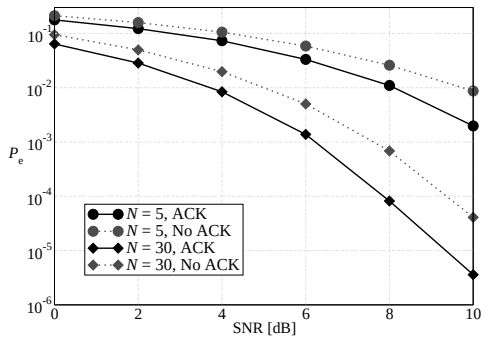


Figure 4.33: Probability of decision error in scenarios with $N = 5$ and $N = 30$ RFDs, respectively, in the absence of clustering. Both the presence (solid lines) and the absence (dotted lines) of ACK messages are considered.

in the 14-1-1 scenario, i.e., with 3 FCs and non-uniform clustering. In this case, in fact, the information collected by the RFDs associated with the largest cluster is very reliable. On the other hand, the information collected by the other two clusters is more likely to be noisy, and the final decision is thus likely to be wrong. Observing the results in Fig. 4.34, one can conclude that, in the presence of non-uniform clustering, the best performance is obtained in the case of less unbalanced clusters. In this case, in fact, decisions made by intermediate FCs are more reliable, so that it is more likely that the final decision made by the coordinator is correct. In the case of uniform clustering, instead, the probability of decision error is *not* affected by the number of clusters in the network, as long as the number of RFDs remains the same. In this case, in fact, observing Fig. 4.34 one can note that the curves corresponding to the scenarios with 4 4-RFD clusters and 2 8-RFD clusters are almost overlapped. This is due to the fact that a smaller number of clusters is compensated by a higher quality of the intermediate decisions. This result is in agreement with the theoretical analysis presented in [81].

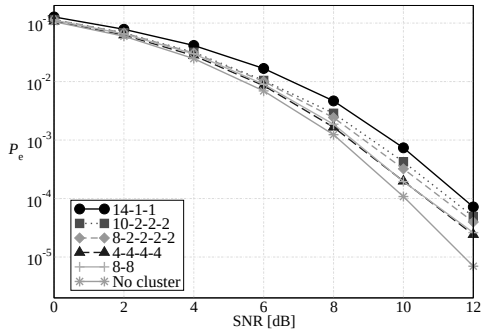


Figure 4.34: Probability of decision error at the AP in scenarios with $N = 16$ RFDs both in uniform and non-uniform clustering configurations. The considered topologies are the following: (i) no clustering, (ii) 2 8-RFD clusters and 2 FCs, (iii) 4 4-RFD clusters and 4 FCs, (iv) 14-1-1 clustering configuration with 3 FCs, (v) 10-2-2-2 clustering configuration with 4 FCs, and (vi) 8-2-2-2-2 clustering configuration with 5 FCs.

4.5.6 Design Guidelines

On the basis of the performance analysis carried out in Section 4.5.2, we now try to derive useful guidelines for the design of Zigbee networks with a desired performance level. In Fig. 4.35, we present combined results of network transmission rate, delay and network tolerable death level in different (both uniform and non-uniform clustering) clustering configurations with $N = 16$ RFDs. The purpose of this figure is to show the unavoidable trade-offs to deal with in order to guarantee a specific performance level. All network configurations considered in Subsection 3.4.2 are considered as well in Fig. 4.35.

In Fig. 4.35 (a), the delay is shown as a function of both network tolerable death level and network transmission rate. In addition, on the $\chi_{\text{net}} - S$ plane the contour curves are shown. The shape of the 3D surface is similar to a skewed half-tube: in fact, there is a peak (in terms of delay), with maximum value in correspondence to $\chi_{\text{net}} = 20$ % and $S = 6000$ b/s. This means that the 1-cluster configuration leads to the highest delay. In an application where a low delay is required, the acceptable

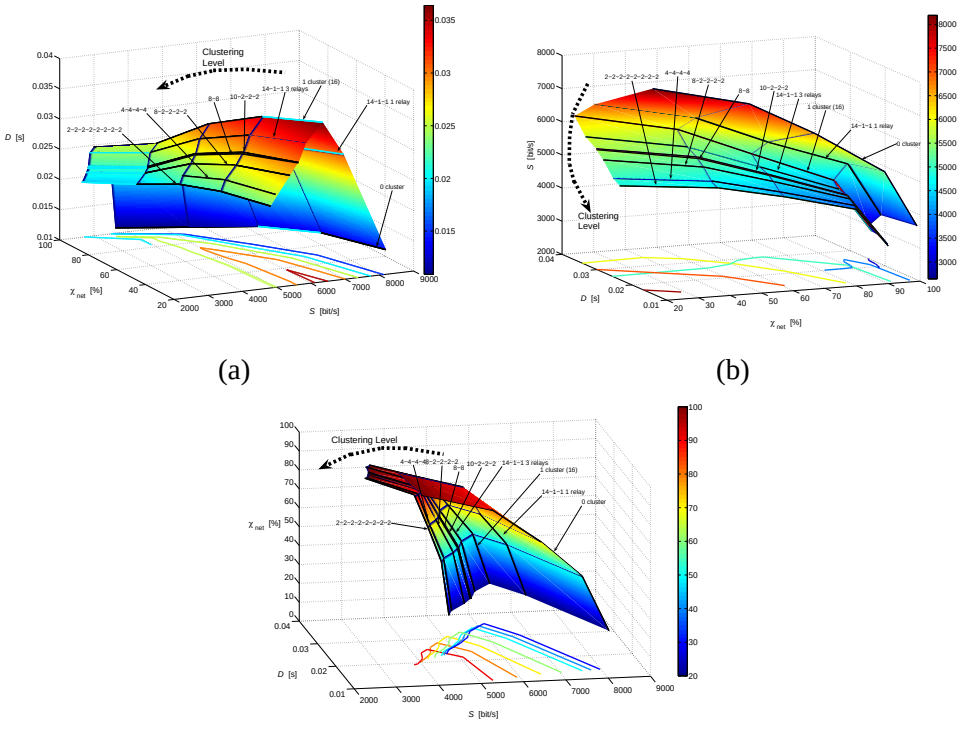


Figure 4.35: Performance evaluation in a scenario with $N = 16$ RFDs and various uniform and non-uniform clustering configurations; (a) delay as a function of network transmission rate and χ_{net}, (b) network transmission rate as a function of delay and χ_{net}, and (c) χ_{net} as a function of network transmission rate and delay. The considered packet generation rate is $g = 2$ pck/s.

configurations are either the ones with no relay or the ones with many relays (i.e., high clustering level). In the former case the network transmission rate is high, as shown in Fig. 4.25 (a) and Fig. 4.25 (b); in the latter case, instead, the network transmission rate is limited by the presence of the relays.

In Fig. 4.35 (b), the network transmission rate is shown as a function of χ_{net} and D. From the results in this figure it is evident that the network transmission rate

reduces as the number of relays in the network increases. In particular, for each configuration and for specific values of χ_{net} and D, there exists a maximum achievable network transmission rate and a suitable network configuration which guarantees the best performance. For a specific value of the network transmission rate, there exists at least a pair of values of D and χ_{net}—and, therefore, a specific pair of network configurations—in correspondence to which the required network performance is met. In Fig. 4.35 (c), χ_{net} is shown as a function of the network transmission rate and the delay. As in the previous figures as well, in this figure, it can be seen that, for a fixed value of χ_{net}, there exists a configuration that guarantees the best trade-off between network transmission rate and delay.

Summarizing, one can conclude that:

- the best network configuration is always without any intermediate relay;
- in clustering configurations where there is at least one RFD directly connected to the AP, the network transmission rate is higher than in fully clustered configurations; the previous conclusion does not hold in terms of delay, even if acceptable values of delay are obtained;
- both the network transmission rate and the delay are decreasing functions of the number of relays in the network: therefore, in the presence of a high clustering level both the network transmission rate and the delay are low, whereas in the presence of a low clustering level the opposite holds.

In Fig. 4.36, the same results of Fig. 4.35 are presented in scenarios with $N = 64$ RFDs. The considerations, carried out for scenarios with $N = 16$ RFDs, are still valid, the only difference between the two network configurations being the fact that the network transmission rate and the delay in the scenario with $N = 64$ RFDs are higher than in the scenario with $N = 16$ RFDs. As before, the most important observation is that the network performance trend is strictly related to the number of relays in the network.

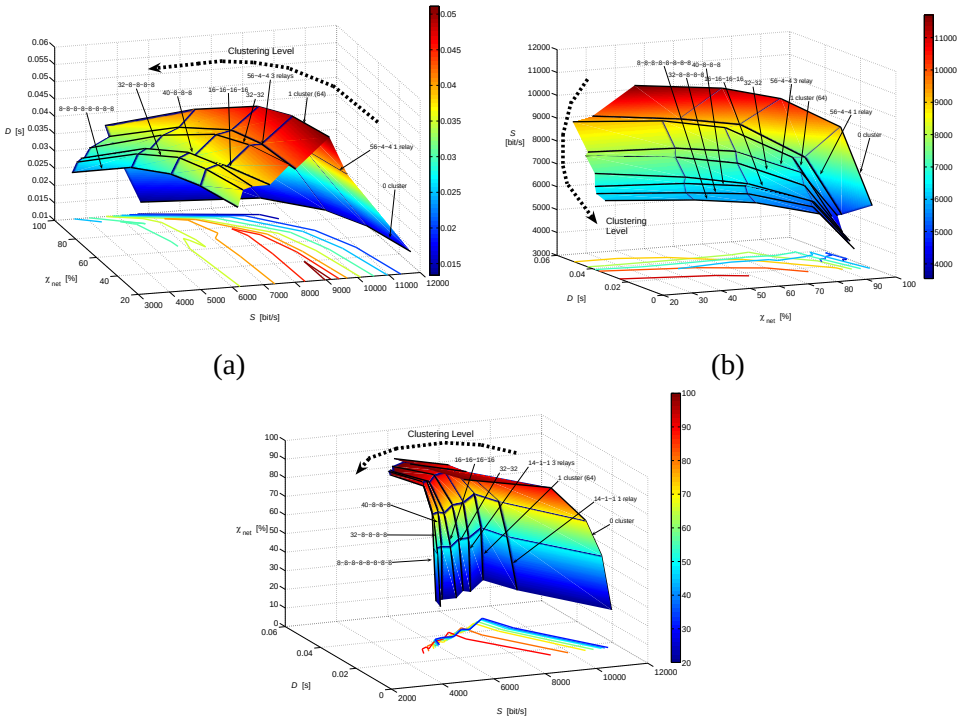
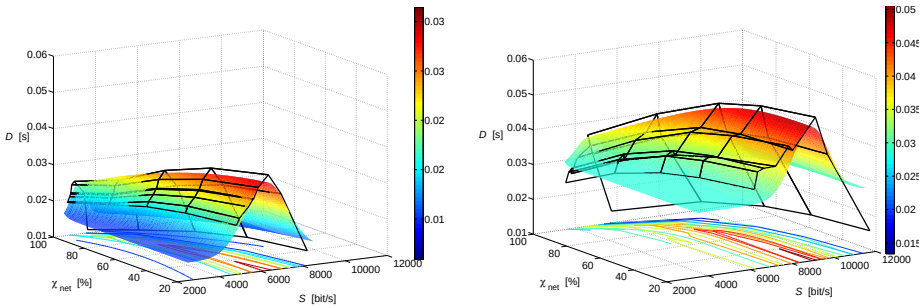


Figure 4.36: Performance evaluation in a scenario with $N = 64$ RFDs and various uniform and non-uniform clustering configurations; (a) delay as a function of network transmission rate and χ_{net}, (b) network transmission rate as a function of delay and χ_{net}, and (c) χ_{net} as a function of network transmission rate and delay. The considered packet generation rate is $g = 2$ pck/s.

4.5.7 A Simple Analytical Model for the Zigbee Performance Surface

In this subsection, we provide a simple approach for deriving a closed-form approximation of the multidimensional network performance surface introduced in Subsection 4.5.6. The derivation of an analytical expression for this surface is very useful, since simulation results may be difficult to obtain, especially in the case of large-scale



(a) (b)

Figure 4.37: Performance evaluation in a scenario with (a) $N = 16$ RFDs and (b) $N = 64$ RFDs. Both the realistic surface (transparent) and its approximation (filled) are shown.

Zigbee WSNs.

We propose a simple fitting model of the simulation results with a pre-determined function of two variables $\widehat{D} = \widehat{D}(S, \chi_{\text{net}})$. In particular, the considered fitting function can be expressed as

$$\widehat{D} = \alpha_1 f_1(S, \chi_{\text{net}}) + \alpha_2 \tag{4.1}$$

where the values of α_1 and α_2 are obtained through the minimization of the mean square error (MSE) between simulation and analytical data. An expression for f_1 is provided in Subsection 4.5.8, where the details of its derivation are also are given.

In Fig. 4.37, the "true" Zigbee performance surface (the transparent one) and its analytical approximation (the filled one) are compared in scenarios with (a) $N = 16$ RFDs and (b) $N = 64$ RFDs, respectively. As one can observe, for both considered scenarios the heuristically derived surfaces approximate well the real surfaces. However, some differences between the true surfaces and their approximations can be observed, and this is due to the non-Gaussian shape of the true surface, which is instead more irregular. In Table 4.5, the optimized values of α_1 and α_2, and the corresponding delay minimum mean square error (MMSE) are shown, in scenarios with $N = 16$ and $N = 64$ sensors, respectively. More precisely, the normalized MMSE is

Table 4.5: Coefficients of the linear combination in (4.1) and corresponding normalized MMSEs in the cases with $N = 16$ and $N = 64$, respectively.

	$N = 16$	$N = 64$
α_1	0.154	0.128
α_2	0.015	0.0028
MMSE	1.53%	39%

computed as

$$\text{MMSE} = \frac{|D - \widehat{D}|^2}{|D|^2}$$

where the notation $|\boldsymbol{x}|$ stands for the norm of $\boldsymbol{x}$. From the results in Table 4.5, the reader should observe that the normalized MMSE is small in scenarios with $N = 16$, whereas it becomes larger in scenarios with $N = 64$. Therefore, the proposed approximation needs to be refined for large-scale WSNs.

The derivation of an accurate approximation for the Zigbee surface allows to simplify the analysis and design of WSNs. In particular, we are currently working on the derivation of a general analytical model for the coefficients in (4.1), in order to have an accurate approximation of the performance surface for a generic value of N. Our simple analytical model could also help in analyzing the network behavior in dynamical scenarios, where the network configuration (e.g., the clustering structure) may change due to sensors' failures or in order to prolong the WSN lifetime [80]. Finally, through a better understanding of the properties of the multidimensional network performance surface, one could design the WSN in order to maintain a given performance level, and this goal could be achieved by properly "moving" the operating point on the surface.

4.5.8 Analytical Approximation of the Performance Surfaces

In order to derive an analytical model of the three-dimensional surfaces depicted in Fig. 4.35, we first extract an equation which describes the "peak" (in terms of delay) of the half tube. Since the maximum of each surface corresponds to the configuration

with 1 cluster and 1 relay, we first obtain an expression for the 1-cluster projection curve on the $S-\chi_{\text{net}}$ plane. The approximating curve, for the case with $N=64$ RFDs,[6] is

$$S = a_2 \cdot \chi_{\text{net}}^{b_2} + c_2 \triangleq g(\chi_{\text{net}}) \tag{4.2}$$

where a_1, b_1, c_1, and d_1 are proper constants whose values are obtained by minimizing the MSE with respect to simulation-based points.

The same procedure has been repeated to approximate the shape of the projection of the "peak" on the $D-\chi_{\text{net}}$ plane. The fitting expression is similar to (4.2) and has the following form:

$$D = a_1 \cdot e^{b_1 \cdot \chi_{\text{net}}} + c_1 \cdot e^{d_1 \cdot \chi_{\text{net}}} \triangleq \ell(\chi_{\text{net}}) \tag{4.3}$$

where a_2, b_2, and c_2 are other suitable constants obtained with the MMSE-based approach used for (4.2).

Finally, we have approximated the $D-S$ curve obtained over the section of the surface by fixing the value of the tolerable network death level. From our analysis, it comes out that the best approximating function, for each value of χ_{net}, is a Gaussian function with varying (optimized) variance. Therefore, the final expression for f_1 is

$$f_1(S,\chi_{\text{net}}) = \ell(\chi_{\text{net}}) \cdot \exp\left\{ -\left[\frac{S-g(\chi_{\text{net}})}{c(\chi_{\text{net}})} \right]^2 \right\}$$

where $\ell(\chi_{\text{net}})$ and $g(\chi_{\text{net}})$ are, respectively, given by (4.2) and (4.3), and $c(\chi_{\text{net}})$ is a proper function which characterizes the standard deviation of the Gaussian approximation and can be expressed as

$$c(\chi_{\text{net}}) = p_1 \chi_{\text{net}}^2 + p_2 \chi_{\text{net}} + p_3. \tag{4.4}$$

By using the Curve Fitting Toolbox of Matlab, all the coefficients involved in (4.2), (4.3), and (4.4) have been computed with a confidence interval equal to 95%. The obtained values are shown in Table 4.6.

[6] Similar considerations can be carried out also for the case with $N=16$ RFDs.

Table 4.6: Coefficients used for the computation of the approximating function in the case with $N = 16$ and $N = 64$, respectively.

	$N = 16$	$N = 64$
a_1	-0.002132	-0.004571
b_1	0.03209	0.03423
c_1	0.1355	0.1585
d_1	-0.001969	0.002149
a_2	-0.001302	-2.943×10^{-12}
b_2	3.187	7.579
c_2	6256	8887
p_1	-0.03429	-0.09405
p_2	-7.48	6.577
p_3	1576	1119

4.6 Concluding Remarks

In this chapter, we have first provided a quick overview of the Opnet simulator. Then, we have presented the performance results for (i) the WSN-based target detection system, (ii) the adjacency-based power allocation strategy, and (iii) the clustered networks with and without data fusion. In each case, we have highlighted the importance of the use of an "smartly" configured network. For such a reason, in each case we have provided for some insights which helps in the optimal tuning of the nodes/network parameters, especially according to given performance requirements.

Chapter 5

RFID-Controlled Zigbee Networks

5.1 Introduction

In this chapter, we focus on the integration between two different technologies in order to obtain a two layers network [88]. One of the two layers will act as “control” of the other layer, which will be used to exchange data. Since nodes may be positioned in non-easily accessible places, a high energy efficiency is required in order to maximize the network lifetime and minimize the maintenance costs.

A simple and straightforward solution to maximize the network lifetime consists in turning off all nodes which are not needed, e.g., when the node spatial density is higher than that required to satisfy the sensing requirements. In such a case, selective (and cyclic) use of the sensing nodes allows to extend the network lifetime, still guaranteeing an acceptable performance level. A possible solution is based on the use an RFID network to control a Zigbee data network, in order to dynamically and cyclically put into a “deep sleep” state nodes with low battery energies.

The structure of this chapter is the following. In Section 5.2, we provide for an overview of the RFID technology. In Section 5.3, we introduce the considered system model and the slotted model of the IEEE 802.15.4 Opnet model. Then, we analyze the proposed approaches to extend the network lifetime. In Section 5.4, we present the performance results related to the use of integrated networks. Finally, Section 5.5

concludes this paper.

5.2 RFID Technology

The RFID technology specifies two different devices: (i) the *reader* and (ii) the *transponder* or *tag*. The tag is an electronic device that stores data useful for identification and is placed on the object to be tracked. On the opposite, the reader is the device used to interrogate the tags. The tags can be classified into three main categories, according to the power source they are equipped with: (i) *active*, (ii) *semi-passive*, and (iii) *passive* [46]. The active tags, in order to respond to the interrogation of the reader, use their own internal batteries for processing operations and signal transmission. The semi-passive tags, instead, use their batteries only to power the internal processor and not to broadcast the return signal. Finally, the passive tags, being not equipped with batteries, exploit the energy of the received signal to respond to the reader (*backscattering* technique). Obviously, the tags with longer lifetime (virtually infinite) are the passive ones, whereas the active tags have the longest transmission range. Once the reader has interrogated a tag and this has replied correctly without incurring into collisions with other tags (proper anti-collision mechanisms are used [89]), the reader sends an *acknowledgment* message to the tag to confirm correct message reception.

The functionalities provided by the RFID technology are mainly related to error-free communication between reader and tags (bidirectional), identification and communication with multiple tags, selection of a target set of tags, lock of data stored in the tags, and writing and overwriting operations in the tags. The communication from the reader to the tags, according to the backscattering technique, is realized through amplitude modulation of the interrogation signal. On the other hand, after the transmission of the interrogation signal, the reader emits a constant-power signal which allows the tags to respond to the former signal through a load modulation technique [46]. The reader receives data from the tags as a variation of the reflection of the constant-power signal.

When the reader interrogates a set of tags, these may respond immediately, leading to a collision. One of the most important standards for the RFID technology is the

ISO/IEC 18000-6 standard [90]. This standard defines a reference model for identification systems which operate in the range of Ultra High Frequencies (UHF). In particular, this standard regulates RFID systems in the frequency band around 868 MHz in Europe. The 18000-6 standard defines two different transmission types: (i) *type A* and (ii) *type B*. Both transmission types make use of a rate equal to 40 kbps and use a binary phase modulation. The difference between these two transmission types resides in the medium access control (MAC) protocol. If a collision is detected by the reader, the Type A tags retransmit the interrogation request according to the *Aloha* MAC protocol [91]. According to this protocol, a tag transmits as soon as a packet is generated and it assumes that a collision has happened if an acknowledgment message is not received after a given period of time. Type B tags, instead, use the Binary Tree Protocol (BTP) [89] as anti-collision mechanism. The idea of the BTP is that, upon a collision, the set of tags is divided into two subsets: one set tries to retransmit their messages immediately, whereas the other set waits till a new interrogation request is generated by the reader. Eventually, only one tag will retransmit, and its identity (or any other stored information) will be successfully acquired by the reader. This recursive procedure is then repeated, under the constraint that the already censed tags do not respond to the interrogation signal.

5.3 Hybrid Zigbee-RFID Networks

5.3.1 System Model

Our system model consists of a network where N nodes are deployed to monitor a particular phenomenon of interest. Under the assumption that a minimum spatial density of observations is required, i.e., a minimum number $N_{\min}$ of RFDs needs to be used to monitor a given surface (with area A) of interest, and that the number N of deployed RFDs is larger than $N_{\min}$, our approach consists in implementing a *selective wake-up* of the RFDs. More precisely, our strategy consists in selectively turning on and off the RFDs in order to equalize the energy consumption of the nodes in the network. In order to maximize the network lifetime, as soon as the residual energy of an active RFD becomes lower than a fixed threshold E_{th}, the RFD is switched into

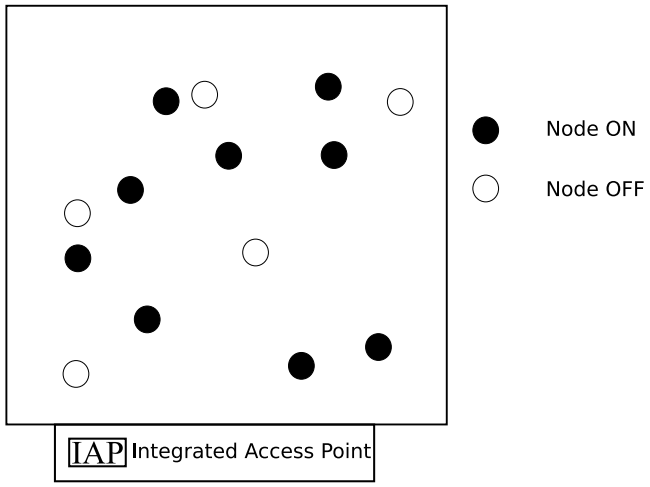


Figure 5.1: Network scenario with $N = 14$ nodes, out of which and $N_{\min} = 9$ are active and $N - N_{\min} = 5$ have been turned off by the deep sleep algorithm.

the sleep state. At the same time, one of the remaining $N - N_{\min}$ RFDs (previously switched off) is woken up, so that an overall minimum spatial density of observations of the phenomenon of interest is guaranteed. This procedure is referred to as *deep sleep algorithm*. An illustrative example of the status, at a given time, of a network where the deep sleep algorithm is used, is shown in Figure 5.1. We point out that the RFDs are only switched into the sleep state, instead of being turned off, in order to prevent them from losing synchronization with the coordinator and the other RFDs in the network. In fact, according to the Zigbee standard, the network joining operations performed when an RFD is turned on introduce a delay longer than the length of the wake up operations, since the RFD must wait for a new beacon in order to synchronize with the network before starting its transmission. On the opposite, the RFDs in the sleep state switch to the active state only when a beacon transmission is scheduled and subsequently, if not required, return into the sleep state without losing their synchronization with the network. The goal of the deep sleep algorithm is energy equalization among the RFDs, so that the residual energies of the RFDs are

balanced. When an RFD is in the sleep state, in fact, its energy consumption is four orders of magnitude lower than in the active state [92, 93].

5.3.2 Integrated Opnet Structure

Our hybrid Zigbee-RFID model is based on an IEEE 802.15.4 Opnet model developed at the University of Porto, Portugal [94] and on an RFID Opnet model developed at the University of Parma, Italy [95]. The IEEE 802.15.4 model is beaconed in order to synchronize the nodes in the network and save energy. In addition, this model contains a battery module which is used to evaluate the energy consumption of the devices.[1] On the other side, the Opnet RFID model implements the ISO/IEC 18000-6 standard, considering faded wireless channels. Since the Zigbee and RFID models are disjoint, we have developed a hybrid Opnet Zigbee-RFID model where the subcomponents communicate and cooperate.[2] Obviously, the beaconing mechanism is modified in order to take into account also the RFDs which are not transmitting during a frame. Note that in our system the BTP (embedded in the ISO/IEC 18000-6 standard) is not used, since the RFID tags are used only as switches for the (Zigbee) RFDs and do not need to be identified. In Figure 5.2 (a), the logical scheme of the integrated Zigbee-RFID network is shown. More precisely, the RFID network lays on top of the Zigbee network. In our hybrid system, the information is transferred through the IEEE 802.15.4 (logical) network which is, in turn, controlled by the RFID network. In Figure 5.2 (b), instead, we show the integrated devices used in our network: the *integrated AP*, which is obtained from the combination of an RFID reader and a Zigbee coordinator; and *integrated node*, which is obtained from the combination of an RFID tag and a Zigbee RFD. We point out that the integrated node is battery powered, whereas the integrated AP is supposed to be connected to the electrical system, so that battery exhaustion is not an issue for the latter, but only for the former.

[1] We remark that the energy consumption values considered in this model are typical of MICAz devices [92].

[2] We point out that the realization of an experimental prototype physically connecting a Zigbee RFD with an RFID tag is currently under study.

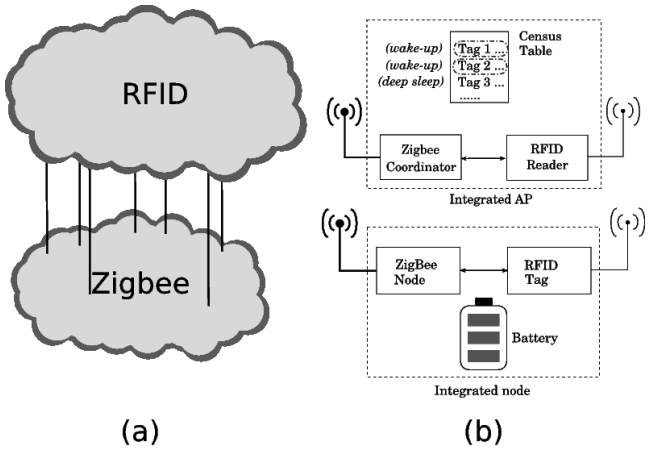


Figure 5.2: Logical scheme of the integrated network: (a) the Zigbee and RFID networks are combined together; (b) the integrated nodes and AP are obtained from the integration of Zigbee and RFID devices.

5.3.3 Deep Sleep Algorithm

We now describe the basic functionalities of an hybrid Zigbee-RFID network. When an RFD is in the sleep state and is selected by the coordinator, the associated RFID tag receives the signaling message from the reader (through the logical RFID network) and switches on its RFD, which enters into the active state and starts communicating (through the logical Zigbee network). Similarly, when an RFD is active and its residual energy becomes lower than a threshold value, the RFD communicates it to the Zigbee coordinator. The coordinator informs its associated RFID reader, which, by sending a proper message to the tag, forces the selected RFD into the sleep state. Since the deep sleep algorithm is managed by the coordinator of the Zigbee network, the active RFDs embed their residual energies inside the data packets sent to the coordinator. In this way, the coordinator is constantly aware of the residual energy of each active RFD. On the other hand, since an RFD in the sleep state does not transmit any packet, the coordinator also uses an estimation algorithm to predict the energy consumed by an RFD during the sleep state.

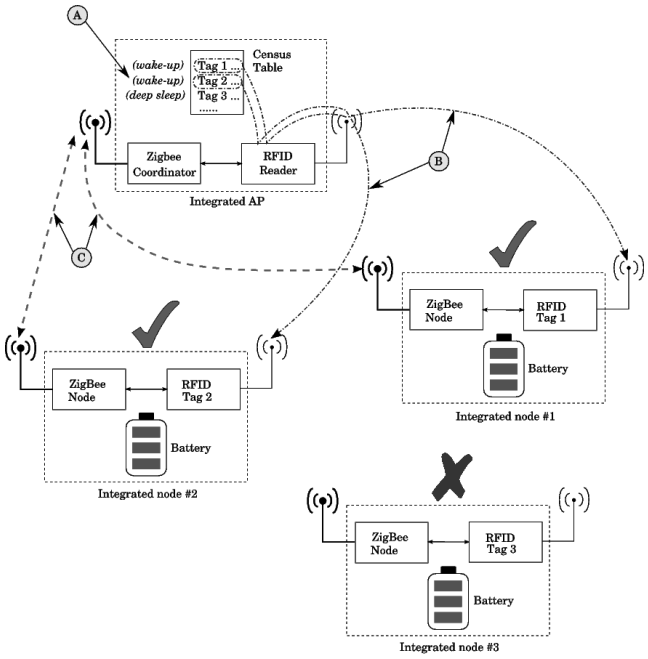


Figure 5.3: Initial steps of the deep sleep algorithm for selective wake-up of remote nodes: (A) $N_{\min}$ out of N RFDs are selected by the AP; (B) the selected RFDs are notified through the associated RFID tags; (C) the activated RFDs start communicating.

We now describe the main steps of the deep sleep algorithm through illustrative diagrams. Figure 5.3 refers to the operations involved when the network is started. All the RFDs are equipped with identical batteries and we refer to the initial energy level as 100%. At the network start-up, only $N_{\min}$ RFDs are turned on, whereas the remaining $N - N_{\min}$ nodes are turned into the sleep state. The integrated AP creates a *census table*, where information about the position, the residual energy, and the status of each RFD is stored. In our simulations, we assume that the position of each RFD is preliminarily available at the integrated AP, whereas the entry of the census table relative to the residual energy of an RFD is updated as soon as a packet (containing this information) is received from the node. If an RFD is in the sleep state, its residual

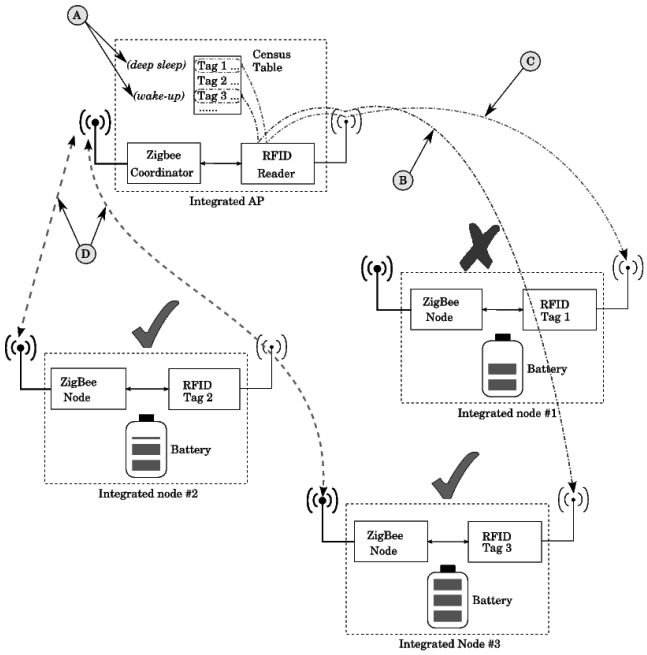


Figure 5.4: Steps of the replacement of an active node with residual energy under threshold and a sleeping node with higher residual energy: (A) the AP identifies an RFD with low residual energy; (B) a sleeping node with high residual energy is turned on and (C) the selected active RFD with low residual energy is turned off; (D) the set of active nodes communicate.

energy value is estimated by the AP before scheduling the transmission of a new beacon.

As soon as an RFD is turned on, its residual energy E_{r} reduces and, consequently, its value in the census table is updated. The evolution of the network according to the deep sleep algorithm is illustrated in Figure 5.4. When the energy of an RFD becomes lower than an initial threshold $E_{\mathrm{th}}^{(0)}$, the integrated AP forces the RFD into the sleep state through its associated RFID tag. At the same time, the integrated AP forces one of the sleeping nodes with higher residual energy to wake up. This procedure is repeated until there are no more RFDs with residual energies higher than $E_{\mathrm{th}}^{(0)}$.

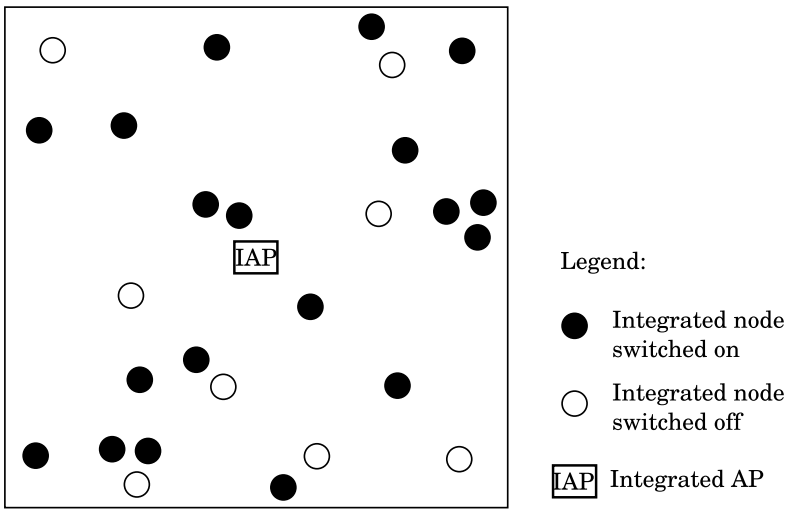


Figure 5.5: Network scenario with $N = 27$ nodes, out of which and $N_{\min} = 19$ are active and $N - N_{\min} = 8$ have been turned off by the deep sleep algorithm. The monitored surface is not divided into cells..

At this point, the threshold value of the residual energy is decreased of a pre-defined energy step E_{s}. More precisely, the new energy threshold is set to $E_{\mathrm{th}}^{(1)} = E_{\mathrm{th}}^{(0)} - E_{\mathrm{s}}$. This procedure is repeated until the energy threshold is so low that data communications in the Zigbee logical network are no longer possible.[3] At this point, as soon as the minimum spatial density of observations is no longer guaranteed, the network is declared dead and the simulation stops. More details on the energy consumption at the nodes will be given in Subsection 5.4.1.

In general, N integrated nodes could be deployed randomly over a given surface. As an illustrative scenario, in Figure 5.5 $N = 27$ nodes are deployed randomly over a square surface. The integrated AP is placed in the center of the monitored surface, whereas the $N = 27$ integrated nodes are deployed according to a 2-D Poisson distribution.

[3]Denoting by n_{fin} the final updating step, $E_{\mathrm{th}}^{(n_{\mathrm{fin}})} = E_{\mathrm{th}}^{(0)} - n_{\mathrm{fin}} \cdot E_{\mathrm{s}}$ is too low to support communication in the Zigbee network. The value of n_{fin} depends on the specific parameters of the Zigbee network.

5.3.4 Deep Sleep Algorithm with Virtual Spatial Grid

While the deep sleep algorithm takes into account the *overall* observation spatial density (considering the total number of nodes deployed over the entire network surface), in several applications it might be of interest to monitor *homogeneously* the network surface, i.e., to guarantee an approximately constant *local* observation spatial density across the surface. In order to do this, we follow an approach based on the use of a *virtual spatial grid* over the network.

If the monitored area is virtually partitioned into observation cells (with a local minimum required node spatial density per cell), then the "standard" deep sleep algorithm can no longer be applied, since it could lead to turn off all nodes of the same cell. However, there might be scenarios where particularly critical phenomena need to be observed with high accuracy over the entire monitored surface. In Figure 5.6, an illustrative example of the network in Figure 5.5 with an overlaid virtual spatial grid is shown. As one can see, different cells may contain different numbers of nodes. In the following, we will denote the cells where there is only one integrated node as *secondary*, whereas the cells with more than one integrated node will be denoted as *primary*. In the presence of a virtual spatial grid, the deep sleep algorithm is applied in each primary cell, where the observations are redundant. In the secondary cells, no deep sleep algorithm is applied, since the integrated nodes cannot be turned off. The cells with no node are not relevant for monitoring purposes.

The network is declared dead as soon as all the RFDs of one of the primary cells die. This criterion comes from the assumption that primary cells are assumed to be associated with critical zones of the phenomenon under observation, whereas the secondary cells may be associated with less critical zones. Thus, when a primary cell contains no more RFD, we assume that the observations are no longer reliable and the network is declared dead. We remark that our approach is valid also in scenarios where $N > N_{\min}$, but it is more effective in scenarios where $N \gg N_{\min}$.

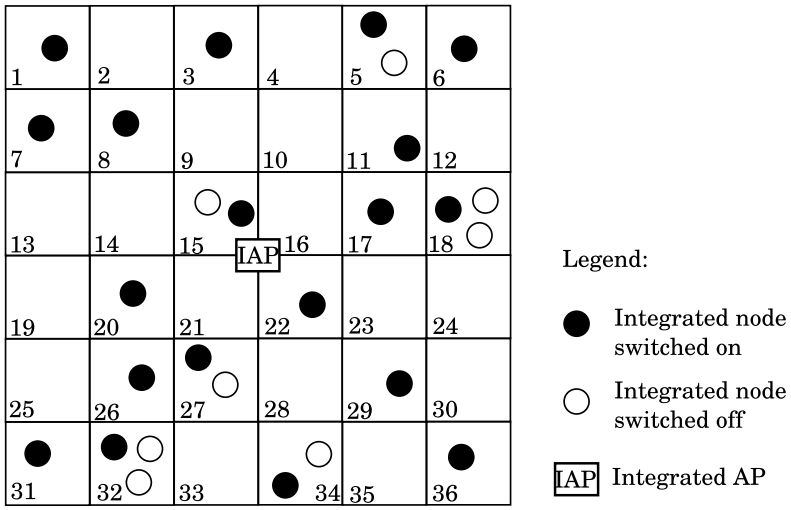


Figure 5.6: Network scenario with $N = 27$ nodes, out of which and $N_{\min} = 19$ are active and $N - N_{\min} = 8$ have been turned off by the deep sleep algorithm. The monitored surface is divided into cells and only one node per cell is active, according to the spatial grid procedure.

5.4 Performance Analysis

The simulations have been carried out considering networks with $N = 40$ nodes deployed over a square surface with 6 m long sides. In all cases, the RFDs send packets directly to the integrated AP, i.e., all network realizations have star topologies. The packet interarrival time T_{int} is either fixed to 0.25 s or Poisson-distributed with average value equal to 0.25 s. The considered packet length is $L = 200$ bit/pck. We assume that the network is dead when the energy level of all RFDs becomes 80% of the initial value. Each performance simulation result is the average of the results obtained considering 6 realizations of the network topology (according to a 2-D Poisson distribution).

Before starting the simulation-based performance analysis of hybrid Zigbee-RFID networks using the deep sleep algorithm, we provide the reader with some

details about the energy consumed by each integrated node. The average residual energy[4] at a generic instant t can be expressed as follows:

$$E_{\mathrm{r}}(t) = E_{\mathrm{i}} - E_{\mathrm{c}}(t) \tag{5.1}$$

where E_{i} is the initial energy of a node and $E_{\mathrm{c}}(t)$ is the energy consumed at the instant t. In particular,

$$E_{\mathrm{c}}(t) = E_{\mathrm{c\_TX}}(t) + E_{\mathrm{c\_RX}}(t) + E_{\mathrm{c\_idle}}(t) + E_{\mathrm{c\_sleep}}(t)$$

where:

$$\begin{aligned}
E_{\mathrm{c_TX}}(t) &= \lambda \cdot E_{\mathrm{c_TX_1_pck}} \cdot t = K_{\mathrm{TX}} \cdot t \\
E_{\mathrm{c_RX}}(t) &= \lambda \cdot (N-1) \cdot E_{\mathrm{c_RX_1_pck}} \cdot t = K_{\mathrm{RX}} \cdot t \\
E_{\mathrm{c_sleep}}(t) &\simeq E_{\mathrm{c_sleep_1_s}} \cdot \frac{\Delta t_{\mathrm{passivezone}}}{\Delta t_{\mathrm{superframe}}} \cdot t = K_{\mathrm{sleep}} \cdot t \\
E_{\mathrm{c_idle}}(t) &\simeq E_{\mathrm{c_idle_1_s}} \cdot \frac{\Delta t_{\mathrm{activezone}}}{\Delta t_{\mathrm{superframe}}} \cdot \\
&\quad \cdot \left[t - \Delta t_{\mathrm{TX_1_pck}} \cdot \lambda \cdot t - \Delta t_{\mathrm{RX_1_pck}} \cdot \lambda \cdot (N-1) \cdot t \right] \\
&= K_{\mathrm{idle}} \cdot t
\end{aligned}$$

λ is the packet transmission rate[5]; $E_{\mathrm{c\_TX\_1\_pck}}$ and $E_{\mathrm{c\_RX\_1\_pck}}$ are the energies consumed per packet transmission and reception acts, respectively; $E_{\mathrm{c\_sleep\_1\_s}}$ and $E_{\mathrm{c\_idle\_1\_s}}$ are the energies consumed during one second of persistence in the sleep and idle states, respectively; and $\Delta t_{\mathrm{activezone}}$, $\Delta t_{\mathrm{passivezone}}$, and $\Delta t_{\mathrm{superframe}}$ are the durations of the active zone, the passive zone, and the superframe, respectively. Therefore, we can approximate $E_{\mathrm{c}}(t)$ as

$$E_{\mathrm{c}}(t) \simeq \left[K_{\mathrm{TX}} + K_{\mathrm{RX}} + K_{\mathrm{sleep}} + K_{\mathrm{idle}} \right] \cdot t = K_{\mathrm{tot}} \cdot t \tag{5.2}$$

[4] The residual energy $E_{\mathrm{r}}(t)$ is the exact, not only the average, residual energy in the case of Poisson-distributed packet generation.

[5] In the case of constant packet interarrival time T_{int}, $\lambda = 1/T_{\mathrm{int}} = 4$ pck/s, whereas in the case of Poisson-distributed packet generation, $\lambda = 4$ pck/s is the *average* packet transmission rate.

Using (5.2) into (5.1), the residual energy at time t can be rewritten as

$$E_{\mathrm{r}}(t) \simeq E_{\mathrm{i}} - K_{\mathrm{tot}} \cdot t \ . \tag{5.3}$$

The final expression (5.3) for the residual energy is obtained under the assumption of no use of the deep sleep algorithm. However, if the deep sleep algorithm is used, only the values of K_{sleep} and K_{idle} change, but the linear dependence of E_{r} from t still holds, as will be shown in the following.

5.4.1 Deep Sleep Algorithm

In Figure 5.7, we evaluate the impact of the energy step E_{s} on the average (over the integrated nodes) residual energy in three different scenarios, considering $N = 40$ nodes in all cases: (i) with $N_{\mathrm{min}} = 40$ (i.e., the deep sleep algorithm is not used), (ii) with $N_{\mathrm{min}} = 30$, and (iii) with $N_{\mathrm{min}} = 25$. The packet generation rate is constant where packet interarrival time equal to 0.25 s is used. Observing the curves related to the scenarios with the deep sleep algorithm, it can be noted that the performance for a given value of N_{min} does not depend on E_{s}. This is to be expected, provided that the energy step is sufficiently smaller than the initial energy. In fact, when the deep sleep algorithm is used, N_{min} nodes are active at a time, and the value of E_{s} determines only the rate at which nodes get activated and deactivated. In particular, the following considerations can be carried out.

- If E_{s} is very small, then the energy levels of the nodes are equalized in a very fine way. However, this implies that nodes will cycle between on and off states very often, i.e., the integrated AP will spend energy in sending control messages and in the processing required to track the network topology changes.

- On the other hand, if E_{s} is not too small, then energy equalization is “rougher” but the integrated AP will spend less energy in network management operations. Since energy consumption is typically not an issue for the integrated AP, “finer” energy equalization should be chosen.

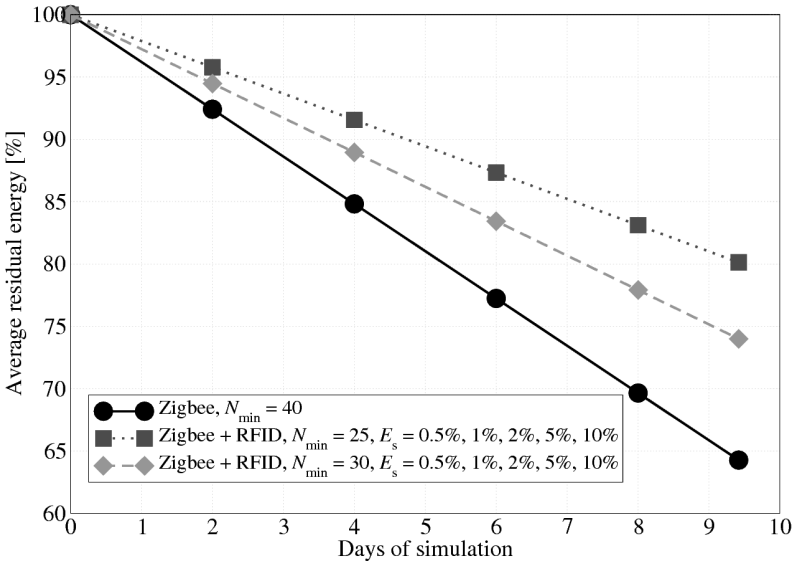


Figure 5.7: Average residual energy as a function of the days of simulation, in scenario with $N = 40$ nodes. Various values (25, 30, 40) of $N_{\min}$ are considered for the application of the deep sleep algorithm. In the case with $N_{\min} < N$, various values of the energy step are used.

To summarize, the network behavior of a hybrid Zigbee-RFID network using the deep sleep algorithm is very similar to a "standard" Zigbee network with $N_{\min}$ RFDs, regardless of the value of E_s.

In Figure 5.8, the average residual energy is shown as a function of the number of days of simulation, for various values of the number $N_{\min}$ of active RFDs. We have considered only $E_s = 1\,\%$ because, as shown in Figure 5.7, the average residual energy is independent of E_s. Observing the curves in Figure 5.8, one can conclude that:

- the smaller $N_{\min}$, the higher the energy saving, regardless of the simulation duration;
- the performance of the network without the deep sleep algorithm (i.e., the sce-

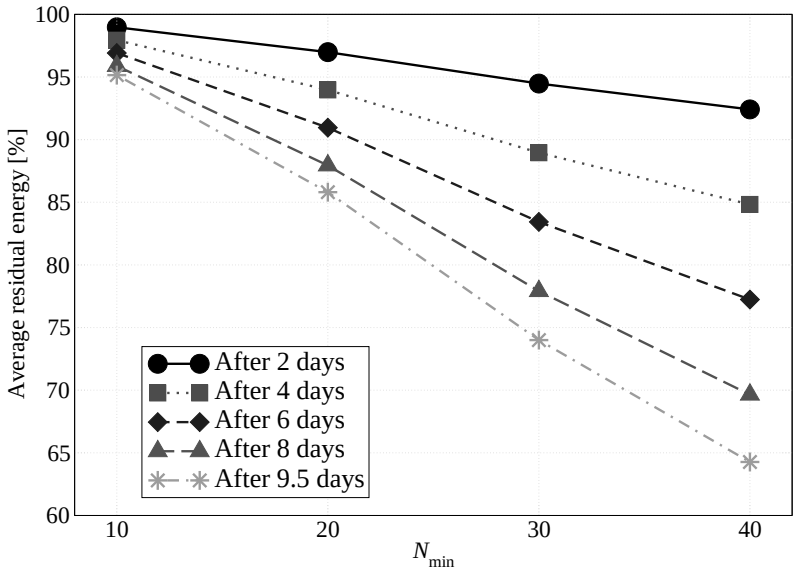


Figure 5.8: Average residual energy, as a function of $N_{\min}$, considering various values for the number of days of simulation. In all cases, $N = 40$.

nario with $N_{\min} = 40$) is the worst.

According to the deep sleep algorithm, in fact, when $N_{\min}$ is small, there is a large number of RFDs in the sleep state, and the energy consumption (averaged over all RFDs in the network) is, therefore, low. On the opposite, when $N_{\min}$ is large or, as a limiting case, the deep sleep algorithm is not used ($N = N_{\min}$), there is a large number of RFDs active in the network, and the energy consumption is high.

In Figure 5.9, the average residual energy is shown as a function of the number of days of simulations, considering constant and Poisson packet generation distributions. As mentioned at the beginning of this section, the packet generation rate is $\lambda = 4$ pck/s—this is the *average* packet generation rate with Poisson distribution. As one can see from results in Figure 5.9, the average residual energy is a linearly decreasing function of the time *regardless* of the packet generation distribution. However, for a given value of $N_{\min}$, it can be observed that a Poisson distributed packet

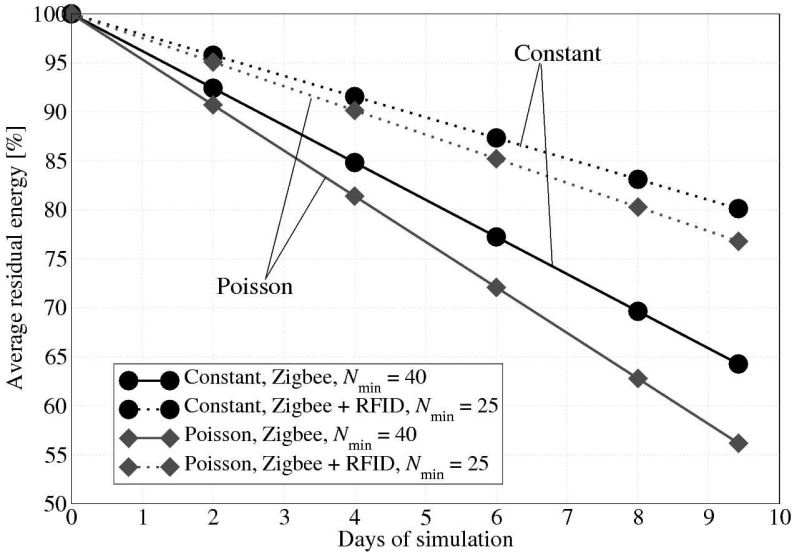


Figure 5.9: Average residual energy, as a function of the number of days of simulation, for various packet generation distributions: (i) constant and (ii) Poisson. In all cases, $N = 40$, whereas N_{min} is set to either 25 or 40.

generation leads to a performance degradation with respect to the case with constant packet generation, and this degradation is more pronounced the larger is N_{min}. This is due to the higher traffic generated with the Poisson distribution. Each active node receives all the packets transmitted in the network and processes only those with a destination address equal to its own address. On the other hand, when a node is in the sleep state, no packet is received and energy is preserved.

Finally, we point out that, also with Poisson-distributed packet generation, different values of the energy step E_{s} do not influence the observed performance.

5.4.2 Impact of the Virtual Spatial Grid

In this section, we present performance results in scenarios where a virtual spatial grid is considered and the deep sleep algorithm is applied locally, as described in

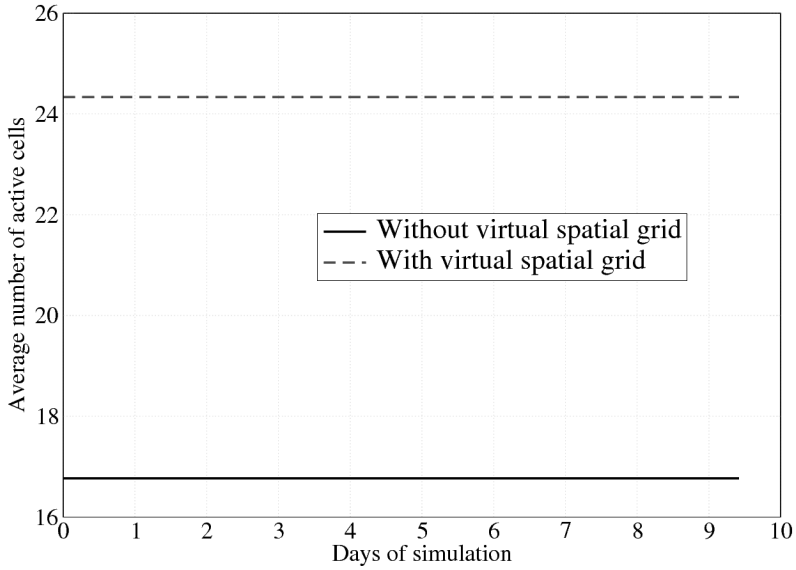


Figure 5.10: Average cells coverage with and without the virtual spatial grid. In all cases, the deep sleep algorithm is used ($N_{\min} = 25$) and $N = 40$.

Subsection 5.3.4. The same network topology of Subsection 5.4.1 is considered and the monitored surface is divided into 36 1 m^2 cells. We remark that, according to our algorithm, in the secondary cells the nodes are turned on according to the traditional superframe structure. In the primary cells, where the deep sleep algorithm is applied cell by cell, instead, there is at least one active node per cell, whereas the other nodes of the cell are cyclically switched into the sleep state.

We have considered different network configurations and we have compared the results in the presence of the virtual spatial grid with those of scenarios without it. For different types of networks, we have considered the same network topology realizations, in order to make the comparison fair.

In Figure 5.10, we present the average number of active cells guaranteed by the two different node management mechanisms. On average, using the different topologies considered in our simulation framework, the virtual spatial grid procedure gua-

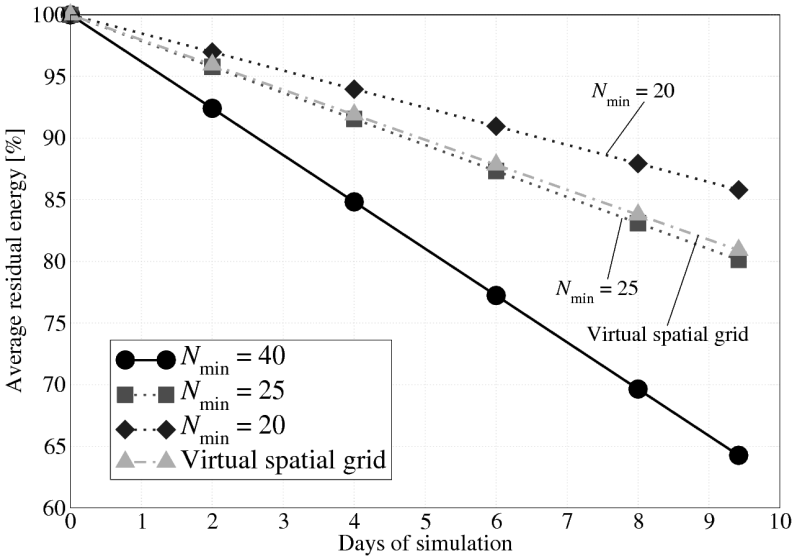


Figure 5.11: Average energy consumption with energy step equal to 1% of the initial energy level. The number of nodes is $N = 40$. The deep sleep algorithm is applied considering $N_{\min} = 20$, $N_{\min} = 25$, respectively. The performance with the virtual spatial grid is also shown.

rantees the coverage of 24 cells, whereas a scenario without virtual spatial grid procedure leads to an average of 17 active cells. As one can see, the average number of active cells remains constant over the simulation: this is due to the fact that in both cases the battery energy is not depleted.

In Figure 5.11 the average energy consumption performance is shown as a function of the number of simulation days, considering scenarios without the deep sleep algorithm ($N_{\min} = 40$), with the deep sleep algorithm without virtual spatial grid ($N_{\min} = 20$, or $N_{\min} = 25$), and with the virtual spatial grid. The curves presented in this figure are obtained in a scenario with $E_s = 1\%$, because we have experienced that, as in the case with deep sleep algorithm, the energy step has no impact on the performance of networks with the virtual spatial grid. The energy consumption of a traditional Zigbee network is larger than that of networks with deep sleep algorithm

both with and without virtual spatial grid. In particular, in scenarios where the virtual spatial grid is applied, on average there are 24 active nodes, and the average residual energy trend is similar to that of the network with deep sleep algorithm and without virtual spatial grid with $N_{\min} = 25$.

5.5 Conclusions

We have shown an approach to integrate Zigbee and RFID networks in order to create an energy-efficient network which allows to selectively turn on and off the remote nodes. Therefore, we have considered a deep sleep algorithm which selects the nodes to be activated according to their residual energy, so that the energy consumption of the nodes in the network is equalized. Finally, we have introduced a procedure which selects the integrated nodes to be activated according not only to their residual energy, but also to their spatial positions. This selection is based on the introduction of a virtual spatial grid over the network surface, and *local* application of the deep sleep algorithm. For the configurations without virtual spatial grid, we have evaluated through the Opnet simulator the average residual energy performance as a function of $N_{\min}$, E_s, highlighting the energy saving guaranteed by the hybrid Zigbee-RFID network. For the configurations with virtual spatial grid, instead, we have analyzed both the area effectively monitored by the sensor network and the energy consumption. In this case, the virtual spatial grid not only provides the same area coverage than traditional Zigbee networks, but also allows to extend the network lifetime. This solution, which should be applied in scenarios where local observation spatial density is relevant, could be adopted in order to create a very energy-efficient wireless sensor network based on totally passive components with addressing capabilities.

Chapter 6

Concluding Remarks and Future Work

In this thesis, we have investigated how to optimally tune the parameters of the nodes of a Wireless Sensor Network (WSN) in order to maximize its performance, possibly satisfying some given constraints. In particular, four different optimization directions have been considered.

First, we have addressed the problem of engineering energy-efficient target detection applications using WSNs. In particular, we have proposed an engineering toolbox which consists of a set of models for describing the probability of missed detection, the alert transmission latency, and the energy consumption under the assumption of random node deployment. By leveraging on this toolbox, we have then characterized the trade-offs faced by the WSNs with respect to energy consumption and quality of service, in terms of detection capabilities and latency. Finally, we have illustrated the use of the toolbox to optimally configure a given WSN under a variety of quality of service requirements. As such, the engineered toolkit gives the possibility to an operator to set up efficiently an unattended WSN for a wide range of scenarios like counting applications with poor latency requirements (e.g., counting animals in a given area using passive infra-red sensing), as well as live monitoring applications with a strong latency and medium P_{md} requirements (e.g., tracking emer-

gency or panic situations in public subway, using audio sensing).

We have presented an optimized transmit power allocation strategy which allows to minimize the Packet Error Rate (PER) at the Access Point (AP) of a WSN. First of all, we have derived a simplified analytical model which describes the performance of a Zigbee WSN, in terms of PER and delay, under the assumption of low offered traffic load. Then, we have presented the proposed transmit power control framework, developed under the assumption of finite overall network transmit power. In particular, we have shown that the performance basically depends on the number of ones in the adjacency matrix: this number represents the active connections between the nodes and is thus an index of the network connectedness. Our analytical model has been validated through the use of the Opnet simulator, underlying the impact, on relevant network performance indicators (PER, network transmission rate, delay, and network lifetime), of the sparsity index of the adjacency matrix, the offered traffic load, and the transmit power allocation strategy. In particular, we have verified that the proposed transmit power control approach, by maximizing the sparsity index of the adjacency matrix, allows to minimize the PER for a given total network transmit power, without reducing the lifetime of the network.

We have then analyzed, first through Opnet-based simulations, the performance of Zigbee WSNs, using physical (probability of decision error at the AP) and network layer (network transmission rate S, throughput, aggregate throughput, and delay D) performance indicators. In non-clustered scenarios, the presence of a large number of transmitting RFDs has a positive effect on the probability of decision error performance, at the price of throughput and delay performance degradation. The probability of decision error can be improved by using ACK messages: in fact, the number of packet losses reduces (i.e., the throughput increases) and the AP can make a final decision on the basis of a larger number of observations. This further improvement is obtained at the expense of a higher delay. We have also analyzed the performance of Zigbee networks in clustered configurations with and without data fusion. In the presence of uniform clustering, our simulation-based results confirm the theoretical predictions (valid for generic distributed detection schemes) obtained in [81]: for a given number of RFDs, the probability of decision error remains the same regard-

less of the clusters' common dimension. On the other hand, the more unbalanced the clustering configuration (e.g., one large cluster and remaining smaller clusters), the higher the probability of decision error. In addition, through a simulation-based analysis of the impact of the network tolerable death level χ_{net}, it appears that the best network configuration is always the one without any relay node. Finally, we have drawn a few simple guidelines for the design of clustered Zigbee wireless sensor networks. A three-dimensional analytical characterization of the network performance in terms of D, S, and χ_{net}, shows that the network operating point lays over a characteristic surface. Given the number of nodes and the required performance level, one can identify, over this surface, the network configuration which guarantees the best trade-off.

Finally, we have proposed an approach to integrate Zigbee and Radio Frequency IDentification (RFID) networks in order to create an energy-efficient radio-controlled WSN for selectively turning on and off the remote nodes. In particular, we have proposed a deep sleep algorithm which selects the nodes to be activated according to their residual energies, so that the energy consumption of the nodes in the network is equalized. Then, we have introduced a procedure which selects the integrated nodes to be activated according not only to their residual energy, but also to their spatial positions. This selection is based on the introduction of a virtual spatial grid over the network surface, and *local* application of the deep sleep algorithm. For the configurations without virtual spatial grid, we have evaluated, through the Opnet simulator, the average residual energy performance as a function of the minimum number of nodes which remain active (N_{min}) and the energy step (E_{s}) used for the equalization, highlighting the energy saving guaranteed by the hybrid Zigbee-RFID network. For the configurations with virtual spatial grid, instead, we have analyzed both the area effectively monitored by the sensor network and the energy consumption. In this case, the virtual spatial grid not only provides the same area coverage than traditional Zigbee networks, but also allows to extend the network lifetime. This solution, which should be applied in scenarios where local observation spatial density is relevant, could be adopted in order to create a very energy-efficient WSN based on totally passive components with addressing capabilities.

Bibliography

[1] R. Abileah and D. Lewis, "Monitoring high-seas fisheries with long-range passive acoustic sensors," in *Proceedings of the International Conference* OCEANS'96*: Prospects for the 21th century*, vol. 1, Fort Lauderdale, FL, USA, September 1996, pp. 378–382.

[2] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A Survey on Sensor Networks," *IEEE Communication Magazine*, vol. 40, no. 8, pp. 102–114, August 2002.

[3] S. Barberis, E. Gaiani, B. Melis, and G. Romano, "Performance evaluation in a large environment for the AWACS system," in *Proceedings of the International Conference on Universal Personal Communications* (ICUPC'98), vol. 1, Florence, Italy, October 1998, pp. 721–725.

[4] C. Chong, S. Kumar, and B. Hamilton, "Sensor networks: Evolution, opportunities, and challenges," *Proceedings of the IEEE*, vol. 91, no. 8, pp. 1247–1256, 2003.

[5] S. N. Simic and S. Sastry, "Distributed environmental monitorning using random sensor networks," in *Proceedings of the 2nd International Workshop on Information Processing in Sensor Networks*, Palo Alto, CA, USA, April 2003, pp. 582–592.

[6] M. Madou, *Fundamentals of Microfabrication*. Boca Raton, FL: CRC Press, 1997.

[7] Zigbee Alliance Website, "`http://www.zigbee.org`."

[8] R. Verdone, D. Dardari, G. Mazzini, and A. Conti, *Wireless sensor and actuator networks: Technologies, Analysis and Design*. London, UK: Academic Press, 2008.

[9] T. He, S. Krishnamurthy, L. Luo, T. Yan, L. Gu, R. Stoleru, G. Zhou, Q. Cao, P. Vicaire, J. A. Stankovic, T. F. Abdelzaher, J. Hui, and B. Krogh, "Vigilnet: An integrated sensor network system for energy-efficient surveillance," *ACM Trans. Sen. Netw.*, vol. 2, no. 1, pp. 1–38, February 2006.

[10] J. Jeong, Y. Gu, T. He, and D. Du, "VISA: Virtual Scanning Algorithm for Dynamic Protection of Road Networks," in *Proc. of 28th IEEE Conference on Computer Communications* (INFOCOM 09), Rio de Janeiro, Brazil, April 2009.

[11] Q. Cao, T. Yan, J. Stankovic, and T. Abdelzaher, "Analysis of target detection performance for wireless sensor networks," *Lecture Notes in Computer Science*, vol. 3560, pp. 276–292, 2005.

[12] L. Lazos, R. Poovendran, and J. A. Ritcey, "Analytic evaluation of target detection in heterogeneous wireless sensor networks," *ACM Trans. Sensor Networks*, vol. 5, no. 2, pp. 1–38, March 2009.

[13] W.-J. Huang, F.-H. Chiu, C.-C. Kuo, and Y.-W. Hong, "Comparison of Power Control Schemes for Relay Sensor Networks," *IEEE International Conference on Acoustics, Speech and Signal Processing (*ICASSP 2007*)*, vol. 3, pp. 477–480, April 2007.

[14] J. Fang and H. Li, "Power Constrained Distributed Estimation Over Noisy Channels in WSNs," *42nd Annual Conference on Information Sciences and Systems (*CISS 2008*)*, pp. 1053–1057, March 2008.

[15] W. Yu and J. Yuan, "Joint Source Coding, Routing and Resource Allocation for Wireless Sensor Networks," *IEEE International Conference on Communications (*ICC 2005*)*, vol. 2, pp. 737–741, May 2005.

[16] J. Matamoros and C. Anton-Haro, "Opportunistic Power Allocation schemes for the maximization of network lifetime in wireless sensor networks," *IEEE International Conference on Acoustics, Speech and Signal Processing (*ICASSP 2008*)*, pp. 2273–2276, April 2008.

[17] B. Z. Ares, P. G. Park, C. Fischione, A. Speranzon, and K. H. Johansson, "On Power Control for Wireless Sensor Networks: System Model, Middleware Component and Experimental Evaluation," *Proceedings of European Control Conference (ECC'07)*, July 2007, 8 pages.

[18] A. Bonivento, C. Fischione, L. Necchi, F. Pianegiani, and A. Sangiovanni-Vincentelli, "System level design for clustered wireless sensor networks," vol. 3, no. 3, pp. 202–214, August 2007.

[19] B. Yin, H. Shi, and Y. Shang, "Analysis of Energy Consumption in Clustered Wireless Sensor Networks," in *Proceedings on the 2nd International Symposium on Wireless Pervasive Computing* (ISWPC'07), San Juan, Porto Rico, February 2007, pp. 77–82.

[20] Z. Zhou, S. Zhou, S. Cui, and J. Cui, "Energy-Efficient Cooperative Communication in Clustered Wireless Sensor Networks," in *Proc. IEEE Military Comm. Conf.* (MILCOM'06), Washington, DC, USA, October 2006, pp. 1–7.

[21] R. R. Tenney and N. R. Sandell, "Detection with distributed sensors," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 17, no. 4, pp. 501–510, December 1981.

[22] W. Shi, T. W. Sun, and R. D. Wesel, "Quasi-convexity and optimal binary fusion for distributed detection with identical sensors in generalized Gaussian noise," *IEEE Trans. Inform. Theory*, vol. 47, no. 1, pp. 446–450, January 2001.

[23] Q. Yang and J. Sun, "An underwater autonomous robot based on multi-sensor data fusion," in *Proc. 6th World Congress on Intelligent Control and Automation(*WCICA'06*)*, vol. 2, Hong Kong, China, June 2006, pp. 9139–9143.

[24] B. Krishnamachari, D. Estrin, and S. B. Wicker, "The impact of data aggregation in wireless sensor networks," in *Proc. of the 22nd Int. Conf. on Distributed Computing Systems* (ICDCS'02), Vienna, Austria, July 2002, pp. 575–578.

[25] A. Kansal, A. Ramamoorthy, M. Srivastava, and G. Pottie, "On sensor network lifetime and data distortion," in *Proc. IEEE Symposium on Information Theory* (ISIT), Adelaide, Australia, September 2005, pp. 6–10.

[26] S. Arnon, "Deriving an upper bound on the average operation time of a wireless sensor network," *IEEE Commun. Lett.*, vol. 9, no. 2, pp. 154–156, February 2005.

[27] F. Ordonez and B. Krishnamachari, "Optimal information extraction in energy-limited wireless sensor networks," *IEEE J. Select. Areas Commun.*, vol. 22, no. 6, pp. 1121–1129, August 2004.

[28] H. Zhang and J. Hou, "On deriving the upper bound of lifetime for large sensor networks," in *Proc. ACM Int. Symp. on Mobile Ad Hoc Network. and Comput.* (MOBIHOC), Tokyo, Japan, May 2004, pp. 121–132.

[29] Z. Hu and B. Li, "On the fundamental capacity and lifetime limits of energy-constrained wireless sensor networks," in *Proc. 10th IEEE Real-Time and Embedded Technology and Applications Symp.* (RTAS'04), Toronto, Canada, May 2004.

[30] D. M. Blough and P. Santi, "Investigating upper bounds on network lifetime extension for cell-based energy conservation techniques in stationary ad-hoc networks," in *Proc. ACM Intern. Conf. on Mobile Comput. and Networking* (MOBICOM), Atlanta, Georgia, USA, September 2002, pp. 183–192.

[31] M. Bhardwaj, T. Garnett, and A. P. Chandrakasan, "Upper bounds on the lifetime of sensor networks," in *Proc. IEEE International Conf. on Commun.* (ICC), vol. 119, Helsinki, Finland, June 2001, pp. 785–790.

[32] M. Bhardwaj and A. P. Chandrakasan, "Bounding the lifetime of sensor networks via optimal role assignments," in *Proc. IEEE Conf. on Computer Commun.* (INFOCOM), vol. 3, New York, NY, USA, June 2002, pp. 1587–1596.

[33] V. Rai and R. N. Mahapatra, "Lifetime modeling of a sensor network," in *Proc. Design, Automation and Test in Europe 2006* (DATE'05), vol. 1, Munich, Germany, March 2003, pp. 202–203.

[34] Y. Chen and Q. Zhao, "On the lifetime of wireless sensor networks," *IEEE Commun. Lett.*, vol. 9, no. 11, pp. 976–978, November 2005.

[35] Q. Zhao, A. Swami, and T. Long, "The interplay between signal processing and networking in sensor networks," *IEEE Signal Processing Mag.*, vol. 23, no. 4, pp. 84–93, 2006.

[36] K. Kalpakis, K. Dasgupta, and P. Namjoshi, "Maximum lifetime data gathering and aggregation in wireless sensor networks," University of Maryland, Baltimore, Tech. Rep., 2002, available at `http://www.csee.umbc.edu/~kalpakis/`.

[37] S. Coleri, M. Ergen, and T. J. Koo, "Lifetime analysis of a sensor network with hybrid automata modelling," in *Proc. First Int. Workshop on Wireless Sensor Networks and Applications 2002* (WSNA'02), Atlanta, GA, USA, September 2002, pp. 98–104.

[38] M. Franceschetti and R. Meester, "Critical node lifetime in random networks via the Chen-Stein method," *IEEE Trans. Inform. Theory*, vol. 52, no. 6, pp. 2831–2837, June 2006.

[39] N. F. Timmons and W. G. Scanlon, "Analysis of the performance of IEEE 802.15.4 for medical sensor body area networking," in *Proc. IEEE Conference on Sensor and Ad Hoc Communications and Networks* (SECON'04), Santa Clara, CA, USA, October 2004, pp. 16–24.

[40] IEEE 802.15.4 Std: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low-Rate Wireless Personal Area Networks (LR-WPANs), *IEEE Computer Society Press*, pp. 1–679, October 2003.

[41] O. Hyncica, P. Kacz, P. Fiedler, Z. Bradac, P. Kucera, and R. Vrba, "The Zigbee experience," in *Proc. Int. Symposium on Communications, Control and Signal Processing* (ISCCSP'06), Marrakech, Morocco, March 2006.

[42] J. S. Lee, "An Experiment on Performance Study of IEEE 802.15.4 Wireless Networks," in *In Proc. 10th Conference on Emerging Technologies and Factory Automation 2005*, (ETFA'05), vol. 2, Catania, Italy, September 2005, pp. 451–458.

[43] G. Ferrari, P. Medagliani, S. Di Piazza, and M. Martalò, "Wireless Sensor Networks: Performance Analysis in Indoor Scenarios," *EURASIP Journal on Wireless Communication and Networking*, Special Issue on *Novel Techniques for Analysis & Design of Cross-Layer Optimized Wireless Sensor Networks*, vol. 2007, 2007, 20 pages.

[44] B. Bougard, F. Catthoor, D. C. Daly, A. Chandrakasan, and W. Dehaene, "Energy Efficiency of the IEEE 802.15.4 Standard in Dense Wireless Microsensor Networks: Modeling and Improvement Perspectives," in *Proc. in Design, Automation and Test in Europe (DATE'05)*, Munich, Germany, March 2005, pp. 196–201.

[45] I. Ramachandran, A. K. Das, and S. Roy, "Analysis of the Contention Access Period of IEEE 802.15.4 MAC," *ACM Trans on Sensor Networks (TOSN)*, vol. 3, no. 1, pp. 1–29, March 2007.

[46] K. Finkenzeller, *RFID Handbook - Fundamentals and Applications in Contactless Smart Cards and Identification*. New York, NY, USA: John Wiley & Sons, 2003.

[47] B. Nath, F. Reynolds, and R. Want, "RFID Technology and Applications," *IEEE Pervasive Computing*, vol. 5, no. 1, pp. 22–24, March 2006.

[48] H. Vogt, "Efficient Object Identification with Passive RFID Tags," in *Proc. of the 1st Int. Conf. on Pervasive Computing (*PERVASIVE'03*)*, Denver, CO, USA, July 2002, pp. 98–113.

[49] R. Weinstein, "RFID: a Technical Overview and its Application to the Enterprise," *IT Professional*, vol. 7, no. 3, pp. 27–33, May 2005.

[50] L. M. Ni, Y. Liu, Y. C. Lau, and A. P. Patil, "LANDMARC: indoor location sensing using active RFID," *Pervasive Computing and Communications, 2003. (*PerCom 2003*). Proc. of the First IEEE Int. Conf.*, pp. 407–415, March 2003.

[51] L. Zhang and Z. Wang, "Integration of RFID into Wireless Sensor Networks: Architectures, Opportunities and Challenging Problems," in *Proc. of 5th Int. Conf. on Grid and Cooperative Computing Workshops (GCCW'06)*, Changsha, Hunan, China, October 2006, pp. 463–469.

[52] J. Sung, T. Sanchez Lopez, and D. Kim, "The EPC Sensor Network for RFID and WSN Integration Infrastructure," *Pervasive Computing and Communications Workshops, 2007. (*PerCom Workshops '07*). Fifth Annual IEEE Int. Conf.*, pp. 618–621, March 2007.

[53] A. Tanenbaum, *Computer networks*. Upper Saddle River, NJ, USA: Prentice-Hall, 2002.

[54] "IEEE Standards for Local Area Networks: Logical Link Control," *IEEE Computer Society Press*, pp. 1–116, December 1984.

[55] N. Standard, "Announcing the ADVANCED ENCRYPTION STANDARD (AES)," *Federal Information Processing Standards Publication*, vol. 197, 2001.

[56] A. Sikora and V. Groza, "Coexistence of IEEE 802. 15.4 with other Systems in the 2.4 GHz-ISM-Band," in *Proc. of IEEE Instrumentation and Measurement Technology Conference* (IMTC), vol. 22, no. 2. Ottawa, ON, Canada: IEEE Computer Society Press, May 2005, pp. 1786–1791.

[57] A. B. Carlson, P. B. Crilly, and J. Rutledge, *Communication systems*. McGraw-Hill, 2002.

[58] M. Petrova, J. Riihijarvi, P. Mahonen, and S. Labella, "Performance Study of IEEE 802.15. 4 Using Measurements and Simulations," in *Proc. of Wireless Communications and Networking Conferenc* (WCNC'06), New Orleans, FL, USA, April 2006, pp. 487–492.

[59] Z. Tao, S. Panwar, D. Gu, and J. Zhang, "Performance analysis and a proposed improvement for the IEEE 802.15. 4 contention access period," in *Proc. of IEEE Wireless Communications and Networking Conference (*WCNC 2006*)*, Las Vegas, NV, USA, April 2006, pp. 1811–1818.

[60] F. Tobagi and L. Kleinrock, "Packet switching in radio channels: part II–the hidden terminal problem in carrier sense multiple-access and the busy-tone solution," *IEEE Trans. on Communications*, vol. 23, no. 12, pp. 1417–1433, 1975.

[61] J. Misic, S. Shafi, and V. Misic, "Avoiding the bottlenecks in the MAC layer in 802.15. 4 low rate WPAN," in *Proc. of the 11th Int. Conf. on Parallel and Distributed Systems (*ICPADS 05*)*, vol. 2, Washington, DC, USA, 2005.

[62] ZigBee Specification, "Zigbee Standards Organization Std. r13," December 1st 2006.

[63] C. Perkins, E. Belding-Royer, and S. Das, "Ad hoc on-demand distance vector (AODV) routing," *RFC 3561*, July 2003.

[64] P. Medagliani, J. Leguay, V. Gay, M. Lopez-Ramos, and G. Ferrari, "Engineering energy-efficient target detection applications in wireless sensor networks," in *8th Ann. IEEE Int. Conf. on Pervasive Computing and Communications* (PERCOM 2010), Mannheim, Germany, February 2010.

[65] M. Buettner, G. Yee, E. Anderson, and R. Han, "X-mac: a short preamble mac protocol for duty-cycled wireless sensor networks," in *Proc. of 4th ACM*

Int. Conf. on Embedded networked sensor systems (SenSys '06), Boulder, CO, USA, November 2006, pp. 307–320.

[66] L. A. Santalò and M. Kac, *Integral geometry and geometric probability*. Cambridge, UK: Cambridge University Press, 2004.

[67] H. Solomon, *Geometric probability*. Philadelphia, PA, USA: SIAM, 1978.

[68] A. Papoulis, *Probability, Random Variables and Stochastic Processes*. New York, NY: McGraw-Hill, 1991.

[69] J. Polastre, J. Hill, and D. Culler, "Versatile Low Power Media Access for Wireless Sensor Networks," in *Proc. of 2nd ACM Int. Conf. on Embedded networked sensor systems* (SenSys '04), Baltimore, MD, USA, November 2004, pp. 95–107.

[70] L. Consolini, P. Medagliani, and G. Ferrari, "Adjacency Matrix-Based Transmit Power Allocation Strategies in Wireless Sensor Networks," *Sensors*, Special Issue on *State-of-the-Art Sensors Technology in Italy*, vol. 9, no. 7, pp. 5390–5422, 2009.

[71] T. Rappaport, *Wireless communications: principles and practice*. Upper Saddle River, NJ, USA: Prentice-Hall, 1996.

[72] D. Bertsekas and R. Gallager, *Data networks*. Upper Saddle River, NJ, USA: Prentice-Hall, 1987.

[73] D. Cox and V. Isham, *Point processes*. Boca Raton, FL, USA: Chapman & Hall/CRC, 1980.

[74] P. Medagliani, M. Martalò, and G. Ferrari, "A Multi-Dimensional Characterization of Clustered Zigbee Networks," in *Proceedings of the 10th IEEE International Symposium on Spread Spectrum Techniques and Applications (*ISSSTA'08*)*, Bologna, Italy, Aug 2008, pp. 12–17.

[75] P. Sinha and A. Zoltners, "The multiple-choice knapsack problem," *Operations Research*, pp. 503–515, 1979.

[76] Mosek Website, "`http://www.mosek.com`."

[77] J. Löfberg, "Yalmip : A toolbox for modeling and optimization in MATLAB," in *Proc. of the IEEE Symp. on Computer Aided Control System Design* (CACSD2004), Taipei, Taiwan, September 2004. [Online]. Available: http://control.ee.ethz.ch/~joloef/yalmip.php

[78] P. Medagliani, M. Martalò, and G. Ferrari, "Clustered Zigbee Networks with Data Fusion: Characterization and Performance Analysis," *Ad-Hoc Networks*, p. 34 pages, 2009.

[79] R. E. Ziemer, *Elements of Engineering Probability and Statistics.* Upper Saddle River, NJ, USA: Prentice-Hall, 1997.

[80] G. Ferrari and M. Martalò, "Extending the lifetime of sensor networks through adaptive reclustering," *EURASIP Journal on Wireless Communications and Networking*, Special Issue on *Novel Techniques for Analysis & Design of Cross-Layer Optimized Wireless Sensor Networks*, vol. 2007, 2007, 20 pages.

[81] G. Ferrari, M. Martalò, and R. Pagliari, "On multi-level decentralized detection in sensor networks," in *International Conference on Intelligent Systems And Computing: Theory And Applications* (ISYC'06), Ayia Napa, Cyprus, July 2006.

[82] Opnet Website, `http://www.opnet.com`.

[83] National Institute of Standards and Technology (NIST) Opnet model Website, `http://w3.antd.nist.gov/Health.shtml`, Washington, DC, USA.

[84] D. Bertsimas and J. Tsitsiklis, *Introduction to linear optimization*. Belmont, MA, USA: Athena Scientific, 1997.

[85] K. Deb, S. Agrawal, A. Pratap, and T. Meyarivan, "A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II," in *Proc. of the 6th Int. Conf. on Parallel Problem Solving From Nature* (PPSN VI), Paris, France, September 2000, pp. 849–858.

[86] N. Golmie, D. Cypher, and O. Rebala, "Performance Evaluation of Low Rate WPANs for Medical Applications," in *Proceedings of the IEEE Military Communucation Conference* (MILCOM '04), vol. 2, November 2004, pp. 927–933.

[87] R. Ziemer and R. Peterson, *Introduction to digital communication*. Prentice Hall Upper Saddle River, NJ, 2001.

[88] P. Medagliani, G. Ferrari, and M. Marastoni, *Remote Instrumentation and Virtual Laboratories*. Springer, 2010, ch. Hybrid Zigbee-RFID Networks for Energy Saving and Lifetime Maximization, pp. 473–491.

[89] J. Capetanakis, "Tree algorithms for packet broadcast channels," *IEEE Trans. on Information Theory*, vol. 25, no. 5, pp. 505–515, 1979.

[90] ISO/IEC 18000-6: Information Technology – Radio Frequency Identification for Item Management – Part 6: Parameters for Air Interface Communications at 860 MHz to 960 MHz, *International Organization for Standardization (ISO)*, pp. 1–134, 2004.

[91] D. Bertsekas and R. Gallagher, *Data Networks*. Upper Saddle River, NJ, USA: Prentice-Hall, 1992.

[92] Crossbow Technology, Inc., *MICAz OEM Module 2.4GHz*, San Jose, CA, USA, 2005, datasheet, website: `http://www.xbow.com/Products/productdetails.aspx?sid=191`.

[93] P. Jurcik and A. Koubaa, "The IEEE 802.15.4 OPNET simulation model: reference guide v2.0," IPP HURRAY! Research Group, Polytechnic Institute of Porto (ISEP-IPP), Porto, Portogallo, Tech. Rep., May 2007, website: `http://www.open-zb.net/research.php`.

[94] Polytechnic Institute of Porto, IPP-HURRAY! Research Group, "`http://www.hurray.isep.ipp.pt:8080/opnet.html`."

[95] G. Ferrari, F. Cappelletti, and R. Raheli, "A simple performance analysis of RFID networks with binary tree collision arbitration," *International Journal on Sensor Networks*, vol. 4, no. 3, pp. 61–75, 2008.

Acknowledgments

I am really at the end of my school course. I am just finished my third thesis and, as usual, I want to thank all the people that, for some reasons have helped me during these years. Unfortunately, every list must have a head and a tail, so there will be someone who will be acknowledged before the others. However, I want to point out that the list is not in order of importance because each person, who I am going to mention, has been, in his way, really really important to me.

I first want to thank my advisor Dr. Gianluigi Ferrari. Thanks to him I managed to fulfill my research attitude and I had been able to work on such interesting topics (at least for me...). During these three years of PhD (four years considering the “Specialistic” thesis) we have had wonderful times together discussing, working hard and having fun. The important thing is that Gian for me has been more than a boss. He has been a friend, so good luck for your future, Gian!!! Of course good luck also to Anna and to the little Sofia, who has brought a lot of changes...

Switching to the personal side, I have to thank a lot my father, my mother, and Isotta (Isotta, do not complain for being the third, I wrote them in order of age). I want to thank them together because they have been and they will be forever the fulcrum of my life. You supported me when I got discouraged, you trusted in me every time and you took part in every moment of happiness that I had experienced. I know it’s not really a “deep” concept, but trust me, my thanks are really sincere and they come from the deep of my heart. I can’t also forget my grandmother who “supported” and rewarded my pains with delicious tasty dishes. A special mansion must be done to my mother-in-law who helped me to be more self confident and more conscious of

my capabilities. In the end, as I am here writing the acknowledgments, she was right.

Let's now start with the colleagues; in Parma I had to share my office with Marco and Busa, but I can't forget Marcone, Muzz, Ace, Roberto M., Vito, Nicola, and Marco il tesista. For some reasons you all contributed to create such a wonderful working group and I am sure I won't ever be able to spend such a pleasant time in a office. Anyway, I think of being quite lucky because also during my internship in France at Thales I met really amazing people that made my staying in Paris wonderful: Vincent, Jérémie, Mario, Marcel, Farid, Nicholas, Stephane. I am sure that I am omitting someone but I just wanted to mention all that people that "help" more. The others will excuse me for my negligence. I can proudly say that all the colleagues that I have had during these years have now friends of mine.

Finally, it's the turn of the other friends. I can't forget Simone, Matteo, Ross, Paolo, Giuseppe, Igor, and Michele, who made me have wonderful fun during these years. Thanks a lot for your unconditioned friendship. When I have been in trouble, you have always been on my side, listening to, supporting, and advising me when needed. The last mention must be done to all the other friends, especially in Cremona, with whom I spent time relaxing and having fun. I don't want to cite all of them because it would be a long list, but a special remark must be given to Roberto, Giorgio, Valentina, and Marco M.

With these last words I conclude my thesis, my PhD. and my experience as a student. Once again thanks to all!!!